



**Politecnico  
di Torino**

# POLITECNICO DI TORINO

**Corso di Laurea Magistrale  
In Ingegneria Gestionale (LM-31)**

**Tesi di Laurea Magistrale**

**Sharing economy e Airbnb: analisi della domanda e delle  
differenze di genere tramite i feedback dei consumatori**

**Relatore:**

Prof.ssa Laura Rondi

**Candidato:**

Sara Chiapello

A.a. 2020/2021

Sessione di Laurea Dicembre 2021

## Sommario

INTRODUZIONE .....	5
DOMANDE DI RICERCA .....	8
1 SHARING ECONOMY .....	9
1.1 Definizione.....	9
1.2 Possesso o accesso .....	13
1.3 Principali fattori che ne hanno favorito il successo .....	18
1.3.1 <i>Il prezzo</i> .....	18
1.3.2 <i>Chiunque può diventare un microimprenditore</i> .....	19
1.3.3 <i>La diffusione di massa delle nuove tecnologie</i> .....	19
1.3.4 <i>L'attenzione alle tematiche ambientali</i> .....	20
1.4 Cosa sancisce il successo di una società di sharing economy .....	24
1.5 Problematiche riscontrate.....	24
1.6 Possibili soluzioni presenti e future dei problemi.....	26
1.6.1 <i>Creazione sentimento di fiducia</i> .....	26
1.6.2 <i>Definire norme di regolamentazione</i> .....	30
2 AIRBNB .....	34
2.1 Storia.....	34
2.2 Growth hacking.....	35
2.3 Business model .....	36
2.4 Problematiche riscontrate.....	37
2.4.1 <i>Discriminazione</i> .....	37
2.4.2 <i>Gentrificazione</i> .....	39
2.4.3 <i>Sicurezza ospiti/host</i> .....	41
2.5 Costruzione del sentimento di fiducia di un utente.....	42
2.5.1 <i>Creazione account</i> .....	43

2.5.2	<i>Sistema di feedback</i> .....	44
2.5.3	<i>Superhost</i> .....	46
3	LETTERATURA ESISTENTE .....	48
3.1	Recensioni.....	48
3.1.1	<i>Overcoming the J-shaped distribution of product reviews</i> .....	48
3.1.2	<i>A first look at online reputation on Airbnb, where every stay is above average</i> .....	49
3.1.3	<i>If nearly all Airbnb reviews are positive, does that make them meaningless?</i> .....	51
3.1.4	<i>Large-Scale Sentiment Analysis on Airbnb Reviews from 15 Cities</i> .....	52
3.1.5	<i>What do Airbnb users care about? An analysis of online review comments</i> .....	53
3.1.6	<i>Neutrality may matter: sentiment analysis in reviews of Airbnb, Booking, and Couchsurfing in Brazil and USA</i> .....	54
3.1.7	<i>Gender differences in the perceived risk of buying online and the effects of receiving a site recommendation</i> .....	57
3.1.8	<i>Exploring gender-based influences on key features of Airbnb accommodations</i> .....	58
3.2	Prezzo.....	59
3.2.1	<i>Key factors affecting the price of airbnb listings: a geografically weighted approach</i> ....	59
3.2.2	<i>Do airbnb host listing attributes influence room pricing homogenously?</i> .....	60
3.2.3	<i>Hedonic pricing and the sharing economy: how profile characteristics affect Airbnb accommodation prices in Barcelona, Madrid, and Seville</i> .....	61
3.2.4	<i>Drivers of Airbnb prices according to property/room type, season and location: a regression approach</i> .....	62
3.2.5	<i>Spatial price mimicking on Airbnb: Multi-host vs single-host</i> .....	63
3.2.6	<i>Location, location and professionalization: a multilevel hedonic analysis of Airbnb listing prices and revenue</i> .....	64
4	ANALISI DATI.....	66
4.1	Descrizione database disponibili .....	66
4.2	Data exploration.....	69
4.3	Classificazione variabili disponibili.....	73
4.3.1	<i>Lingua</i> .....	73
4.3.2	<i>Maschio/femmina</i> .....	74

5	ANALISI FEEDBACK .....	76
5.1	Punteggio rating recensioni .....	76
5.1.1	<i>Parametri disponibili</i> .....	76
5.1.2	<i>Distorsione recensioni troppo positive</i> .....	76
5.2	Recensioni testuali .....	78
5.2.1	<i>Lunghezza recensioni</i> .....	78
5.2.2	<i>Classificazione delle recensioni</i> .....	79
5.2.3	<i>L'host è chiamato con il proprio nome</i> .....	83
5.2.4	<i>Parole più utilizzate (in differenti situazioni)</i> .....	83
6	DIFFERENZE COMPORTAMENTALI E STRUTTURALI SIGNIFICATIVE..	89
6.1	Esiste una differenza di genere nel mercato Airbnb? .....	90
6.1.1	<i>C'è una differenza tra i recensori?</i> .....	91
6.1.2	<i>C'è differenza nelle recensioni rilasciate?</i> .....	91
6.1.3	<i>C'è una differenza nei possibili ricavi?</i> .....	93
6.1.4	<i>C'è una differenza negli host?</i> .....	93
6.1.5	<i>C'è una differenza nei feedback degli annunci?</i> .....	94
6.1.6	<i>La scelta dell'host è influenzata dal genere?</i> .....	95
6.2	Esistono differenze tra host amatoriali e professionisti .....	96
6.2.1	<i>Relazione genere ospite e host professionista</i> .....	99
6.3	Esistono delle differenze significative tra gli alloggi condivisi e quelli privati ...	100
7	ANALISI REGRESSIONE .....	103
7.1	Concetti teorici.....	103
7.2	Variabili .....	106
7.3	Regressione .....	108
7.3.1	<i>Elementi strutturali dell'alloggio</i> .....	109
7.3.2	<i>Notti</i> .....	111
7.3.3	<i>Posizione</i> .....	111
7.3.4	<i>Host</i> .....	112

7.3.5	<i>Disponibilità</i> .....	112
7.3.6	<i>Recensioni</i> .....	113
7.3.7	<i>Risultati</i> .....	114
7.3.8	<i>Il genere dell'host influenza il prezzo?</i> .....	118
7.3.9	<i>Differenza tra sistemazioni private e condivise</i> .....	121
7.3.10	<i>Contenuto dei feedback e prezzo</i> .....	125
7.3.11	<i>Le donne host ricevono recensioni migliori dalle donne guest?</i> .....	129
7.3.12	<i>Gli ospiti preferiscono soggiornare da host dello stesso sesso?</i> .....	130
8	CONCLUSIONI .....	135
9	LIMITI DELL'ANALISI.....	138
	INDICE FIGURE .....	139
	INDICE GRAFICI.....	139
	INDICE TABELLE .....	140
	BIBLIOGRAFIA .....	142
	SITOGRAFIA.....	149
	RINGRAZIAMENTI.....	150

## INTRODUZIONE

La “sharing economy” è una moderna tipologia di economia che si è sviluppata grazie alla rapida diffusione di Internet, nella quale chiunque voglia affittare o vendere un prodotto o un servizio può farlo caricando direttamente un’inserzione sulle apposite piattaforme ed il consumatore interessato può usufruirne in cambio di un compenso.

Il presente lavoro di tesi analizza “Airbnb”, la principale piattaforma di sharing economy che ha rivoluzionato in poco tempo il mercato degli affitti a breve termine: nata nel 2008, è cresciuta rapidamente, grazie alla competitività del prezzo e all’interazione diretta con la popolazione locale, raggiungendo nel dicembre del 2020 circa 5,6 milioni di annunci attivi in tutto il mondo.

Nei primi capitoli dell’elaborato, dopo aver illustrato le principali caratteristiche, le maggiori problematiche emerse, nonché i punti di forza e di debolezza relativi alla sharing economy il discorso si è poi focalizzato su Airbnb.

Successivamente si sviluppa la parte operativa del lavoro in cui sono stati analizzati i dati scaricati il 7 luglio 2021 dalla piattaforma InsideAirbnb relativi alla città di Barcellona; in particolare, sono stati considerati i listing attivi in quella data e i feedback rilasciati dagli utenti al termine del loro soggiorno presso alloggi di Airbnb della città tra il primo gennaio 2019 e i primi di luglio 2021. Si è iniziato esplorando i dati disponibili per individuare eventuali errori di battitura, si sono classificati i generi dei recensori e degli host a seconda del loro nome e si è individuata la lingua principale del commento per poi tradurli tutti in inglese. Dopo un primo esame dei punteggi numerici, tramite alcuni algoritmi di Sentiment Analysis e Text Mining sono stati studiati i commenti testuali individuando il sentimento che esprimevano, con l’obiettivo di trovare quali attributi e argomenti influenzano principalmente la “Airbnb experience”. Lo studio di analisi si è poi concentrato su fattori come il genere dei recensori, il periodo temporale e la professione dell’host per valutare se condizionassero i commenti rilasciati. Il lavoro statistico si è successivamente focalizzato sull’utente per esaminare se il suo comportamento fosse influenzato dal genere: esaminando per i recensori le caratteristiche delle opinioni espresse, per gli ospiti la tipologia di annunci scelto ed infine se il genere dell’host desse un “premio” di prezzo. In particolare, sono state svolte delle regressioni lineari (OLS) per valutare quali fossero i fattori chiave ad

influenzare il prezzo degli annunci considerando le variabili strutturali dell'appartamento, la sua posizione e la sua disponibilità, le caratteristiche degli host e le recensioni ottenute. Infine, utilizzando una funzione probit si è valutato se il sesso degli ospiti fosse significativo nella scelta di un annuncio con determinate caratteristiche. Le diverse operazioni e le analisi sono state svolte utilizzando il linguaggio di programmazione Python codificato su un notebook Jupyter e il software statistico Stata.

Nel dettaglio, il lavoro è strutturato come segue. Il primo capitolo presenta la sharing economy dandone la definizione, soffermandosi sul concetto di accesso rispetto a quello di possesso, presentando inoltre i fattori che ne hanno sancito il successo nonché le principali problematiche riscontrate che in genere possono essere risolte con apposite norme di regolamentazione o con la creazione di fiducia tra gli utenti. Quest'ultimo sentimento è generato solitamente dai feedback che si scambiano gli utenti attraverso punteggi numerici o recensioni scritte e sarà l'argomento chiave della tesi.

Nel secondo, invece, il focus passa su Airbnb, dopo averne presentato la storia e il modello di business, si descrivono i principali problemi che deve affrontare la piattaforma legati alla discriminazione, gentrificazione e alla sicurezza degli utenti, per poi presentarne le principali soluzioni soffermandosi sull'importanza della costruzione del sentimento di fiducia tra gli utenti che è resa possibile, oltre che dalla precisa creazione dell'account e dal titolo di superhost, principalmente dal sistema di feedback che mette a disposizione la piattaforma, che, in quanto fulcro della tesi, sarà descritto nel dettaglio.

Nel terzo capitolo sono riportati alcuni studi esistenti in letteratura significativi sul tema delle recensioni e del prezzo in quanto saranno poi gli argomenti chiave dell'analisi empirica.

Nel quarto capitolo invece sono presentati i dati disponibili e sono svolte le operazioni per creare i database per le analisi successive: si classifica il genere degli utenti attraverso il loro nome, si individua la lingua con cui hanno scritto il commento per procedere con la traduzione in inglese.

Nel quinto capitolo si analizzano i feedback numerici e testuali, in particolare questi ultimi sono classificati a seconda del sentimento che esprimono in positivi, negativi o neutri per poi studiarne i termini più utilizzati.

Nel sesto capitolo con il test di ipotesi si studiano le differenze comportamentali e strutturali, le prime legate al genere degli utenti, agli host amatoriali e professionisti, mentre le seconde in relazione alla tipologia di alloggio privato o condiviso.

Nel settimo capitolo si presentano le regressioni più significative che sono emerse. Con il metodo di regressione lineare OLS si sono cercate di individuare le variabili più significative che influiscono sul prezzo degli annunci variando i campioni utilizzati focalizzandosi sul genere degli host per valutare se questo permettesse un premio di prezzo. Sempre con il modello OLS si è valutato se le recensioni rilasciate fossero più critiche al variare del genere degli host e degli ospiti. Successivamente utilizzando un modello di funzione probit si è provato a rispondere all'interrogativo se gli ospiti preferiscano soggiornare da host dello stesso sesso, individuando una funzione che dati alcuni parametri permette di stabilire la probabilità che l'ospite sia maschio o femmina.

Negli ultimi due capitoli sono stati infine raccolte le conclusioni e i limiti dell'analisi.

I risultati che emergono da questi studi possono interessare: gli host, Airbnb e il settore alberghiero. In quanto, gli host possono capire meglio il mercato e individuare una migliore strategia di prezzi inoltre possono adattare i propri servizi in base alle richieste e alle aspettative dei clienti, ad esempio nelle inserzioni possono inserire fotografie o scrivere nei titoli o nelle descrizioni gli aspetti che interessano maggiormente ai consumatori. La stessa piattaforma, individuati i parametri chiave, può migliorare i filtri e le caratteristiche del sito affinché un utente trovi quello che cerca. Infine, visto il notevole sviluppo di Airbnb a discapito del settore alberghiero, questi ultimi possono individuare spunti per migliorare i propri servizi offerti, scegliendo quando competere e quando differenziarsi.

## DOMANDE DI RICERCA

Come descritto nel titolo della tesi “Sharing economy e Airbnb: analisi della domanda e delle differenze di genere tramite i feedback dei consumatori”, il lavoro consiste nel presentare le piattaforme di sharing economy soffermandosi su Airbnb, per poi analizzarne la domanda quindi le scelte dei consumatori e il prezzo dei soggiorni utilizzando i listing e i feedback relativi a Barcellona.

Le domande di ricerca possono così essere riassunte in:

- Esiste un bias positivo nei feedback rilasciati?
- Cosa trattano le recensioni rilasciate dagli utenti? Esse possono differire a seconda di determinate condizioni, come ad esempio il genere dei recensori, il sentimento che esprimono, se sono state rilasciate nel periodo prima o durante la pandemia di COVID-19, se sono rivolte a host professionisti o amatoriali?
- Esistono delle differenze comportamentali significative degli utenti a seconda del loro genere? Tra host professionisti e amatoriali?
- Esistono delle differenze significative negli annunci di alloggi privati e condivisi?
- Quali fattori influenzano il prezzo degli annunci? Essi variano a seconda della tipologia dell'alloggio?
- Gli ospiti preferiscono soggiornare da host dello stesso sesso?

# 1 SHARING ECONOMY

Nella stesura di questo capitolo si è utilizzato principalmente il paper “Navigating the Landscape of the Sharing and Gig Economy Literature: a systematic and interdisciplinary review” di Volker Stocker, Aaron Kolleck, Saba Rebecca Brause e Nadine Schawe del 2021 che raccoglie e descrive gli articoli più significativi esistenti sull’argomento. Ciò ha permesso di individuare alcuni concetti fondamentali che caratterizzano questa nuova economia, infatti dopo averne dato la definizione sono trattati i seguenti temi: inizialmente si descrive come il concetto di possesso sia sostituito da quello di accesso seguendo l’articolo “Owning, Using and Renting: Some Simple Economics of the “Sharing Economy”” di Filippas, Horton e Zeckhauser del 2019, poi si analizzano i fattori che ne hanno aiutato la diffusione come i benefici economici e sull’impatto ambientale, ma al contempo sono evidenziati i principali rischi che hanno dovuto affrontare e come la fiducia tra le parti e la creazione di norme ad hoc rivestano un ruolo fondamentale per diminuirli.

## 1.1 Definizione

Una definizione univoca di sharing economy non esiste, infatti numerosi studiosi si sono scontrati su questo tema. La condivisione non è un’invenzione recente; da sempre l’uomo ha messo in comune i propri prodotti con familiari e amici, ma questa nuova economia sfruttando le potenzialità e la diffusione della rete Internet, ha permesso di collegare persone che vivono molto distanti tra loro, estendendo così la condivisione anche agli estranei. Questa nuova modalità di commercio permette la riduzione delle spese di transazione come i costi di ricerca e contrattuali; i consumatori, infatti possono trovare più facilmente i beni e confrontarli autonomamente, mentre i piccoli venditori hanno a disposizione dei contratti standard che permettono di regolare le transazioni e numerosi sistemi di pagamento online.

Una delle definizioni di Sharing Economy ricorrenti è quella proposta da Frenken e al. (2015) “*I consumatori che si concedono a vicenda l'accesso temporaneo a beni fisici sottoutilizzati ("capacità inutilizzata"), possibilmente per denaro*”. Tali beni, infatti, possono essere utilizzati per qualche minuto al giorno, al mese o all’anno e diversamente rimanere inutilizzati, il proprietario può sfruttare questa capacità in eccesso offrendola ad alcuni consumatori interessati solitamente in cambio di un compenso.

Le principali caratteristiche della sharing economy sono quindi:

1. Interazione tra consumatori (c2c), ossia sono gli stessi consumatori ad offrire ad altri l'accesso ai beni, essi possono quindi essere al contempo venditori di alcuni beni e acquirenti di altri.
2. Accesso invece di proprietà: i consumatori per poter utilizzare un bene non devono necessariamente comprarlo e quindi possederlo, ma possono noleggiarlo per un periodo e quindi avere accesso al prodotto di qualcun altro. Si supera così il concetto del possesso con l'accesso e l'economia circolare.
3. Utilizzo migliore delle risorse sottoutilizzate

Altre definizioni sono quelle rilasciate dai principali dizionari, sono più generiche rispetto alla precedente, ma analoghe tra loro. L'enciclopedia Treccani scrive ad esempio: *“sharing economy: Economia collaborativa, che consiste nella condivisione delle risorse di spazio, tempo, beni e servizi, soprattutto tramite l'uso di piattaforme digitali.”*

Mentre il Cambridge Dictionary: *“an economic system that is based on people sharing possessions and services, either for free or for payment, usually using the internet to organize this”*.

Infine, l'Oxford Learners dictionaries: *“an economic system in which people can share possessions, services, etc., usually by means of the Internet”*.

Per descrivere in modo semplice i principi della sharing economy la ricercatrice Rachel Botsman durante una conferenza a Sidney domanda al pubblico di TED <sup>1</sup> se fosse servito un trapano o un buco nel muro. La risposta sebbene automatica, permette di vedere le cose da un'altra prospettiva, infatti le persone erano solite comprare un trapano e usarlo per quei secondi necessari per fare il foro e poi chiuderlo nella scatola per lungo tempo. Si è calcolato

---

<sup>1</sup> Ted è l'acronimo di Technology Entertainment Design, una serie di conferenze organizzate dalla no-profit Sapling Foundation, inizialmente i temi erano la tecnologia e il design, mentre negli anni ha raggiunto il mondo scientifico, culturale e accademico.

che in media tale oggetto viene usato per 13 minuti nella sua vita e praticamente ogni famiglia ne possiede uno, ma questo possesso è proprio necessario?

Numerosi studi sui cambiamenti climatici indicano la necessità di ridurre gli sprechi, gli acquisti e i consumi di nuovi prodotti, ricorrendo invece ad altre possibilità. Proseguendo con l'esempio del trapano, una persona può scegliere tra tre alternative invece di acquistarne uno nuovo: di prenderne uno di seconda mano, di noleggiarne uno o dato che il suo obiettivo è il buco nel muro chiedere direttamente a qualcuno che faccio il lavoro.

Per diminuire i rifiuti è possibile utilizzare l'economia circolare seguendo la regola delle 5 r: ridurre, riusare, riciclare, fare raccolta (differenziata) e recuperare.

Un rimedio a questi sprechi è proprio la condivisione, infatti tramite lo sharing economy, un consumatore quando non utilizza un bene, come ad esempio il trapano, può noleggiarlo ad un altro consumatore a cui serve per un limitato periodo di tempo senza che quest'ultimo ne debba acquistare uno nuovo, ma semplicemente avendone accesso.

Lo stesso discorso può essere applicato ai beni più costosi come gli appartamenti e le automobili, queste ultime dall'articolo "L'automobile? Inutilizzata per circa il 95% del suo tempo" del 23/11/2017 di La Repubblica emerge che restano inutilizzate in media per il 95% del loro tempo, mentre se fossero trasformate in un mezzo di car sharing sarebbero attive per il 40%; in questi casi i benefici economici che ne possono derivare incoraggiano notevolmente a tale economia.

In generale, la sharing economy (economia di condivisione) avviene grazie a una piattaforma web che permette di interfacciare persone che offrono beni/servizi con coloro che li stanno cercando anche se questi non si conoscono e abitano a diversi chilometri di distanza; la transazione economica solitamente è gestita dal sito che generalmente ne trattiene una percentuale. Tale sistema incentiva forme di consumo più consapevoli basate sul riuso invece che sull'acquisto e sull'accesso al posto del possesso, inoltre permette a chiunque di diventare un microimprenditore.

Ad oggi il termine sharing economy è molto ampio e racchiude al suo interno servizi che secondo alcuni studiosi non dovrebbero rientrare. Se si considera la definizione sull'utilizzo della capacità inutilizzata, ad esempio il servizio offerto da Uber è differente da quello

offerto da BlaBlaCar, infatti nel primo caso il viaggio viene effettuato appositamente per il passeggero (*on-demand economy*), mentre nel secondo esso sfrutta un posto sull'auto che il conducente non avrebbe utilizzato nel tragitto. Sul sito stesso di BlaBlaCar è spiegata la differenza dei tre termini car Sharing, ride sharing e car pooling che in genere sono confusi tra loro, ma che in ottica legale hanno diverse implicazioni:

- Car sharing: una persona può noleggiare per un periodo limitato una vettura che è di proprietà di terzi che in Italia, a differenza di altri paesi, non possono essere privati ma solo aziende (Car2Go)
- Ride hailing: corrisponde all'esempio fatto in precedenza di Uber (on demand), ossia un privato proprietario di un'automobile trasporta i passeggeri con o senza finalità di lucro.
- Car pooling: è ad esempio BlaBlaCar in cui il conducente con il proprio mezzo mentre effettua un viaggio utilizza i posti liberi trasportando altre persone. Non sono presenti fini di lucro, ma si dividono semplicemente le spese del viaggio.

Anche nel settore alberghiero è utile sottolineare il concetto di capacità inutilizzata in quanto sarebbe opportuno differenziare le case che sono utilizzate dai proprietari da quelle che sono acquistate solo per essere affittate, infatti nel primo caso quando non sono utilizzate e sono affittate a terzi è possibile parlare di sharing economy, mentre nel secondo la gestione sarebbe analoga agli hotel o B&B.

Un'altra caratteristica della sharing economy è l'utilizzo temporaneo del bene; infatti, piattaforme come Ebay<sup>2</sup> dovrebbero essere classificate come *economia dell'usato*, in quanto effettuando la transazione i consumatori diventano i nuovi proprietari del prodotto

---

<sup>2</sup> Per questa affermazione si considera la piattaforma Ebay iniziale, quando è nata infatti era principalmente un sito di aste online in cui un venditore poteva caricare un oggetto e un consumatore poteva acquistarlo, in questa transazione c'era anche il passaggio di proprietà, negli anni invece ha ampliato il proprio business agli oggetti nuovi diventando uno dei maggiori siti di e-commerce.

senza limitarsi al solo accesso.

Di seguito sono presentati gli argomenti principali relativi alla sharing economy: iniziando con i cambiamenti dovuti al passaggio dalla mentalità del possesso a quella dell'accesso, successivamente si sottolineano i fattori che hanno portato alla diffusione di questa nuova economia.

## **1.2 Possesso o accesso**

L'articolo "Owning, Using and Renting: Some Simple Economics of the "Sharing Economy"" di Filippas, Horton e Zeckhauser (2019) analizza come i consumatori scelgono tra queste due alternative e mostra come il mercato P2P abbia modificato il mercato tradizionale in quanto:

- Ha permesso un aumento dei consumi
- Lo sviluppo del noleggio ha incentivato le aziende produttrici a modificare il loro business, invece di cercare di aumentare la domanda riducendo la vita utile del prodotto hanno scelto di renderlo durevole e condivisibile, ma al contempo stanno provando ad entrare nel mercato dei servizi complementari. Ad esempio, le case automobilistiche e coloro che costruiscono gli appartamenti hanno iniziato ad investire nei servizi che rendono le modalità di condivisione più agevoli, come lo sblocco automatico senza la necessità di fare incontrare gli utenti per scambiare le chiavi.
- Prima di scegliere di comprare un oggetto coloro che ne sono interessati possono noleggiarlo e facendo "esperienza diretta" possono effettuare la scelta d'acquisto con un numero maggiore di informazioni disponibili.
- Gli utenti effettuano con più facilità la scelta di acquistare un prodotto anche costoso in quanto sono consapevoli di poter recuperare in parte il capitale investito nei momenti di non utilizzo personale, ad esempio per quanto riguarda beni come le stampanti 3D.

- Se le aziende scelgono di aumentare il prezzo in presenza di mercato P2P non implica automaticamente che la domanda diminuisca, ad esempio incrementando il prezzo di abbonamento allo stadio e dando la possibilità di subaffittare il posto in cambio di una fee aggiuntiva si innesca un mercato secondario parallelo, che al contempo aumenta il numero di abbonamenti venduti.

Dopo una prima parte più descrittiva del fenomeno l'articolo si concentra su analisi più empiriche su come i consumatori scelgono tra il possesso e l'accesso. Il primo modello rappresentato descrive come l'utilità del consumatore nell'utilizzare un bene sia data dalla differenza tra il beneficio che ottiene consumandolo e il costo opportunità della migliore alternativa.

*Equazione 1.1*

$$u(x; \alpha) = 2\alpha x - x^2$$

Dove  $x \in [0; 1]$  è il tempo di utilizzo, mentre  $\alpha \in (0; 1)$  è un coefficiente che varia da individuo a individuo. L'ottimo in questo modello si avrà con  $x^*(\alpha) = \alpha$  e di conseguenza  $u^*(\alpha) = \alpha^2$ .

Dato  $p$  il prezzo di acquisto del bene, se questo valore risulta minore di  $\alpha^2$ , ossia minore della propria utilità, il consumatore acquisterà il bene diventando proprietario.

Successivamente è stata inserita la possibilità di affittare l'oggetto a un canone di locazione  $r$ , ossia si considera il mercato P2P, l'utilità del proprietario diventerà

*Equazione 1.2*

$$u(x; \alpha) = 2\alpha x - x^2 + r(1 - x)$$

dove all'utilità del consumare il bene si aggiunge il guadagno derivante dell'affittarlo nel tempo restante. In questo caso sarà incentivato ad utilizzare il bene per un tempo minore per poter ricevere un guadagno dall'affitto. In ogni caso il proprietario otterrà un beneficio uguale o superiore rispetto al caso precedente  $x^*(\alpha) = \max \{ \alpha - \frac{r}{2}; 0 \}$  ottenendo un'utilità di

Equazione 1.3

$$u^*(r, \alpha) = \alpha^2 - \alpha r + \frac{r^2}{4} + r - p$$

Coloro che non sono i proprietari possono scegliere di affittare i prodotti

Equazione 1.4

$$u(x; \alpha) = 2\alpha x - x^2 - rx$$

Ottenendo  $x^*(\alpha) = \max \{\alpha - \frac{r}{2}; 0\}$  e un'utilità di  $u^*(r, \alpha) = \alpha^2 - \alpha r + \frac{r^2}{4}$ , ciò implica che consumeranno tutti coloro che hanno  $\alpha > \frac{r}{2}$

Dopo questa prima analisi semplificata l'articolo introduce alcuni costi che emergono dal nuovo mercato

- Bring to market (BTM): possono essere la manodopera, materiali consumabili complementari e ammortamenti dei beni dovuti all'uso (ad esempio nel caso di Uber sarebbero la manodopera, il combustibile e l'aumento dei km auto; nel caso di Airbnb sono invece il trovare e trattare con cliente, la pulizia e la distribuzione chiavi). Se questi costi sono molto alti rispetto prezzo acquisto bene, il noleggio non è supportato o aumenta notevolmente il canone di noleggio.
- Tempo ripristino bene: rappresenta l'intervallo di tempo in cui non può essere venduto (l'obiettivo è ridurlo)

Se il prezzo del bene è molto basso mentre sono presenti significativi costi BTM la gente tende all'acquisto dello stesso come, ad esempio, nel caso delle forbici, mentre un prezzo elevato ne incentiva il noleggio.

I costi BTM sono influenzati da due importanti fattori:

- La durata della singola sessione di noleggio ("chunkiness")
- La prevedibilità dell'utilizzo, che rappresenta quanto tempo in anticipo un consumatore sa di aver bisogno di un bene

Se questi parametri sono bassi implicano maggiori costi BTM, come ad esempio nel caso

della domanda dell'ultimo minuto in cui sono richiesti maggiori costi per trovare una corrispondenza.

Successivamente gli autori suddividono gli effetti nel breve e nel lungo periodo, la principale differenza è che nel primo caso i proprietari e gli affittuari non possono cambiare la loro posizione, mentre nel secondo sì. Nel lungo periodo si è notato come la scelta di possedere un bene solo per affittarlo non produce profitto se i BTM sono nulli, ma addirittura una perdita se tali costi sono positivi; infatti, più questi risultano elevati minore è la profittabilità dell'affitto e quindi i proprietari sono incentivati a utilizzare maggiormente il prodotto (come emerge dalle equazioni 1.5 e 1.6). L'utile che ricavano le società di acquisto tradizionale è dovuto a fattori di economie di scale più eventuali conoscenze che permettono di diminuire questi costi.

Alle precedenti formule si inserisce  $\gamma$ =tasso costo BTM, l'utilità dei proprietari diventa:

*Equazione 1.5*

$$u(r, \alpha) = 2\alpha - x^2 + (r - \gamma)(1 - x)$$

Da cui si ricava il tasso di utilizzo  $x_0$  :

*Equazione 1.6*

$$x_0(r, \alpha) = \operatorname{argmax}_{x \in [0,1]} 2\alpha - x^2 + (r - \gamma)(1 - x) = \max \left\{ 0, \alpha - \frac{r - \gamma}{2} \right\}$$

L'utilità è invece:

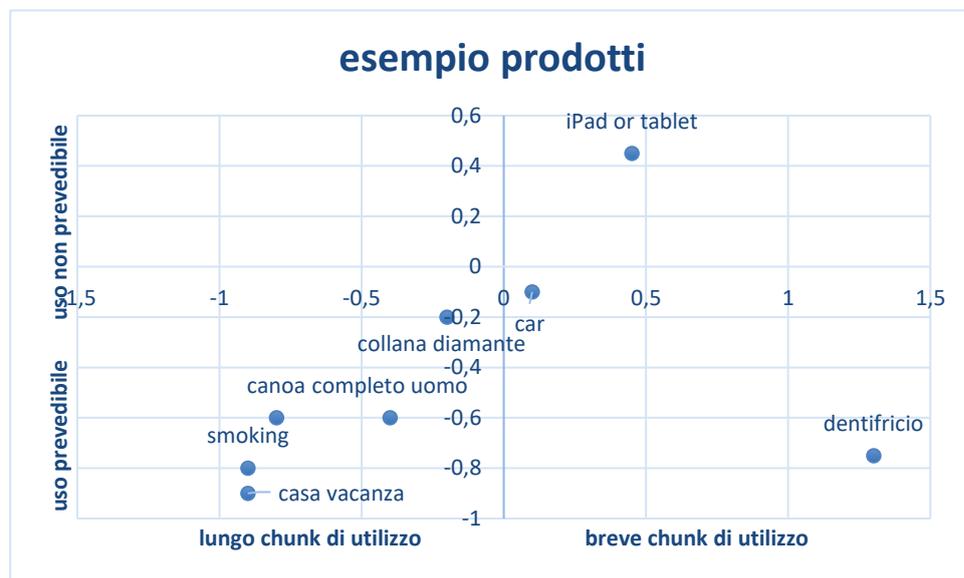
*Equazione 1.7*

$$u_0(r, \alpha) = \alpha^2 - p + (r - \gamma) \left( 1 - \alpha + \frac{r - \gamma}{4} \right)$$

In seguito, gli autori dimostrano tramite un sondaggio condotto su “Amazon Mechanical Turk” come la probabilità di acquistare un bene sia correlata all'utilizzo. Data una lista di beni si è richiesto agli utenti se ne sono in possesso e di quantificare con un valore tra 1 e 5 la prevedibilità e la frequenza di utilizzo. Con i risultati ottenuti, sono state svolte delle analisi di regressione per individuare le caratteristiche fondamentali, inoltre normalizzando

i dati si è rappresentato il grafico di dispersione 1.1, si nota che i beni nel quadrante in basso a sinistra sono quelli per cui un mercato degli affitti è già esistente.

Grafico 1.1 Predicibilità utilizzo verso chunkiness di alcuni beni



La prima analisi, rappresentata dall'equazione 1.8, dimostra che la proprietà è correlata positivamente all'utilizzo; infatti, un utilizzo doppio implica un aumento di 2.5 punti percentuali la probabilità che il bene sia posseduto.

Equazione 1.8

$$Own_{ig} = \beta_0 + \beta_1 \log x_{ig} + c_g + \varepsilon_g$$

Dove la variabile dipendente ( $Own_{ig}$ ) rappresenta la proprietà del bene  $g$  dell'intervistato  $i$ , mentre  $x_{ig}$  è il valore che emerge dal sondaggio che indica la frazione di tempo di utilizzo stimata e  $c_g$  è un effetto fisso del bene  $g$ .

Successivamente è stata aggiunta la variabile  $\log y_i$  che rappresenta il reddito familiare. Da ciò emerge che se quest'ultimo aumenta del 10%, produce un aumento della probabilità di possesso del 1%, infatti sia il reddito che l'utilizzo sono significativi, ma raccogliendo le ragioni chiave per la non proprietà di un bene emerge che solo pochi le attribuiscono al reddito, mentre la maggioranza indica il non utilizzo e solo una piccola percentuale la mancanza di spazio. Emergono però alcune eccezioni in questa suddivisione per

determinate categorie come le case vacanza e le cuffiette di alta fascia in cui la motivazione del reddito è la più quotata.

Lo studio successivamente cerca una correlazione tra la probabilità di possedere il bene e l'imprevedibilità di utilizzo (Equazione 1.9), trovando che questa è positiva e altamente significativa.

*Equazione 1.9*

$$Own_{ig} = \beta_0 + \beta_1 UnpredictabilityScore_{ig} + c_g + \varepsilon_i$$

Inserendo una variabile dipendente che rappresenti la “chunkiness score” i valori confermano il risultato precedente ed emerge che anche quest'ultimo parametro è correlato positivamente ed è altamente significativo. Ciò conferma quanto detto in precedenza relativo alla maggiore probabilità di possedere un bene se il suo utilizzo è imprevedibile e limitato a brevi intervalli.

I risultati che emergono da questo studio confermano che il possesso dei beni è correlato a quanto e a come sono utilizzati, alla capacità di prevederne l'utilizzo e infine al reddito degli intervistati.

### **1.3 Principali fattori che ne hanno favorito il successo**

Il successo della sharing economy è legato a numerosi aspetti, i principali sono analizzati nel seguente capitolo e sono: prezzo, possibilità che chiunque possa diventare microimprenditore, la diffusione di massa delle nuove tecnologie e la crescente attenzione delle persone al tema del cambiamento climatico.

#### *1.3.1 Il prezzo*

Queste piattaforme hanno permesso di connettere direttamente gli utenti interessati riducendo e in alcuni casi eliminando gli intermediari che comportavano un aumento del prezzo complessivo del prodotto/servizio, passando da una catena del valore verticale a una che oggi è possibile definire orizzontale. Questo fenomeno ha quindi reso l'esperienza

d'acquisto online sia meno costosa che più veloce.

### *1.3.2 Chiunque può diventare un microimprenditore*

Chiunque può essere il fornitore del bene o servizio semplicemente registrandosi sulla piattaforma e specificando l'oggetto in questione, consentendo così la nascita di alcuni microimprenditori. Questi ultimi come riportato da Filippas, Horton & Zeckhauser (2019), a differenza dei venditori tradizionali, in generale non dispongono di esperienza, di budget e di competenze nel marketing, di contratti, di assicurazioni adeguate, di tempi e strutture per riposizionare merci dopo l'uso; quindi, per colmare queste lacune si affidano alle piattaforme che provvedono a tutto.

### *1.3.3 La diffusione di massa delle nuove tecnologie*

Inoltre, grazie alla diffusione di massa di PC e smartphone, al calo di costi e alla crescente capacità di Internet tale economia sta raggiungendo un successo originariamente inimmaginato. Infatti, come emerge dalla presentazione "Digital in 2021" condotta da "We are social" e "Hootsuite", Internet collega 4,66 miliardi di persone ossia circa il 59,5% della popolazione mondiale. La pandemia ne ha accelerato la diffusione incrementando questo valore del 7,3% tra gennaio 2020 e gennaio 2021. La distribuzione però non è uniforme nei vari stati come risulta dallo studio del 2020 in cui emerge che in Europa l'84% della popolazione ha accesso a Internet mentre in Africa tale percentuale risulta solo del 34%. Delineare un andamento temporale del valore risulta complesso in quanto lo studio Digital 2021 ha variato la modalità di calcolo di alcuni parametri che quindi non permettono la comparazione con gli stessi studi precedenti. Per comprendere però lo sviluppo di internet si riportano i dati dell'analisi del 2020 (dove tale comparazione è possibile) in cui emerge che si è passati da 2,831 milioni di utenti connessi nel 2015 a 4,538 milioni nel 2020, tali numeri includono alcuni valori che sono stati trascurati per una maggiore correttezza nell'ultimo report. Inoltre, il numero di utenti che possiedono un telefono risultano 5,22 miliardi, ma solo 4,32 miliardi hanno uno smartphone connesso alla rete, che corrisponde al 92,6% degli utenti che utilizza Internet.

La sharing economy sfruttando il mondo digitale e lo sviluppo di queste nuove tecnologie collegate alla rete può raggiungere un pubblico potenzialmente infinito di consumatori con costi espositivi nulli (o comunque molto limitati). Questi fattori le permettono di soddisfare sia la domanda di beni di grande popolarità che quelli di nicchia; infatti, i bassi costi menzionati rispetto a quelli che si avrebbero nel mercato classico permettono di vendere anche quest'ultima tipologia di prodotti che vista la loro scarsità generalmente consentono di generare utili maggiori rispetto a ciò che si otterrebbe da quelli popolari.

#### *1.3.4 L'attenzione alle tematiche ambientali*

Ad una prima rapida analisi si afferma che tale economia permetta di diminuire le emissioni di carbone e i relativi impatti sul pianeta, questo sistema infatti utilizza in modo più efficiente le risorse, diminuisce l'inquinamento e gli sprechi. Vista l'importanza di tale tematica nella società attuale e l'interesse dei governi, sono stati finanziati numerosi studi i cui risultati sono trattati nel presente lavoro con un certo dettaglio.

Per valutare l'effettivo impatto ambientale si devono analizzare i vari aspetti positivi e negativi. I fattori favorevoli possono essere:

- La massimizzazione dell'utilizzo
- Il prolungamento della durata di vita del prodotto non in termini di usura, ma di migliore attenzione e mantenimento (manutenzione)
- La riduzione della domanda di oggetti nuovi

I contro invece:

- Più gente può accedere al prodotto/servizio, che sebbene sia positivo dal punto di vista sociale, dal lato ambiente ciò risulta problematico in quanto, ad esempio, nel caso dei mezzi di trasporto le persone che non possedevano una macchina dovevano scegliere altre alternative come i mezzi pubblici o la bicicletta, mentre ora possono utilizzare l'automobile per tutto il periodo in cui ne hanno bisogno. Un discorso analogo vale per gli alloggi, infatti avendo costi più bassi più gente può permettersi

di viaggiare e di allungare i periodi di ferie.

- Se un consumatore cede un vecchio prodotto con il ricavato può acquistarne uno nuovo anche a maggiore impatto ambientale, mentre se non avesse potuto fare lo scambio avrebbe probabilmente tenuto il prodotto ancora per qualche anno

Andando più nel dettaglio, l'analisi può variare a seconda dei settori che si osservano. La mobility sharing che racchiude al suo interno car sharing, carpooling, bike sharing, scooter sharing è sempre più sviluppata nelle grandi città. I servizi di carpooling come, ad esempio, BlaBlaCar permettono agli utenti di trovare altre persone che effettuano lo stesso viaggio e che quindi possono scegliere di muoversi con un'unica automobile privata, sfruttando i posti vuoti presenti, dividendo le spese e al contempo le emissioni di CO<sub>2</sub>. I servizi di car sharing e scootersharing (scooter e monopattini) invece stanno cercando di mettere a disposizione veicoli più nuovi rispetto a quelli privati e per lo più elettrici e quindi a basso impatto ambientale. Infine il bike sharing è l'alternativa sicuramente più salutare e sostenibile vista l'assenza di necessità di carburante, essa permette con l'acquisto di un abbonamento di utilizzare la bicicletta negli spostamenti in città senza doverne possedere necessariamente una e al contempo di eliminare il timore che venga rubata, infatti una persona può prenderne una di quelle dislocate nella città e parcheggiarla, a seconda delle regole del servizio, ovunque o negli appositi spazi una volta arrivato a destinazione. Queste soluzioni portano ad una riduzione della circolazione di mezzi propri, permettendo così di diminuire il traffico e le aree destinate ai parcheggi.

L'accomodation sharing invece include la condivisione di camere, alloggi e spazi per un breve periodo temporale. Questa permette utilizzando gli alloggi già esistenti di diminuire la necessità di costruire nuove enormi strutture alberghiere che talvolta danneggiano le aree naturali e storiche, inoltre i risultati dello studio di Cleantech "Environmental Impact of Home-Sharing" nel giugno 2014 mostrano come i servizi offerti dalle piattaforme di sharing economy abbiano permesso di ridurre del 78% il consumo energetico in Europa e del 63% negli Stati Uniti rispetto ai normali soggiorni alberghieri, oltre al risparmio delle risorse idriche; questo può essere dovuto al risparmio nei servizi di pulizia, lavanderia, cucina, ristoranti, spa, piscine, campi da golf... I risultati dello studio sono elencati nella tabella 1.1, mentre nella tabella 1.2 si riportano i dati relativi all'attenzione degli host alle tematiche ambientali.

Tabella 1.1 Risparmio risorsa a notte rispetto hotel

Risorsa	Nord America	Europa
Energia	63% (uso di 19000 case)	78% (uso di 68000 case)
GHG	61% (consumo 33000 macchine su strada)	89% (consumo 200000 macchine su strada)
Acqua	12% (270 piscine olimpioniche)	48% (1100 piscine olimpioniche)
Immondizia	32%	0-28%

Tabella 1.2 Comportamenti host sharing economy

Comportamenti sostenibili degli host	Nord America	Europa
Appartamento con almeno un apparecchio ad alta efficienza energetica	83%	79%
Host che danno la possibilità di riciclare agli ospiti	95%	89%
Host che danno prodotti monodose	<50%	<50%
Host che utilizzano prodotti per la pulizia green e sostenibili	82%	72%

Inoltre, il 10-15% degli ospiti in più rispetto a coloro che soggiornano negli hotel riferisce di preferire i mezzi di trasporto pubblici, camminare o muoversi con la bicicletta.

Al contempo però, come già evidenziato in precedenza, il minor costo permette a un maggior numero di persone di viaggiare, allungare il periodo del loro soggiorno e aumentarne la frequenza, tutti aspetti che sebbene dal punto di vista personale siano positivi, incidono negativamente sull'impatto ambientale. Nel complesso però lo studio sostiene che i benefici della sharing economy sono maggiori degli impatti negativi.

I benefici sull'impatto ambientale non sono limitati alla sola riduzione di produzione e dei volumi di scarti da smantellare, ma anche all'efficienza dell'utilizzo del suolo pubblico; infatti, l'accomodation sharing permetterebbe di diminuire il numero di alloggi vuoti dando la possibilità di ottenere migliori efficienze energetiche. Inoltre, questi servizi permettono di gestire al meglio i picchi di domanda e richiedere minori investimenti per eventi "straordinari". Ad esempio, come è riportato nell'articolo de La Repubblica "AirBnb

sponsor delle Olimpiadi, partnership speciale con il CIO” di Antonello Guerrera del 18/12/2019 e sul sito stesso di Airbnb, durante le Olimpiadi di Rio de Janeiro del 2016 l’azienda ha ottenuto il titolo di “Official alternative accomodation service supplier” ospitando numerosi visitatori e famigliari degli atleti ha evitato la costruzione di circa 257 nuovi hotel. Questo evento ha permesso ad Airbnb di diventare sponsor ufficiale delle Olimpiadi: tale collaborazione con il CIO (Comitato Olimpico Internazionale) e il CPI (Comitato Paralimpico Internazionale) ha portato a una riduzione dei costi organizzativi dei Giochi, limitando la costruzione di nuove strutture per ospitare tutte le persone coinvolte e generando al contempo un guadagno per la popolazione locale. Tutto ciò consente di evitare la realizzazione di numerosi palazzi che dopo l’evento solitamente vengono abbandonati rappresentando quindi uno spreco di risorse e un’inutile fonte di inquinamento a causa della CO<sub>2</sub> rilasciata durante la produzione e dell’occupazione del suolo stesso.

Un altro settore interessato alla sharing economy è quello del fashion che consuma una quantità molto elevata di acqua e di emissioni, questo nuovo modello permette anche in questo caso di avere un impatto ambientale positivo perché diminuisce i volumi di produzione. Di conseguenza quando un utente non indossa più un capo può affittarlo o venderlo a qualcuno interessato dando così una seconda vita al prodotto e allungandone la durata.

La condivisione può anche coinvolgere beni di consumo non durevoli come il cibo, si parla in questo caso di food sharing, questo ha un’importanza enorme in quanto permette di combattere lo spreco alimentare che ogni anno si stima intorno a un terzo di tutto il cibo prodotto per il consumo.

In generale, lo sviluppo di moderne economie come la sharing economy, la on-demand economy e quella di seconda mano permette agli utenti di non dover necessariamente acquistare un oggetto nuovo, contribuendo così alla riduzione della domanda di beni nuovi a favore di un’economia circolare, questi fattori permettono un migliore utilizzo del suolo pubblico e di ridurre la produzione e il volume di prodotti da smaltire, di conseguenza le emissioni dei gas serra (GHG) e i consumi energetici.

## **1.4 Cosa sancisce il successo di una società di sharing economy**

Il fattore che però sancisce il successo o meno della sharing economy è il raggiungimento di una massa critica, ossia una soglia oltre la quale il sistema è in grado di autosostenersi. Tale valore permette di ridurre il rischio di scarsità, sia lato domanda che lato offerta; infatti, il consumatore può scegliere quale prodotto soddisfi maggiormente le proprie aspettative e al contempo se la domanda aumenta il venditore ha maggiore probabilità di vendere il proprio prodotto/servizio. Ad esempio, nel caso del car sharing per incentivare una persona a utilizzare il servizio bisogna garantirgli un numero adeguato di auto disponibili in diversi punti della città, stesso discorso nel caso di Airbnb che dovrà assicurare la scelta tra più alloggi in varie aree e con molteplici tipologie di prezzo e dimensioni in modo che un utente riesca a trovare quella che rispecchi le sue esigenze. Infatti, le piattaforme devono incentivare le iscrizioni di nuovi utenti da entrambe le parti, perché maggiore è la presenza dei venditori più elevata è la probabilità che i potenziali acquirenti trovino ciò che cercano e al contempo un alto numero di clienti implica che più venditori riescono a trarne profitto. Il beneficio che ne emerge incentiva nuovi utenti ad accedere. Inoltre, la soddisfazione delle persone che operano sulla piattaforma è fondamentale, in quanto queste possono rilasciare commenti positivi e invitare amici ad iscriversi al servizio, in questo modo i siti possono beneficiare di una pubblicità gratuita che viene percepita dalla gente come più veritiera e verso la quale provano maggiore fiducia. Si innesca così un circolo che si autoalimenta, anche il sito stesso ne trarrà beneficio in quanto un numero maggiore di utenti implica più transazioni e quindi ricavi più elevati.

## **1.5 Problematiche riscontrate**

La sharing economy soffre di alcuni rischi che devono essere tenuti in considerazione e mitigati.

Nell'articolo "Optimal risk management for the sharing economy with stranger danger and service quality" del 2019 gli autori Hong Ji Hyun, Kim Byung Cho e Park Kyung Sam distinguono due tipologie di rischi:

- Di prestazione: sono quelli legati alla bassa qualità dei prodotti o dei servizi offerti,

si può manifestare come “rammarico” dell’acquirente dopo un acquisto in quanto il valore effettivo risulta minore di quello previsto, una possibile motivazione può essere legata alla mancanza di professionalità.

- Fisici: sono quelli relativi alla sicurezza della salute fisica e mentale delle persone coinvolte, queste solitamente sono dovute da un pericolo estraneo e generano un generico sentimento di insicurezza. Alcuni esempi possono essere le aggressioni subite dagli autisti di Uber o dai proprietari di casa Airbnb o dalla controparte.

Gli studiosi evidenziano come le recensioni che rilasciano gli utenti dopo aver utilizzato il prodotto/servizio sono essenziali per ridurre il rischio di prestazione perché guida la scelta dei consumatori in fase di acquisto, mentre per i rischi legati alla sicurezza è richiesto l’intervento dello stato nella creazione di norme ad hoc. In seguito, l’articolo costruisce alcuni modelli per individuare alcune operazioni da compiere per diminuire quest’ultimo rischio e aumentare i profitti.

Se le piattaforme investono nella sicurezza degli utenti emerge che questi ultimi sono disposti a pagare anche un prezzo maggiore, ciò permette quindi di ottenere ricavi più elevati.

L’attenzione degli stessi governi sul tema è alta in quanto l’elevata diffusione della sharing economy richiede un intervento per la protezione dei consumatori. Le possibili soluzioni individuate sono:

- Un intervento fisso diretto in cui il governo cerca di ridurre il rischio fisico e il sostegno che dà è fisso e indipendente dalla piattaforma di condivisione.
- Un intervento a variabile diretta o indiretta: l’importo della sovvenzione del governo è direttamente proporzionale all’investimento effettuato dalla piattaforma per la sicurezza, nel primo caso permette alla piattaforma di aumentare la propria utilità di autoselezione, mentre nel secondo no.

Oltre ai rischi presentati dall’articolo precedente è opportuno aggiungere le seguenti problematiche:

- Sfruttamento dei lavoratori da parte delle società di sharing economy

- Mancanza di chiare norme sulla sharing economy quali ad esempio tassazione, licenze... che fanno sì che gli operatori presenti sul mercato tradizionale non riescano ad essere competitivi. Questi ultimi, vista la mole di affari, chiedono una rivalutazione del quadro normativo a riguardo per tutelare i propri interessi.
- Discriminazione razziale e sessuale: tale fenomeno dipende dalle informazioni che la singola piattaforma decide di condividere tra gli utenti, salvaguardando il sentimento di fiducia che sta alla base della sharing economy. Alcune società indicano per il singolo profilo nome, fotografie e recensioni avute, ma, com'è emerso da alcuni studi come "Trust and Reputation in the Sharing Economy: The Role of Personal Photos on Airbnb" di Ert, Fleischer e Magen del 2016, ciò può provocare comportamenti discriminatori, come sarà descritto in seguito in relazione alla politica di Airbnb. In altre piattaforme questa problematica non emerge in quanto si limitano a rendere disponibili i profili in modo anonimo, cioè pubblicando le sole recensioni ad esso associate.
- Responsabilità in caso di danni, per tale motivo le piattaforme hanno prestato maggiore attenzione e hanno fatto ricorso a numerose assicurazioni per limitarne gli effetti.

Questi ultimi problemi hanno catturato l'attenzione dei legislatori come sarà descritto nello specifico nel capitolo "norme".

## **1.6 Possibili soluzioni presenti e future dei problemi**

### *1.6.1 Creazione sentimento di fiducia*

Nella sharing economy la fiducia è la chiave per il successo della piattaforma, essa serve per superare il timore che l'oggetto condiviso possa essere danneggiato o rubato, che un utente possa comportarsi scorrettamente o che la piattaforma o i prodotti-servizi non siano sicuri. La mancata fiducia, infatti, rappresenta una delle maggiori barriere per lo sviluppo di questo modello di business.

Nella sharing economy tale sentimento si sviluppa tra le diverse parti coinvolte:

- tra i membri della community tramite la condivisione delle esperienze e delle opinioni
- tra chi condivide e il consumatore che è fondamentale affinché lo scambio avvenga
- verso la piattaforma ossia in termini assicurativi e di sicurezza dei dati e dei pagamenti

Quest'ultimo livello è quello di maggiore importanza perché permette a una persona di sentirsi sicura ad utilizzarlo e che quindi diventi un consumatore.

I concetti di reputazione e fiducia sono molto simili, ma differiscono in quanto il primo rappresenta la valutazione collettiva effettuata dal pubblico di una entità o di una persona, tali informazioni sono rilasciate dagli utenti; mentre il sentimento della fiducia è soggettivo e implica che il fiduciario si comporterà nel modo promesso. L'utilizzo delle recensioni online facilita l'innescò di questo sentimento, la reputazione è solo uno degli elementi per costruirlo, ma non è necessaria.

La fiducia permette di creare meccanismi di passaparola positivi ossia pubblicità gratuita, incrementare la frequenza d'acquisto e aumentare la disposizione dell'utente a comprare prodotti sempre più costosi.

Questo sentimento è fondamentale nella sharing economy perché l'individuo agisce in condizioni di incertezza, grazie alle condivisioni delle esperienze di acquisto degli utenti tale asimmetria informativa può diminuire. Ad esempio, nel mondo online le vendite sono cambiate, infatti inizialmente l'impresa catturava l'attenzione del consumatore attraverso la pubblicità e le informazioni che rilasciava, ora invece hanno maggiore potere i racconti di esperienza diretta di altri consumatori condivise sugli appositi blog a discapito delle pubblicità creata dall'azienda stessa che essendo direttamente interessata risulta meno reale.

### *EWOM*

In inglese il passaparola “faccia a faccia” che porta le persone a descrivere spontaneamente ad altri, senza fini commerciali, i prodotti/servizi che hanno acquistato prende il nome di Word of Mouth (WOM). Con la nascita del mondo digitale, questo concetto si è evoluto in eWOM (Electronic Word-of-Mouth), rispetto a quello tradizionale permette di sfruttare la

potenza di internet e di raggiungere quindi un pubblico notevolmente più ampio. Come è sottolineato nell'articolo del 2008 "Electronic word-of-mouth in hospitality and tourism management" di Litvin, Goldsmith, e Pan le informazioni che emergono dalle recensioni influenzano notevolmente la scelta d'acquisto di un consumatore, in modo particolare nel settore turistico e dell'ospitalità, perché i beni offerti sono intangibili e non possono essere valutati direttamente prima del loro utilizzo, inoltre i consumatori hanno un alto rischio di acquisto dovuto alle asimmetrie informative tra le parti. Le eWOM possono riguardare i prodotti stessi, le aziende o i marchi.

Il confronto tra WOM e eWOM riassunto da Huete-Alcoer in "A Literature Review of Word of Mouth and Electronic Word of Mouth: Implications for Consumer Behavior." (2017) ne fa emergere le principali differenze:

- **Credibilità:** l'anonimato delle eWOM può diminuire la credibilità dell'informazione rilasciate
- **Privacy:** le WOM sono trasmesse in forma privata attraverso conversazioni private, mentre quelle digitali sono pubbliche ed essendo scritte chiunque può vederle in qualsiasi momento e non possono essere distorte come avviene nel passaparola tradizionale.
- **Velocità di diffusione:** la rapidità che permette alle eWOM di raggiungere un numero potenzialmente infinito di consumatori è uno dei motivi per cui numerosi accademici stanno studiando questo fenomeno. Quando un consumatore vuole acquistare qualche prodotto può cercare recensioni direttamente online senza la necessità di dover trovare/attendere un amico o familiare che abbia acquistato direttamente il bene o che abbia ricevuto da un terzo informazioni relative.
- **Accessibilità:** come già indicato precedentemente, l'accessibilità del WOM era molto limitata rispetto a quella dell'eWOM.

Negli anni l'uomo ha acquisito consapevolezza sull'importanza del proprio tempo, infatti quando, ad esempio, effettua una ricerca sul web esige di trovare ciò che sta cercando nel minor tempo possibile. Le eWOM devono sottostare a questa sua esigenza, ciò fa emergere le caratteristiche essenziali che influenzano positivamente l'utilità delle informazioni e sono

rappresentate da uno studio condotto da Ronnie Cheung descritte nell'articolo del 2014 "The influence of electronic word-of-mouth on information adoption in online customer communities" e sono:

- Affidabilità: indica se l'informazione è percepita come credibile e affidabile dai consumatori
- Tempestività e completezza: la tempestività indica che le informazioni sono attuali e aggiornate
- Qualità: richiede che le informazioni fornite siano oggettive

Dallo studio emerge che l'utilità delle informazioni che un consumatore può trovare online influenzano positivamente l'intenzione di acquistare un prodotto.

Analizzando più nel dettaglio l'argomento centrale della tesi, le piattaforme che permettono di prenotare i viaggi direttamente da Internet negli ultimi anni stanno raggiungendo numeri e successo impressionanti, prendendo il posto delle vecchie agenzie e dei relativi travel agent. Il vecchio modo di viaggiare implicava che un turista si affidasse direttamente a un'agenzia che godeva di elevata fiducia ed essa pensava a tutto, l'offerta era limitata ai pacchetti che aveva in catalogo. Al contrario, la nascita di queste piattaforme ha permesso di ampliare notevolmente l'offerta disponibile e di eliminare i costi di agenzia dell'intermediario, ma il consumatore aveva bisogno di sentirsi sicuro, per tale motivo l'eWOM ha avuto un ruolo chiave nello sviluppo. I principali mezzi con cui esso si manifesta sono i siti web, i blog e l'e-mail.

I commenti rilasciati infatti hanno un'alta influenza per la propensione all'acquisto, ciò implica che i venditori devono prestare molta attenzione a come vengono recensiti i propri prodotti, controllando anche i contenuti rilasciati dai profili anonimi o fake, in quanto una recensione negativa ha un peso notevolmente superiore rispetto a una positiva sul lettore (Litvin, Goldsmith, & Pan, 2008). Per adattarsi alle esigenze del mercato molte aziende hanno inserito sul sito una sezione apposita per le recensioni, così da poterle monitorare più facilmente ed instaurare un dialogo diretto con i consumatori; questo permette ai venditori di analizzare la reazione del pubblico ai propri prodotti e di valutarne eventuali migliorie, cosa che diversamente non poteva fare con il passaparola tradizionale.

In passato i venditori costruivano la loro reputazione attraverso le relazioni umane, oggi invece le vendite online necessitano delle review e del sistema di rating che in generale sono nominate feedback, questi infatti influenzano le scelte dei consumatori e il sentimento di fiducia che essi maturano nei confronti dell'azienda.

Come sarà presentato in seguito nel capitolo della letteratura esistente e dimostrato nel lavoro d'analisi, nei siti di e-commerce e sharing economy le recensioni positive sono nettamente superiori rispetto a quelle negative, le principali motivazioni possono essere: un utente è influenzato dalle review precedenti, l'azienda manipola le recensioni o un utente soddisfatto ha maggiore propensione a rilasciare un commento. Affinché la community valuti affidabili le recensioni, è necessario che esse non risultino tutte "troppo positive", diversamente potrebbe pensare che siano scritte dall'azienda stessa o comunque dietro compenso.

In conclusione, i principali motivi per cui chi è interessato all'acquisto di un prodotto confronta le diverse eWOM presenti è per risparmiare tempo nel processo decisionale, per fare scelte d'acquisto migliori, per diminuire il rischio e per capire com'è realmente un prodotto (Hennig-Thurau & Walsh, 2003). Inoltre, le recensioni essendo rilasciate spontaneamente dagli utenti beneficiano di una maggiore credibilità rispetto alle pubblicità tradizionali. Capita l'influenza dell'eWOM sulle decisioni d'acquisto dei consumatori, i venditori devono riuscire a controllarla senza esserne passivamente controllati (Litvin, Goldsmith, & Pan, 2008).

### *1.6.2 Definire norme di regolamentazione*

L'articolo "Navigating the Landscape of the Sharing and Gig Economy Literature: a systematic and interdisciplinary review" di Volker Stocker, Aaron Kolleck, Saba Rebecca Brause e Nadine Schawe (2021) raccoglie numerosi papers relativi alla sharing economy, nel seguente paragrafo si esaminano quelli relativi alla regolamentazione.

In questo ambito l'argomento sharing economy è molto dibattuto; gli stessi regolatori, infatti, hanno difficoltà a scegliere come agire in quanto l'inserimento di norme potrebbe comprometterne il carattere innovativo di questi servizi, ma al contempo c'è la necessità

fondamentale di tutelare tutti gli attori interessati: i consumatori, i lavoratori che svolgono attività sulla piattaforma e gli operatori storici.

In “Does Sharing Mean Caring? Regulating Innovation in the Sharing Economy.” Ranchordàs (2015) sottolinea la difficoltà di inquadrare l’attività del venditore che oscilla tra la sfera personale e commerciale, inoltre evidenzia come questa economia di condivisione necessiti di poche regole appositamente studiate che non soffochino l’innovazione, ad esempio potrebbero essere richiesti requisiti minimi di competenze, norme sulla responsabilità e contratti obbligatori tra le parti. Diversamente si rischierebbe di eliminarli dal mercato come nel caso di Uber che in alcune città è stato vietato.

Nell’articolo “The Sharing Economy and Consumer Protection Regulation: The Case for Policy Change.” Koopman, Mitchell e Thierer (2015) condividono la precedente conclusione indicando come le leggi possano essere delle barriere all’innovazione, all’ingresso e all’imprenditorialità. Inoltre, sottolineano la necessità di tutelare i consumatori nei confronti dei fattori di rischio come le esternalità, l’aumento dei prezzi, l’asimmetria informativa, la differenza del potere contrattuale e la concorrenza inadeguata. Soffermandosi sul benessere del consumatore finale mostrano come l’ascesa dei mercati digitali abbia permesso di ampliare la disponibilità di servizi, prodotti e informazioni e di dar voce ai loro pensieri attraverso i feedback. Le piattaforme raccogliendo questi dati si autoregolano migliorando i servizi offerti adattandoli alle esigenze degli utenti finali; questo comportamento implica una minore necessità di una ferrea regolamentazione. Al contrario, il successo delle aziende storiche dipende principalmente della capacità di compiacere i regolatori stessi, ciò comporta un basso interesse ad assecondare i desideri dei consumatori. Questi ultimi, infatti, invece di investire le risorse per aumentare il valore al cliente o per innovarsi preferiscono utilizzarli per ottenere privilegi esclusivi. Un esempio è come le società tipo Uber si sono adattate immediatamente ai pagamenti con carte di credito a differenza dei taxi che si sono allineati solo quando è emersa la norma.

In generale, come indicato da Miller nell’articolo “First Principles for regulating the sharing economy” del 2016 (Miller, 2016, p. 147-148), i principi su cui si basa la regolamentazione della sharing economy possono essere i seguenti:

- “è differenziata e richiede una risposta normativa differenziata”

- “deve essere fatta chiarezza (illuminata)”
- “richiede il giusto tipo di informazioni”
- “è qui per restare (ed è una buona cosa)”
- “sconvolge e reinventa mercati consolidati”
- “crea nuovi mercati”
- “sconvolge e reinventa strutture normative consolidate”
- “richiede una risposta oltre al regolamento tradizionale”
- “il danno e il relativo rimedio sono una sfida unica da determinare”
- “coinvolge diverse parti, ognuna delle quali dovrebbe essere interessata da apposite norme”

Lo stesso autore indica come i venditori già presenti sul mercato rispondono a questi nuovi servizi: cercano di mantenere la loro quota di mercato e al contempo di trovare modi per entrarci.

Gli stessi stati, visto l'aumento del numero di lavoratori coinvolti nei nuovi servizi, si sono accorti di dover prestare attenzione che non sia sfruttata e degradata la loro dignità umana. È emerso, infatti, che molte volte i rischi ricadono direttamente sugli stessi lavoratori invece che sulle imprese (Aloisi, 2016). Queste persone devono essere tutelate, da qui emerge la necessità di sviluppare nuovi strumenti per garantire i diritti fondamentali del lavoro; infatti, si stanno creando fondi per avere ad esempio una copertura sanitaria, un'assicurazione per gli infortuni, le ferie e tutti i diritti relativi ai dipendenti ordinari.

Infine, gli studiosi Edelman e Geradin nell'articolo “Efficiencies and Regulatory Shortcuts: How Should We Regulate Companies like Airbnb and Uber?” del 2015 affermano che l'intervento normativo è richiesto quando gli utenti o i fornitori dei servizi non sono in grado di valutare adeguatamente i rischi e quindi possono non essere in grado di prendere le dovute precauzioni, ad esempio la richiesta di ispezione del veicolo con una certa frequenza o nel caso di un alloggio di dispositivi di sicurezza per incendi o di protezione

verso eventuali intrusi esterni. I consumatori, infatti, rischiano di sottovalutare questi parametri salvavita rinunciandoci in cambio di piccoli risparmi. In aggiunta anche questi autori sottolineano che i regolatori dovrebbero incoraggiare le efficienze offerte da queste piattaforme adattando i requisiti e le regole esistenti che diversamente avvantaggerebbero gli operatori storici, prestando attenzione che queste piattaforme non risultino al di sopra della legge. Anche in questo caso la richiesta di una licenza potrebbe essere un modo per imporre gli standard di qualità minimi sul mercato, ma questo spinge i detentori di licenza ad escludere i nuovi entranti per evitare la concorrenza e rischiare che il valore delle licenze diminuisca.

In conclusione, nei diversi articoli emerge la necessità di regolamentare il mercato della sharing economy per evitare che prevalga sulla legge stessa e per tutelare tutti gli attori coinvolti, ma devono essere create leggi ad hoc affinché non sia compromesso il suo carattere innovativo e autoregolatore.

## 2 AIRBNB

### 2.1 Storia

Airbnb è un portale di annunci che, tramite computer o cellulare, permette di mettere in contatto persone che vogliono affittare per un breve periodo delle camere o degli alloggi con altre che lo cercano. Gli host possono affittare stanze private, appartamenti interi, ville, castelli, baite, barche... e qualsiasi altra tipologia di alloggio.

L'idea nacque nel 2007 a San Francisco da Joe e Brian che notarono che l'importante conferenza di Design industriale locale ("Industrial Design Society of America") stava attirando molti visitatori, gli hotel della città erano quasi al completo e la gente aveva difficoltà a trovare posti per dormire. I due ragazzi comprarono allora alcuni materassi ad aria e crearono il sito "air Bed and Breakfast", che mostrava l'opportunità ai turisti di dormire e fare colazione nell'appartamento, trasformarono così il loro loft in un bed & breakfast e ospitarono i primi tre ospiti.

In seguito, coinvolsero Nathan Blecharczyk un ingegnere informatico laureato ad Harvard e loro ex compagno di stanza, che divenne il terzo co-fondatore.

Nel 2008 dopo alcuni cambiamenti la piattaforma venne rilanciata una prima volta a marzo durante il festival SXSW ma ricevette solo 2 prenotazioni e una seconda volta ad agosto in occasione della "Democratic National Convention" in Denver, dove ottenne 80 prenotazioni, ma il successo non ebbe vita lunga.

A causa dei problemi finanziari sopraggiunti, i fondatori per finanziare il proprio progetto iniziarono a vendere cereali in confezioni limitate a tema elettorale riuscendo a raccogliere 30.000\$.

Il 2009 fu l'anno di svolta, vennero seguiti dall'incubatore Y Combinator grazie al quale ottennero un primo finanziamento di 20.000\$. Capendo l'importanza estetica della presentazione dell'inserzione e notando che le fotografie caricate erano di scarsa qualità, iniziarono a visitare le case degli host a New York per conoscerli e raccogliere le loro esigenze, scattare fotografie professionali e scrivere recensioni, infine cambiarono il nome

in “Airbnb”. Nello stesso anno Sequoia finanziò la compagnia per 58.500\$ permettendole una forte espansione.

Nel tempo la piattaforma ha migliorato e incrementato i servizi offerti, coinvolto sempre più investitori, e raggiunto numerosi paesi anche oltre oceano; ma la società dovette anche affrontare numerose battaglie legali e problemi causati da ospiti indisciplinati e dal settore alberghiero che sentì minata la propria autorità.

Nel dicembre 2020 la società si è quotata in borsa.

Oggi, secondo quanto è riportato direttamente sul sito in data 30 settembre 2020, conta 5,6 milioni di annunci attivi in tutto il mondo, 100.000 città e oltre 220 paesi; mentre gli utenti coinvolti sono più di 4 milioni gli host e oltre 800 milioni i check-in di ospiti.

Il viaggiatore che prenota su Airbnb lo fa sia per il prezzo più competitivo rispetto a quello offerto dagli hotel, sia per l’esperienza di viaggio unica che gli viene offerta; infatti, riesce a immergersi più facilmente nella reale vita locale anche grazie al contatto diretto con l’host. Al contrario gli hotel, in modo particolare le catene, ormai si sono standardizzati e offrono gli stessi servizi indipendentemente dal paese in cui si trovino, che sia Roma, New York o Pechino. La filosofia di Airbnb è quella di far sentire i viaggiatori a casa ovunque si trovino vivendo il luogo non solo come semplici turisti (“belong anywhere”).

## **2.2 Growth hacking**

La growth hacking è una metodologia in cui, per aumentare il proprio traffico e trasformare i visitatori in utenti, le startup del web sviluppano tecniche e strategie di marketing e hacking. Ad esempio, per poter aver successo Airbnb doveva riuscire a raggiungere un pubblico molto ampio di utenti, per fare ciò sfruttò una falla nel sistema del suo competitor Craigslist. Quest’ultimo era il più importante portale di annunci degli Stati Uniti e contava già un elevato numero di utenti. L’host in fase di compilazione del form per creare l’annuncio da inserire sul portale di Airbnb poteva scegliere di caricarlo anche su Craigslist. Quando un utente navigando sul sito di Craigslist selezionava uno di questi annunci era direzionato sulla pagina di Airbnb che a livello estetico era molto più accattivante e bella; infatti, essendo i fondatori due designer sapevano quanto fossero importanti questi fattori.

I turisti nonostante l'errore di indirizzamento colpiti dalla semplicità e dalla bellezza della piattaforma continuavano a navigare sul nuovo sito.

Quando Craigslist si è accorta del baco e lo ha corretto, Airbnb era già riuscita ad attirare numerosi nuovi clienti.

### **2.3 Business model**

Airbnb mette a disposizione una piattaforma su cui interagiscono due tipologie di utenti: gli host e gli ospiti. L'obiettivo è quello di migliorare e incentivare le interazioni in modo che si crei valore per sé stessi e per gli altri. L'aumento del numero degli utenti appartenenti a un gruppo crea vantaggi per l'altra parte, ossia se aumentano gli host che offrono un appartamento gli utenti avranno maggiore possibilità di scelta, mentre se cresce il numero di ospiti ci sarà una maggiore domanda e quindi più case saranno prenotate.

Per incentivare l'utilizzo della piattaforma e quindi per aumentare il proprio valore Airbnb ha delle pagine molto semplici e intuitive, inoltre per ridurre le barriere d'ingresso permette ai nuovi host di caricare il proprio annuncio gratuitamente, mentre la difficoltà che devono fronteggiare è il fatto che non hanno recensioni e quindi risulta complicato generare un sentimento di fiducia verso i potenziali turisti.

In cambio del servizio offerto la piattaforma si prende una commissione di intermediazione, ossia quando si verifica una transazione ne trattiene una percentuale del subtotale della prenotazione composta dal prezzo medio per notte, dalle spese di pulizia e dai costi per ospiti aggiuntivi escluse le tasse. Dal dicembre del 2020 ha modificato la propria ricompensa per renderla più semplice, ossia Airbnb si trattiene il 15% della transazione a carico dell'host, mentre prima i costi del servizio erano condivisi tra l'host e l'ospite, rispettivamente il 3% per il primo e fino a 14,2% per il secondo. Questo nuovo meccanismo di costi permette una più semplice strategia di prezzo per gli host e rendono gli annunci più competitivi perché al turista non risultano spese ulteriori da sostenere.

## 2.4 Problematiche riscontrate

Le principali problematiche che emergono sulla piattaforma sono legate alla discriminazione (principalmente di tipo razziale), alla gentrificazione e alla sicurezza degli utenti.

### 2.4.1 *Discriminazione*

Sono numerosi gli studi che dimostrano come le piattaforme di sharing economy soffrano di questo fenomeno. Lo stesso Airbnb è colpito da questa problematica sia lato host che lato ospite.

Il comportamento dei primi è analizzato nel paper del 2017 “Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment” di Edelman, Luca e Svirsky, lo studio condotto negli Stati Uniti aveva infatti l’obiettivo di quantificare l’eventuale discriminazione razziale al momento della richiesta di locazione nella piattaforma di Airbnb.

L’esperimento si è sviluppato su 6400 annunci in 5 città (Baltimora, Dallas, Los Angeles, St. Louis e Washington) con 20 potenziali ospiti identici che presentavano nomi indicanti chiaramente la loro origine bianca o afroamericana (rispettivamente 5 maschili e 5 femminili), non sono state invece inserite le fotografie personali per evitare che subentrassero altri parametri nella scelta. È emerso che nel 16% dei casi l’host è meno propenso a concludere l’accordo con un potenziale ospite che rivela nel nome origini afroamericane, tale risultato riguarda sia che i padroni di casa siano piccoli o grandi proprietari, siano uomini o donne, siano bianchi o afroamericani, mentre si riduce significativamente solo se l’host ha già avuto in passato ospiti afroamericani.

Negli ultimi decenni le società si sono impegnate a combattere questo comportamento nei mercati tradizionali introducendo anche apposite leggi, ma non hanno ancora prestato particolare attenzione allo sviluppo di questo fenomeno nel mondo online lasciando le piattaforme quindi libere di operare.

Tale discriminazione non emerge nelle situazioni di mercato online in cui le transazioni

avvengono in modo più anonimo senza essere condizionate da elementi di tipo razziale, per ovviare al problema anche la società di sharing economy potrebbe non passare l'informazione del nome e la foto degli utenti, utilizzando uno pseudonimo come avviene con eBay, oppure incentivare gli host a utilizzare la “prenotazione istantanea”.

Dall'altro lato, il comportamento discriminatorio degli ospiti ha una ripercussione significativa anche sul prezzo come emerge dagli articoli “Digital Discrimination: The Case of Airbnb.com” di Edelman e Luca del 2014 e “The impact of host race and gender on prices on Airbnb” di Marchenko del 2019.

Nel primo articolo emerge, come già riportato in precedenza, che sulle piattaforme online un acquirente ha a disposizione oltre alle informazioni dei prodotti anche quelle sul venditore che in alcuni casi ha il profilo collegato alle proprie pagine social, questo serve per aumentare la fiducia nei suoi confronti ma fa emergere un grave problema legato ai comportamenti discriminatori di vario tipo, nel paper in particolare si analizza quello di tipo razziale collegato al colore della pelle. I dati dello studio sono del 2012 relativi alla città di New York, inizialmente gli autori hanno chiesto a degli operatori di classificare gli host in neri e non neri in base alle foto del profilo, hanno poi calcolato il prezzo medio dei listing per le due categorie ottenendo 107€ per quelli di colore e 144 € per gli altri. Successivamente hanno valutato l'effetto stimato della razza utilizzando il metodo della regressione: una prima analisi è stata fatta considerando solamente la tipologia dell'appartamento (se intero o una stanza) e il numero di stanze da letto, mentre nella seconda sono stati inseriti anche gli altri parametri disponibili in fase di acquisto dei consumatori, ma tale differenza è stata persistente sebbene il divario si sia praticamente dimezzato passando da 31 a 16. Lo studio ha confermato la presenza discriminatoria trovando che in media a parità dei valori disponibili un host nero guadagna 12% in meno.

Anche nel secondo articolo è stato fatto un controllo visivo umano per classificare il genere, la razza (bianchi, asiatici o di colore) e l'età di 70000 immagini di host di 7 centri urbani americani nel 2015 e 2016. Gli autori hanno trovato che in media gli host maschi asiatici e di colore guadagnano rispettivamente il 4,5 e il 3% in meno rispetto a quelli di bianchi su proprietà analoghe, inoltre hanno visto che nonostante il prezzo più economico la domanda è minore, questi aspetti sottolineano il problema della discriminazione. In generale infine hanno stimato una perdita annuale di 350\$ per le host di colore e di 300\$ per i maschi neri

rispetto agli host bianchi. La stessa piattaforma Airbnb incentivata dai media ha introdotto una nuova politica antidiscriminatoria inserendo la possibilità della prenotazione immediata e pubblicando le foto profilo in un formato minore.

#### 2.4.2 *Gentrificazione*

L'articolo "Regulating Airbnb: how cities deal with perceived negative externalities of short-term rentals" del 2018 di Shirley Nieuwland & Rianne van Melik nella parte empirica analizza e raccoglie gli studi esistenti su questa problematica (tra i principali ricercatori sul tema emerge Còcola Gant con due articoli "Tourism and commercial gentrification" e "Holiday rentals: The new gentrification battlefront"). I turisti odierni sono diversi da quelli del passato, infatti non si limitano più a visitare e soggiornare nelle classiche mete turistiche, ma cercano esperienze anche in altre zone. Questo ha permesso a numerose nuove aree di diventare centri turisti, ma nonostante il profitto che ne deriva dagli affitti a breve termine (STR) sono emerse numerose esternalità negative. I residenti locali hanno infatti riportato numerose problematiche legate al rumore e agli schiamazzi provocati dai turisti di giorno e di notte, alla sicurezza, alla disponibilità dei parcheggi e al traffico. Inoltre, in numerosi quartieri residenziali i proprietari hanno iniziato ad affittare gli appartamenti come case vacanza per brevi periodi e visti i maggiori ricavi ottenuti li hanno tolti dal mercato per i residenti per affittarli ai turisti. Questo ha obbligato molti residenti visto l'aumento del prezzo ad abbandonare l'appartamento e a trasferirsi in altre zone oppure ad affittare una parte della casa per poter dividere le spese, ma così facendo si è innescato un circolo vizioso (Holm 2016) in cui aumentano i numeri di turisti e di conseguenza aumentano i prezzi degli affitti. Tutto questo sta portando a una perdita culturale di alcuni quartieri che hanno trasformato molte attività commerciali locali in attrazioni turistiche come bar, catene di ristorazione e negozi di souvenir.

In seguito, è presentato come gli Stati stiano agendo in modo diverso per affrontare il fenomeno:

- tramite restrizioni: che possono essere classificate di tipo quantitativo (è imposto ad esempio un numero massimo di strutture STR, di giorni per l'affitto, di numero di visitatori o la quantità di soggiorni), qualitativo (solo determinate tipologie di

alloggio), di localizzazione (solo in alcune aree) o di densità (un massimo numero in un quartiere).

- con divieti: i governi scelgono di vietare le STR, ma così facendo rischiano di perdere numerose entrate economiche e di incentivare un mercato secondario.
- con il “Laissez-faire”: non sono create delle norme per contrastare il fenomeno, ma in alcuni casi si sono fatti degli accordi diretti con le società di affitti per tassare le transazioni.

In particolare, gli autori Ayoub, Breuille, Grivault e Le Gallo (2019) hanno condotto uno studio sull’impatto della sharing economy sugli affitti tradizionali in 8 città francesi utilizzando i dati Airdna tra il 2015 e il 2016 e hanno ottenuto che solo su 3 città questo effetto risulta significativo, in particolare nella capitale. Per tale motivo hanno sottolineato l’importanza di un intervento normativo differente a seconda della zona considerata.

Inoltre, in “Do short-term rental platforms affect housing markets? Evidence from Airbnb in Barcelona” gli autori Garcia-López, Jofre-Monseny, Martínez-Mazza e Segú (2020) studiano l’impatto dell’affitto a breve termine delle case a Barcellona, si è scelto di presentare l’articolo in quanto la parte finale dell’elaborato svolgerà delle analisi sugli annunci e sulle recensioni di questa città.

Anche in quest studio si evidenziano le paure delle amministrazioni locali sulla gestione di queste attività, preoccupate delle conseguenze negative che potrebbero emergere, come ad esempio la piattaforma potrebbe compromettere l’accessibilità degli alloggi negli affitti di lungo periodo ai residenti a causa della maggior profittabilità di quelli a breve termine per i turisti.

Lo studio analizza l’impatto di Airbnb sul mercato immobiliare della città svolgendo numerose regressioni considerando diverse tempistiche e quartieri, ciò che è emerso è stato che in un quartiere con attività media di Airbnb in città, gli affitti sono aumentati dell’1,9%, mentre i prezzi delle transazioni del 4,6% e i prezzi pubblicati del 3,7%; tale incremento è risultato molto più elevato nei quartieri più turistici portando ad un aumento del 7% degli affitti e del 17% e del 14% rispettivamente delle transazioni e dei pezzi pubblicati. Il mercato degli affitti in questi anni ha aumentato notevolmente il proprio prezzo, anche se

tale variazione è però attribuibile solo in parte alle attività di Airbnb.

### 2.4.3 *Sicurezza ospiti/host*

Sebbene negli ultimi anni si stiano cercando dei nuovi metodi per le attività di login e logout per rendere più semplici l'affitto ai padroni di casa, la maggior parte dei soggiorni richiede un incontro diretto tra host e ospite almeno per la consegna fisica delle chiavi dell'appartamento. Questi incontri faccia a faccia tra persone sconosciute sono rischiosi per la sicurezza, in particolare quando l'esperienza del soggiorno prevede una stanza condivisa in cui le parti diventano quindi coinquilini. Le recensioni hanno in questo caso un ruolo fondamentale per incentivare le persone a fidarsi tra di loro.

Il tema della sicurezza è molto delicato, esso compare infatti in molte recensioni, in particolare in quelle rilasciate da ospiti donne. Inoltre, il rischio percepito è più elevato nel caso in cui le persone viaggino da sole, come è presentato dallo studio “Airbnb’s Reputation System and Gender Differences Among Guests: Evidence from Large-Scale Data Analysis and a Controlled Experiment” di Choi e Horvát del 2019. Nell’articolo gli autori analizzano circa 150000 commenti di Airbnb relativi alle città statunitensi di New York, Los Angeles e Chicago tra il 2009 e il 2018 suddividendoli in maschili e femminili (con il database GenderChecker), hanno trovato che la maggioranza degli ospiti è maschile, mentre quella degli host è femminile, inoltre il numero di commenti disponibili per gli appartamenti con padrone di casa donna è inferiore rispetto a quelli della controparte, i risultati indicano la presenza del sentimento di “omofilia” che indurrebbe gli ospiti a scegliere host dello stesso sesso. Successivamente analizzando i contenuti dei commenti individuano quelli che contengono temi relativi alla sicurezza ed emerge come il 6,6% delle donne e il 5,3% degli uomini ne abbia parlato, tale differenza risulta statisticamente significativa. Nella seconda parte dell’articolo è condotto un esperimento controllato in cui sono stati creati degli host fittizi con storico e profilo chiaramente maschile o femminile. Inizialmente si è analizzata la probabilità di prenotazione in caso di presenza di un’unica recensione negativa ottenendo 1,3% per le ospiti femminili e 6,2% per quelli maschili, tale differenza risulta statisticamente significativa, diversamente la presenza di un’unica recensione positiva aumenta notevolmente la probabilità e annulla questa differenza legato al sesso, per questo

motivo per i nuovi host è necessario ottenere la prima recensione positiva. Anche l'esperimento conferma la presenza di comportamenti omofili in fase di scelta di prenotazione, ma solo per le femmine questo valore è statisticamente significativo. Infine, è stato richiesto agli utenti che hanno fatto il test cosa scelgono tra hotel e Airbnb in caso dovessero viaggiare da soli, il 70% ha risposto hotel e l'80% ha dato come motivazione la sicurezza, in particolare i risultati indicano che l'80% delle donne scelgono l'hotel di cui il 73,3% per la sicurezza, mentre per i maschi tali percentuali sono rispettivamente del 65,9% e 69,6%. Gli autori con i risultati ottenuti dalle analisi presentate confermano la maggiore avversione al rischio delle donne e la necessità quindi di Airbnb di trovare soluzioni per contrastarla.

Nella parte sperimentale "La scelta dell'host è influenzata dal genere?" del seguente elaborato si andrà ad effettuare uno studio analogo e si confronteranno i risultati ottenuti.

## **2.5 Costruzione del sentimento di fiducia di un utente**

Come emerge dai capitoli precedenti la fiducia gioca un ruolo fondamentale nella sharing economy, anche la piattaforma Airbnb risente di questo fattore, per tale motivo negli anni ha sperimentato numerose strategie per ottenerla, le più significative sono:

- Obbligo di registrarsi sulla piattaforma fornendo i dati personali richiesti sia per l'host prima di poter caricare un annuncio, sia per l'ospite per poter usufruire di un appartamento/servizio
- Sistema di feedback
- Titolo di superhost

Tutti questi metodi permettono agli utenti di crearsi una propria reputazione all'interno della community.

Lo studio condotto da Abrate e Viglia in "Personal or product reputation? Optimizing revenues in the sharing economy" del 2017 dimostra l'importanza della reputazione del venditore per ottimizzare i ricavi, affermando che le stesse piattaforme possono aumentarne

l'efficacia favorendo la trasparenza attraverso il rilascio di un maggior numero di informazioni personali del venditore e di attributi qualitativi del prodotto. L'articolo dimostra però come abbia maggiore importanza la reputazione personale del venditore rispetto a quella del prodotto nel diminuire il gap informativo e il divario tra i ricavi potenziali e quelli reali. Infatti, il titolo di superhost e la sua data di inizio collaborazione sul sito, permettendo di migliorare il sentimento di fiducia nei confronti del padrone di casa, influenzando positivamente i ricavi, mentre il numero di recensioni e l'inserimento di fotografie professionali, sebbene siano significative nello sviluppo della reputazione del prodotto, hanno una bassa efficienza in termini economici.

### *2.5.1 Creazione account*

Per tutelare la sicurezza della community per caricare un annuncio e per affittare un appartamento è necessario avere un account: nella fase di registrazione sono richieste alcune informazioni fondamentali per l'identificazione. Per creare un account (sia come host che come ospite) è necessario avere almeno un indirizzo e-mail, un account Facebook o Google, un ID Apple o un numero di telefono. Successivamente, per effettuare una prenotazione è necessario caricare l'indirizzo e-mail, il nome e il cognome, il numero di telefono verificato, una foto di un documento di identità, l'accettazione delle regole dell'appartamento, informazioni di pagamento (i dati sensibili non saranno accessibili agli host) e, a seconda degli host, un messaggio e una foto del profilo, inoltre, nella procedura di verifica, può essere richiesto di scattarsi un selfie.

Quando un utente ha creato un proprio account può prenotare o caricare un annuncio in maniera molto semplice; infatti, è sufficiente rispondere alle domande mostrate dalla piattaforma.

L'host ha accesso al nome e alla foto caricati sul profilo dell'ospite, al nome presente sul documento d'identità, se ha più di 25 anni e se il documento è stato caricato in modo corretto.

Questa fase è di fondamentale importanza perché la fotografia e il nome compaiono sia nel listing in fase di ricerca dell'ospite, sia quando un utente chiede all'host di poter prenotare.

Come emerge dall'articolo "Trust and reputation in the sharing economy: the role of personal photo on Airbnb" di Ert, Fleischer, & Magen del 2016 la fotografia dell'host influenza la sua affidabilità e quindi, dimostrata la correlazione positiva, il prezzo dell'annuncio e la probabilità che sia scelto. Gli autori indicano inoltre come la decisione d'acquisto degli ospiti avvenga considerando sia gli attributi dei prodotti che quelli degli host, questi ultimi comprendono sia informazioni non visive come la reputazione, sia visive come le fotografie personali che generano sentimenti di fiducia e interesse. Questo implica che la scelta d'acquisto non sia influenzata dalla bellezza o meno dell'host, ma dall'affidabilità che suscita nell'interlocutore. Infine, sottolineano come la reputazione sia un elemento chiave per la costruzione della fiducia, senza però essere una condizione necessaria, talvolta infatti gli host si affidano a terzi senza avere informazioni sulla loro reputazione.

### 2.5.2 *Sistema di feedback*

L'ospite al termine del proprio soggiorno ha 14 giorni di tempo per valutare in modo quantitativo con l'utilizzo delle stelle il proprio soggiorno e rilasciare una recensione testuale. I temi su cui è invitato ad esprimere un voto tra 0 (molto negativa) e 5 (molto positiva) sono: l'esperienza complessiva, la pulizia, la comunicazione, il check-in, la precisione, la posizione e il rapporto qualità-prezzo.

Quando almeno tre ospiti hanno rilasciato queste valutazioni, il risultato compare nell'annuncio: in primo piano, sotto il titolo è indicata la valutazione media di tutte le categorie, mentre in fondo alla pagina compaiono i risultati per ogni singolo parametro. Le figure 2.1 e 2.2 mostrano un esempio di come è visualizzato un listing attivo da un telefono.

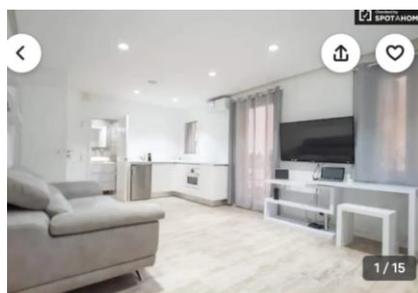
Nello stesso intervallo di tempo anche l'host può scrivere una propria recensione sull'ospite.

Per evitare che una parte sia influenzata da quello che scrive l'altra, le recensioni sono pubblicate solo allo scadere dei 14 giorni o dopo che entrambe hanno rilasciato il loro giudizio.

Le recensioni possono avere al massimo 1000 parole e le migliori sono quelle che

comprendono informazioni utili per la scelta dei futuri ospiti e per eventuali azioni correttive degli host. Gli utenti possono anche rilasciare un feedback privato direttamente all'altra parte.

Figura 2.1 Esempio annuncio1 (screenshot della pagina)



### New central loft 50 mt from the BEACH

★ 4.80 (11 recensioni) · Superhost

Barcelona, Catalunya, Spagna

Intero alloggio: unità in affitto  
affittato da Gabriele



3 ospiti · 1 camera da letto · 1 letto · 1 bagno

Figura 2.2 Esempio annuncio2(screenshot della pagina)

★ 4.80 · 11 recensioni

Cerca recensioni

Pulizia	4.5
Precisione	5.0
Comunicazione	5.0
Posizione	5.0
Check-in	5.0
Qualità/prezzo	4.8

Traduci le recensioni in Italiano (IT)



Fiona  
giugno 2021

I loved this apartment so much, I spent 6 weeks here and everything was perfect. I thought the apartment was a great design, loads of storage space, bright, comfortable and clean. It has everything you need, and I loved the balcony overlooking the street. Great location, super close to the beach, friendly neighbourhood. Gabriele was a great host, 10/10.

I commenti sono utili quando forniscono informazioni imparziali, pertinenti, veritieri e senza pregiudizi. Gli utenti stessi non possono rilasciare giudizi direttamente sulla loro stessa attività e al contempo non è possibile minacciare una recensione negativa o incentivarne una positiva, non è accettata la falsa prenotazione o utilizzare un secondo account per avere recensioni positive.

La piattaforma controlla che le informazioni contenute nelle recensioni non danneggino la community ma ne accrescano il contenuto informativo; infatti, se viene segnalata una violazione delle linee guida, gli operatori di Airbnb dopo un'attenta analisi, possono scegliere di rimuoverla, nel caso in cui questo comportamento scorretto venga ripetuto l'account può essere sospeso o disattivato in modo permanente.

Anche in caso di cancellazione della prenotazione è possibile lasciare un commento dopo le 00:00 del giorno del check-in, questa deve essere pertinente e rilevante per la community, diversamente viene rimossa.

### 2.5.3 Superhost

Il badge di superhost è rilasciato agli host in modo automatico se nella data di valutazione trimestrale sono superati alcuni requisiti; esso, appare sull'annuncio vicino al punteggio complessivo e sulla foto profilo con il simbolo seguente:  (come mostrato nelle figure 2.1 e 2.2). I superhost sono gli host più apprezzati, offrono esperienze straordinarie ai propri ospiti e quindi rappresentano un modello da seguire.

I prerequisiti sono:

- Valutazione complessiva superiore o uguale a 4,8 rispetto ai punteggi rilasciati dagli ospiti nell'anno precedente
- Tasso risposta entro le 24 ore maggiore del 90%
- Aver avuto oltre 10 soggiorni o 100 notti prenotate tra almeno 3 persone diverse nell'anno precedente.
- Tasso di cancellazione minore dell'1% con eccezione di apposite circostanze.

Ogni 3 mesi si valuta se l'host continua a svolgere un ottimo servizio rispecchiando gli standard richiesti, diversamente gli viene tolto il titolo. Per conservare il badge è richiesto un elevato investimento in termini di energia e attenzione non solo nella pulizia e nel buono stato della casa, ma anche e soprattutto nei dettagli nei listing e nella tempestività della risposta alle richieste dei clienti.

Il badge permette di ottenere una maggiore visibilità, aumentando quindi le prenotazioni e i relativi guadagni, ed inoltre di accedere ad esclusive ricompense. I turisti, infatti, possono filtrare i risultati degli annunci dei superhost verso i quali provano maggiore fiducia. Il titolo inoltre offre un coupon da 100 USD ogni anno e un ulteriore bonus del 20% se si invita un nuovo host a registrarsi.

Airbnb riconoscendo gli effetti che ha avuto la pandemia del Covid-19 nel settore turistico e a livello societario ha allentato i prerequisiti per questi mesi lasciando inalterato il titolo di superhost a coloro che lo possedevano prima di febbraio 2020 e che soddisfino anche

solo i criteri della valutazione complessiva e del tasso di risposta.

Uno studio condotto da Sai Liang, Markus Schuckertb, Rob Law e Chih-Chien Chenc riportato nell'articolo "Be a "Superhost": The importance of badge systems for peer-to-peer rental accommodations" del 2017 dimostra che lo sforzo richiesto per il badge è pienamente ricompensato in quanto un annuncio di un superhost ha maggiore probabilità di ricevere recensioni, le valutazioni sono più alte, i clienti sono disposti a pagare un prezzo più alto per questi alloggi e ha un effetto pubblicitario che porta gli ospiti a prestare maggiore attenzione aumentando così la possibilità di ottenere maggiori prenotazioni. Tale badge è una "gamificazione" della figura degli host, mentre l'utilizzo di un'analogia strategia per i clienti non produrrebbe nessun risultato utile.

Anche un'analisi di Ulrich Gunter riportata nel paper "What makes an Airbnb host a superhost? Empirical evidence from San Francisco and the Bay Area." del 2018 riporta che tra i benefici indiretti del badge c'è l'aumento delle prenotazioni e delle entrate, inoltre rivela che gli host che hanno più annunci hanno una probabilità maggiore di mantenere un profilo di superhost in quanto solitamente sono persone che ricavano da questa attività degli elevati ricavi e quindi la svolgono in maniera più professionale.

### 3 LETTERATURA ESISTENTE

Nel seguente capitolo si presentano alcuni articoli che trattano il tema delle recensioni e del prezzo con obiettivi analoghi a quelli che saranno trattati nella parte di analisi così da poterne poi confrontare i risultati ottenuti.

#### 3.1 Recensioni

Come detto in precedenza un acquirente dopo l'acquisto può lasciare un proprio feedback in merito al prodotto/servizio, questo può essere espresso con un punteggio o con un commento testuale, vista la difficoltà di esprimere la qualità percepita con un valore numerico le recensioni testuali stanno acquistando sempre più valore perché riescono ad esprimere maggiori contenuti, per questo si stanno sviluppando numerosi studi sul text mining e la Sentiment Analysis che permettono di codificare le informazioni rilasciate come testi. La piattaforma Airbnb richiede agli utenti di rilasciare un feedback numerico per valutare alcuni parametri e uno testuale per descrivere la propria esperienza. Di seguito sono presentati prima due articoli relativi ai punteggi e in seguito alcuni che riguardano anche i contenuti dei commenti testuali.

##### *3.1.1 Overcoming the J-shaped distribution of product reviews*

Nell'articolo "Overcoming the J-shaped distribution of product reviews" di Nan Hu, Paul A. Pavlou e Jie Zhang del 2009 gli autori analizzano i valori delle recensioni fatte tra febbraio e luglio 2005 di alcuni prodotti presenti sul sito di Amazon appartenenti alle categorie libri, dvd e video e trovano che rispettivamente il 78%, il 73% e il 72% di esse ha un punteggio maggiore o uguale a 4 stelle, ciò conferma le loro aspettative che i prodotti fossero valutati eccezionali. Rappresentando graficamente i risultati ottenuti trovano, oltre al picco positivo, che una piccola parte di utenti attribuisce un punteggio minimo pari a 1, mentre sono pochi i risultati ottenuti per i valori intermedi; ciò produce una forma a J che significa che la distribuzione è binomiale e asimmetrica e quindi non è statisticamente significativo parlare di valore medio.

Successivamente è stato svolto un esperimento controllato per individuare le cause della forma di questa distribuzione, preso un campione di individui si è chiesto di valutare alcuni prodotti analoghi ai precedenti; il risultato che si è ottenuto è stato una distribuzione unimodale con media significativamente più bassa, ciò a conferma che generalmente le persone hanno dei gusti più moderati. Le cause che sono state individuate sono:

- Purchasing bias: gli acquirenti che hanno dato un basso punteggio nell'esperimento controllato sono meno incentivati a comprare il prodotto e quindi non scriveranno il loro giudizio negativo sulla piattaforma. Questo permette di spiegare il basso numero di recensioni non positive.
- Under-reporting bias: sono più propensi a rilasciare valutazioni coloro che sono estremamente soddisfatti o insoddisfatti dei prodotti, questo spiega la mancanza di punteggi intermedi che ci si sarebbe aspettati per avere una distribuzione unimodale centrata su valori moderati.

Quanto trovato porta gli autori ad affermare che per poter prevedere meglio la domanda futura il valore della media semplice non è significativo, mentre sarebbero necessarie altre informazioni come la deviazione standard, le due modalità di distribuzione binomiale e il prezzo del prodotto, che permettono di ottenere una rappresentazione più reale della distribuzione.

### *3.1.2 A first look at online reputation on Airbnb, where every stay is above average*

Gli studiosi Zervas, Proserpio e Byers in “A first look at online reputation on Airbnb, where every stay is above average” del 2015 hanno analizzato le recensioni di Airbnb confrontandole con quelle presenti sulla piattaforma di TripAdvisor. Ciò che è emerso è stato che nella prima piattaforma le valutazioni tra 4,5 e 5 stelle coprono il 95% delle proprietà considerate, mentre un numero molto ridotto ha punteggi inferiori a 3,5. Del secondo sito sono stati studiati 700.000 listing di hotel, b&B e case vacanza, ottenendo un punteggio rispettivamente di 3,8 e 4,1 entrambi più bassi e con maggiore varianza rispetto alla controparte, mentre le case hanno un punteggio più elevato ma solo l'85% è

nell'intervallo 4.5-5, ciò può essere spiegato da fattori sociologici che fanno sì che gli host valutino in modo differente le strutture alberghiere rispetto agli alloggi. L'esperimento si è svolto su due periodi (2015 e 2018) e i risultati emersi sono stati coerenti.

Considerando la letteratura esistente sono state presentate alcuni possibili motivazioni per giustificare la distorsione positiva delle recensioni:

- effetto gregge, cioè è difficile che un utente scriva la propria recensione se non è in linea con quelle esistenti
- i recensori in molti casi preferiscono non esprimere il loro giudizio se non sono rimasti completamente soddisfatti del loro acquisto.
- manipolazione strategica delle recensioni

Gli autori giustificano parte di questa distorsione positiva confermando quanto dimostrato da Fradkin e al (2015) e Proserpio et al (2017) che nelle transazioni di Airbnb l'aspetto sociale è importante nel modo in cui si valuta l'altra parte in quanto il carattere personale riveste un ruolo importante dal momento che quantificano la loro esperienza in casa di altri. Inoltre, è sottolineata la necessità delle piattaforme di creare fiducia tra gli utenti, in quanto è un metodo fondamentale per contrastare l'asimmetria informativa presente tra gli host e ospiti, in generale nei mercati peer-to-peer, in cui si sa poco uno dell'altro; in particolare dal lato host le recensioni sono uno dei pochi modi per crearsi una buona reputazione, perché a differenza delle maggiori aziende già presenti sul mercato, essi non hanno budget da dedicare per farsi pubblicità.

Lo studio ha successivamente cercato di trovare gli alloggi presenti su entrambe le piattaforme per poter effettuare un confronto incrociato diretto, nonostante la complessità nell'individuare la corrispondenza dal momento che negli annunci di Airbnb non è presente l'indirizzo preciso, è emerso che su Airbnb gli annunci valutati con un punteggio maggiore o uguale a 4,5 sono 14% in più rispetto a quelli presenti su TripAdvisor. Questa discrepanza può essere spiegata sulla base dei lavori preesistenti con la logica dei sistemi di revisione bilaterale che incentivano le parti a scambiarsi feedback positivi per paura di ritorsioni, fattore che nel 2014 ha portato la piattaforma a modificare i propri sistemi pubblicando le recensioni solo dopo che entrambe le parti le avevano scritte o allo scadere dell'intervallo

di 14 giorni disponibile.

### 3.1.3 *If nearly all Airbnb reviews are positive, does that make them meaningless?*

Lo studio “If nearly all Airbnb reviews are positive, does that make them meaningless?” di Bridges e Vásquez del 2016 individua alcuni fattori che favoriscono la positività delle recensioni:

- Aspettative degli ospiti sono inferiori rispetto ai soggiorni negli hotel
- Più difficile e imbarazzante rilasciare feedback negativi dopo l’incontro faccia a faccia con l’altra parte
- Reciprocità del sistema di recensione, in passato c’era il problema che il primo che scriveva la recensione temeva la ritorsione dell’altra parte che dopo aver letto il feedback poteva scrivere la propria in risposta.
- Non sono anonime, anche perché questa condizione implicherebbe una minore credibilità, ma ciò fa sì che alcuni utenti non scrivano recensioni negative perché i loro profili sono collegati a quelli presenti su altre piattaforme e quindi temono che “la propria identità online sia legata a commenti negativi”.
- Non si sa se in effetti Airbnb pubblichi tutte le recensioni e quante siano quelle che vengono rifiutate in quanto non conformi alle politiche aziendali, inoltre non è noto se effettivamente Airbnb rimuova gli alloggi con un elevato numero di recensioni negative.

L’analisi è stata svolta su 400 recensioni tra agosto e settembre 2015, 100 per quattro città (50 rilasciate dagli host e 50 dagli ospiti). Si è trovato che la lunghezza media delle recensioni degli ospiti è di 75 parole, mentre quelle degli host è circa di 28.

Successivamente tramite un apposito software di concordato AntConc si sono trovati i valori di TTR relativo alle recensioni degli ospiti e degli host, questo numero indica il grado di variazione delle parole e va tra 0% (una sola parola che viene ripetuta più volte) a 100%

(nessuna parola è mai stata ripetuta); nella lingua parlata tale valore è circa 45-55% mentre in quella scritta 65-75%. Nell'esperimento si è rilevato un TTR pari a 14,2% per gli ospiti e 16,9% per gli host; ciò significa che tali recensioni utilizzano un vocabolario ristretto di termini. Il risultato ottenuto può essere condizionato al fatto che la stessa piattaforma Airbnb suggerisce un modello di recensione.

L'analisi continua con la classificazione manuale delle recensioni in positive (93%) e negative. Con il software AntConc si sono rilevate le parole più utilizzate: host, pulizia e confort. In particolare, l'host compare nel 89,5% dei commenti e nel 79,5% con il nome proprio. Le recensioni negative riscontrate sono state poche, ma anche in questo caso le lamentele interessano la mancanza di confort, di comunicazione e di pulizia e non l'altra parte, inoltre spesso sono racchiuse in commenti che descrivono anche sentimenti positivi.

Gli autori hanno poi prestato attenzione alle analisi "tiepide" che sebbene a una prima lettura possono sembrare positive ad una più attenta analisi è possibile far emergere che il soggiorno non è stato soddisfacente, infatti manca il linguaggio più entusiastico ma compaiono parole più neutre come "bene" e "ok". Per avere conferma di quanto ipotizzato si sono analizzate le recensioni fatte dagli stessi recensori per altre esperienze ed è emerso che negli altri casi utilizzano un linguaggio differente e chiaramente positivo.

In conclusione, la risposta alla domanda posta nel titolo dell'articolo "il fatto che le recensioni sono quasi tutte positive le rende prive di significato?" è no, in quanto emerge che sebbene sembrino tutte molto simili in superficie, permettono di aumentare la quantità di informazioni utili per altri utenti, diminuendo l'incertezza degli stessi.

#### 3.1.4 *Large-Scale Sentiment Analysis on Airbnb Reviews from 15 Cities*

Nell'articolo "Large-Scale Sentiment Analysis on Airbnb Reviews from 15 Cities" di Abdulkareem Alsudais e Timm Teubner sono analizzate 2.686.354 recensioni (corrispondente a 12.353.382 frasi) di Airbnb relative a 15 principali città statunitensi, dopo una fase di preelaborazione sono state eliminate 179.671 recensioni problematiche corrispondenti a righe vuote, lingua non inglese e quelle generate automaticamente. Attraverso il Vader Sentiment che è un metodo che rileva 4 misure:

- negatività del testo
- positività del testo
- neutralità del testo
- aggregazione e normalizzazione delle precedenti misure

Gli autori utilizzano quest'ultimo valore per classificare le recensioni utilizzando i valori soglia indicati dai creatori del modello:

- positive: se il numero è maggiore o uguale a 0,05
- negativo: se è minore o uguale a -0,05
- neutro: se è compreso tra -0,05 e 0,05

I risultati che si ottengono, allineati agli studi sull'argomento, mostrano come ben il 98,1% delle recensioni sono positive (76,4% delle frasi), mentre solo l'1,6% sono negative (4,7% delle frasi). La differenza che emerge tra le recensioni e le frasi è principalmente dovuta al fatto che nel complesso le recensioni possono essere classificate positive nonostante la presenza di alcune frasi negative.

Per valutare le performance del classificatore hanno fatto classificare un campione di recensioni ad un giudice umano, successivamente hanno confrontato questi risultati con quelli ottenuti automaticamente dal modello, ciò che è emerso è stato che quelle negative hanno una maggiore probabilità d'errore. Analogamente si è fatto con le frasi, i risultati ottenuti sono stati analoghi, ma in questo caso la precisione del classificatore è più elevata anche in caso di frasi negative, ciò è in linea con quanto affermato in precedenza dell'insieme di più frasi per fare una recensione.

### 3.1.5 *What do Airbnb users care about? An analysis of online review comments*

Nell'articolo del 2019 gli autori Cheng e Jin utilizzando l'analisi dei "Big Data" studiano le recensioni di Airbnb a Sydney per individuare i fattori chiave che interessano gli utenti.

Riportano come le caratteristiche distintive dei Big Data sono:

- volume: la quantità dei dati
- velocità: la rapidità di elaborazione dei dati
- varietà: la diversità della tipologia dei dati

Dopo aver pulito il database dai dati non inglesi con il programma OpenRefine si sono analizzati i dati con il software Leximancer, che ha permesso di ottenere risultati oggettivi che diversamente avrebbero potuto essere compromessi da pregiudizio se effettuati con un'analisi manuale. Leximancer crea come output una “mappa termica” che codifica con i colori l'importanza dei temi e raggruppa i concetti con forti significati semantici.

I risultati ottenuti indicano che il termine host per il 66% è espresso positivamente, mentre argomenti come rumore, pavimento, doccia, parcheggio e porta sono collegati a sentimenti più negativi. I temi più citati sono la posizione, i servizi dell'alloggio (strutture del luogo, la camera e l'ambiente notturno), l'host (disponibilità), la comunicazione tra host e ospiti (per arrivo e check-in/out) e gli animali. In particolare, le recensioni contano una media di 34 parole, mentre i temi principali sono legati alla “localizzazione”, ai “servizi”, al “host” e alla “raccomandazione”, quest'ultimo termine è collegato agli altri tre. Si è inoltre cercato cosa producesse la raccomandazione utilizzando “likelihood score” trovando alcune espressioni come “posizione”, “bel posto per la famiglia”, “host cordiale e disponibile” e “bella casa”.

Lo studio ha trovato che gli attributi utilizzati da Airbnb sono analoghi a quelli utilizzati dagli ospiti degli hotel.

### *3.1.6 Neutrality may matter: sentiment analysis in reviews of Airbnb, Booking, and Couchsurfing in Brazil and USA*

Nell'articolo del 2020 gli autori Santos, Mota, Benevenuto e Silva valutano il problema relativo alla distorsione positiva delle recensioni nei servizi di sharing economy confrontando i commenti rilasciati su alcune città degli Stati Uniti e del Brasile tra ottobre

2016 e marzo 2017 sulle tre principali piattaforme di alloggi: Airbnb, Booking e Couchsurfing analizzando rispettivamente 115760, 648030 e 8589 recensioni .

Per rilevare il sentimento di un testo si possono utilizzare tre metodi:

- apprendimento automatico (train classifier)
- lessicali (usa liste con parole del vocabolario quantificate numericamente)
- ibridi

È stato scelto quest'ultimo con lo strumento SentiStrength, che sfruttando un vocabolario lessicale implementato dagli uomini e migliorato dall'apprendimento automatico dà un punteggio tra -4 e +4 (assolutamente negativo e assolutamente positivo) che indica il sentimento del commento.

I risultati mostrano che sulle piattaforme di Airbnb e Couchsurfing i valori della polarità delle recensioni sono più elevati rispetto a quelli di Booking, le percentuali dei commenti negativi sono rispettivamente 4%, 2% e 17%; ciò conferma gli studi esistenti relativi alla maggiore propensione degli utenti nel mercato della sharing economy di rilasciare commenti positivi rispetto a quelli dell'economia tradizionale. Successivamente gli autori spiegano le possibili cause di questa distorsione raccogliendo i dati della letteratura esistente sul tema, come emerge dagli studi di Fradkin et altri (2015) e da Bulchand-Gidumal & Melin-Gonzlez (2020) esse possono riassumersi:

- l'incontro personale tra host e ospite, che in alcuni casi si trovano proprio a vivere come coinquilini in quanto non si affitta la casa intera ma solo un letto o una stanza.
- timore di ritorsioni nel commento dell'host (dato che il sistema di rating in queste piattaforme è bilaterale)
- paura di danneggiare l'host, infatti una recensione non positiva potrebbe disincentivare altri ospiti a prenotare.
- gli utenti di Airbnb preferiscono non rilasciare un commento se il soggiorno non è stato soddisfacente piuttosto di scrivere una recensione negativa

Su Booking invece gli utenti si sentono più liberi di scrivere l'esperienza reale in quanto il sistema di rating è unilaterale, solitamente il soggiorno avviene in hotel o in case vacanza in cui c'è più distacco tra lo staff e l'ospite.

Con ulteriori analisi dimostrano inoltre che la polarità del sentimento:

- non è influenzata dalla città analizzata
- non è influenzata dalla popolarità dell'alloggio
- è influenzata positivamente dalla vicinanza tra ospite e host: limitando la scelta del campione su Booking alle sole "case" escludendo quindi gli hotel hanno trovato che le recensioni negative sono del 4,9% in linea con quelle della sharing economy

Gli autori riferiscono quindi che, data la mancanza di recensioni negative, è necessario prestare attenzione a quelle neutre; infatti, è come se la polarità non fosse espressa su una scala tra -4 e +4, ma fosse limitata tra 0 e +4 quindi i valori vicino a 0 sono quelli da tenere maggiormente in considerazione perché descrivono qualche problematica del soggiorno.

Lo studio dei pochi commenti negativi effettuato con la tecnica LDA (Latent Dirichlet Allocation che permette di modellare alcuni argomenti presenti nei testi) ha evidenziato 10 argomenti in base ai commenti di Airbnb e altrettanti per Booking, per ognuno l'algoritmo ha rilevato le 10 parole che lo identificano maggiormente, ciò che emerge è che tutti questi topics nel caso di Booking hanno una connotazione negativa, mentre per Airbnb 4 su 10 sono positivi, questo sta a significare che sebbene il commento nel complesso sia negativo gli utenti della sharing economy molte volte rilasciano anche alcune esperienze positive per determinati argomenti.

Nell'ultima parte dell'analisi, si sono suddivise i listing di Airbnb in quattro diversi quantili a seconda della loro polarità: preso un campione di 4 possibili soggiorni (1 per ogni quantile) si è chiesto a un campione di trenta volontari con differenti caratteristiche di valutarli avendo solo a disposizione le recensioni, per confrontare i risultati ottenuti con quelli individuati automaticamente. Ne è emerso un lieve scostamento: ciò implica che questa tendenza positiva delle recensioni può compromettere la valutazione umana.

In conclusione, lo studio suggerisce di inserire nei listing di Airbnb un valore strategico che

mette in relazione la polarità media ( $P \in [-4,4]$ ) e il numero ( $C$ ) delle recensioni per aiutare gli utenti a individuare la scelta migliore:

$$score = \log C + (4 + P)^2$$

Gli autori, con i risultati ottenuti, sottolineano la necessità di valutare le recensioni appartenenti a piattaforme diverse con sistemi di interpretazione diversi per poter prevedere in maniera più corretta il processo decisionale degli utenti.

### *3.1.7 Gender differences in the perceived risk of buying online and the effects of receiving a site recommendation*

Nell'articolo "Gender differences in the perceived risk of buying online and the effects of receiving a site recommendation" del 2004, gli autori Garbino e Strahilevitz presentano 3 studi in cui si analizza come i maschi e le femmine percepiscono i rischi legati all'e-commerce e se le raccomandazioni da parte di amici permettano di ridurli. La percezione della probabilità che qualcosa vada storto unito alla percezione della gravità delle conseguenze definisce la percezione del rischio.

Nel primo studio inizialmente sono stati individuati i 5 principali rischi percepiti nel mercato online: utilizzo non autorizzato dei dati sensibili della carta di credito, acquisto su un sito non sicuro, pubblicazione di altri dati personali sensibili, problemi legati alla spedizione o alla consegna e di caratteristiche del prodotto diverse da quelle dichiarate. A un campione di 260 utenti 54% uomini e 46% donne di età differenti sono stati chiesti di valutare la probabilità che qualcosa vada storto con un punteggio da 1 a 7 (improbabile-molto probabile) e di quantificarne l'impatto in caso esso si verifici con un valore da 1 (per nulla gravi) a 7 (molto gravi). Attraverso un'analisi MANOVA è emerso che le donne percepiscono i rischi in maniera maggiore sia per la probabilità che per le conseguenze.

Il secondo studio si concentra su come si potrebbero mitigare questi rischi intervistando un campione di 276 utenti composto per il 51% da maschi e 49% da femmine, in particolare si è prestata attenzione all'efficienza delle raccomandazioni da parte di amici per diminuire il rischio percepito. Il risultato ottenuto ha evidenziato che questo fattore permette ad entrambi i generi di diminuire il rischio percepito, ma l'impatto per le donne risulta

maggiore.

Nel terzo studio ad un campione di 220 utenti si è chiesto di valutare con una scala da -5 (sicuramente online) a +5 (sicuramente in negozio) la probabilità di acquistare online rispetto ad un negozio tradizionale un prodotto il cui prezzo con l'opzione e-commerce risultasse di poco inferiore e sono state monitorate situazioni in cui i prodotti erano stati raccomandati da amici e altre no. Ciò che è emerso è stato che queste raccomandazioni hanno avuto impatti significativi sulle donne, ma non sugli uomini.

I risultati di queste analisi portano gli autori a consigliare alle aziende che vogliono vendere i propri prodotti ad un pubblico femminile di utilizzare dei programmi che incentivino gli utenti a raccomandare i siti ad altri, in particolare tra la cerchia dei propri amici.

### *3.1.8 Exploring gender-based influences on key features of Airbnb accommodations*

Nell'articolo del 2020 gli autori Sánchez-Franco e Alonso-Dos-Santos conducono un'analisi con approcci predittivi ed esplorativi dei commenti e degli annunci della città di New York raccolti tra il 2015 e il 2017 in cui valutano l'esistenza di segnali distintivi del genere.

Lo studio si limita alle recensioni in lingua inglese di cui il 54% è stata scritta da ospiti donne. Sebbene gli studiosi siano consapevoli che le recensioni siano influenzate, oltre che dal sesso, da fattori come età e reddito, ciò che è emerso è che le donne sono più motivate e menzionano principalmente la posizione dell'alloggio, le esperienze locali, i vantaggi per i viaggiatori disabili e per i bambini, l'efficienza e la tempestività, mentre i maschi prestano più attenzione alla pulizia e alla struttura dell'appartamento. Infine, anche in questo studio emerge la maggiore attenzione e aversità al rischio del genere femminile.

In generale, gli autori sottolineano l'importanza di analizzare i feedback in quando un utente che rilascia una recensione comunica agli altri ciò che avrebbe voluto sapere in fase di scelta, per tale motivo lo studio del contenuto delle recensioni è fondamentale sia per gli host quando scrivono i contenuti degli annunci, sia per la stessa piattaforma che può migliorare il servizio offerto.

## 3.2 Prezzo

Il prezzo molto competitivo ha contribuito notevolmente al successo della piattaforma, vista la sua importanza numerosi studiosi si sono concentrati sul tema e hanno individuato i principali fattori che lo influenzano. In particolare, si utilizza il concetto di modelli di “prezzo edonico” in cui si esaminano gli insiemi dei fattori interni ed esterni che determinano il prezzo di mercato di un prodotto; questa definizione è applicabile anche per gli annunci di Airbnb in quanto il prezzo complessivo può essere visto come un insieme di più caratteristiche interne, come il numero di ospiti e di camere, ed esterne come i quartieri e le recensioni. Di seguito sono riportati in sintesi alcune parti degli articoli che sono risultate più significative durante l’attività di ricerca e che hanno permesso di scegliere i parametri da considerare nell’analisi di regressione finale, si confronteranno successivamente i risultati di questi studi con quelli ottenuti.

### 3.2.1 *Key factors affecting the price of airbnb listings: a geographically weighted approach*

Nell’articolo di Zhang, Chen, Han e Yang del 2017 (2017) è presentato uno studio sui parametri che influenzano il prezzo utilizzando due modelli: uno lineare generale (GLM) e l’altro di regressione ponderata geograficamente (GWR), quest’ultimo permette di analizzare la variazione spaziale nei singoli distretti. Il campione scelto è formato da 794 annunci di superhost a Metro Nashville in Tennessee. Le variabili scelte seguendo la letteratura esistente sono state:

- price: considerato come il rapporto tra prezzo di listino di una notte e il massimo numero di ospiti
- reviews: numero di recensioni del listing
- h\_distance: distanza euclidea (km) dall’autostrada più vicina
- c\_distance: distanza euclidea (km) dal centro congressi
- age: quando è stato caricato sulla piattaforma l’annuncio (mesi)
- rating: punteggio complessivo di valutazione

Mentre non sono state considerate variabili relative alla tipologia dell'appartamento.

Nel modello GLM si trova che il numero di recensioni, la loro valutazione e la distanza dal centro congressi sono negativamente correlate al prezzo, l'ultima in particolare ha la maggiore influenza; mentre l'età e la distanza dall'autostrada non hanno un'influenza significativa. In particolare, gli autori motivano che l'influenza negativa della variabile numero di recensioni possa essere dovuta al fatto che Airbnb è da poco comparso in questa area, quindi gli host per incentivare gli ospiti a soggiornare da loro devono puntare su un prezzo molto competitivo: quindi, più gente affitta maggiore sarà il numero di recensioni. Nel caso di GWR invece le variabili distanza dal centro congressi e il numero di recensioni sono sempre negativamente correlate al prezzo, mentre gli altri parametri variano a seconda del distretto.

Confrontando i risultati del modello GLM con GWR gli studiosi affermano che quest'ultimo risulta avere una migliore accuratezza, in quanto permette di valutare l'impatto delle diverse variabili nei quartieri, confermando che ad esempio il coefficiente negativo della distanza dal centro diminuisce nelle zone più lontane dal centro.

### *3.2.2 Do airbnb host listing attributes influence room pricing homogenously?*

Il seguente studio di Chattopadhyay e Mitra (2019) ha individuato i principali fattori che influenzano il prezzo degli annunci di Airbnb applicando 3 diversi modelli: OLS, Random Forest e albero decisionale su un campione di 151955 listing di 11 città statunitensi scaricati da InsideAirbnb. Sono stati considerati 143 attributi differenti: oltre ai più comuni (bagni e numero di letti, il numero di ospiti, il numero di recensioni mensili e i loro punteggi medi) è stata inserita una lista di 137 dummies che rappresentano i servizi offerti dagli host, mentre sono stati esclusi i fattori qualitativi come i parametri sociali.

I risultati mostrano che le variabili più significative nei tre modelli sono simili, in particolare per OLS risultano essere: Bathrooms, Accommodates, Review scores rating, Elevator, Reviews per month, Beds, Paid parking off premises, Indoor fireplace, Cable TV, and Family kid friendly. Analogamente per RF sono: Accommodates, Bathrooms, Beds,

Reviews per month, Number of reviews, Review scores rating, Elevator, Free parking on premises, Family kid friendly, Smoking allowed. Infine, per CT: Accommodates, Bathrooms, Reviews per month, Review scores rating, Elevator, Wifi, Pool, Coffee maker, Beds and Internet.

Gli autori hanno inoltre controllato la non correlazione delle variabili attraverso la matrice di correlazione, inoltre hanno verificato che i valori di VIF fossero minori di 10 affinché non emergesse il problema della multicollinearità.

Confrontando infine le performance dei 3 modelli nelle 11 città gli autori indicano che RF è il modello migliore in quanto ha un valore notevolmente più elevato di R2 mentre ha i minori punteggi di RMSE, MAPE e MAE.

L'individuazione di queste caratteristiche può aiutare gli host a migliorare o aggiungere alcuni servizi al proprio noleggio.

### *3.2.3 Hedonic pricing and the sharing economy: how profile characteristics affect Airbnb accommodation prices in Barcelona, Madrid, and Seville*

Nel seguente articolo sono studiate le tre principali città spagnole: Barcellona, Madrid e Seville, tra il 2015 e il 2016. Gli autori Tong e Gunter (2020) scelgono di utilizzare per la regressione di  $\ln(\text{price})$  le variabili: numero di camere da letto e di bagni, numero di recensioni, punteggio complessivo, numero massimo ospiti, tasso risposta, tempo risposta, numero di foto, distanza dal centro della città, prenotazione istantanea, host possiede più di un annuncio, superhost, politica di cancellazione non flessibile e appartamento intero. I risultati ottenuti con il metodo OLS per la città di Barcellona permettono di considerare tutte le variabili significative, in particolare sono tutte correlate positivamente ad eccezione di host con più annunci, della distanza dal centro (tale punto è rappresentato da Plaça de Catalunya) e del numero di recensioni. Anche in questo articolo gli autori indicano che quest'ultima relazione può essere spiegata come un prezzo minore permetta di catturare una domanda maggiore. I risultati ottenuti dalle altre due città risultano analoghi.

Successivamente hanno calcolato i valori di VIF (fattori di inflazione della varianza) che

risultano minori di 10, questo significa che le variabili non sono fonte di multicollinearità. Inoltre, è stata verificata l'eteroschedasticità. L'articolo utilizza i modelli OLS e WLS per controllarne la robustezza, ottenendo risultati analoghi, anche se WLS può dare risultati più efficienti con errori standard robusti.

Gli studiosi definiscono gli alloggi come prodotti non perfettamente sostituibili in quanto essi sono differenziabili tra loro; quindi, non parlano di concorrenza perfetta ma di concorrenza monopolistica, questo significa che sono gli host che devono essere in grado di attribuire delle tariffe differenti ai propri annunci e di distinguere il loro prodotto da quello degli altri.

#### *3.2.4 Drivers of Airbnb prices according to property/room type, season and location: a regression approach*

Nel seguente articolo gli autori Voltes-Dorta e Sánchez-Medina (2020) prestano particolare attenzione a come la variabile spaziale, la stagionalità e la tipologia di proprietà influenzano il prezzo utilizzando i dati di 2056 annunci di InsideAirbnb della città di Bristol nelle date del 9 febbraio e 14 luglio 2019. In particolare, è emerso che i prezzi di tipologie differenti di camere sono influenzati da fattori diversi come anche la differente posizione produce risultati diversi. Nello studio sono state inserite più variabili per individuare i parametri che rappresentano meglio la posizione: come rispetto ai punti di interesse, alla fermata degli autobus/treni più vicina, ai luoghi di shopping cibo e alloggio, ma molte di queste ultime sono state scartate data l'elevata correlazione lineare con la distanza dal centro.

Si sono così stimate alcune equazioni di prezzo edoniche con i metodi OLS (Ordinary Least Squared) e GWR (Geographically-Weighted Regression) applicando il logaritmo alla variabile dipendente prezzo variando i sottocampioni in analisi combinando le tipologie di proprietà e la stagionalità. Le variabili considerate sono distanza dalle fermate dei mezzi pubblici, distanza dalle stazioni, distanza dal centro, da quanto l'host è sulla piattaforma e il numero di listing che possiede, il numero di ospiti che può ospitare l'appartamento, il numero di bagni e di camere, la disponibilità annuale, il numero recensioni e il loro punteggio, la distanza dalle principali attrazioni turistiche.

Dai risultati emergono differenze relative all'host, nelle caratteristiche delle stanze e nella posizione tra le case e stanze, in particolare risultano significative le variabili bathroom, bedroom e degli ospiti aggiuntivi. Il modello GWR indica che la distanza dal centro influisce solamente sul prezzo delle case intere (principalmente nelle aree occidentali), nella periferia invece la disponibilità annuale influenza il prezzo, i quartieri del nord forniscono invece un maggior mercato agli host con più annunci.

Gli autori suggeriscono che sia meglio differenziare i database di stanze e case invece di inserire una semplice variabile dummy.

### 3.2.5 *Spatial price mimicking on Airbnb: Multi-host vs single-host*

In questo documento gli autori Boto-García, Mayor, e De la Vega (2021) analizzano i comportamenti degli host di Madrid nella definizione dei prezzi distinguendoli tra “professionisti” se hanno più annunci e “non professionisti” e tenendo in considerazione le dipendenze spaziali degli alloggi, ossia le distanze dai mezzi di trasporto e dai luoghi turistici oltre alle caratteristiche strutturali degli annunci. Hanno così analizzato 4000 annunci scaricati da DataHippo. Inizialmente gli autori cercano di individuare una soglia per distinguere gli host non professionisti da quelli professionisti, provano prima con 2 listing e poi con 10, trovando che quest'ultimo valore sia migliore perché permette di rifiutare l'ipotesi  $h_0$  che i prezzi dei due parametri siano uguali, cosa che non accade considerando il valore 2. Ciò che emerge è che gli host professionisti fissano i prezzi in maniera più indipendente e potrebbero cercare di raggiungere condizioni di equilibrio di monopolio competitivo per poter massimizzare i propri profitti a discapito degli altri proprietari che li adattano a quelli di analoghe proprietà nei quartieri (concorrenza di misura). Per limitare tale effetto, i responsabili politici hanno scelto di seguire l'esempio di altre città come Amsterdam, Parigi e Londra, approvando un piano che impedisce agli host di affittare i loro annunci per più di 90 giorni all'anno, inoltre potrebbero scegliere di introdurre una apposita tassa per gli host professionisti ma devono prestare attenzione alle conseguenze che potrebbero ottenere in quanto produrrebbero degli effetti non neutri sul mercato, tali host potrebbero infatti incrementare i prezzi dei propri annunci e di conseguenza anche gli stessi “non professionisti” seguendo il loro comportamento

potrebbero alzarli, portando così ad un aumento collettivo dei prezzi del servizio di condivisione.

I risultati delle regressioni svolte con il metodo dei prezzi edonici (con il prezzo in log) considerando variabili relative alle caratteristiche dell'alloggio e della sua posizione (rispetto ai luoghi turistici, alla stazione ferroviaria e alla metropolitana) indicano che gli appartamenti interi hanno un premio di prezzo di quasi 107%, inoltre emerge una correlazione positiva con la capacità e il numero di camere da letto, mentre è negativa rispetto al numero di recensioni. In generale i prezzi medi degli host professionisti sono più bassi ad eccezione per interi appartamenti, questo può essere dovuto al fatto che essi possono sfruttare le economie di scala. Per quanto riguarda la posizione, la vicinanza con musei e monumenti entro i 500 m e l'accessibilità alle fermate della metropolitana sono valutate positivamente, mentre la vicinanza alla stazione ferroviaria no.

### *3.2.6 Location, location and professionalization: a multilevel hedonic analysis of Airbnb listing prices and revenue*

L'articolo di Deboosere, Kerrigan, Wachsmuth e El-Geneidy (2019) analizza quali variabili definiscono il prezzo degli annunci utilizzando le transazioni di Airbnb effettuate da agosto 2014 a settembre 2016 nella città di New York City. Sono state inserite la stagionalità, variabili strutturali (come camere, bagni, stanza condivisa, alloggio intero, numero di ospiti), variabili relative agli host (superhost, multihost che possiedono tra 2 e 10 annunci, multihost con più di 10 annunci, numero foto, prenotazione immediata, recensioni, valutazione generale) e variabili di posizione e quartiere (metropolitana, accessibilità trasporti pubblici, distanza Times Square, diversità, densità demografica, reddito famiglie, percentuale di bianchi, densità inserzioni).

Dai risultati ottenuti gli autori sottolineano in primo luogo l'importanza della distanza dai mezzi e da Time Square, in secondo luogo gli attributi relativi all'esperienza degli ospiti e degli host come il numero di recensioni di un annuncio e il titolo di superhost permettono di ottenere un premio di prezzo; anche la prenotazione istantanea che è presente maggiormente negli host professionisti, in quanto quelli "amatoriali" preferiscono approvare manualmente una richiesta, dà in media entrate significativamente più elevate.

Inoltre, il prezzo degli host che possiedono meno di 10 annunci è analogo a quello dei proprietari con un unico listing, mentre coloro che possiedono più di 10 annunci hanno un prezzo più basso che permette loro di ottenere una maggiore domanda e quindi ricavi più elevati.

In conclusione, gli autori sottolineano che i responsabili politici devono prestare attenzione ai multiproprietari professionisti in quanto questi sembrano utilizzare la piattaforma non in ottica di sharing economy part-time ma come degli hotel full-time, portando a ridurre gli alloggi destinati ai residenti in favore dei turisti da cui possono ottenere migliori ricavi, portando quindi al problema della gentrificazione di alcune aree della città.

## 4 ANALISI DATI

L'obiettivo di questi capitoli finali è quello di studiare le recensioni rilasciate dagli utenti e gli annunci caricati sulla piattaforma di Airbnb, per fare ciò si inizia costruendo le statistiche più significative, successivamente attraverso la Sentiment Analysis e il text mining si individuano i sentimenti che emergono dalle recensioni ed infine è svolta un'analisi di regressione sul prezzo per valutare quali fattori impattano su questo. Visto l'elevato numero di dati da trattare si è scelto di utilizzare il linguaggio di programmazione Python che mette anche a disposizione numerose librerie e tools gratuiti per effettuare queste operazioni e il software statistico Stata che permette le attività di manipolazione dei dati, visualizzazione grafica, costruzione delle statistiche e di creazione dei report; per ogni operazione è indicata il mezzo con cui è stata svolta.

### 4.1 Descrizione database disponibili

I dati per l'analisi sono stati scaricati liberamente e gratuitamente dalla piattaforma Inside Airbnb, indipendente da Airbnb, che raccoglie e aggiorna mensilmente sotto forma di database in formato csv le informazioni pubbliche rilasciate dalla piattaforma di sharing economy. Questi database sono disponibili solo per un numero limitato delle principali città turistiche mondiali e sono:

- a) Listing.csv.gz: dati dettagliati delle inserzioni
- informazioni generiche relative all'appartamento (id, listing\_url, scrape\_id, last\_scraped, name, description, neighborhood\_overview, picture\_url)
  - informazioni sull'host (host\_id, host\_url, host\_name, host\_since, host\_location, host\_about, host\_response\_time, host\_response\_rate, host\_acceptance\_rate, host\_is\_superhost, host\_thumbnail\_url, host\_picture\_url, host\_neighbourhood, host\_listings\_count, host\_total\_listings\_count, host\_verifications, host\_has\_profile\_pic, host\_identity\_verified, calculated\_host\_listings\_count, calculated\_host\_listings\_count\_entire\_homes, calculated\_host\_listings\_count\_private\_rooms,

calculated\_host\_listings\_count\_shared\_rooms)

- informazioni relative alla posizione (neighbourhood, neighbourhood\_group, neighbourhood\_cleansed, neighbourhood\_group\_cleansed, latitude, longitude)
  - informazioni sulla tipologia e i servizi dell'appartamento (property\_type, room\_type, accommodates, bathrooms, bathrooms\_text, bedrooms, beds, amenities)
  - informazioni tecniche sulla prenotazione del soggiorno (price, minimum\_nights, maximum\_nights, minimum\_minimum\_nights, maximum\_minimum\_nights, minimum\_maximum\_nights, maximum\_maximum\_nights, minimum\_nights\_avg\_ntm, maximum\_nights\_avg\_ntm, instant\_bookable)
  - informazioni sulla disponibilità (calendar\_updated, has\_availability, availability\_30, availability\_60, availability\_90, availability\_365, calendar\_last\_scraped)
  - informazioni sulle recensioni rilasciate degli ospiti (number\_of\_reviews, number\_of\_reviews\_ltm, umber\_of\_reviews\_l30d, first\_review, last\_review, review\_scores\_rating, review\_scores\_accuracy, review\_scores\_cleanliness, review\_scores\_checkin, review\_scores\_communication, review\_scores\_location, review\_scores\_value, reviews\_per\_month)
  - altro (license)
- b) calendar.csv.gz: dati dettagliati del calendario
- Listing\_id
  - Date
  - Availabòe
  - Price
  - Adjusted price
  - Minimum\_nights

- Maximum\_nights
- c) reviews.csv.gz: dati dettagliati delle recensioni rilasciate per gli annunci relativi alla città
- Listing\_id
  - Id
  - Date
  - Reviewer\_id
  - Reviewer\_name
  - Comments
- d) listings.csv: informazioni di riepilogo delle inserzioni
- informazioni generiche relative all'appartamento (id, name)
  - informazioni sull'host (host\_id, host\_name, calculated\_host\_listings\_count)
  - informazioni relative alla posizione (neighbourhood, neighbourhood\_group, latitude, longitude)
  - informazioni sulla tipologia e i servizi dell'appartamento ( room\_type)
  - informazioni tecniche sulla prenotazione del soggiorno (price, minimum\_nights)
  - informazioni sulla disponibilità ( availability\_365)
  - informazioni sulle recensioni rilasciate degli ospiti (number\_of\_reviews, reviews\_per\_month, last\_review)
- e) reviews.csv: informazioni di riepilogo delle recensioni
- listing\_id
  - date

- f) neighbourhoods.csv: elenco dei quartieri
- neighbourhood\_group
  - neighbourhood
- g) neighbourhoods.geojson: file GeoJSON dei quartieri della città

Il seguente lavoro di tesi analizza i dati della città di Barcellona per continuità di studi precedenti, inoltre la città è una meta che richiama turisti e lavoratori da tutto il mondo, ciò permette di avere a disposizione un elevato numero di listings e commenti rilasciati da utenti di diverse culture. In particolare, per lo studio sono stati utilizzati i database più dettagliati Listing.csv.gz e reviews.csv.gz con data di compilazione il 7 luglio 2021 contenenti rispettivamente 17079 annunci e 555889 recensioni.

## 4.2 Data exploration

Inizialmente si sono esplorati i dati disponibili nei due database salvati come file csv: per leggerli è stato importato “pandas” che è uno strumento open source per l’analisi e manipolazione dei dati che si basa sul linguaggio Python, si è quindi utilizzata la funzione `“.read_csv(“nome_del_file”)` che legge automaticamente ogni riga del file csv e la separa in colonne a seconda dell’elemento scelto come separatore creando un DataFrame che è una struttura tabellare di righe e colonne.

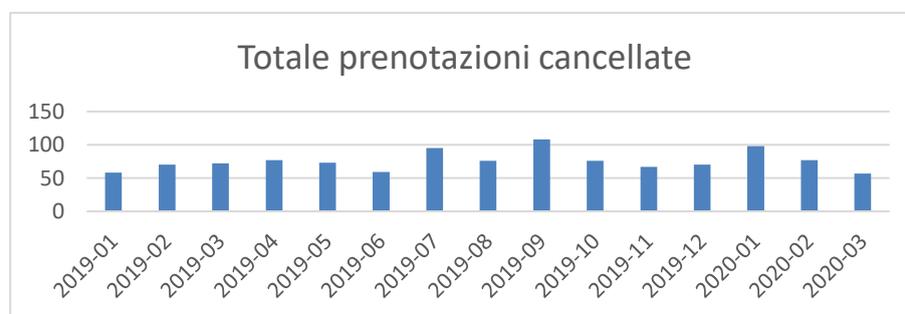
- a) reviews.csv.gz

Il database completo ha 555889 recensioni, le prime risalgono a luglio del 2010, visto l’elevato numero e il cambiamento dei servizi offerti nei soggiorni e le esigenze dei consumatori si è scelto di limitare l’analisi alle recensioni successive al 1° gennaio 2019, ottenendo così 230006 valori, la finestra temporale permette di analizzare anche gli effetti della pandemia del Covid-19 che da febbraio 2020 ha compromesso la vita della popolazione mondiale e ha influenzato notevolmente il settore turistico.

Dopo aver individuato 261 recensioni che presentavano caratteri nulli si è scelto di eliminarli visto il basso impatto sul totale. Ad una più attenta analisi si sono trovati dei

commenti come “The host canceled this reservation 2 days before arrival. This is an automated posting.” che indicano quindi che il soggiorno non è stato svolto in quanto l’host ha cancellato la prenotazione, per evitare che influenzassero i risultati si sono eliminati le 1133 righe che contenevano la stringa “This is an automated posting”. Successivamente si è provato a vedere se erano presenti informazioni relative alle cancellazioni legate alla pandemia, ma Airbnb ha aggiornato le proprie politiche da marzo 2020 come si può notare dal grafico 4.1 per permettere agli host di annullare i soggiorni con più tranquillità senza che ne rimanesse traccia, in quanto la presenza di tali recensioni può non essere ben vista da un utente che sta scegliendo un alloggio perché può percepire l’host come inaffidabile.

*Grafico 4.1 Distribuzione delle recensioni di prenotazioni cancellate*

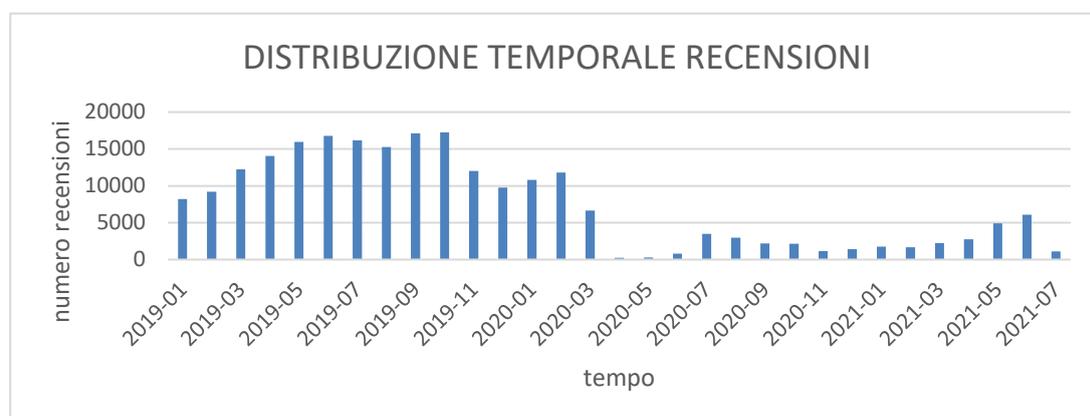


Con le 228612 righe rimaste si è costruito un grafico in cui è evidenziato l’andamento mensile dei commenti, questi corrispondono in parte al numero degli utenti che hanno usufruito del servizio, in quanto al termine del soggiorno è richiesto di rilasciare un feedback, ma non essendo obbligatorio qualcuno può non averlo scritto e quindi di conseguenza il numero reale di coloro che hanno effettuato una prenotazione sulla piattaforma potrebbe essere maggiore. Inoltre, potrebbe esserci una discrepanza di un mese tra quando un ospite ha effettuato il proprio soggiorno rispetto a quando ha pubblicato la recensione visto l’intervallo temporale di 14 giorni concesso della piattaforma o eventualmente se la durata dell’occupazione è stata su due o più mesi differenti (ad esempio tra il 28 luglio e il 3 agosto).

Dal grafico 4.2 si può notare, come atteso, che nei mesi in cui è scoppiata la pandemia e sono stati fatti i lockdown nei diversi stati per contenere la diffusione del virus il numero di recensioni si sia drasticamente ridotto, mentre nel periodo estivo del 2020 c’è stata una lieve ripresa, per poi decrescere nuovamente in autunno con il secondo lockdown,

successivamente ha ricominciato a salire gradualmente. Il basso valore del mese di luglio 2021 diversamente da quanto immaginato è dovuto alla mancanza di dati successivi al 7 luglio, data in cui sono stati scaricati i dati.

Grafico 4.2 Distribuzione temporale delle recensioni



Un esempio di una riga del database iniziale è quella mostrata nella tabella:

Tabella 4.1 Esempio tabella reviews

Listing_id	id	date	Reviewer_id	Reviewer_name	comments
18674	447343763	2019-05-03	124511137	Natalia	Un appartamento bonito y bien ubicado

b) Listing.csv.gz

I soggiorni disponibili nel database sono 17079 e sono distribuiti nella città di Barcellona come raffigurato dalla cartina nella Figura 4.1. Le colonne presenti sono 74, per cui si è scelto di esaminare quelle più importanti che sono utilizzate nelle successive analisi. In particolare, in questo paragrafo sono presentati quelle più generali, mentre gli altri parametri sono descritti man mano che sono utilizzati.

Le tipologie di proprietà disponibili sono molteplici, ma 43 di queste (tra cui ad esempio boat, camper, private room in chalet, room in apartment...) se considerate singolarmente impattano sul complessivo meno dell'1%, 10 intorno all'1% e 2 sul 2%, mentre "private room" ha un valore del 37% e "entire apartment" del 46% coerentemente con il business chiave di Airbnb, per cui InsideAirbnb aggrega alcuni valori passando da 57 valori della

tipologia di proprietà a 4 nel Grafico 4.3 della tipologia di camera. Per fare ciò, sotto la voce “private room” sono racchiuse tutte le proprietà che comprendono “private room”, stesso discorso vale per le “Shared room” e per “Entire room” solo pochi altri valori sono comprese in questi parametri, infine con “hotel room” sono identificati i restanti (‘Room in serviced apartment’, ‘Room in aparthotel’, ‘Room in boutique hotel’, ‘Room in bed and breakfast’, ‘Room in hotel’, ‘Room in hostel’).

Figura 4.1 Distribuzione appartamenti Barcellona

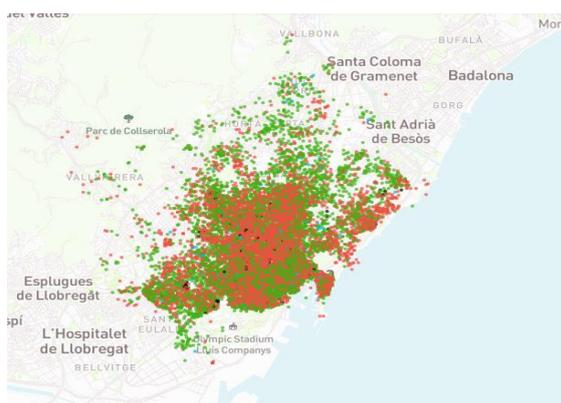
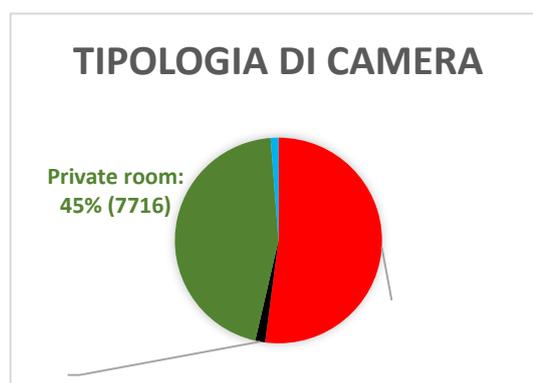


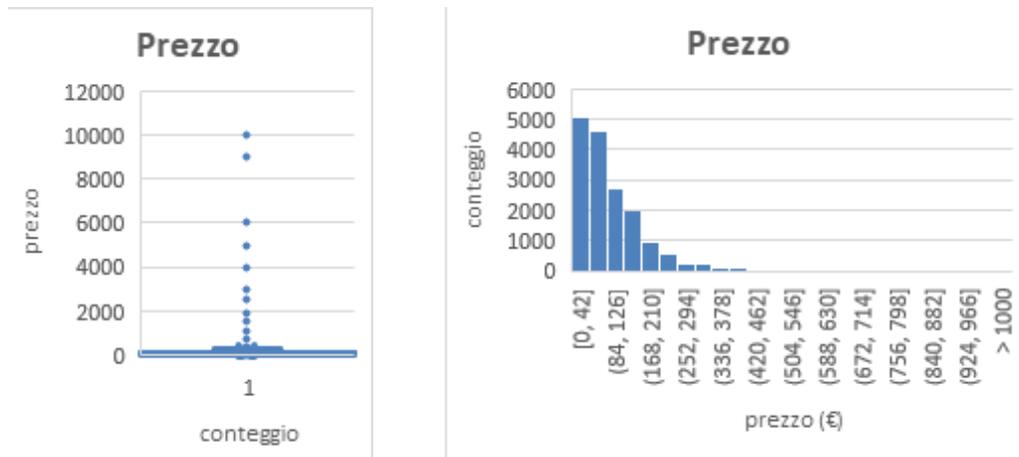
Grafico 4.3 Tipologia di camere



Quando si eseguirà la regressione non saranno considerati i listing con tipologia “hotel room” perché potrebbero compromettere i risultati dal momento che non sembrano rispecchiare le idee basi della piattaforma.

La rapida espansione di Airbnb è da attribuire principalmente al prezzo molto più basso rispetto ai competitor tradizionali, vista la sua importanza si è scelto di individuare i principali fattori che lo caratterizzano come verrà svolto nel capitolo della regressione. La sua distribuzione risulta molta ampia anche se è possibile individuare la presenza di outliers nel caso di valori oltre i 2000 euro a notte; infatti, verificando manualmente alcuni di essi per la maggior parte il prezzo è esorbitante rispetto a quanto varrebbe nella realtà; quindi, ipotizzando errori di battitura si sono trascurati tali listing da quelli che sono utilizzati dalla regressione.

Grafico 4.4 Andamento prezzo listing a notte



### 4.3 Classificazione variabili disponibili

#### 4.3.1 *Lingua*

Per poter effettuare una Sentiment Analysis più precisa sia a livello di quantità di elementi analizzata sia a livello di eterogeneità e veridicità dei dati, non si è ridotto il campione alle sole recensioni inglesi che avrebbero limitato l'influenza dei fattori culturali ma si è proceduto con la traduzione dei commenti in lingua inglese. Questa fase ha richiesto molto tempo, in primo luogo ciclando sui commenti si è cercato contemporaneamente di rilevare la lingua del commento con la funzione "detect" della libreria "langdetect" e di tradurlo. Nonostante l'utilizzo delle principali API e librerie gratuite come "google-trans" (API gratuita di Google Traduttore per Python), Traslator (che permette di tradurre il testo usando Google Translate) e la funzione "translate" del pacchetto di TextBlob (anche questa libreria utilizza Google Translate), si sono scatenate delle eccezioni o in alcuni casi sebbene dicesse che l'operazione era stata effettuata correttamente la funzione si era limitata a copiare la recensione senza tradurla, tali problemi si sono verificati in parte a causa della limitazione nell'utilizzo ad un basso numero di dati; quindi è stato necessario trovare un'ulteriore soluzione. Dopo una attenta ricerca sul web si è trovato un tool meno utilizzato "deep-translator" che però è illimitato e gratuito. Si è quindi cercato di procedere come descritto

in precedenza, ma compresa la lunga tempistica richiesta per questa fase che ha portato il computer a girare per un paio di giorni senza dare il risultato atteso, si è scelto di iniziare rilevando la lingua dei commenti con la libreria “langdetect” (ottenendo i dati rappresentati nel grafico 4.5 in cui sotto la voce altro sono stati raggruppati le lingue che avevano valori inferiori al 0,9%), successivamente si è creato un nuovo database con le sole recensioni non in lingua inglese (85762 righe invece di 228612) e si è applicato il tool di traduzione solo su questi ultimi dati. Si può notare che se si fossero esclusi i commenti in lingua non inglese di sarebbe eliminato più di un terzo del database (37%)

Grafico 4.5 Distribuzione delle lingue dei commenti



Osservando i dati ottenuti dal processo di traduzione, confrontando i commenti che erano stati classificati come non inglesi originali con quelli tradotti, è emerso che in alcuni casi (3551) essi fossero uguali anche se ciò non sarebbe dovuto succedere. Ad una più attenta analisi si è potuto individuare che tali errori commessi dall’algoritmo fossero legati molte volte al fatto che i commenti fossero molto brevi (una, due o tre parole), altre volte invece essi erano limitati a una lettera (“h”/“j”) o ancora contenevano errori di battitura. infine, alcuni commenti sono stati classificati con “non\_rilevata” a causa del loro contenuto che si limitata a numeri (come ad esempio “10”, “7”, ...) o a segni di punteggiatura (come ad esempio “: )”, “\*\*\*\*\*”,...) che quindi non permette all'algoritmo di identificarne la lingua.

#### 4.3.2 Maschio/femmina

Successivamente si è cercato di classificare il genere dei recensori e degli host, utilizzando il pacchetto “gender” di Python che passando un nome (nel caso in analisi “reviewer\_name”

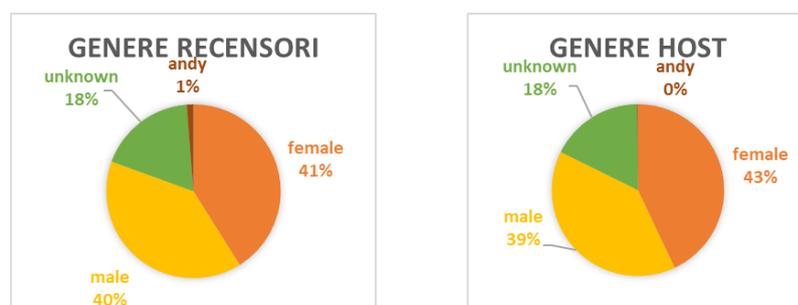
e “host\_name”) restituisce: “unknown” se non trova il nome, “andy” (androgino), “male”, “female”, “mostly\_male”, “mostly\_female”. Per poter effettuare l’analisi “mostly\_male” e “mostly\_female” sono stati classificati rispettivamente come “male” e “female”

Sfortunatamente è emerso che un’alta percentuale (18% dei recensori e 27% degli host) è stato “unknown”, quindi non è stato possibile classificarli come maschio e femmina. Osservando i dati presenti nelle righe così etichettate si nota che questo spesso è dovuto al fatto che numerosi account non hanno utilizzato il loro nome proprio ma soprannomi, cognomi, nomi composti da due o più persone o nomi di agenzie come ad esempio “Mireia And Maria” o “The Onsider” o i “Bcn Clois Apartments, S.L.” che non permettono all’algoritmo di classificarli.

Analizzando il database listing si possono rilevare 8099 host. Un algoritmo classifica i nomi in maschili, femminili, unknown e andy, in seguito all’elevato numero degli host (circa 27%) etichettati con gli ultimi due attributi che rappresentano il 41% degli annunci, si è eseguita una più attenta analisi su questi ultimi.

Per tale motivo si è individuata una nuova classe chiamata “agenzia” in cui sono raccolti tutti gli host il cui “host\_name” conteneva stringhe come “apartm”, “barcell”, “hotel”, “hostel”, “room”.. che celavano agenzie o enti del settore alberghiero. Questa etichetta ha permesso di spiegare 215 degli host “unknown”(corrispondenti a 2010 listing). In altri 380 nomi sono stati individuati caratteri come “&”,”/”,”|” o parole come “e”,”and”, “y” che implicavano la presenza di più persone, questi sono definiti come “coppie\_trii” e corrispondono a 1564 listing. Dei restanti, tramite un’analisi visiva, si sono definiti i generi di altri 694 host corrispondenti a 1095 listing.

Grafico 4.6 Percentuali genere dei recensori e degli host



## 5 ANALISI FEEDBACK

### 5.1 Punteggio rating recensioni

#### 5.1.1 Parametri disponibili

Nel database “listing” sono presenti le medie dei punteggi rilasciati dagli ospiti al termine del loro soggiorno che possono variare tra 1 e 5:

- review\_scores\_accuracy: i dati indicati nel listing sono veritieri e precisi?
- review\_scores\_cleanliness: la pulizia è adeguata agli standard indicati da Airbnb?
- review\_scores\_checkin: il checkin è stato semplice?
- review\_scores\_communication: la comunicazione con l’host è stata semplice e rapida?
- review\_scores\_location: l’ospite era consapevole delle caratteristiche della zona dell’alloggio?
- review\_scores\_value: com’è il rapporto qualità-prezzo?

Tali valori sono resi pubblici solo quando almeno tre utenti hanno rilasciato la loro valutazione, diversamente è presente il valore nullo.

Inoltre, c’è un ulteriore parametro “review\_scores\_rating” che raccoglie i precedenti valori per una più veloce rappresentazione.

#### 5.1.2 Distorsione recensioni troppo positive

Analizzando i punteggi complessivi dei listing presenti nella città si sono ottenuti risultati in linea con gli studi del 2020 di Zerpas, Proserpio e Byers in “A First Look at Online Reputation on Airbnb, Where Every Stay is Above Average” che indicano la presenza di

una forte distorsione verso i punteggi positivi. Come si può notare dai Grafici 5.1 e 5.2 essi non hanno distribuzione normale con media nei punteggi intermedi, ma hanno il picco verso l'estremo destro, non è stata però riscontrata una distribuzione a forma di J (bimodale) come descritto nello studio sui mercati online in “Overcoming the J-shaped distribution of product reviews” di Nan Hu, Paul A. Pavlou e Jie Zhang del 2009.

Grafico 5.1 Distribuzione punteggi complessivi

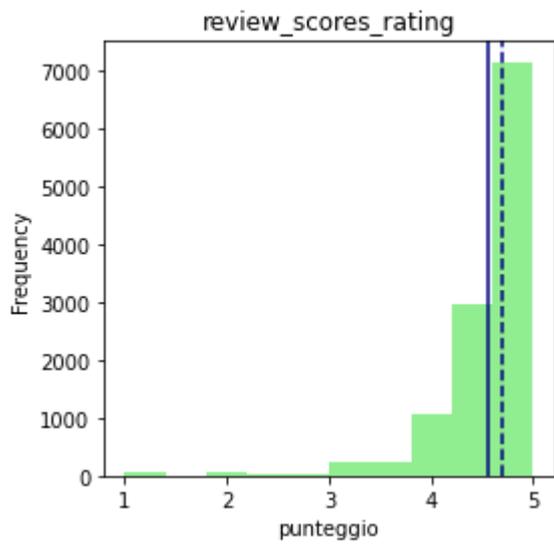
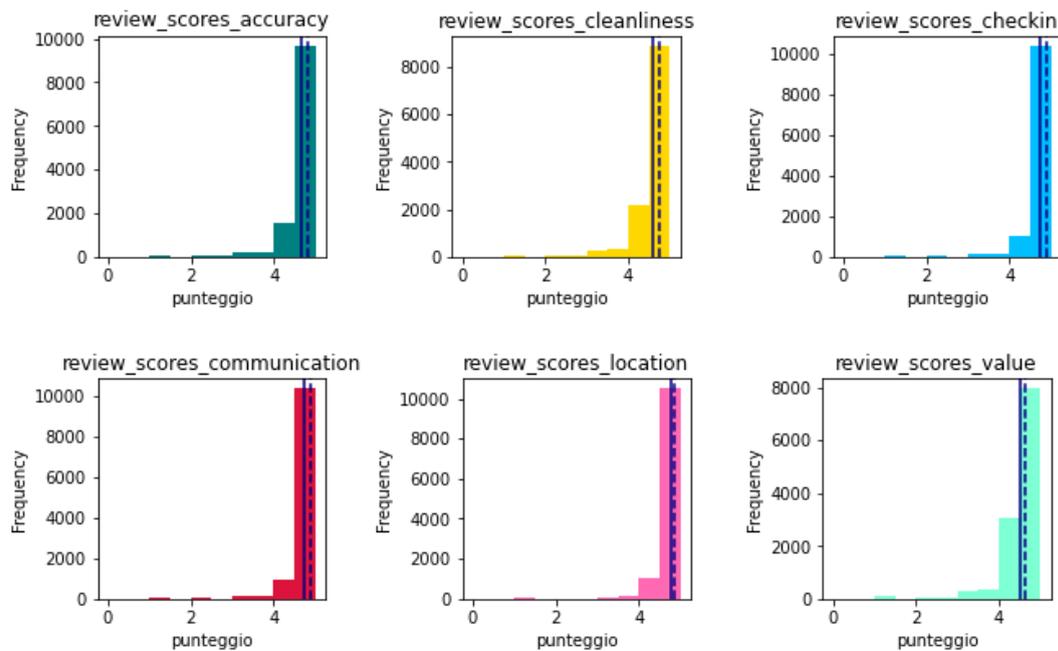


Grafico 5.2 Distribuzione variabili rilasciate dai feedback



Le statistiche riportate in tabella 5.1 dimostrano come i punteggi medi siano molto elevati,

infatti risultano tutti superiori a 4,49.

Tabella 5.1 Statistiche punteggi recensioni

parametro	mediana	media	varianza
review_scores_rating	4,68	4,49	0,60
review_scores_accuracy	4,8	4,66	0,27
review_scores_cleanliness	4,75	4,59	0,31
review_scores_checkin	4,89	4,75	0,22
review_scores_communication	4,89	4,75	0,23
review_scores_location	4,86	4,76	0,15
review_scores_value	4,63	4,52	0,29

## 5.2 Recensioni testuali

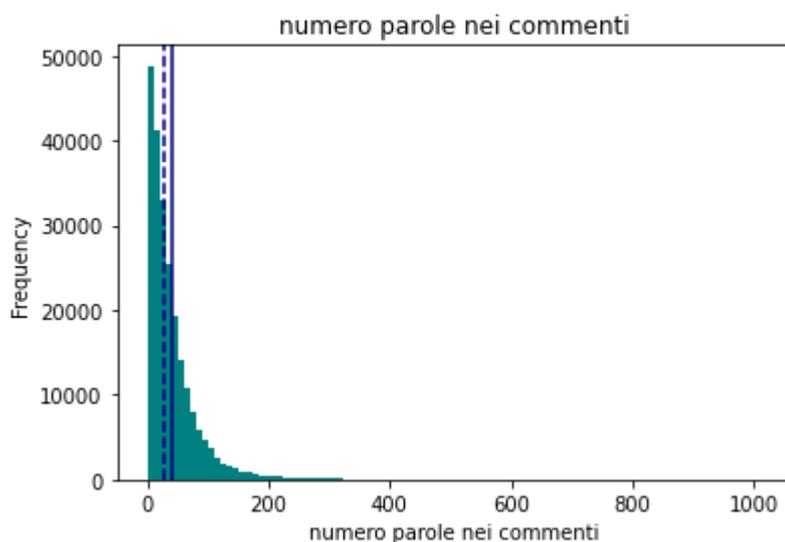
### 5.2.1 Lunghezza recensioni

Successivamente, si sono calcolati il numero di parole presenti in ogni recensione, calcolando la lunghezza della lista di parole ottenuta con il metodo `".split()"` che divide una stringa. I risultati ottenuti sono rappresentati nell'istogramma seguente, per policy aziendale le recensioni hanno un limite di 1000 parole ma come si può notare tendono ad essere meno di 200. In questo campione la media di 40,7 risulta inferiore rispetto al valore ottenuto nello studio Bridges e Vásquez che era 75.

Tabella 5.2 Statistiche relative al numero di parole nei commenti

media	mediana	varianza	numerosità campione
40,7	28,00	2099,75	228612

Grafico 5.3 Distribuzione numero parole nei commenti



### 5.2.2 Classificazione delle recensioni

In questi ultimi anni sempre più acquirenti dopo aver acquistato un prodotto o servizio hanno iniziato a rilasciare la loro opinione su Internet, questo sta portando sempre più aziende a cercare tali feedback e ad analizzarli per capire come sono percepiti i propri prodotti e come possono essere migliorati. Attraverso la sentiment analysis da una frase è possibile quantificare un contenuto testuale e comprenderne il significato.

Per eseguire questo tipo di analisi ci sono tre metodi:

- apprendimento automatico
- basato sul lessico senza supervisione
- ibrido

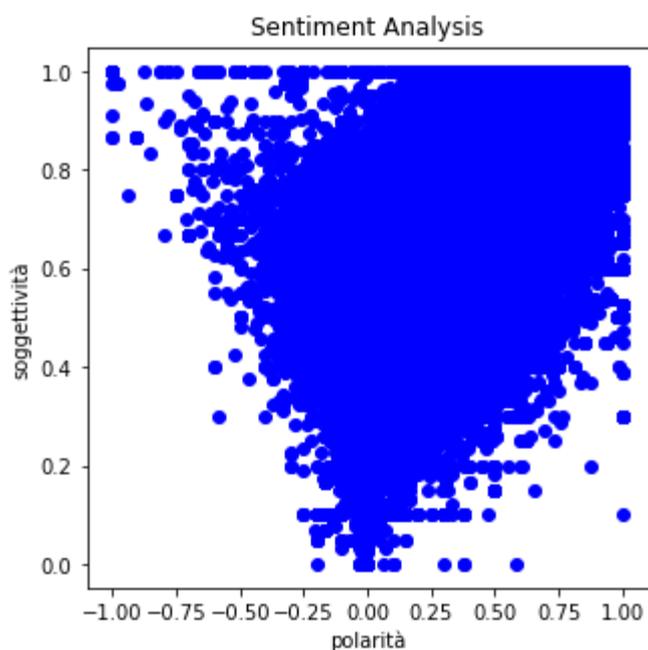
Nel caso in esame si è scelto di utilizzare il secondo metodo in quanto non si hanno a disposizione dati pre-etichettati da dare al primo per svolgere un'attività di training che permetterebbe al modello di imparare a classificare. Si è così impiegata la libreria di Python TextBlob che permette di accedere a numerose operazioni di elaborazione del linguaggio naturale (sfruttando la libreria NLTK che permette l'accesso a molte risorse lessicali).

Questo classificatore si basa sul lessico presente nella frase, sommando il valore attribuito al set di addestramento NLTK ad ogni parola presente nella frase e considerando l'intensità di ogni vocabolo.

TextBlob permette di ricavare la polarità e la soggettività della frase passata. La prima ha un valore float compreso tra -1 (sentimento negativo) e +1 (sentimento positivo), mentre la seconda varia tra 0 (oggettiva) e 1 (soggettiva). Bisogna però prestare attenzione alla precisione di questo modello che ha migliori prestazioni nell'analizzare frasi singole rispetto a testi composti da più frasi.

Come si può notare dal Grafico 5.4 i commenti si concentrano nell'angolo in alto a destra, questo significa che la maggioranza sia positiva e soggettiva.

*Grafico 5.4 Relazione tra soggettività e polarità*



Con i dati di polarità numerici ottenuti per una più semplice analisi si sono etichettate le recensioni in positive se il valore era maggiore di zero, neutre se era uguale a zero e negative se minore, ottenendo i valori riportati nella tabella successiva. I risultati ottenuti confermano i papers presenti sull'argomento che indicano che su Airbnb i commenti hanno una distorsione positiva, infatti, il 95,8% di questi risultano positivi (Alsudais & Teubner, 2019) (Bridges & Vásquez, 2018).

Grafico 5.5 Sentimento dei commenti

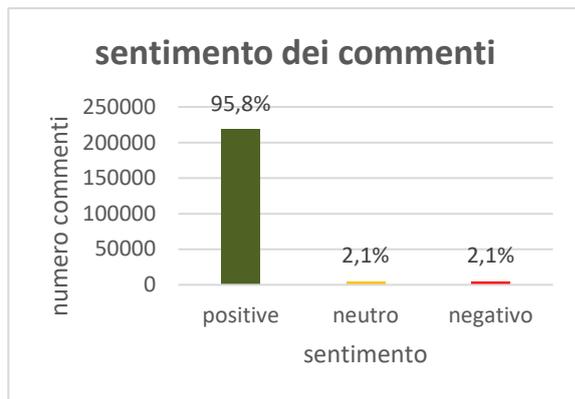


Tabella 5.3 Ripartizione del sentimento dei commenti

sentimento	quantità
positivo	218985
neutro	4714
negativo	4913

I dati così ottenuti sono stati confrontati con un altro strumento di analisi VADER (Valence Aware Dictionary and Sentiment Reasoner) che è basato sul lessico e regole legati principalmente ai sentimenti espressi sui social media ma funziona anche bene in situazioni differenti. Da “nlk.sentiment.vader” si è importato “SentimentIntensityAnalyzer” per utilizzare la funzione “polarity\_scores” che restituisce quattro punteggi: “pos”, “neu” e “neg” che indicano la frazione del peso del punteggio che è racchiuso in ogni categoria e “compound” che rappresenta la somma dei pesi dei punteggi anche esso normalizzato e compreso tra -1 molto negativo e +1 molto positivo e descrive il sentimento complessivo della recensione. I risultati ricavati con questo metodo hanno confermato la distorsione molto positiva dei commenti rilasciati. Esaminando i punteggi ottenuti con i due tool, sebbene risultassero analoghi nella maggior parte dei casi, si è preferito memorizzare e usare nelle successive classificazioni il primo in quanto è emerso adattarsi meglio alla tipologia di testo da classificare.

Seguendo l’esempio dell’articolo di Alsudais e Teubner (2019) si sono suddivise le recensioni in frasi, anche con questo campione le distribuzioni sebbene rimangano alte diminuiscono, questo può significare che alcune recensioni che l’algoritmo classifica positivamente, comprendono al loro interno delle frasi neutre o negative. Come, ad esempio, il commento “.... To improve, the double bed.”, sebbene nel complesso sia positivo, analizzando le singole frasi si nota come “To improve, the double bed.” sia classificata neutra dall’algoritmo. In generale ogni commento è composto in media da 3,48 frasi.

Grafico 5.6 Sentimento delle singole frasi delle recensioni

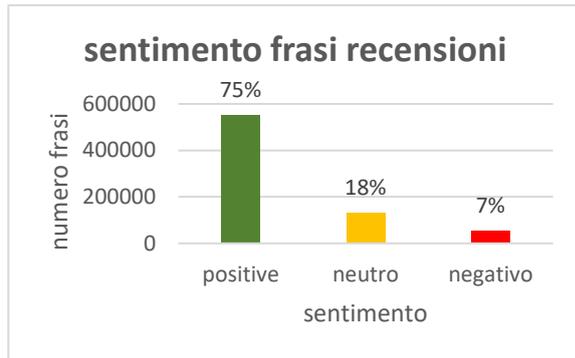
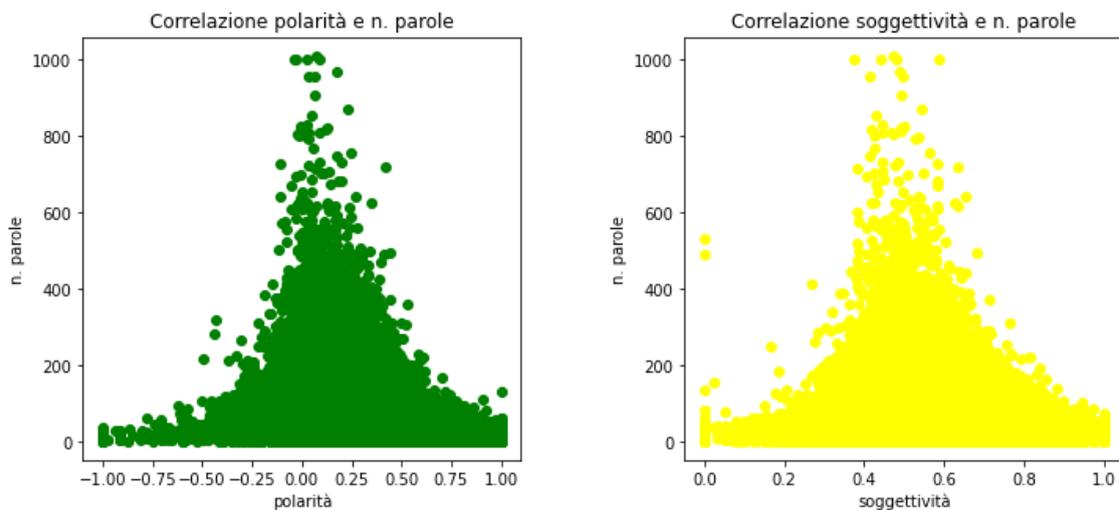


Tabella 5.4 Ripartizione del sentimento delle singole frasi nei commenti

sentimento	quantità
positivo	592650
neutro	146188
negativo	56146

Si è successivamente provato ad indagare l'esistenza di una correlazione della polarità e della soggettività con la lunghezza dei commenti attraverso il comando di Stata "`corr polarit lunghezza`" e "`corr soggettivit lunghezza`" ottenendo rispettivamente -0,3453 e -0,1885, in entrambi i casi la correlazione risulta negativa e non particolarmente significativa dal momento che tale valore è compreso tra -1 e +1 e gli estremi indicano la massima correlazione (negativa e positiva) come si nota anche dal Grafico 5.7 le recensioni contenenti i sentimenti neutri hanno un numero di parole più elevato, mentre quando essi hanno un valore medio di soggettività i commenti sono più lunghi.

Grafico 5.7 Correlazioni del numero di parole con la polarità e la soggettività



### 5.2.3 *L'host è chiamato con il proprio nome*

Per analizzare il rapporto diretto tra ospite e host e quindi valutare il livello di confidenza raggiunto, si è cercato il numero di commenti in cui comparisse il nome proprio del padrone di casa e in quanti fosse citato il termine “host” trasformando tutte le stringhe analizzate in caratteri minuscoli in quanto il confronto è “case-sensitive” (sensibile ai caratteri maiuscolo e minuscolo), ottenendo i dati riportati in Tabella 5.5. Per fare ciò è stato necessario unire le due tabelle “reviews” e “listings” in un’unica tabella “join” mettendo in chiave il numero del listing che è presente in entrambe.

Dall’analisi emerge che nonostante il problema legato alla scelta del nome dell’host spesso “inventato”, che non permette di individuarne il genere, esso compare nel 29% delle recensioni considerate, mentre il termine “host” ha una percentuale 25%. La presenza dei nomi propri sottolinea la differenza rispetto ai soggiorni negli hotel tradizionali in cui l’interazione con i membri dello staff in genere è più formale.

Come sarà dimostrato in seguito i recensori donna sono più propense a rilasciare il nome degli host nei commenti come ad indicare che riescono a creare una maggiore confidenza.

*Tabella 5.5 Incidenza nome proprio e termine host sulle recensioni*

valore	Nome proprio	Termine “host”
True	65539 (29%)	57605 (25%)
False	162907 (71%)	171007 (75%)
Errore	166 (0%)	0(0%)

### 5.2.4 *Parole più utilizzate (in differenti situazioni)*

Si analizzano con maggiore dettaglio i contenuti dei commenti. Su Python si è proceduto nel seguente modo: inizialmente si è creato un set che contenesse tutte le parole trasformate in minuscolo presenti nelle recensioni tradotte, da queste sono stati eliminati i vocaboli con caratteri non alfanumerici e le “stop words” che sono parole come articoli, congiunzioni,



Tabella 5.6 Percentuale delle recensioni in cui compaiono le parole più citate confrontando quelle generali, con il sentimento e il genere

	comprehensive		positive		negative		femmine		maschi	
	parole	%	parole	%	parole	%	parole	%	parole	%
1	locat	51	locat	52	veri	44	veri	55	locat	49
2	veri	51	veri	52	apart	40	locat	54	veri	47
3	apart	38	apart	38	locat	30	apart	43	apart	36
4	great	35	great	36	room	27	great	37	great	36
5	stay	31	stay	32	stay	21	stay	35	place	29
6	place	29	place	29	place	20	clean	31	stay	29
7	clean	27	clean	28	bathroom	19	place	29	clean	24
8	recommend	25	recommend	25	small	19	recommend	27	recommend	24
9	host	23	host	24	host	18	host	24	host	23
10	good	22	good	23	well	18	nice	22	good	22
11	nice	21	nice	22	clean	17	barcelona	21	nice	21
12	barcelona	20	barcelona	20	day	16	good	21	barcelona	19
13	everyth	17	everyth	17	one	16	everyth	19	well	16
14	well	17	well	17	bad	16	well	18	everyth	15
15	room	16	room	16	bed	15	room	17	perfect	14
16	perfect	15	perfect	15	time	15	would	17	would	13
17	close	14	close	15	night	14	perfect	17	room	13
18	would	14	would	15	recommend	14	comfort	16	close	13
19	comfort	14	comfort	14	dirty	14	metro	16	metro	12
20	metro	14	metro	14	would	13	close	16	walk	11
21	walk	13	walk	14	even	13	walk	16	comfort	11
22	help	12	help	13	kitchen	13	help	14	help	11
23	need	11	need	11	arriv	12	love	13	need	10
24	realli	10	friend	11	shower	12	need	13	excel	10
25	friend	10	thank	11	also	12	thank	12	friend	9
26	thank	10	realli	11	airbnb	12	realli	12	realli	9
27	time	10	love	10	good	12	time	11	time	9
28	love	10	time	10	close	11	friend	11	citi	9
29	also	10	also	10	work	11	also	11	thank	9

30	citi	9	excel	9	like	11	easi	10	high	8
31	excel	9	citi	9	door	10	restaur	10	communic	8
32	high	9	high	9	littl	10	citi	10	also	8
33	restaur	9	restaur	9	peopl	10	area	9	easi	8
34	easi	8	easi	9	use	10	super	9	restaur	7
35	communic	8	communic	8	go	10	high	9	area	7
36	area	8	area	8	barcelona	10	communic	9	love	7
37	super	8	super	8	could	10	definit	9	super	6
38	definit	7	definit	7	floor	10	beauti	9	amaz	6
39	get	7	beauti	7	everyth	10	bed	8	quiet	6
40	lot	7	lot	7	get	9	lot	8	definit	6

*Tabella 5.7 Parole più utilizzate nelle recensioni confrontando quelle complessive, con quelle pre-durante covid e professionisti e amatoriali*

	complessive		pre-covid		durante-covid		professionisti		amatoriali	
	parole	%	parole	%	parole	%	parole	%	parole	%
1	locat	51	locat	53	veri	51	locat	54	veri	54
2	veri	51	veri	51	locat	44	veri	47	locat	50
3	apart	38	apart	38	apart	36	apart	44	apart	34
4	great	35	great	36	stay	28	great	35	great	34
5	stay	31	stay	32	great	28	stay	30	stay	32
6	place	29	place	30	clean	24	clean	27	place	30
7	clean	27	clean	28	recommend	24	place	26	host	27
8	recommend	25	recommend	25	place	23	good	23	clean	27
9	host	23	host	24	good	23	recommend	22	recommend	27
10	good	22	good	22	nice	21	nice	20	good	22
11	nice	21	nice	21	host	20	barcelona	17	nice	22
12	barcelona	20	barcelona	21	well	19	well	17	barcelona	21
13	everyth	17	everyth	17	everyth	18	host	16	room	19
14	well	17	well	16	perfect	16	everyth	16	everyth	18
15	room	16	room	16	barcelona	15	would	15	well	16
16	perfect	15	close	15	room	14	perfect	15	close	15
17	close	14	would	15	comfort	12	close	13	metro	15
18	would	14	metro	15	need	10	walk	13	perfect	15

19	comfort	14	walk	15	would	10	comfort	13	help	14
20	metro	14	perfect	15	close	10	metro	11	comfort	14
21	walk	13	comfort	14	thank	10	need	11	would	14
22	help	12	help	13	help	10	room	11	walk	14
23	need	11	need	11	realli	9	realli	10	friend	12
24	realli	10	realli	10	friend	9	love	9	need	12
25	friend	10	friend	10	super	9	time	9	thank	11
26	thank	10	love	10	time	9	easi	9	realli	11
27	time	10	time	10	communic	8	help	9	time	11
28	love	10	thank	10	high	8	restaur	9	also	10
29	also	10	also	10	excel	8	also	9	love	10
30	citi	9	citi	10	also	7	excel	8	citi	10
31	excel	9	restaur	9	love	7	communic	8	can	9
32	high	9	excel	9	walk	7	thank	8	excel	9
33	restaur	9	easi	9	metro	7	high	8	ha	9
34	easi	8	high	9	beauti	7	area	8	high	9
35	communic	8	area	8	accommod	7	can	8	restaur	9
36	area	8	communic	8	area	6	citi	8	super	8
37	super	8	definit	8	citi	6	ha	8	easi	8
38	definit	7	get	8	quiet	6	bed	7	area	8
39	get	7	super	7	amaz	6	spacious	7	welcom	8
40	lot	7	lot	7	day	6	definit	7	communic	8

Ad una prima analisi si possono individuare tre macro-concetti relativi alla posizione, all'host e alla pulizia. Il primo è rappresentato direttamente da attributi come "locat" e "area" e da altri come "walk" e "metro", si nota come invece la parola "car" non compaia in queste prime parole questo dovrebbe implicare che un maggior numero di ospiti preferisca spostarsi con i mezzi pubblici o camminando, considerazioni che sembrano confermare lo studio di Cleantech "Enviromtal Impact of Home-Sharing" per cui chi sceglie Airbnb è più propenso a muoversi con i mezzi pubblici o a piedi. L'host invece è descritto un elevato numero di volte con gli aggettivi "help" e "friend" (prima di normalizzare si notava che la parola friendly occupava le prime posizioni). Infine, emerge l'attenzione degli ospiti alla pulizia, che in media infatti occupa la sesta posizione, in particolare sembra che le viaggiatrici prestino più attenzione a questo parametro citandolo nel 31% dei commenti rispetto agli uomini in cui compare nel 24% delle recensioni, risultato opposto rispetto allo

studio di Sánchez-Franco & Alonso-Dos-Santos (2020). Si può notare inoltre come i recensori siano propensi a raccomandare agli altri la propria esperienza come si può notare dal posizionamento di “recommend” e a ringraziare il proprietario per l’ospitalità (“thank”). Interessante risulta invece la parola “restaur” che potrebbe essere legata al fatto che i turisti essendo in vacanza apprezzino andare a mangiare fuori. La comunicazione con i proprietari come indicato da “communic” si rivela molto importante per l’ospite. Il termine “veri” che compare tra le prime due posizioni in tutte le suddivisioni e “realli” sembrano confermare l’ipotesi che siano più propensi a rilasciare un commento coloro che sono rimasti molto soddisfatti o molto delusi dall’esperienza.

L’elevata percentuale del 95% di commenti positivi fa sì che le prime 40 parole per i commenti positivi siano praticamente le stesse di quelle generali, mentre rivela più informazioni ciò che emerge dai commenti negativi. Le principali problematiche riscontrate sembrano essere legate alla pulizia (“clean” e “dirti”), alla dimensione dell’alloggio (“small”, “littl”), ai servizi disponibili (“bathroom”, “kitchen”, “shower”, “bed”, “water”, “door”) e alla notte (“night”, probabilmente a causa della sicurezza o del rumore).

Nel complesso, sembra che il lessico del recensore non sia influenzato né dal proprio sesso, né dal tipo di proprietario (mono o multiproprietario) da cui soggiorna, né dal periodo in cui è avvenuto; infatti, queste colonne sono quasi identiche.

In conclusione, quest’analisi ha evidenziato che in generale coloro che hanno soggiornato a Barcellona sono rimasti soddisfatti vista la presenza tra le prime 40 parole più utilizzate di termini come “great”, “good”, “well”, “perfect”, “comfort”, “excel”, “easi”, “super”, “love” e “beauti”, e l’assenza di parole negative. Inoltre, si può vedere come le recensioni siano strutturate su 6 macro-argomenti che risultano analoghi al modello di recensione suggerito dalla stessa piattaforma: la posizione, i servizi dell’appartamento, l’esperienza con l’host, la pulizia della casa, la raccomandazione del soggiorno e il ringraziamento del servizio.

## 6 DIFFERENZE COMPORTAMENTALI E STRUTTURALI SIGNIFICATIVE

Nel seguente capitolo si analizza l'esistenza di differenze comportamentali a seconda del genere degli host e degli ospiti, che l'host sia o meno un professionista e che si tratti di un appartamento privato o condiviso. Per valutare se queste differenze siano significative si è scelto di applicare il test di ipotesi, considerando come ipotesi nulla ( $H_0$ ) che la differenza delle medie dei parametri dei due gruppi (A e B) sia uguale a zero, ossia che le due medie siano uguali, mentre come ipotesi alternativa ( $H_1$ ) che la differenza sia diversa da zero, quindi che le media siano diverse, come mostrato nell'equazione 6.1.

*Equazione 6.1*

$$\begin{cases} H_0: \mu_A - \mu_B = 0 \\ H_1: \mu_A - \mu_B \neq 0 \end{cases}$$

Il lavoro è strutturato in due modi: nel caso del genere si è applicato direttamente il comando “ttest” di Stata, mentre per gli altri si è scelto di scaricare le medie, le varianze e la numerosità dei campioni dei due gruppi e applicare la formula per ricavare il parametro t.

Con il programma Stata è possibile effettuare automaticamente questo test digitando:

```
ttest nome_variabile, by(variabile_di_gruppo)
```

o

```
ttest nome_variabile1= nome_variabile2
```

Il programma restituisce come output una tabella contenente i valori di t e dell'intervallo di confidenza del 95%.

Ci sono due possibilità per valutare se rifiutare  $H_0$ :

- il valore di t (che è approssimabile con z nel caso di grandi campioni) soddisfa la disequazione:  $|Z_{test}| > z_{1-\frac{\alpha}{2}}$
- nell'intervallo di confidenza non è presente lo 0

$$- p_{value} < \alpha$$

Diversamente si è svolta l'equazione 6.2 trovando  $Z_{test}$  e lo si è confrontato con i valori teorici presenti sulle tavole con  $\alpha$  pari a 1%, 5% e 10% per attribuirne la significatività, rispettivamente se la disequazione  $|Z_{test}| > z_{1-\frac{\alpha}{2}}$  con  $\alpha=1\%$  è presente \*\*\*, con  $\alpha=5\%$  \*\*, con  $\alpha=10\%$  \* nel caso restante non è presente nessun \*.

Vista l'elevata dimensione dei campioni ( $N \gg 30$ ) è possibile approssimare la distribuzione t-student a una normale, calcolando quindi la statistica:

Equazione 6.2

$$Z_{test} = \frac{(\bar{x}_A - \bar{x}_B) - d}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} \simeq \frac{(\bar{x}_A - \bar{x}_B) - d}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}}$$

Si utilizza la formula di sinistra in caso di varianza nota, diversamente quella di destra.

## 6.1 Esiste una differenza di genere nel mercato Airbnb?

Nel seguente sottocapitolo si valuta se emergono delle significative differenze comportamentali legate al genere degli host e degli ospiti. Per farlo si sono create due variabili dummy “fem\_h” e “fem\_r” che se uguali a 1 indicano rispettivamente host femmina e recensore femmina, diversamente se il genere è maschile valgono 0. Il lavoro è strutturato applicando il comando “ttest” di Stata, per valutare l'esistenza di una differenza statisticamente significativa tra maschio e femmina, prima ai dati relativi alle recensioni (numero parole, polarità e soggettività), poi ai listings (prezzo, numero recensioni, disponibilità, super\_host, numero listing, punteggi) e infine all'unione dei due per la confidenza con gli host e per valutare se la loro scelta di prenotazione fosse influenzata dal genere dell'host.

### 6.1.1 C'è una differenza tra i recensori?

Di seguito sono descritti in modo dettagliati i primi risultati ottenuti dal ttest, negli output successivi saranno solo indicati i valori riassuntivi per una migliore visualizzazione. Scritto l'input "ttest n\_r\_f=n\_r\_m" in cui si considera come H0 l'ipotesi che il numero di recensori maschili e femminili sia uguale, mentre H1 indica che sono diversi, il programma Stata restituisce come output:

```
Paired t test
-----+-----
Variable |      Obs      Mean   Std. Err.   Std. Dev.   [95% Conf. Interval]
-----+-----
n_r_f | 184191   .5102638   .0011648   .499896   .5079809   .5125468
n_r_m | 184191   .4897362   .0011648   .499896   .4874532   .4920191
-----+-----
diff | 184191   .0205276   .0023296   .999792   .0159617   .0250935
-----+-----
      mean(diff) = mean(n_r_f - n_r_m)                t =      8.8118
Ho: mean(diff) = 0                                degrees of freedom = 184190

Ha: mean(diff) < 0                Ha: mean(diff) != 0                Ha: mean(diff) > 0
Pr(T < t) = 1.0000                Pr(|T| > |t|) = 0.0000                Pr(T > t) = 0.0000
```

Si nota che considerando un intervallo di confidenza del 95% ( $\alpha=5\%$ ) si rifiuta H0 in quanto:

- $t=8,8118 > z_{97,5\%}=1,96$
- l'intervallo di confidenza  $[0,0159617;0,0250935]$  non comprende lo 0
- $p_{value} = 0 < \alpha = 0,05$

Questo significa che la differenza del numero dei recensori maschili e femminili è significativa: il numero di recensori femmine (51%) risulta prevalente, questo può essere una conseguenza che le donne siano più propense a rilasciare un commento dopo il soggiorno o può essere legato al fatto che più donne utilizzano Airbnb in quanto, ad esempio, sono loro ad occuparsi della prenotazione degli alloggi in caso di viaggi tra più persone.

### 6.1.2 C'è differenza nelle recensioni rilasciate?

Analizzando le recensioni rilasciate, come emerge da altri studi (Sánchez-Franco & Alonso-Dos-Santos, 2020), le donne e gli uomini si comportano e valutano diversamente le

esperienze, si nota infatti che le donne tendono a scrivere dei feedback più lunghi fornendo quindi un maggior numero di informazioni agli altri utenti della piattaforma: questo emerge considerando la variabile conteggio che indica il numero di parole presenti nel commento.

Le donne risultano inoltre più critiche come si nota dalla *t* positiva e pari a 19,515 per il test sulla polarità in cui tale parametro è compreso tra -1 e +1 (da molto negativo a molto positivo). In particolare, come sarà descritto in seguito, entrambi i generi dei recensori rilasciano commenti meno positivi nei confronti degli appartamenti di host donne.

Non si può invece rifiutare  $H_0$  che i valori della soggettività siano uguali nei due generi.

Mentre le donne rispetto alla controparte comunicano un maggiore grado di confidenza con l'host, chiamandolo più volte per nome come se fossero riuscite a creare un legame maggiore, per calcolarlo si è creata una variabile dummy *c\_host* che assume valore 1 nel caso in cui nel commento compaia il nome proprio dell'host.

Dal momento che l'output generato da Stata come mostrato in precedenza ha un elevato numero di informazioni che non servono al fine della seguente analisi e per una migliore lettura sono stati sintetizzati nella tabella 6.1 i dati dei test effettuati con le variabili della prima colonna attraverso il "`ttest VARIABILE, by(fem_r)`", si sono così indicate le informazioni relative alle medie dei due gruppi (maschi e femmine), la *t*, l'intervallo di confidenza al 95% e il *p\_value*.

Tabella 6.1 Differenza nelle recensioni rilasciate

VARIABILE	$\mu_{MAS}$	$\mu_{FEMM}$	<i>t</i>	Intervallo(95%)	<i>p_value</i>
conteggio	36,568	46,382	-45,994	(-10,231;-9,395)	0
polarità	0,445	0,434	19,515	(-0,192;0,023)	0
soggettività	0,635	0,636	-1,100	(-0,002;0,007)	0,27
<i>c_host</i>	0,264	0,332	-31,307	(-0,071;-0,063)	0

Per verificare se i recensori fossero più critici nei confronti di appartamenti di host dello stesso sesso si è proceduto svolgendo prima un test di ipotesi i cui risultati sono riportanti in seguito e successivamente svolgendo una regressione come è presentato nel capitolo successivo. Si sono quindi eseguiti i comandi "`ttest polarità if(fem_r==1),`

`by(fem_h)`” e `“ttest polatità if(fem_r==0), by(fem_h)”` ottenendo come t rispettivamente i valori 3,0445 e 3,6414, questi indicano che i recensori di entrambi i sessi rilasciano commenti più critici nei confronti delle donne. Inoltre, con i comandi `“ttest polatità if(fem_h==1), by(fem_r)”` e `“ttest polatità if(fem_h==0), by(fem_r)”` si sono ottenuti rispettivamente 9,8662 e 11,1784 che significa che sono le donne a scrivere recensioni più critiche per entrambe le tipologie di host.

### 6.1.3 C'è una differenza nei possibili ricavi?

Se si considera la domanda degli alloggi come il numero di recensioni rilasciate, sebbene non sia molto corretto in quanto non è presente un'indicazione sulla durata del soggiorno e non sono considerati coloro che non rilasciano una recensione nonostante abbiano alloggiato nell'appartamento, si può vedere che emerge una differenza significativa tra la variabile “number\_of\_reviews” agli annunci maschili e a quelli femminili. Questo significa che i listing maschili catturano più domanda probabilmente grazie alla loro maggiore disponibilità e al numero più elevato di annunci con il titolo di superhost, nonostante il loro prezzo che risulta più elevato. I risultati ottenuti sono presentati nella seguente tabella 6.2 eseguendo `“ttest VARIABILE, by(fem_h)”` e indicano che i ricavi delle host femmine possano essere considerati inferiori rispetto alla controparte sia per la minore domanda che per il minor prezzo degli annunci; infatti, nonostante il prezzo più elevato i listing maschili riescono a catturare una domanda superiore.

Tabella 6.2 Differenza nei ricavi

VARIABILE	$\mu_{MAS}$	$\mu_{FEMM}$	t	Intervallo(95%)	p_value
number_of_reviews	37,31	34,67	1,9815	(0,028;5,234)	0,05
availability_365	157,116	151,567	2,1639	(0,522;10,575)	0
s_host	0,2	0,163	5,096	(0,023;0,052)	0
price	90,427	85,09	2,556	(1,244;9,429)	0

### 6.1.4 C'è una differenza negli host?

Lo stesso sito sottolinea che “le donne sono la forza trainante di Airbnb” in quanto le host

femmine rappresentano il 54% degli host della comunità, si è valutato se questo valore fosse conforme con i risultati ottenuti nella città di Barcellona. Nel caso in analisi, trascurando i parametri a cui non è stato possibile attribuire un genere, tale percentuale è del 52% e la differenza tra il numero di host maschili e femminili è significativa. Inoltre, il numero dei listing che possiede un host non risulta influenzato dal genere, mentre le host femmine risultano da più tempo sulla piattaforma.

Tabella 6.3 Differenza tra gli host

VARIABILE	$\mu_{MAS}$	$\mu_{FEMM}$	t	Intervallo(95%)	p_value
n_h_m=n_h_f	0,478	0,522	-3,669	(-0,069;-0,021)	0
n_listing	1,733	1,617	1,42	(-0,044;0,276)	0,142
anni_host	5,141	5,303	-2,555	(-0,286;-0,038)	0

#### 6.1.5 C'è una differenza nei feedback degli annunci?

Per valutare invece se l'attenzione e la precisione degli host varia a seconda del genere si analizza la differenza nei punteggi di review\_score (`ttest Review_scores_`, `by (fem_h)`); ciò che emerge è che non è presente una differenza significativa nei punteggi complessivi, ma analizzando i singoli parametri si nota che le host donne ottengono punteggi statisticamente più elevati nella pulizia e nell'accuratezza.

Tabella 6.4 Differenza nei feedback degli annunci

Review_scores_	$\mu_{MAS}$	$\mu_{FEMM}$	t	Intervallo(95%)	p_value
Rating	4,512	4,509	0,159	(-0,033;0,038)	0,87
Accuracy	4,674	4,697	-2,235	(-0,048;-0,003)	0,03
Cleanliness	4,563	4,635	-5,552	(-0,097;-0,046)	0
Checkin	4,774	4,778	-0,397	(-0,024;0,016)	0,70
Communication	4,773	4,766	0,658	(-0,014;0,027)	0,51
Location	4,768	4,763	0,59	(-0,012;0,023)	0,56
Value	4,545	4,565	-1,685	(-0,043;0,003)	0,09

### 6.1.6 La scelta dell'host è influenzata dal genere?

Per analizzare se il genere dei recensori influenzi la scelta dell'host si sono svolti due esperimenti: nel primo caso si è fatto un test d'ipotesi, come seconda cosa invece si è scelto di utilizzare una funzione probit. Di seguito è riportata la prima analisi in cui si sono ottenuti i risultati riportati in tabella che confermano tale influenza. In particolare, si può notare che gli uomini siano più propensi a prenotare un appartamento da altri uomini, mentre per le donne questa preferenza ad interagire con host dello stesso sesso non emerge.

```
ttest fem_r, by (fem_h)
```

$\mu_{MAS}$	$\mu_{FEMM}$	t	Intervallo(95%)	p_value
0,504	0,522	-6,267	(-0,023;0,012)	0

Tabella 6.5 Distribuzione recensioni genere recensori e host

genere_recensore:host	n. recensioni
rec_femm:host_femm	32130
rec_masc:host_femm	29414
rec_femm:host_masc	32858
rec_masc:host_masc	32277

La preferenza ad interagire con utenti dello stesso sesso dovrebbe essere maggiore se si tratta di viaggiatori soli in quanto in questo caso l'interazione diretta con l'host è più esposta come rivela lo studio di Choi & Horvát del 2019 "Airbnb's reputation system and gender difference among guest: Evidence from large-scale data analysis and a controlled experiment". Tra i dati disponibili di Inside Airbnb però non è presente il numero di persone che hanno effettuato il soggiorno, ma per ogni listing è indicato il numero di ospiti ("accommodates") che l'alloggio può accogliere. Gli autori avevano eliminato tutti le recensioni che i termini come "we", "us" e "our" che fanno intendere di aver soggiornato in più viaggiatori. Nella seguente analisi invece si sono filtrati i listing con accommodates pari a 1 ottenendo i seguenti valori:

```
ttest fem_r if(accommodates==1), by (fem_h)
```

$\mu_{MAS}$	$\mu_{FEMM}$	t	Intervallo(95%)	p_value
0,427	0,493	-7,211	(-0,083;0,047)	0

Tabella 6.6 Risultati esperimento viaggiatori solitari

genere_recensore:host	conteggio
rec_femm:host_femm	3395
rec_masc:host_femm	3493
rec_femm:host_masc	2275
rec_masc:host_masc	3048

I risultati mostrano un maggiore numero di recensori maschili rispetto all'altro genere in accordo con Choi & Horvát (2019), come possibile spiegazione del fenomeno essi avevano segnalato che le donne che viaggiano da sole preferiscono soggiornare in hotel per ragioni principalmente dovute alla sicurezza o diversamente sono più propense ad affidarsi ad host dello stesso sesso. Per gli uomini invece questa inclinazione non compare, infatti il numero di prenotazioni in alloggi femminili risulta più elevato. Inoltre, anche in questo caso, l'output di Stata fa rifiutare  $H_0$ , confermando le conclusioni di Choi & Horvát che indicano che la scelta dell'host è influenzata dal genere dell'ospite.

## 6.2 Esistono differenze tra host amatoriali e professionisti

Ad una prima osservazione dei database disponibili si è notato che alcuni host avevano più listing nella città, in particolare la colonna "host\_total\_listing" avrebbe dovuto indicare esattamente quanti annunci possedeva ogni host, ma osservando attentamente i numeri presenti è emerso che molte volte erano incongruenti, in quanto ad esempio alcuni avevano valore nullo ma ciò non aveva senso in quanto avrebbe dovuto essere pari o superiore a 1 dal momento che compariva nel database oppure in alcuni era presente il punteggio 1 ma nella realtà esso avrebbe dovuto valore 52. Per questo motivo si è scelto di inserire una

nuova variabile “n\_listing” che indica il numero di listing per un “host\_id” presenti nella città di Barcellona.

Osservando i parametri inseriti si è notato che il 73% degli host presenti nella città possiede un unico annuncio, mentre altri anche più di 100, come è mostrato dal grafico.

Grafico 6.1 Distribuzione numero listing



Si è quindi ipotizzato che se un host possiede più appartamenti potrebbe avere un comportamento diverso rispetto a uno che ne possiede solo uno, in quanto potrebbe farlo in maniera professionale e gli stessi recensori potrebbero esprimere giudizi più severi superando l’ostacolo dovuto all’incontro con il padrone di casa, il maggior spirito di adattamento e le minori aspettative del cliente rispetto ad un soggiorno in hotel. Per valutare tale comportamento si è scelto di introdurre una variabile dummy “professionista” che distinguesse questi host da quelli amatoriali, è stato però studiato quale potesse essere un valore soglia per suddividere queste tipologie. A livello logico si è pensato che anche un host amatoriale potesse gestire un paio di annunci, mentre può essere più complicato gestirne di più per questioni di tempo e attenzione richiesti, questo ragionamento è stato confermato da alcuni test di ipotesi svolti confrontando tra loro chi ha un unico annuncio, chi ne possiede più di uno, chi ne ha un numero maggiore o uguale a cinque e chi tra uno e cinque esclusi. Si è utilizzato 5 in quanto si è preferito scegliere un numero inferiore rispetto a 10 usato da Deboosere, Kerrigan, Wachsmuth, & El-Geneidy (2019) e Boto-García, Mayor, & De la Vega (2021) perché si è valutato che tale comportamento intervenisse già per numeri inferiori.

Con i risultati ottenuti e con le considerazioni descritte precedentemente si è scelto di considerare 5 come limite; quindi, si è introdotta una dummy “professionista” che assumesse valore 1 in caso di host con più di 5 listing o classificati come agenzie.

Tabella 6.7 Riassunto statistiche host amatoriali e professionisti

VARIABILI	PROFESSIONISTI			AMATORIALI			DIFF	T	Signific
	N	mean	sd	N	mean	sd			
host_response_rate	523	90,41	19,51	3911	80,49	34,49	9,92	9,77	***
host_acceptance_rate	519	87,99	22,18	3559	77,93	33,92	10,06	8,92	***
anni_host	595	4,96	2,99	7493	5,18	2,57	-0,22	-1,73	*
response_fast	523	0,63	0,49	3911	0,48	0,50	0,15	6,49	***
s_host	595	0,17	0,38	7493	0,18	0,39	-0,01	-0,81	
host_id_verified	595	0,86	0,34	7493	0,63	0,48	0,24	15,60	***
host_has_pic	595	1,00	0,06	7493	1,00	0,06	0,00	0,00	
price	7100	138,10	130,90	9690	78,18	101,10	59,92	32,18	***
availability_365	7100	199,60	120,50	9690	141,80	137,20	57,80	28,94	***
number_of_reviews	7100	23,70	49,97	9690	39,42	74,11	-15,72	-16,40	***
review_scores_rating	4613	4,44	0,68	7176	4,52	0,84	-0,08	-5,84	***
review_scores_accuracy	4582	4,58	0,57	7012	4,71	0,49	-0,12	-11,98	***
review_scores_cleanliness	4582	4,55	0,57	7015	4,62	0,56	-0,07	-6,37	***
review_scores_checkin	4580	4,66	0,51	7011	4,80	0,42	-0,14	-15,46	***
review_scores_communication	4583	4,67	0,53	7013	4,79	0,44	-0,13	-13,42	***
review_scores_location	4580	4,75	0,39	7010	4,77	0,39	-0,01	-2,03	**
review_scores_value	4580	4,41	0,57	7009	4,59	0,50	-0,18	-17,46	***
alloggio_privato	7100	0,82	0,38	9690	0,31	0,46	0,51	78,23	***
has_instant_bookable	7100	0,52	0,50	9690	0,38	0,49	0,14	18,56	***
Inrec	4613	2,42	1,67	7176	2,73	1,77	-0,32	-9,82	***
Inprice	7100	4,61	0,82	9690	4,01	0,75	0,60	48,41	***

Confrontando i padroni di casa “amatoriali” con i “professionisti” emerge che i punteggi dei singoli parametri, sebbene siano in entrambi i casi molto alti, sono statisticamente più elevati per i primi in tutti i casi; ciò sembra confermare una possibile inclinazione degli ospiti a sentirsi più vincolati nel rilascio delle recensioni maggiori nei confronti di host amatoriali, mentre tale preoccupazione è in parte attenuata con padroni di casa più professionali.

Dall'altra parte i professionisti hanno più listing con la prenotazione istantanea, questo può confermare che gli host amatoriali siano più attenti alle persone che devono ospitare per cui preferiscono scegliere di confermare o meno la richiesta di prenotazione.

Diversamente da quanto ci si sarebbe aspettati, si può constatare che l'attribuzione del titolo di superhost non è statisticamente significativa: questo è in disaccordo con quanto indicato nella letteratura (Gunter, 2018) dal momento che questo badge richiede molto impegno sia per acquisirlo che per mantenerlo, perché bisogna sottostare a degli appositi target molto severi, quindi se un host lo fa di mestiere ha più tempo e attenzione da dedicarci come confermato dal loro più veloce tempo di risposta (si è creata una dummy "response\_fast" che assume valore pari a 1 se l'host risponde entro l'ora), una possibile causa legata a questo è che sebbene per gli host amatoriali sia più complicato questo titolo sono consapevoli dell'importanza che riveste sulla piattaforma e quindi incentivati ad ottenerlo.

L'analisi inoltre permette di osservare che il numero di recensioni rilasciate nei listing di host amatoriali siano più elevate, questo può essere tradotto come una maggiore domanda e può essere dovuto sia dal fatto che essi siano presenti sulla piattaforma da più tempo che dal prezzo significativamente più basso dei loro soggiorni, nonostante la minore disponibilità dell'appartamento probabilmente legato al fatto che in determinati periodi dell'anno l'host amatoriale non riuscirebbe a seguire l'attività e quindi non lo mette a disposizione, oppure è una sistemazione che affitta solo nei periodi in cui non la utilizza lui stesso direttamente.

### *6.2.1 Relazione genere ospite e host professionista*

Il genere degli ospiti influenza la loro scelta di acquisto, tenendo in considerazione il fatto che il numero delle recensioni degli alloggi di host amatoriali è quasi il doppio di quelli professionali, si può notare che gli ospiti maschi siano più propensi delle donne ad affittare da professionisti, questo può confermare quanto riportato da Santos, Mota, Benevenuto, & Silva (2020) in cui si presenta che dai loro soggiorni gli uomini prestano maggiore attenzione ai servizi disponibili quasi come se si aspettassero un'esperienza analoga agli hotel, mentre le donne un'esperienza locale diretta e quindi prediligono i padroni di casa amatoriali.

*ttest professionista, by (fem\_r)*

$\mu_{MAS}$	$\mu_{FEMM}$	t	Intervallo(95%)	p_value
0,369	0,346	10,569	(0,019;0,028)	0

*Tabella 6.8 Genere recensore e tipologia di host*

genere_recensore:host	n. recensioni
recensore_femminile:host_professionista	31873
recensore_maschile:host_professionista	32737
recensore_femminile:host_amatoriale	60332
recensore_maschile:host_amatoriale	55862

### 6.3 Esistono delle differenze significative tra gli alloggi condivisi e quelli privati

Il core business di Airbnb è l'affitto di alloggi interi o camere, anche se negli ultimi anni si è data la possibilità anche agli hotel di caricare le proprie camere, ma queste sono state escluse dall'analisi in quanto sono differenti sotto numerosi aspetti e quindi avrebbero potuto compromettere lo studio. Il fatto però che l'annuncio riguardi un alloggio intero ("entire home/apt") o un letto in un appartamento condiviso (sia esso di tipo "private room" o "shared room") può avere effetti dissimili sulle variabili, per questo motivo si è svolto un test di ipotesi sulle caratteristiche più significative per individuarne il comportamento, ottenendo la seguente tabella.

*Tabella 6.9 Riassunto statistiche appartamenti privati e condivisi*

VARIABILI	sharing			privato			t_test		
	N	mean	sd	N	mean	sd	differenza	t	sign
price	7916	56,28	74,13	8874	145,60	133,60	-89,32	-54,30	***
lnprice	7916	3,75	0,65	8874	4,73	0,70	-0,97	-93,61	***
accommodates	7916	1,90	1,07	8874	4,69	2,17	-2,79	-107,58	***
bedrooms	7709	1,08	0,39	8511	2,21	1,09	-1,13	-89,43	***
beds	7606	1,29	1,04	8855	3,17	2,05	-1,88	-75,71	***

bathroom_private	7902	0,34	0,48	8872	1,00	0,00	-0,66	-123,14	***
wifi	7916	0,97	0,18	8874	0,98	0,14	-0,01	-5,67	***
tv	7916	0,54	0,50	8874	0,95	0,23	-0,41	-66,91	***
washer	7916	0,69	0,46	8874	0,93	0,26	-0,24	-40,61	***
airconditioning	7916	0,37	0,48	8874	0,90	0,30	-0,53	-84,73	***
pool	7916	0,02	0,13	8874	0,05	0,22	-0,03	-11,63	***
gym	7916	0,01	0,12	8874	0,02	0,13	0,00	-1,39	
beachfront	7916	0,02	0,13	8874	0,02	0,16	-0,01	-3,32	***
sauna	7916	0,00	0,00	8874	0,00	0,05	0,00	-4,47	***
parcheggio_gratuito	7916	0,12	0,32	8874	0,08	0,27	0,03	7,20	***
minimum_nights	7916	8,79	30,64	8874	15,48	27,01	-6,70	-14,94	***
maximum_nights	7916	662,30	523,20	8874	692,60	469,70	-30,30	-3,93	***
distanza_centro	7916	2,11	1,34	8874	1,79	1,08	0,33	17,14	***
ciutat_vella	7916	0,23	0,42	8874	0,23	0,42	0,00	-0,15	
gracia	7916	0,09	0,28	8874	0,09	0,28	0,00	-0,58	
horta_guinardo	7916	0,05	0,21	8874	0,02	0,15	0,02	8,42	***
les_corts	7916	0,02	0,16	8874	0,02	0,14	0,01	2,28	**
nou_barris	7916	0,03	0,16	8874	0,00	0,06	0,02	11,73	***
sant_andreu	7916	0,02	0,15	8874	0,01	0,09	0,02	7,95	***
sant_marti	7916	0,10	0,30	8874	0,09	0,29	0,01	1,96	*
sants_montjuic	7916	0,13	0,34	8874	0,10	0,31	0,03	5,80	***
sarria_sant_gervasi	7916	0,04	0,19	8874	0,05	0,22	-0,01	-3,96	***
n_listing	7916	4,86	13,91	8874	29,97	44,46	-25,12	-50,51	***
professionista	7916	0,16	0,37	8874	0,66	0,47	-0,50	-77,17	***
s_host	7909	0,14	0,35	8832	0,18	0,39	-0,04	-6,87	***
host_id_verified	7909	0,61	0,49	8832	0,90	0,30	-0,29	-46,39	***
host_has_pic	7909	1,00	0,07	8832	1,00	0,03	0,00	-3,70	***
anni_host	7909	4,85	2,56	8832	5,86	2,98	-1,01	-23,49	***
fem_h	6433	0,53	0,50	4672	0,48	0,50	0,05	5,10	***
has_instant_bookable	7916	0,40	0,49	8874	0,48	0,50	-0,09	-11,26	***
availability_365	7916	138,70	143,00	8874	190,90	119,10	-52,20	-25,53	***
host_response_rate	4212	78,24	34,68	7904	92,13	16,87	-13,89	-24,50	***
host_acceptance_rate	3781	78,93	34,04	7840	85,62	22,26	-6,69	-11,00	***
lnrec	5459	2,43	1,67	6330	2,77	1,77	-0,34	-10,77	***
number_of_reviews	7916	25,96	55,44	8874	38,85	72,73	-12,89	-12,99	***
review_scores_rating	5459	4,46	0,92	6330	4,52	0,63	-0,05	-3,51	***
review_scores_accuracy	5306	4,67	0,56	6288	4,65	0,50	0,02	1,52	
review_scores_cleanliness	5309	4,57	0,62	6288	4,60	0,51	-0,03	-3,11	***
review_scores_checkin	5304	4,78	0,47	6287	4,72	0,46	0,06	7,14	***
review_scores_communication	5307	4,76	0,49	6289	4,73	0,47	0,04	3,90	***
review_scores_location	5303	4,76	0,41	6287	4,76	0,36	-0,01	-0,82	
review_scores_value	5302	4,58	0,56	6287	4,47	0,52	0,11	10,88	***

Come atteso gli alloggi privati sono più grandi, infatti hanno un numero maggiore di letti,

camere e ospitano più persone, hanno un bagno privato, possiedono in percentuale più servizi ad eccezione della palestra e del parcheggio privato, mentre sono generalmente più vicino al centro. Questo posizionamento può essere una conseguenza del fatto che le case più centrali sono destinate ai turisti in quanto più caotiche e costose, mentre in quelle condivise spesso l'host ci vive; quindi, l'alloggio è situato nei quartieri più residenziali. Inoltre, anche tutti i parametri relativi all'host, alle recensioni e ai punteggi risultano "migliori" per gli alloggi privati, infine le host femminili hanno più listing di alloggi condivisi.

## 7 ANALISI REGRESSIONE

Nel seguente capitolo si definiscono dei modelli di regressione multipla in cui si analizzano quali siano i fattori ad influenzare il prezzo di un listing: utilizzando la teoria del prezzo edonico che indica che il prezzo di un prodotto può essere collegato a una funzione che misura l'utilità degli attributi di prodotti eterogenei. Infatti, la qualità di un alloggio è rappresentata da un insieme di caratteristiche di un annuncio di Airbnb, per cui la presenza o meno di alcuni specifici attributi può contribuire alla definizione del suo prezzo. Il capitolo si struttura nel seguente modo: si inizia con un riassunto dei concetti teorici utilizzati, una successiva spiegazione delle variabili considerate, suddivise in dipendenti e indipendenti, e una finale individuazione delle possibili regressioni sotto condizioni differenti.

### 7.1 Concetti teorici

Dal momento che il prezzo è generalmente calcolato considerando più fattori si è utilizzato un modello di regressione multipla che permette di mettere in relazione la variabile dipendente (prezzo) con diverse variabili indipendenti chiamati predittori. L'obiettivo di queste operazioni è quello di riuscire a ricavare la variabile dipendente dati alcuni valori delle variabili indipendenti.

Il modello può essere rappresentato dalla seguente equazione lineare:

*Equazione 7.1*

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u$$

Dove  $u$  rappresenta l'errore che si verifica perché alcune variabili (che prendono il nome di variabili omesse) sebbene influenzino  $Y$  non sono state inserite nel modello.

L'analisi permetterà di stimare i coefficienti  $\beta$ , l'errore  $u$  e l'intercetta  $\beta_0$  considerando un campione di  $n$  osservazioni.

Per trovare questi parametri si utilizza il "metodo dei minimi quadrati" (OLS) che calcola la funzione che rende minima la somma dei quadrati delle distanze tra i dati disponibili

della variabile dipendente e quelli derivati dalla funzione stessa.

Equazione 7.2

$$\min_{b_0, b_1, \dots, b_n} \sum_{i=1}^n [y_i - (b_0 + b_1 x_{1i} + \dots + b_n x_{ni})]^2$$

Le assunzioni di questo metodo sono:

- 1) La media della distribuzione di  $u$  condizionata alle  $X$  è nulla  
 $(E(u_i | X_{1i} = x_1, \dots, X_{ki} = x_k) = 0$
- 2) Le variabili  $X_{1i}, X_{2i} \dots X_{ki}, Y_i$ , con  $i = 1, \dots, N$  sono indipendenti e identicamente distribuite
- 3) Gli outliers sono rari
- 4) Non c'è perfetta collinearità

Per svolgere l'analisi è stato utilizzato il programma Stata sfruttando la funzione: "`reg y x1 x2 xn, robust`". L'inserimento finale di `robust` serve affinché gli errori standard siano robusti all'eteroschedasticità.

In ogni regressione l'output restituisce  $R^2$  che è l'indice di determinazione lineare che rappresenta la frazione della varianza di  $Y$  che è spiegata dai regressori  $X$  utilizzati nel modello ed ha un valore compreso tra 0 (nessun adattamento) e 1 (adattamento perfetto).

Equazione 7.3

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS} = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sum_{t=1}^n (y_t - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n \hat{u}_i^2}{\sum_{t=1}^n (y_t - \bar{y})^2}$$

Questo indice però ha un problema: al crescere del numero delle variabili  $X$  tende ad aumentare, se tale indicatore aumenta non è quindi detto che l'aggiunta del nuovo regressore spieghi in modo significativo  $Y$ . Per tale motivo esso può essere un indice "per eccesso". Per correggere  $R^2$  si può ricorrere a  $R^2$  adj ( $R^2$  corretto), che sarà crescente se l'aggiunta di una variabile porta ad un aumento di  $R^2$  che supera la penalità legata alla

presenza di un regressore ulteriore.

*Equazione 7.4*

$$\overline{R^2} = 1 - \left( \frac{n-1}{n-k-1} \right) (1 - R^2)$$

Per vedere se la nuova variabile inserita è statisticamente significativa non è sufficiente un incremento dei parametri  $R^2_{adj}$  e  $R^2$ , ma si utilizza il test  $t$  di Student in cui si considera come ipotesi nulla  $H_0$  che  $\beta$  sia uguale a 0, mentre l'alternativa è che esso sia diverso da 0.

Si svolge poi il calcolo  $t = \frac{\beta}{\sqrt{var(\beta)}}$ . Si rifiuta  $H_0$  se, scelto un livello di significatività  $\alpha$ , la  $t$  calcolata è maggiore del valore che emerge dalle tavole statistiche:

*Equazione 7.5*

$$t \geq t_{\alpha, n-k-1}$$

Nelle analisi che saranno svolte successivamente il numero di campioni analizzati è molto elevato quindi si può approssimare a una distribuzione normale.

Una condizione della regressione multipla è che ogni  $X$  sia altamente correlata con  $Y$ , ma non con le altre variabili indipendenti, idealmente infatti i valori di correlazione tra le variabili indipendenti dovrebbero essere 0. Nella realtà però quest'ultima condizione solitamente non è valida, si parla così di multicollinearità. Questo fenomeno produce un effetto che limita la capacità di previsione di una singola variabile indipendente a seconda della sua correlazione con le altre variabili: infatti, gli effetti delle singole variabili indipendenti sono mescolati e confusi tra loro, compromettendo la "capacità di spiegazione" del modello e la stima dei parametri. Per valutare la presenza di multicollinearità nel modello dell'analisi si è utilizzato l'indicatore VIF (Fattore di Accrescimento della Varianza), esso è calcolato per ogni  $X$  del modello di regressione e corrisponde al rapporto tra la varianza totale e la varianza di un modello con quella variabile come unica variabile indipendente, bisogna prestare attenzione se tale valore risulta superiore a una soglia che in alcuni studi è considerata pari a 10 mentre in altri a 5 in quanto indicherebbe un'alta multicollinearità.

## 7.2 Variabili

Osservando attentamente il database Listing scaricato da InsideAirbnb si sono individuate le variabili non nulle che dovrebbero essere più rilevanti nell'analisi del prezzo, alcune di queste sono numeriche, mentre altre come i quartieri e la tipologia di alloggio sono testuali quindi è stato necessario trasformarle in dummy come sarà descritto nel dettaglio in seguito. Attraverso il comando “sum” di Stata si è ricavata la seguente tabella in cui sono rappresentate per ogni variabile una breve descrizione, il numero di osservazioni non nulle, la media, la deviazione standard, il valore minimo e il valore massimo, inoltre vicino al nome tra parentesi è riportato il segno che ci si aspetta di ottenere dalla regressione. Come si può notare in alcune variabili la presenza di valori nulli risulta molto elevata.

*Tabella 7.1 Statistiche descrittive delle variabili (N=numero, D=dummy, %=percentuale, €=valuta)*

VARIABILI	tipo	DESCRIZIONE	N	mean	sd	min	max
Price (dipendente)	€	Prezzo annuncio una notte	16790	103,50	118,40	8	1893,00
Lnprice(dipendente)	€	Ln(price)	16790	4,27	0,83	2,08	7,55
Accommodates(+)	N	Numero di ospiti	16790	3,38	2,23	1	16
Bedrooms(+)	N	Numero camere da letto	16220	1,67	1,01	1	16
beds(+)	N	Numeri letti	16461	2,30	1,91	0	40
bathroom_private(+)	D	=1 se ha il bagno privato	16774	0,69	0,46	0	1
alloggio_privato(+)	D	=1 se è un alloggio privato	16790	0,53	0,50	0	1
wifi(+)	D	=1 se tra i servizi c'è il wifi	16790	0,98	0,16	0	1
tv(+)	D	=1 se tra i servizi c'è la tv	16790	0,75	0,43	0	1
washer(+)	D	=1 se tra i servizi c'è la lavatrice	16790	0,82	0,39	0	1
airconditioning(+)	D	=1 se tra i servizi c'è il l'aria condizionata	16790	0,65	0,48	0	1
pool(+)	D	=1 se tra i servizi c'è la piscina	16790	0,03	0,18	0	1
gym(+)	D	=1 se tra i servizi c'è la palestra	16790	0,02	0,13	0	1
beachfront(+)	D	=1 se l'alloggio è di fronte alla spiaggia	16790	0,02	0,14	0	1
sauna(+)	D	=1 se tra i servizi c'è la sauna	16790	0,00	0,03	0	1
parcheggio_gratuito(+)	D	=1 se tra i servizi c'è il parcheggio gratuito	16790	0,10	0,30	0	1
minimum_nights(-)	N	Numero minimo notti per l'affitto	16790	12,32	28,97	1	1000,00
maximum_nights(-)	N	Numero massimo notti per l'affitto	16790	678,30	495,90	1	3000,00
distanza_centro(-)	N	Distanza da Plaça de Catalunya	16790	1,94	1,22	0,02	8,47

ciutat_vella	D	=1 se è nel quartiere Ciutat Vella	16790	0,23	0,42	0	1
gracia	D	=1 se è nel quartiere Gràcia	16790	0,09	0,28	0	1
horta_guinardo	D	=1 se è nel quartiere Horta-Guinardò	16790	0,03	0,18	0	1
les_corts	D	=1 se è nel quartiere Les Corts	16790	0,02	0,15	0	1
nou_barris	D	=1 se è nel quartiere Nou Barris	16790	0,01	0,12	0	1
sant_andreu	D	=1 se è nel quartiere Sant Andreu	16790	0,02	0,12	0	1
sant_marti	D	=1 se è nel quartiere Sant Martí	16790	0,10	0,29	0	1
sants_montjuic	D	=1 se è nel quartiere Sants_Montjuïc	16790	0,12	0,32	0	1
sarria_sant_gervasi	D	=1 se è nel quartiere Sarrià-Sant Gervasi	16790	0,04	0,20	0	1
n_listing(+/-)	N	Numero di listing posseduti dall'host a Barcellona	16790	18,13	35,96	1	174
Professionista(+/-)	D	=1 se l'host ha più di 5 annunci nella città	16790	0,42	0,49	0	1
s_host(+)	D	=1 se l'host ha il titolo di superhost	16741	0,16	0,37	0	1
host_id_verified(+)	D	=1 se l'host ha l'identità verificata	16741	0,76	0,43	0	1
host_has_pic(+)	D	=1 se l'host ha l'immagine	16741	1,00	0,05	0	1
anni_host(+)	N	Numero di anni in cui l'host è sulla piattaforma	16741	5,38	2,83	0	13
fem_h	D	=1 se l'host è femmina, 0 nel caso in cui sia maschio, nullo se non è definito	11105	0,51	0,50	0	1
has_instant_bookable(+)	D	=1 se il listing ha la prenotazione istantanea	16790	0,44	0,50	0	1
availability_365(+)	N	Numero di giorni in un anno in cui è disponibile	16790	166,30	133,50	0	365
host_response_rate	%	Tasso di risposta dell'host	12116	87,30	25,44	0	100
host_acceptance_rate	%	Tasso di accettazione dell'host	11621	83,44	26,85	0	100
number_of_reviews(+)	N	Numero recensioni rilasciate dagli ospiti	16790	32,77	65,47	0	831
lnrec(+)	N	Ln(number_of_reviews)	11789	2,61	1,73	0	6,72
review_scores_rating(+)	N	Valore medio dei punteggi rilasciati dagli ospiti	11789	4,49	0,78	0	5
review_scores_accuracy(+)	N	Punteggio medio per l'accuratezza	11594	4,66	0,52	0	5
review_scores_cleanliness(+)	N	Punteggio medio per la pulizia	11597	4,59	0,56	0	5
review_scores_checkin(+)	N	Punteggio medio per il checkin	11591	4,75	0,47	0	5
review_scores_communication(+)	N	Punteggio medio per la comunicazione	11596	4,74	0,48	0	5
review_scores_location(+)	N	Punteggio medio per la posizione	11590	4,76	0,39	0	5

review_scores_value(+)	N	Punteggio medio per il valore	11589	4,52	0,54	0	5
------------------------	---	-------------------------------	-------	------	------	---	---

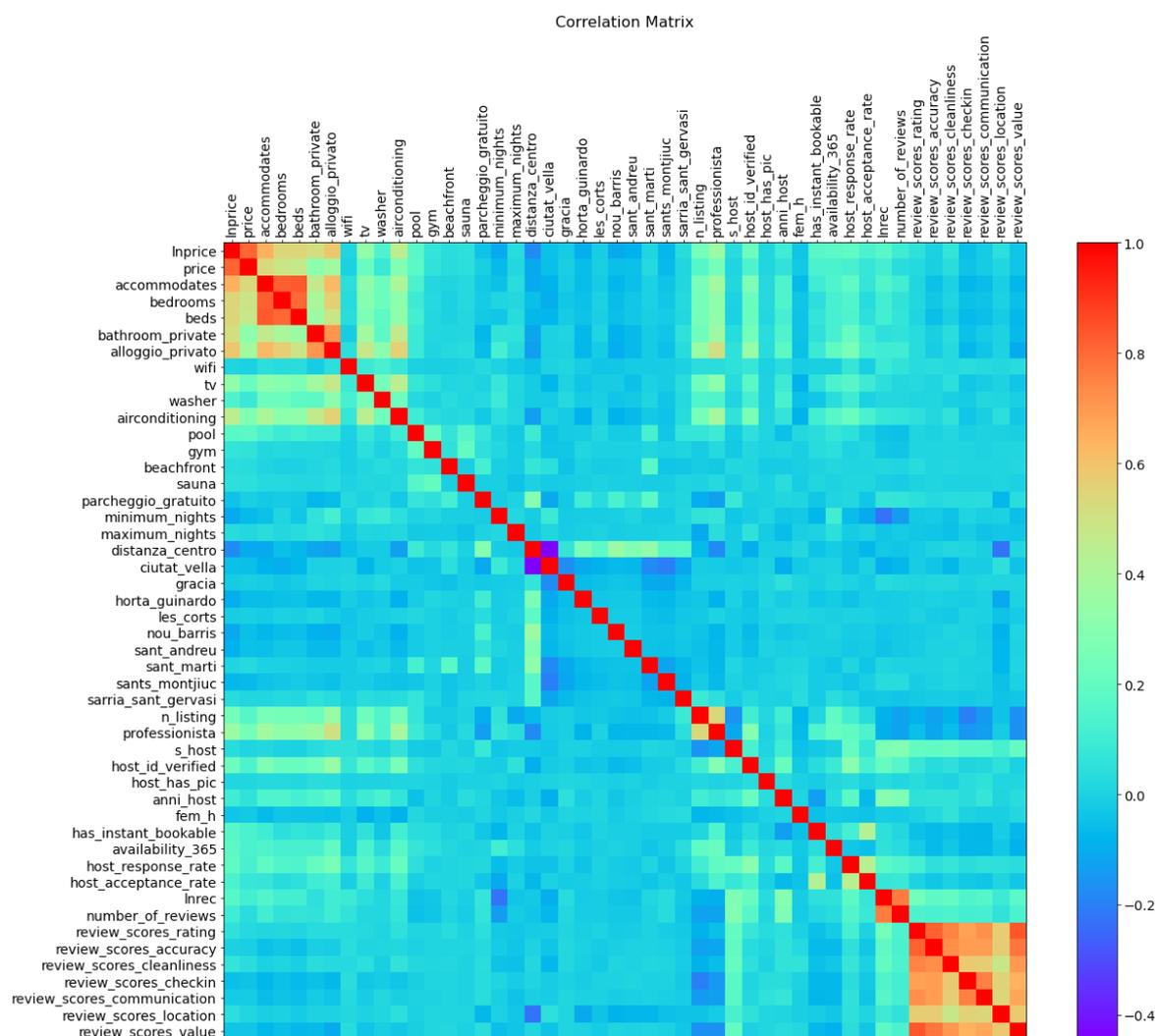
### 7.3 Regressione

L'analisi ha in primo luogo utilizzato un modello di regressione lineare con variabile dipendente il prezzo e successivamente ha utilizzato la sua trasformata logaritmica per valutare eventuali miglurie.

La presenza di un elevato numero di variabili disponibili ha fatto emergere la necessità di individuare quali inserire nel modello, in quanto l'idea di base sarebbe selezionare come variabili indipendenti quelle che sono non correlate tra loro, ma correlate alla variabile dipendente, per tale motivo, inizialmente, è stata calcolata la matrice di correlazione con le variabili a disposizione. Per una più semplice visualizzazione è stata inserita una matrice con le diverse sfumature al variare della correlazione, occorre prestare attenzione alle variabili in cui è presente un quadrato con i colori caldi.

Osservando le correlazioni tra le variabili indipendenti e il prezzo si è scelto di iniziare lo studio dagli elementi strutturali dell'alloggio, in quanto l'alta correlazione può indicare che queste variabili siano tra quelle che influenzino maggiormente il prezzo.

Figura 7.1 Matrice di correlazione delle variabili



### 7.3.1 Elementi strutturali dell'alloggio

Si inizia considerando gli elementi strutturali dell'alloggio che risultano più significativi dalla correlazione con il prezzo: `accommodates`, `bedrooms`, `beds`, `bathroom_private` e `alloggio_privato`. Queste ultime due variabili sono delle dummy che assumo valore 1 rispettivamente se nella in “`bathroom_text`” è presente la parola “private” e se “`type_room`” risulta uguale a “Entire home/apt”. È necessario però tenere in considerazione la loro correlazione perché, come è mostrato in tabella, essa risulta molto elevata ( $\geq 0,8$ ), per cui nella regressione si è scelto di non considerare `bedrooms` e `beds`, mentre le altre variabili (`accommodates`, `bathroom_private` e `hotel_room`), avendo punteggi più bassi e dalla

regressione completa VIF minore di 4, sono state inserite nel modello.

Tabella 7.2 Correlazione variabili struttura alloggio

	accommodates	bedrooms	beds	bathroom_private	alloggio_privato
accommodates	1,00	0,82	0,83	0,47	0,63
bedrooms	0,82	1,00	0,80	0,40	0,56
beds	0,83	0,80	1,00	0,37	0,49
bathroom_private	0,47	0,40	0,37	1,00	0,71
alloggio_privato	0,63	0,56	0,49	0,71	1,00

Nella colonna amenities sono invece elencati una lista di servizi disponibili nella casa, la stessa piattaforma consiglia agli ospiti di mettere in risalto questi elementi, in particolare tra i servizi più richiesti emergono wifi, parcheggio gratuito, accesso agli animali domestici, cucina, piscina, aria condizionata, riscaldamento, tv, lavatrice e vasca idromassaggio. Leggendoli attentamente si possono suddividere in due gruppi: essenziali ed extra. Successivamente sono stati estrapolati quelli che probabilmente possono avere un impatto sul prezzo analogamente a quanto fatto da Chattopadhyay & Mitra (2019) e Falk, Larpin, & Scaglione (2019): tv, wifi, washer, airconditioning, parcheggio\_gratuito (che comprende freeparkingonpremises, freestreetparking e freeparking), pool, gym, beachfront e sauna. Per ognuno di questi servizi è stata creata una dummy che è stata inserita nel modello.

I risultati della regressione con le variabili appena descritte sono riportati nella prima colonna (struttura) della tabella, si nota che solo con queste variabili si riesce a spiegare circa il 30,71% della varianza del prezzo come indica l'R<sup>2</sup>\_adj. Per valutare la scelta di non considerare le variabili bedrooms e beds si è provato ad inserirle nel modello con diverse combinazioni, ma la significatività, il segno dei coefficienti e l'R<sup>2</sup>\_adj risultavano analoghi; quindi, si è scelto di continuare con le stesse condizioni per non aumentare “inutilmente” la complessità del modello.

Come atteso le variabili “accommodates”, “bathroom\_private”, “tv”, “airconditioning”, “pool” e “gym” risultano influenzare positivamente e significativamente il prezzo, mentre diversamente da quanto atteso le variabili più semplici “wifi” e “washer” hanno un’influenza significativa ma negativa probabilmente dovuta all’omissione di qualche variabile significativa, mentre la sauna, il parcheggio gratuito e l’alloggio privato in questo primo modello non risultano significative.

### 7.3.2 *Notti*

Per ogni listing sono inoltre presenti il numero di notti minime e massime *in cui* è possibile prenotare l'alloggio. La prima può significare che un host preferisca affittare per un numero maggiore di giorni perché questo può implicare un minore impegno per lui sia in termini di tempo per la pulizia tra un ospite e un altro, che di energia per comunicare con diversi ospiti, inoltre potrebbero essere minori i periodi in cui l'alloggio rimane vuoto; quindi, si può ipotizzare che l'host sia disposto ad impostare un prezzo più basso in cambio di questi benefici. Mentre se il numero massimo di notti è più contenuto l'host potrebbe avere maggiore interesse a ospitare turisti per soggiorni brevi disposti anche a pagare di più. Quindi per entrambe le variabili si ipotizza una correlazione negativa con il prezzo, che dalla regressione risulta significativa per "minimum nights". L'introduzione di queste variabili ha fatto emergere la correlazione significativa positiva della variabile "alloggio\_privato" a conferma che l'assenza iniziale fosse dovuta a variabili omesse.

### 7.3.3 *Posizione*

Successivamente si è cercato di valutare l'eventuale legame tra la posizione dell'alloggio e il prezzo. Sfruttando le coordinate disponibili degli appartamenti si è calcolata la distanza da Plaça de Catalunya che è considerata il centro della città in km ("distanza\_centro"). Per farlo si è utilizzata la libreria di Python GeoPy che permette di fare operazioni geografiche, in particolare attraverso la funzione "geodesic" passando le coordinate di due punti è possibile individuare la distanza geodetica. Inoltre, nella colonna "neighbourhood\_group\_cleansed" è indicato il distretto in cui è posizionato, dato che il quartiere può dare informazioni relative alla popolazione residente e ai servizi disponibili che influiscono sul prezzo nel mercato degli affitti, si ipotizza che possa esserci una ripercussione anche nei listing analizzati. I quartieri rappresentati in figura della città sono 10, si sono quindi create 9 variabili dummy corrispondenti ( $10-1=9$  per evitare di incorrere nella trappola delle variabili dummy che porta alla collinearità perfetta), il quartiere per cui non si è inserita la dummy è Eixample.

Si è quindi svolta la regressione aggiungendo queste nuove 10 variabili, i risultati sono riportati nella colonna "posizione".

L'introduzione di questi parametri ha confermato i precedenti risultati, inoltre come atteso la variabile distanza dal centro è risultata correlata negativamente e significativa, mentre i quartieri Ciutat Vella e Sants Montjuïc hanno un coefficiente negativo altamente significativo e Les Courts, Nou Barris, Sant Andreu e Sant Marti sono positivi e significativi.

#### 7.3.4 *Host*

Di seguito sono presentate alcune caratteristiche degli host che posso essere utili da aggiungere al modello in quanto il consumatore potrebbe essere disposto a spendere di più per prenotare un alloggio con questi requisiti, ad esempio potrebbe sentirsi più sicuro ad affittare da un proprietario che possieda più listing, che sia sulla piattaforma da più tempo, che abbia una immagine del profilo da poter osservare, che abbia l'identità verificata o che sia un superhost, inoltre potrebbe valutare l'opzione della prenotazione istantanea perché più immediata; non si sono invece considerate le variabili relative al tasso di risposta e di accettazione dell'host perché sono concetti collegati alla prenotazione istantanea e porterebbe ad un dimezzamento di campioni disponibili perché in numerosi listing questo parametro è nullo.

Il risultato conferma che la prenotazione istantanea e la presenza di un host professionista possa portare a un premio di prezzo. Quest'ultimo valore indica che gli host professionisti non applicano un prezzo inferiore sfruttando le economie di scala che potrebbero avere come invece emerge da Boto-García, Mayor e De la Vega (2021), ma al contempo può confermare che gli host amatoriali preferiscono avere un prezzo più basso per assicurarsi la clientela che diversamente probabilmente a parità di condizioni sceglierebbe un host professionista perché ritenuto più affidabile. Invece il parametro "host\_id\_verification" diversamente da quanto atteso risulta negativo e altamente significativo.

#### 7.3.5 *Disponibilità*

Per valutare se la disponibilità abbia un impatto sul prezzo si è inserita la variabile "availability\_365" che rappresenta il numero di giorni che è disponibile in un anno,

ottenendo i risultati della colonna “disponibilità”. Anche questa variabile risulta significativa e positiva.

### 7.3.6 *Recensioni*

Infine, si inseriscono le informazioni relative alle recensioni rilasciate dagli ospiti che hanno soggiornato nella struttura considerando la loro importanza nelle scelte degli utenti che stanno valutando di affittare qualche alloggio, infatti gli host potrebbero sfruttare a loro favore questa tendenza, ad esempio coloro che hanno un numero di recensioni più elevate e alti punteggi possono provare a utilizzare un prezzo più elevato. In particolare, si è scelto di inserire nel modello la variabile “lnrec” che equivale al logaritmo del numero di recensioni perché da alcuni calcoli è risultato migliore rispetto al semplice “number\_of\_reviews”, mentre nonostante un ospite come descritto in precedenza sia invitato a rilasciare un punteggio per i 6 parametri “accuracy”, “cleanliness”, “checkin”, “communication”, “location” e “value”, si è preferito inserire solo la variabile “review\_scores\_rating” che rappresenta un punteggio medio in quanto come dimostrato nel precedente capitolo i punteggi di queste variabili sono molto elevati, simili e correlati, quindi si è scelto di non “appesantire” inutilmente il modello.

Il parametro relativo a “n\_reviews” può essere interpretato come “domanda” del bene e quindi ci si aspetterebbe una correlazione negativa con il prezzo, questa affermazione, come descritto in precedenza, ha però alcuni aspetti da considerare come il fatto che non tutti gli ospiti possono aver rilasciato una recensione e che non è possibile risalire alla durata del soggiorno; quindi, può non essere una buona rappresentazione della domanda. Dall’altro lato si può aspettare una relazione positiva di questa variabile con il prezzo in quanto gli ospiti preferiscono affittare gli alloggi con un numero maggiori di recensioni e che abbiano valutazioni più elevate. In ogni caso i risultati ottenuti dalla regressione non permettono di rifiutare l’ipotesi nulla che queste variabili non influenzino il prezzo, mentre la variabile “Sarria\_sant\_gervasi” è diventata significativa con un livello di significatività del 90%.

### 7.3.7 Risultati

I risultati sono riportati nella seguente tabella. Ogni colonna rappresenta una diversa regressione lineare, per ogni variabile indipendente è riportata la media che corrisponde al coefficiente beta cercato, la varianza nella parentesi e gli asterischi indicano il valore del p\_value secondo cui \*\*\* p<0.01, \*\* p<0.05, \* p<0.1, ciò significa che in presenza di \* non è possibile accettare l'ipotesi nulla che  $\beta$  possa essere uguale a zero con il rispettivo livello di significatività. Inoltre, al fondo sono riportati i numeri delle osservazioni considerate, r2 e r2\_adj.

Tabella 7.3 Risultati regressione con price

VARIABILI	(1)	(2)	(3)	(4)	(5)	(6)
	struttura	notte	posizione	host	disponibilità	recensioni
accommodates	24.95*** (0.906)	24.46*** (0.924)	23.99*** (0.927)	24.05*** (0.929)	23.97*** (0.931)	24.69*** (1.049)
bathroom_private	12.18*** (1.786)	11.79*** (1.803)	11.26*** (1.793)	11.77*** (1.793)	12.41*** (1.791)	8.901*** (1.958)
alloggio_privato	4.032 (2.927)	6.650** (3.040)	6.402** (3.009)	6.700** (3.186)	5.705* (3.197)	7.219** (3.491)
wifi	-16.11** (6.545)	-16.37** (6.532)	-18.30*** (6.508)	-16.61*** (6.423)	-15.99** (6.406)	-11.44** (5.188)
tv	5.617*** (1.926)	6.074*** (1.943)	6.543*** (1.956)	5.838*** (1.958)	5.443*** (1.949)	5.364*** (1.854)
washer	-17.26*** (2.361)	-16.28*** (2.392)	-16.19*** (2.396)	-14.48*** (2.384)	-13.96*** (2.356)	-3.226* (1.719)
airconditioning	13.86*** (1.775)	13.54*** (1.776)	11.93*** (1.765)	10.78*** (1.757)	10.06*** (1.757)	6.886*** (1.711)
pool	66.43*** (8.179)	66.66*** (8.161)	69.15*** (8.162)	71.43*** (8.165)	70.24*** (8.154)	73.10*** (9.111)
gym	21.76*** (8.282)	22.98*** (8.272)	22.41*** (8.229)	21.19*** (8.122)	20.04** (8.176)	21.05** (10.49)
beachfront	16.23*** (5.407)	15.76*** (5.389)	17.78*** (5.477)	17.11*** (5.460)	16.34*** (5.451)	16.32*** (5.694)
sauna	32.76 (26.44)	30.09 (26.58)	27.22 (26.45)	22.24 (26.99)	24.18 (26.51)	67.64 (62.08)
parcheggio_gratuito	0.920 (3.264)	0.745 (3.257)	6.482* (3.334)	7.093** (3.325)	6.536** (3.317)	4.238 (3.036)
minimum_nights		-0.189*** (0.0582)	-0.177*** (0.0571)	-0.171*** (0.0560)	-0.187*** (0.0574)	-0.501*** (0.133)
maximum_nights		0.000139 (0.00153)	1.80e-05 (0.00153)	-0.000809 (0.00153)	-0.00123 (0.00154)	-0.000342 (0.00155)
distanza_centro			-10.19***	-9.709***	-9.786***	-10.84***

		(1.177)	(1.185)	(1.185)	(1.265)	
ciutat_vella		-12.89***	-12.77***	-13.07***	-12.62***	
		(2.129)	(2.149)	(2.142)	(2.313)	
gracia		2.560	2.671	2.148	-0.465	
		(3.089)	(3.092)	(3.089)	(2.985)	
horta_guinardo		1.513	1.708	1.683	0.722	
		(4.641)	(4.604)	(4.611)	(3.750)	
les_corts		14.66**	15.70***	15.96***	13.42**	
		(5.830)	(5.749)	(5.766)	(6.088)	
nou_barris		25.53**	23.03**	22.03**	36.18**	
		(11.06)	(11.04)	(11.08)	(14.84)	
sant_andreu		13.77**	11.81*	11.51*	22.57***	
		(6.974)	(6.886)	(6.889)	(8.516)	
sant_marti		14.20***	14.50***	15.09***	12.11***	
		(3.220)	(3.187)	(3.182)	(3.026)	
sants_montjiuc		-8.136***	-7.264***	-7.371***	-8.116***	
		(2.322)	(2.299)	(2.294)	(2.295)	
sarria_sant_gervasi		7.086	6.760	6.044	9.551*	
		(5.079)	(5.130)	(5.146)	(5.094)	
professionista			4.314**	2.717	9.727***	
			(2.167)	(2.171)	(2.257)	
s_host			0.401	0.325	1.752	
			(1.932)	(1.927)	(1.966)	
host_id_verified			-9.813***	-9.817***	-5.571***	
			(1.839)	(1.836)	(1.907)	
host_has_pic			-0.148	-2.338	5.921	
			(11.75)	(11.57)	(7.095)	
anni_host			-0.165	-0.198	-0.153	
			(0.253)	(0.253)	(0.296)	
has_instant_bookable			15.61***	15.56***	8.400***	
			(1.642)	(1.641)	(1.653)	
availability_365				0.0419***	0.0132*	
				(0.00654)	(0.00716)	
Inrec					0.339	
					(0.656)	
review_scores_rating					2.017	
					(1.275)	
Constant	22.25***	24.33***	49.04***	46.03***	43.14***	16.27
	(6.834)	(6.886)	(7.239)	(13.73)	(13.53)	(10.55)
Observations	16,774	16,774	16,774	16,725	16,725	11,750
R-squared	0.308	0.310	0.316	0.326	0.328	0.429
Adjusted R-squared	0.307	0.309	0.315	0.325	0.327	0.428

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Si può notare che la significatività delle variabili si mantenga relativamente costante nonostante l'inserimento di nuovi parametri, questa operazione contribuisce alla riduzione

delle problematiche relative alle variabili omesse e porta a una diminuzione del numero di osservazioni in quanto il modello in automatico elimina quelle con valori delle variabili utilizzate nulli. Inoltre, sia la R-squared che la Adjusted R-Squared risultano crescenti, per cui le variabili inserite di volta in volta migliorano il modello.

Con gli stessi parametri si è provato a svolgere la regressione considerando come variabile dipendente la trasformata logaritmica del prezzo ossia  $\ln(\text{price})$  analogamente a quanto svolto da Tong e Gunter (2020) e Boto-García, Mayor e De la Vega (2021). Si riportano i valori ottenuti nella successiva tabella. Dal confronto dei risultati completi ottenuti con le due variabili indipendenti, si nota che la trasformazione ha permesso alle variabili relative alle recensioni ( $\ln(\text{rec})$  e  $\text{review\_score\_rating}$ ) di diventare entrambe positive e altamente significative, inoltre ha confermato che caratteristiche come “accommodates”, “bathroom-private”, “alloggio\_privato”, “tv”, “airconditioning”, “pool”, “professionista” e “has\_instant\_bookable” contribuiscano positivamente alla definizione del prezzo, mentre “wifi”, “washer”, “minimum\_nights”, “distanza\_dal\_centro” e “host\_id\_verified” in modo negativo. Per quanto riguarda i quartieri in entrambi emerge che “ciutat\_vella” e “sants\_monjiuc” hanno coefficienti negativi e altamente significativi, mentre risultano significativi e positivi i quartieri di “sant\_marti” e “sarria\_sant\_gervasi”; inoltre i distretti “les\_corts”, “nou\_barris” e “sant\_andreu” perdono la loro significatività, mentre “gracia” e “horta\_guinardo” la acquistano negativa. Infine, si può notare che i valori di  $R^2$  e  $R^2_{\text{adj}}$  sono molto più elevati circa 0,64 rispetto al 0,43 del precedente modello, ciò significa che se si utilizza come variabile dipendente la trasformata logaritmica del prezzo è possibile ottenere un modello che spiega una percentuale maggiore della varianza di Y, approssimando meglio la realtà.

*Tabella 7.4 Risultati regressione con  $\ln(\text{price})$*

VARIABILI	(1) struttura	(2) notti	(3) posizione	(4) host	(5) disponibilità	(6) recensioni
accommodates	0.164*** (0.00355)	0.155*** (0.00393)	0.151*** (0.00382)	0.152*** (0.00366)	0.151*** (0.00369)	0.151*** (0.00417)
bathroom_private	0.258*** (0.0142)	0.251*** (0.0143)	0.243*** (0.0142)	0.248*** (0.0141)	0.253*** (0.0140)	0.221*** (0.0158)
alloggio_privato	0.233*** (0.0176)	0.278*** (0.0204)	0.274*** (0.0199)	0.282*** (0.0205)	0.273*** (0.0205)	0.323*** (0.0235)
wifi	-0.0822**	-0.0864***	-0.113***	-0.110***	-0.104***	-0.119***

	(0.0338)	(0.0334)	(0.0331)	(0.0314)	(0.0310)	(0.0386)
tv	0.0565***	0.0643***	0.0712***	0.0688***	0.0652***	0.0543***
	(0.0125)	(0.0124)	(0.0124)	(0.0123)	(0.0123)	(0.0132)
washer	-0.140***	-0.123***	-0.122***	-0.106***	-0.102***	-0.0291**
	(0.0133)	(0.0134)	(0.0132)	(0.0131)	(0.0130)	(0.0134)
airconditioning	0.209***	0.204***	0.182***	0.171***	0.165***	0.134***
	(0.0121)	(0.0120)	(0.0119)	(0.0117)	(0.0117)	(0.0127)
pool	0.338***	0.342***	0.364***	0.390***	0.379***	0.393***
	(0.0346)	(0.0340)	(0.0339)	(0.0306)	(0.0304)	(0.0318)
gym	0.194***	0.215***	0.208***	0.199***	0.188***	0.152***
	(0.0414)	(0.0418)	(0.0412)	(0.0383)	(0.0384)	(0.0441)
beachfront	0.118***	0.110***	0.132***	0.119***	0.112***	0.0794**
	(0.0308)	(0.0303)	(0.0310)	(0.0307)	(0.0306)	(0.0312)
sauna	0.233***	0.190**	0.153*	0.0955	0.113	0.244*
	(0.0805)	(0.0797)	(0.0791)	(0.0767)	(0.0721)	(0.145)
parcheggio_gratuito	-0.0388**	-0.0418***	0.0352**	0.0310**	0.0259*	0.0135
	(0.0154)	(0.0152)	(0.0155)	(0.0154)	(0.0153)	(0.0156)
minimum_nights		-0.00322***	-0.00313***	-0.00308***	-0.00324***	-0.00677***
		(0.000707)	(0.000695)	(0.000686)	(0.000707)	(0.00117)
maximum_nights		6.32e-06	3.77e-06	-1.14e-05	-1.53e-05*	-3.72e-06
		(9.13e-06)	(9.01e-06)	(8.68e-06)	(8.63e-06)	(9.27e-06)
distanza_centro			-0.0852***	-0.0813***	-0.0820***	-0.0913***
			(0.00705)	(0.00666)	(0.00665)	(0.00726)
ciutat_vella			-0.111***	-0.113***	-0.116***	-0.100***
			(0.0136)	(0.0134)	(0.0133)	(0.0153)
gracia			-0.0304*	-0.0280	-0.0327*	-0.0367*
			(0.0177)	(0.0172)	(0.0171)	(0.0189)
horta_guinardo			-0.0866***	-0.0845***	-0.0848***	-0.0779***
			(0.0297)	(0.0288)	(0.0288)	(0.0274)
les_corts			0.0773**	0.0926***	0.0949***	0.0339
			(0.0372)	(0.0344)	(0.0344)	(0.0360)
nou_barris			-0.0430	-0.0610	-0.0701	0.00962
			(0.0486)	(0.0478)	(0.0476)	(0.0555)
sant_andreu			-0.0502	-0.0664	-0.0691*	-0.00143
			(0.0421)	(0.0410)	(0.0410)	(0.0482)
sant_marti			0.0697***	0.0745***	0.0798***	0.0875***
			(0.0201)	(0.0190)	(0.0189)	(0.0197)
sants_montjiuc			-0.0638***	-0.0570***	-0.0580***	-0.0535***
			(0.0160)	(0.0155)	(0.0154)	(0.0158)
sarria_sant_gervasi			0.0895***	0.0846***	0.0781***	0.127***
			(0.0235)	(0.0231)	(0.0231)	(0.0266)
professionista				0.0122	-0.00233	0.0582***
				(0.0114)	(0.0114)	(0.0129)
s_host				0.0166	0.0159	0.00520
				(0.0104)	(0.0105)	(0.0114)
host_id_verified				-0.0526***	-0.0526***	-0.0392***
				(0.0119)	(0.0118)	(0.0132)
host_has_pic				-0.110	-0.130	-0.0831

				(0.0798)	(0.0797)	(0.0756)
anni_host				-0.00122	-0.00153	-0.00303
				(0.00152)	(0.00151)	(0.00187)
has_instant_bookable				0.133***	0.133***	0.0676***
				(0.00965)	(0.00964)	(0.00985)
availability_365					0.000382***	0.000284***
					(3.64e-05)	(4.13e-05)
review_scores_rating						0.0248***
						(0.00672)
Inrec						0.0157***
						(0.00451)
Constant	3.415***	3.448***	3.690***	3.770***	3.744***	3.580***
	(0.0352)	(0.0362)	(0.0381)	(0.0873)	(0.0869)	(0.0913)
Observations	16,774	16,774	16,774	16,725	16,725	11,750
R-squared	0.509	0.521	0.534	0.558	0.562	0.641
Adjusted R-squared	0.508	0.520	0.533	0.558	0.561	0.640

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

### 7.3.8 Il genere dell'host influenza il prezzo?

Successivamente si è voluto studiare se il genere dell'host influenzi il prezzo dell'annuncio, per farlo si è utilizzata una variabile dummy fem\_h che assume valore pari a 1 nel caso il padrone di casa sia femmina, 0 se maschio. Questo ha portato a utilizzare un database ridotto ai soli 11105 annunci in cui è stato possibile classificare il padrone. Per motivi di spazio si sono riportati in tabella i risultati della regressione semplice e di quella completa con tutte le variabili utilizzando sia price che lnprice come variabile dipendente.

Tabella 7.5 Risultati regressione con genere host con variabili dipendenti price e lnprice

VARIABILI	price		lnprice	
	genere	completa	genere	completa
fem_h	-5.337**	3.857**	-0.0536***	0.0207*
	(2.089)	(1.715)	(0.0152)	(0.0111)
accommodates		24.96***		0.166***
		(1.398)		(0.00559)
bathroom_private		8.522***		0.205***
		(2.003)		(0.0172)
alloggio_privato		4.514		0.292***
		(4.112)		(0.0268)
wifi		-10.72*		-0.109**

	(5.629)	(0.0459)
tv	3.172*	0.0326**
	(1.795)	(0.0147)
washer	0.407	0.0139
	(1.840)	(0.0152)
airconditioning	5.899***	0.116***
	(1.659)	(0.0142)
pool	93.79***	0.507***
	(13.08)	(0.0445)
gym	4.758	0.0325
	(12.74)	(0.0631)
beachfront	18.37***	0.0879**
	(6.857)	(0.0385)
parcheggio_gratuito	5.827*	0.0160
	(3.541)	(0.0183)
minimum_nights	-0.396**	-0.00573***
	(0.155)	(0.00131)
maximum_nights	-0.000604	-1.05e-05
	(0.00167)	(1.09e-05)
distanza_centro	-10.23***	-0.0932***
	(1.519)	(0.00884)
ciutat_vella	-16.42***	-0.118***
	(2.775)	(0.0186)
gracia	-6.361**	-0.0589***
	(2.819)	(0.0220)
horta_guinardo	1.803	-0.0741**
	(4.380)	(0.0323)
les_corts	5.626	-0.0179
	(7.239)	(0.0437)
nou_barris	35.00**	0.0261
	(17.75)	(0.0639)
sant_andreu	22.43**	0.0138
	(10.22)	(0.0571)
sant_marti	8.036**	0.0865***
	(3.252)	(0.0236)
sants_montjiuc	-10.32***	-0.0669***
	(2.467)	(0.0190)
sarria_sant_gervasi	13.33**	0.163***
	(5.552)	(0.0315)
professionista	8.841***	-0.00451
	(2.829)	(0.0169)
s_host	0.348	0.00444
	(2.214)	(0.0132)

host_id_verified	-2.801		-0.0146	
	(2.074)		(0.0144)	
host_has_pic	-13.13		-0.248*	
	(12.95)		(0.132)	
anni_host	-0.0594		-0.00417*	
	(0.367)		(0.00245)	
has_instant_bookable	5.729***		0.0349***	
	(1.924)		(0.0119)	
availability_365	0.0246***		0.000344***	
	(0.00851)		(5.11e-05)	
Inrec	-0.440		0.0122**	
	(0.761)		(0.00524)	
review_scores_rating	1.340		0.0235***	
	(1.437)		(0.00736)	
Constant	90.43***	33.58**	4.128***	3.700***
	(1.525)	(15.47)	(0.0109)	(0.146)
Observations	11,105	7,995	11,105	7,995
R-squared	0.001	0.418	0.001	0.613
Adjusted R-squared	0.000	0.415	0.001	0.611

La prima e la terza colonna rappresentano una regressione lineare semplice in cui l'unica variabile indipendente è il genere da cui si ottiene una correlazione negativa altamente significativa che è conforme a quanto si era ottenuto facendo il test di ipotesi in cui si valutava se il prezzo variasse a seconda del genere considerato. In entrambi i casi si ottiene che il genere influenzi il prezzo, in particolare che un listing di un host maschio abbia un prezzo più elevato.

Introducendo altri elementi nella regressione però si nota che questa variabile perde la propria significatività e il coefficiente diventa positivo, in particolare dopo aver aggiunto il numero e il punteggio delle recensioni esso diventa positivamente significativo (tale significatività risulta più elevata nel caso della variabile price), questo significa che i listing di host femmina hanno un prezzo più elevato e quindi inizialmente il risultato era distorto perché erano state omesse delle variabili significative.

Nel secondo modello si nota che rispetto al precedente di price senza genere la variabile "alloggio\_privato" perde la propria significatività, come "washer", "gym" e "les\_corts", mentre diventano significative "parcheggio\_gratuito" e "gracia", rispettivamente in modo positivo e negativo.

Nella quarta colonna emerge che i coefficienti della maggior parte delle variabili hanno la stessa significatività del modello senza distinzione di genere ad eccezione di servizi “washer” e “gym” e delle caratteristiche degli host “professionista” e “host\_id\_verified” che perdono la propria significatività, mentre “anni\_host” e “has\_host\_pic” ne acquistano una minima correlazione negativa.

Dal confronto diretto tra le colonne 2 e 4 si nota in particolare che anche in questo caso le variabili relative alle recensioni risultano significative solo nel modello con la trasforma logaritmica come anche “host\_has\_pic” e “anni\_host”, inoltre la dummy “alloggio\_privato” con l’inserimento della variabile “fem\_h” nel price perde la propria significatività. Anche in questo caso il modello logaritmico presenta un R2\_adj più elevato.

### 7.3.9 Differenza tra sistemazioni private e condivise

Dal momento che la differenza tra gli alloggi privati e quelli condivisi risulta significativa sul 92% delle variabili utilizzate nella regressione si è voluto cercare di catturare questo effetto: nelle precedenti regressioni presentate si è inserita direttamente una variabile dummy “alloggio\_privato” ottenendo che essa fosse effettivamente significativa, di seguito invece si è provato a dividere il database a seconda che la dummy “alloggio\_privato” risultasse pari a 1 o a 0 e si sono svolte le regressioni complete, i risultati sono riportati nella seguente tabella considerando le diverse combinazioni delle condizioni delle variabili indipendenti price e lnprice, della tipologia di alloggio condiviso e privato, trascurando e inserendo la variabile del genere dell’host (“fem\_h”).

Tabella 7.6 Risultati regressioni separando gli alloggi condivisi da quelli privati

VARIABILI	condiviso		privato		condiviso		privato	
	price	price	price	price	lnprice	lnprice	lnprice	lnprice
fem_h		-3.893** (1.606)		12.23*** (3.205)		-0.00189 (0.0146)		0.0653*** (0.0170)
accommodates	13.35*** (2.054)	13.55*** (1.855)	26.00*** (1.169)	26.15*** (1.625)	0.188*** (0.0164)	0.225*** (0.0153)	0.136*** (0.00404)	0.142*** (0.00601)
bathroom_private		12.99*** (1.887)		12.65*** (1.882)		0.217*** (0.0163)		0.201*** (0.0174)
wifi	-2.785 (4.808)	-1.288 (3.266)	-20.53* (11.33)	-25.13* (14.77)	-0.0781 (0.0476)	-0.0585 (0.0544)	-0.168** (0.0725)	-0.226** (0.0928)
tv	5.141***	2.566	-1.084	-6.148	0.0431***	0.0293*	0.0775**	0.0314

	(1.763)	(1.707)	(6.411)	(6.887)	(0.0144)	(0.0154)	(0.0319)	(0.0427)
washer	-5.292***	-1.064	-9.368**	-7.000	-0.0621***	-0.0167	0.0118	0.0840*
	(1.830)	(1.747)	(3.945)	(5.815)	(0.0147)	(0.0153)	(0.0312)	(0.0464)
airconditioning	6.789***	4.819***	15.82***	17.34***	0.131***	0.104***	0.118***	0.106***
	(1.883)	(1.689)	(3.439)	(3.809)	(0.0158)	(0.0162)	(0.0203)	(0.0269)
pool	35.80**	37.24*	81.80***	114.6***	0.371***	0.347***	0.406***	0.584***
	(14.15)	(19.42)	(10.83)	(16.33)	(0.0692)	(0.0809)	(0.0356)	(0.0529)
gym	18.54	0.421	26.65*	13.51	0.183***	0.0713	0.156***	0.0441
	(13.96)	(6.005)	(14.68)	(25.17)	(0.0701)	(0.0849)	(0.0533)	(0.0916)
beachfront	-5.095*	-6.523**	32.51***	41.85***	-0.0447	-0.0815*	0.180***	0.264***
	(2.774)	(2.946)	(9.325)	(12.21)	(0.0414)	(0.0442)	(0.0422)	(0.0547)
sauna			54.99				0.236	
			(62.99)				(0.151)	
parcheggio_gratuito	6.073**	8.887**	3.850	3.781	0.0538**	0.0667***	-0.0209	-0.0486*
	(3.001)	(3.525)	(5.639)	(7.113)	(0.0213)	(0.0240)	(0.0225)	(0.0267)
minimum_nights	0.0288	0.105	-0.77***	-0.74***	-0.00320*	-0.00234	-0.0081***	-0.00741***
	(0.213)	(0.215)	(0.142)	(0.194)	(0.00165)	(0.00157)	(0.00143)	(0.00185)
maximum_nights	-0.000281	-0.000192	-0.00077	-0.0003	-2.43e-05*	-3.9e-05***	1.79e-05	2.55e-05
	(0.00161)	(0.00149)	(0.00262)	(0.00335)	(1.26e-05)	(1.32e-05)	(1.36e-05)	(1.82e-05)
distanza_centro	-4.405***	-3.451**	-18.30***	-21.15***	-0.0666***	-0.0584***	-0.111***	-0.132***
	(1.481)	(1.617)	(1.993)	(2.729)	(0.0110)	(0.0118)	(0.00957)	(0.0128)
ciutat_vella	1.715	-1.878	-16.26***	-23.08***	0.0344*	0.0111	-0.187***	-0.241***
	(2.327)	(2.224)	(3.531)	(5.110)	(0.0199)	(0.0210)	(0.0214)	(0.0307)
gracia	1.289	-0.228	-0.109	-9.883**	-0.0383	-0.0422	-0.0463*	-0.0911***
	(2.712)	(2.963)	(4.751)	(4.604)	(0.0279)	(0.0293)	(0.0249)	(0.0317)
horta_guinardo	-3.257	-6.686*	5.262	18.79	-0.102***	-0.121***	-0.0467	0.0193
	(3.253)	(3.516)	(7.898)	(11.47)	(0.0364)	(0.0391)	(0.0419)	(0.0569)
les_corts	12.08	7.666	13.46	-1.548	-0.0281	-0.0502	0.0939**	0.00982
	(9.410)	(10.54)	(8.321)	(8.800)	(0.0563)	(0.0577)	(0.0437)	(0.0576)
nou_barris	12.36	9.905	60.15*	77.37	-0.0777	-0.0899	0.0811	0.108
	(17.21)	(19.78)	(35.44)	(48.60)	(0.0706)	(0.0764)	(0.0781)	(0.107)
sant_andreu	10.64	9.442	30.32*	40.28	-0.0440	-0.0419	0.00592	0.0483
	(9.480)	(10.85)	(17.84)	(24.67)	(0.0594)	(0.0661)	(0.0779)	(0.102)
sant_marti	3.401	1.944	25.04***	23.00***	0.0281	0.0268	0.153***	0.166***
	(2.557)	(2.627)	(5.294)	(6.604)	(0.0280)	(0.0303)	(0.0269)	(0.0359)
sants_montjuic	1.117	-2.808	-13.73***	-15.12***	-0.0302	-0.0527**	-0.0717***	-0.0753***
	(3.026)	(2.565)	(3.256)	(4.375)	(0.0254)	(0.0263)	(0.0196)	(0.0278)
sarria_sant_gervasi	12.32*	3.489	13.60*	32.97***	0.173***	0.137***	0.109***	0.198***
	(7.127)	(3.416)	(6.948)	(10.90)	(0.0441)	(0.0392)	(0.0321)	(0.0497)
professionista	3.504	-6.013*	10.22***	6.955*	-0.0435	-0.145***	0.0876***	0.0304
	(3.654)	(3.499)	(2.952)	(3.828)	(0.0293)	(0.0372)	(0.0153)	(0.0202)
s_host	-1.456	-2.697*	4.394	3.914	0.0101	0.00345	0.0195	0.0227
	(1.493)	(1.376)	(3.185)	(3.993)	(0.0163)	(0.0175)	(0.0151)	(0.0186)
host_id_verified	-1.751	-2.319	-5.477	5.496	-0.0199	-0.0170	-0.0548**	0.0337
	(1.847)	(1.934)	(5.184)	(6.402)	(0.0150)	(0.0161)	(0.0274)	(0.0309)
host_has_pic	-3.496	-4.967	26.28	-51.47	-0.258***	-0.237**	0.275	-0.455
	(3.882)	(5.404)	(28.64)	(63.22)	(0.0698)	(0.110)	(0.227)	(0.604)
anni_host	-0.582*	-0.453	-0.447	-0.0116	-0.00759**	-0.00549*	-0.00126	-0.00257

	(0.300)	(0.295)	(0.454)	(0.680)	(0.00308)	(0.00324)	(0.00242)	(0.00366)
has_instant_bookable	6.913***	4.022**	9.519***	7.254*	0.0577***	0.0233	0.0643***	0.0373*
	(1.668)	(1.826)	(2.701)	(3.738)	(0.0142)	(0.0153)	(0.0137)	(0.0194)
availability_365	0.0280***	0.0230***	-0.00522	0.0276	0.0005***	0.00046***	0.000144**	0.000285***
	(0.00775)	(0.00800)	(0.0121)	(0.0178)	(5.65e-05)	(5.82e-05)	(5.78e-05)	(8.63e-05)
review_scores_rating	1.125	1.649*	4.511	1.758	0.0321***	0.0335***	0.0312**	0.0248
	(1.119)	(0.862)	(3.044)	(4.775)	(0.00804)	(0.00819)	(0.0123)	(0.0165)
lnrec	-2.086***	-1.634***	1.127	-0.590	-0.0211***	-0.0188***	0.0336***	0.0299***
	(0.610)	(0.598)	(1.026)	(1.445)	(0.00515)	(0.00551)	(0.00680)	(0.00961)
sauna			54.99				0.236	
			(62.99)				(0.151)	
Constant	29.99***	28.03***	24.29	106.4	3.646***	3.509***	3.818***	4.516***
	(7.756)	(8.696)	(34.93)	(66.23)	(0.0974)	(0.133)	(0.258)	(0.614)
Observations	5,449	4,459	6,301	3,536	5,449	4,459	6,301	3,536
R-squared	0.111	0.088	0.359	0.364	0.285	0.262	0.489	0.482
Adjusted R-squared	0.106	0.081	0.356	0.358	0.280	0.257	0.487	0.477

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Analizzando quindi il modello con lnprice emerge che la significatività dei valori dei coefficienti dell'appartamento privato trovati è molto simile a quella ottenuta dalla regressione completa se non per la perdita di significatività di “washer” e la significatività ottenuta per “horta\_guinardo”, mentre nel caso di alloggi condivisi emergono maggiori differenze: le variabili “wifi”, “beachfront”, “gracia”, “les\_corts”, “sant\_marti”, “professionista”, “host\_id\_verified” hanno perso significatività, mentre sono diventate significative “parcheggio\_gratuito” in modo positivo e “anni\_host”, “host\_has\_pic”, “maximum\_nights” in modo negativo.

L'ipotesi che gli ospiti scelgano di soggiornare in appartamenti condivisi per il minor prezzo sembra confermata dal valore negativo altamente significativo del coefficiente “lnrec”, infatti se si considera il numero delle recensioni come la domanda essa dovrebbe essere elastica e avere una correlazione negativa con il prezzo, quindi al diminuire del prezzo dovrebbe crescere la domanda, nel caso invece di appartamenti privati essa è correlata positivamente e altamente significativa indicando che il numero di recensioni influisce sul prezzo come se l'ospite fosse disponibile a pagare di più se altri membri della comunità hanno già usufruito dell'alloggio.

Inoltre, coloro che soggiornano in appartamenti condivisi sembra che reputino “superflui” elementi come “wifi”, “washer”, “beachfront”, mentre siano disposti a pagare di più per

servizi come la “tv”, il “parcheggio\_gratuito” e il “bathroom\_private”, quest’ultimo invece non compare negli alloggi privati in quanto risulta sempre pari a 1.

Si nota inoltre che sebbene la distanza dal centro permetta in entrambi i database di avere una correlazione negativa significativa, nel caso di alloggio condiviso non risulta essere significativo nessun quartiere se non una minima significatività positiva con “sarria\_sant\_gervasi”.

Nella sesta e nell’ottava colonna è stata inserita la variabile “fem\_h” che risulta significativa e positiva solo nel caso di alloggi privati. Dal confronto con il modello di lnprice in funzione anche del genere emerge che nell’alloggio privato le variabili “tv”, “horta\_guinardo”, “host\_has\_pic”, “anni\_host” e “review\_score\_rating” perdono significatività, mentre la acquista “washer” in modo positivo, nell’alloggio condiviso invece perdono significatività le variabili “wifi”, “minimum\_nights”, “ciutat\_vella”, “gracia”, “sant\_marti”, “has\_instant\_bookable”, mentre la acquistano “maximum\_nights” e “professionista”. Infine, “parcheggio\_gratuito”, “beachfront” e “lnrec” invertono il loro segno.

Considerando invece che dall’analisi sul genere dell’host nel modello con “price” la variabile “alloggio\_privato” non risultava significativa come a significare che la tipologia di alloggio non influenzasse il prezzo, suddividendo i due database è emerso che la significatività di un alto numero di variabili differisce a seconda della tipologia di alloggio considerato, come i coefficienti “wifi”, “minimum\_night”, “ciutat\_vella”, “sant\_marti”, “sants\_monjiuc”, “sarria\_sant\_gervasi” risultano significativi nel caso di alloggio privato, mentre risultano significative solo per la tipologia condivisa “parcheggio\_gratuito”, “horta\_guinardo”, “s\_host”, “availability\_365” e “lnrec”, infine “fem\_h”, “beachfront” e “professionista” invertono il loro segno.

Da questa analisi emerge quindi che i fattori che influenzano il prezzo variano a seconda che si tratti di alloggi privati o condivisi, in particolare si nota che il comportamento che si è osservato nel modello in cui è stata inserita la variabile “alloggio\_privato” è analogo a quello degli alloggi privati, mentre nel caso di alloggi condivisi emergono maggiori differenze.

### 7.3.10 *Contenuto dei feedback e prezzo*

In seguito, considerando l'importanza dei feedback testuali rilasciati dagli utenti si è valutato se il contenuto delle recensioni analizzate potesse influenzare il prezzo degli annunci. Si è scelto di considerare le parole che sono emerse come più utilizzate e che potessero avere effettivamente un impatto, esse sono “help”, “clean” e “recommend”. È però stata riscontrata la necessità di trovare se nel commento esse fossero rilasciate in modo positivo o negativo perché nel primo caso si sarebbe attesa una influenza neutra o positiva sul prezzo, mentre diversamente neutra o negativa. Tale operazione è stata impegnativa, infatti a causa dell'elevato numero di recensioni da analizzare non è stata possibile una classificazione manuale. Dal momento che non era sufficiente vedere il sentimento della recensione e la presenza di tali parole in quanto come si è trovato in precedenza una recensione che nel complesso può risultare positiva, in realtà al suo interno può avere delle parti negative, si è allora provato a suddividere ogni commento in frasi e valutare direttamente il sentimento e la presenza della parola nella frase, ma ispezionando il risultato ottenuto molte volte non coincideva con la realtà a causa solitamente di frasi lunghe composte da più elementi o perché il sentimento era “mascherato” utilizzando parole neutre che non permettevano all'algoritmo una classificazione netta della frase rilasciando un punteggio molto prossimo allo zero; anche questa seconda soluzione è stata scartata. Si è quindi proceduto osservando i risultati ottenuti e costruendo delle variabili dummy apposite che assumevano valore 1 in caso di presenza di certi termini nella recensione, in particolare per ognuna delle tre parole sono state considerate i sinonimi più comuni e le relative negazioni. Ad esempio nel caso di “clean” sono stati ricercati anche i commenti contenenti “tidy” e “tidiness” per i positivi, mentre per i negativi “unclean”, “dirt”, “not clean”, “untid”, “improve clean”, “n't clean”, “n't tidy” e “n't tidiness”; per evitare di conteggiare la presenza della parola clean sia nella dummy positiva che negativa per ogni commento è stato prima verificato l'eventuale presenza del valore negativo e solo successivamente, se tale valore non era presente, si è eseguita la ricerca della parola positiva (if-else). In seguito, si è ricavata la tabella pivot con chiave l'id del listing ottenendo per ogni annuncio il numero di recensioni in cui era trattato l'argomento, si è poi trasformato il valore in percentuale rispetto al totale dei commenti dell'annuncio nel periodo temporale analizzato per poterlo inserire nella regressione del prezzo.

Anche in questo caso si è cercato prima il modello che rappresentasse tutti gli annunci e poi quelli relativi alle due differenti tipologie di alloggi (privati e condivisi).

L'introduzione di queste variabili non ha portato a nette differenze di significatività dei coefficienti ottenuti nelle analisi precedenti, in particolare le host donne continuano ad avere un premio di prezzo nei modelli completi e degli alloggi privati, ma non per gli alloggi condivisi.

I risultati ottenuti in alcuni casi non confermano le ipotesi iniziali, in particolare il concetto di "help" risulta significativo ma negativo negli scenari complessivi e con lnprice in appartamento privato e "recommen" risulta negativa negli alloggi privati con "price". Gli altri coefficienti in concordanza con quanto atteso non risultano significativi, solo "clean" risulta significativo e positivo nel caso di variabile dipendente "lnprice" nel modello completo e negli alloggi condivisi.

Per un'ulteriore valutazione della presenza di un impatto delle recensioni sul prezzo si è sperimentata un'alternativa più semplice: si è considerata se la percentuale delle recensioni positive, negative e neutre impattasse sul prezzo. Per tale motivo sono state create 3 variabili che rappresentassero per ogni annuncio le percentuali di recensioni positive, negative e neutre che avevano. Si attende che queste variabili rispetto al prezzo possono non essere significative o nel caso lo siano che le positive influiscano positivamente, le negative negativamente, mentre il comportamento delle neutre non è facilmente definibile a priori. I risultati della regressione, però, dimostrano che queste variabili non influenzano il prezzo.

Si è provato a modellare le variabili disponibili introducendo nella regressione prima solo le variabili relative alle parole presenti, poi con le sole dummy che indicano il sentimento e infine combinando entrambe, dal momento che la significatività di questi differenti modelli non è variata in tabella è stata rappresentata quella completa.

I risultati contraddittori ottenuti limitano la significatività di questi modelli.

Tabella 7.7 Risultati regressioni con inserimento variabili relative alle recensioni rilasciate

variabili	completa		alloggio_privato		alloggio_condiviso	
	price	lnprice	price	lnprice	price	lnprice
fem_h	4.551** (1.866)	0.0227* (0.0117)	10.25*** (3.257)	0.0499*** (0.0171)	-2.114 (1.717)	0.00758 (0.0156)
accommodates	24.82*** (1.473)	0.160*** (0.00584)	25.93*** (1.681)	0.139*** (0.00658)	12.84*** (2.199)	0.218*** (0.0179)
bathroom_private	7.776*** (2.328)	0.227*** (0.0189)			13.04*** (2.229)	0.223*** (0.0193)
alloggio_privato	7.707* (4.145)	0.318*** (0.0275)				
wifi	-5.427 (6.939)	-0.0841 (0.0525)	-5.395 (16.75)	-0.0297 (0.0930)	-0.815 (4.239)	-0.0906 (0.0638)
tv	5.141** (2.009)	0.0392** (0.0156)	-1.186 (6.663)	0.0503 (0.0420)	4.324** (1.957)	0.0371** (0.0165)
washer	0.0279 (2.150)	0.0187 (0.0170)	-7.924 (6.133)	0.0941* (0.0497)	-1.318 (2.081)	-0.0106 (0.0169)
airconditioning	6.780*** (1.843)	0.121*** (0.0152)	16.00*** (3.848)	0.0934*** (0.0269)	5.904*** (1.922)	0.115*** (0.0177)
pool	93.16*** (12.61)	0.482*** (0.0452)	109.2*** (14.75)	0.539*** (0.0511)	37.22 (23.71)	0.265*** (0.0857)
gym	13.51 (14.29)	0.0968 (0.0712)	24.21 (26.15)	0.0889 (0.0949)	5.048 (7.388)	0.127 (0.101)
beachfront	19.53*** (6.954)	0.0805** (0.0384)	42.60*** (12.29)	0.253*** (0.0536)	-6.374*** (2.438)	-0.0895** (0.0437)
parcheggio_gratuito	5.219 (3.688)	0.00528 (0.0182)	3.501 (7.055)	-0.0324 (0.0256)	8.415** (3.835)	0.0440* (0.0249)
minimum_nights	-0.610*** (0.130)	-0.00800*** (0.00166)	-0.888*** (0.312)	-0.00951*** (0.00320)	-0.204*** (0.0523)	-0.00488*** (0.00131)
maximum_nights	-0.000209 (0.00193)	-3.77e-06 (1.17e-05)	0.000608 (0.00343)	2.96e-05 (1.85e-05)	0.000159 (0.00190)	-3.49e-05** (1.48e-05)
distanza_centro	-12.03*** (1.673)	-0.105*** (0.00893)	-20.98*** (2.810)	-0.127*** (0.0131)	-5.445*** (1.764)	-0.0777*** (0.0120)
ciutat_vella	-15.40*** (2.905)	-0.104*** (0.0191)	-17.47*** (5.516)	-0.193*** (0.0354)	-3.233 (2.655)	0.00335 (0.0235)
gracia	-5.305* (3.179)	-0.0524** (0.0238)	-7.850* (4.641)	-0.0853*** (0.0325)	1.566 (3.783)	-0.0262 (0.0335)
horta_guinardo	6.080 (4.832)	-0.0319 (0.0333)	22.83* (11.72)	0.0416 (0.0565)	-2.846 (3.499)	-0.0653 (0.0405)
les_corts	12.16 (8.365)	0.0454 (0.0465)	-0.171 (8.786)	0.000910 (0.0574)	18.33 (13.32)	0.0461 (0.0660)
nou_barris	45.86** (21.73)	0.0362 (0.0709)	96.67 (59.04)	0.135 (0.114)	18.75 (24.33)	-0.0614 (0.0843)
sant_andreu	32.30** (14.24)	0.0293 (0.0694)	52.47** (26.22)	0.101 (0.107)	20.44 (16.44)	-0.0227 (0.0851)
sant_marti	9.856*** (3.617)	0.113*** (0.0250)	24.20*** (6.826)	0.189*** (0.0366)	1.654 (2.958)	0.0380 (0.0327)

sants_montjiuc	-10.71*** (2.646)	-0.0622*** (0.0197)	-14.12*** (4.417)	-0.0708** (0.0281)	-3.313 (2.728)	-0.0493* (0.0282)
sarria_sant_gervasi	13.21*** (4.618)	0.187*** (0.0329)	26.63*** (7.979)	0.180*** (0.0494)	7.109* (4.281)	0.193*** (0.0436)
professionista	6.563** (2.982)	-0.0133 (0.0169)	5.716 (4.134)	0.0224 (0.0209)	-9.548*** (3.013)	-0.156*** (0.0340)
s_host	-0.0923 (2.293)	0.00193 (0.0139)	1.633 (4.144)	0.0144 (0.0202)	-1.563 (1.386)	0.00849 (0.0179)
host_id_verified	-3.997 (2.492)	-0.0270* (0.0155)	4.969 (7.194)	0.0352 (0.0309)	-2.394 (2.221)	-0.0263 (0.0174)
host_has_pic	-28.45** (14.39)	-0.375** (0.157)	-132.6*** (27.27)	-1.311*** (0.130)	-8.162 (6.473)	-0.301** (0.136)
anni_host	-0.0867 (0.392)	-0.00370 (0.00263)	-0.336 (0.699)	-0.00274 (0.00394)	-0.167 (0.280)	-0.00396 (0.00345)
has_instant_bookable	5.045** (1.997)	0.0281** (0.0127)	6.286* (3.731)	0.0264 (0.0211)	3.369* (1.874)	0.0140 (0.0162)
availability_365	0.0205** (0.00909)	0.000294*** (5.15e-05)	0.0225 (0.0184)	0.000261*** (9.39e-05)	0.0163* (0.00852)	0.000387*** (5.98e-05)
lnrec	-0.550 (0.825)	0.0119* (0.00644)	-0.134 (1.751)	0.0273** (0.0138)	-2.135*** (0.653)	-0.0186*** (0.00618)
review_scores_rating	7.048* (3.772)	0.0475*** (0.0180)	12.18 (8.562)	0.0624* (0.0334)	2.585 (2.132)	0.0441** (0.0202)
clean	3.648 (4.117)	0.0908*** (0.0277)	4.535 (7.389)	0.0710 (0.0448)	1.826 (4.610)	0.102*** (0.0370)
help	-7.864* (4.406)	-0.109*** (0.0324)	-11.35 (7.326)	-0.136*** (0.0508)	-1.511 (5.146)	-0.0452 (0.0416)
not_recommen	-9.676 (11.70)	0.123 (0.103)	0.684 (20.85)	0.116 (0.169)	-10.35 (15.86)	0.160 (0.129)
not_clean	13.14 (26.29)	-0.0153 (0.103)	-10.92 (23.02)	-0.204 (0.156)	36.50 (44.72)	0.139 (0.137)
not_help	-5.062 (30.57)	0.209 (0.341)	-2.327 (30.83)	0.123 (0.291)	15.85 (37.53)	0.398 (0.538)
recommen	-5.354 (4.489)	-0.0147 (0.0286)	-14.31** (6.928)	-0.0560 (0.0408)	2.703 (5.704)	0.00948 (0.0400)
positivo	521.5 (354.5)	2.588 (2.336)	651.2 (621.6)	3.698 (3.085)	181.9 (190.9)	0.497 (3.292)
negativo	542.9 (354.2)	2.555 (2.336)	683.1 (621.5)	3.754 (3.081)	196.6 (190.6)	0.414 (3.296)
neutro	508.3 (354.8)	2.465 (2.337)	632.6 (622.3)	3.592 (3.085)	172.7 (190.7)	0.401 (3.292)
Constant	-499.1 (355.3)	1.130 (2.344)	-526.0 (624.0)	1.344 (3.084)	-151.8 (192.1)	3.078 (3.303)
Observations	6,685	6,685	3,279	3,279	3,406	3,406
R-squared	0.438	0.650	0.380	0.501	0.098	0.319
Adjusted R-squared	0.435	0.648	0.372	0.495	0.087	0.311

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

### 7.3.11 *Le donne host ricevono recensioni migliori dalle donne guest?*

Per rispondere a questo interrogativo si sono svolte alcune regressioni con Stata considerando come variabile dipendente la polarità delle recensioni. Questo valore è stato calcolato come descritto in precedenza con l'apposita funzione che attribuisce a ogni commento un punteggio compreso tra -1 (completamente negativo) e +1 (completamente positivo).

Nella seguente analisi si è inserita la clausola “if” per studiare prima i soli soggiorni in alloggi di host femminili e in un secondo momento solo in quelli maschili, in particolare si sono svolte due regressioni lineari semplici considerando come variabile indipendente solo il genere del recensore (“fem\_r” che assume valore 1 in caso di recensore femmina e 0 per i maschi) e due regressioni lineari multiple in cui sono stati aggiunti i parametri relativi all'appartamento sia strutturali che della posizione oltre ad alcune variabili relative agli host (“esperienza” se è da più di tre anni sulla piattaforma, “host\_id\_verified”, “host\_has\_pic”). Con questo metodo si è verificato il risultato ottenuto con i test d'ipotesi che i recensori donna sono più inclini a rilasciare commenti più critici sia nei confronti di padroni di casa femminili che maschili ed emerge sia dalle regressioni semplici che aggiungendo altre variabili nelle multiple, come è riportato in tabella 7.8. Inoltre, in generale, si può affermare che la distanza dal centro e il posizionamento in alcuni quartieri abbia un'influenza negativa sulla polarità delle recensioni, mentre attributi relativi all'host non risultano significativi se non “host\_id\_verified” per le host donna. Infine, per gli uomini non emerge una correlazione con “alloggio\_privato” e “bathroom\_private”, mentre per le donne sì, risultando rispettivamente negativa la prima e positiva la seconda, al contrario negli alloggi di host uomini le variabili “host\_id\_verified” ed “esperienza” risultano essere positive e significative.

In aggiunta, a conferma di quanto ottenuto dal test d'ipotesi, una ulteriore analisi (i cui risultati per semplicità sono stati omessi) ha permesso di rilevare che nel seguente campione le host donna ottengono recensioni più critiche rispetto alla controparte indipendentemente dal genere del recensore.

Tabella 7.8 Regressione polarità recensioni relative al genere degli host

VARIABILI	Host donne		Host uomini	
fem_r	-0.0183*** (0.00186)	-0.0188*** (0.00186)	-0.0202*** (0.00180)	-0.0204*** (0.00181)
alloggio_privato		-0.0113*** (0.00283)		-0.00253 (0.00280)
bathroom_private		0.0101*** (0.00294)		0.00284 (0.00298)
distanza_centro		-0.00591*** (0.00136)		-0.00456*** (0.00144)
ciutat_vella		-0.0131*** (0.00282)		-0.0134*** (0.00275)
gracia		0.00411 (0.00334)		-0.00891** (0.00349)
horta_guinardo		-0.0145** (0.00570)		-0.0504*** (0.00662)
les_corts		-0.00583 (0.00723)		-0.0443*** (0.0105)
nou_barris		-0.00802 (0.0103)		-0.00143 (0.0123)
sant_andreu		-0.00901 (0.0102)		-0.0287*** (0.00960)
sant_marti		-0.0133*** (0.00382)		-0.00813** (0.00403)
sants_montjuic		-0.0174*** (0.00326)		-0.0160*** (0.00319)
sarria_sant_gervasi		-0.0164*** (0.00574)		-0.0100* (0.00559)
host_id_verified		0.00374 (0.00282)		0.00731** (0.00307)
host_has_pic		0.0701 (0.0947)		0.0568 (0.0475)
esperienza		-0.000919 (0.00352)		0.00702** (0.00330)
Constant	0.440*** (0.00141)	0.386*** (0.0948)	0.447*** (0.00134)	0.394*** (0.0477)
Observations	61,544	61,537	65,135	65,010
R-squared	0.002	0.004	0.002	0.005

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

### 7.3.12 Gli ospiti preferiscono soggiornare da host dello stesso sesso?

In questo sottocapitolo si valuta se le viaggiatrici donna preferiscono soggiornare presso

appartamenti di host femminili, per effettuare l'analisi su Stata si è utilizzato un modello probit.

Il modello probit è un modello di regressione non lineare con variabile dipendente dicotomica (o binaria), ossia che può assumere solo due valori, nel caso in questione 1 se il recensore è femmina e 0 per i maschi. Il modello può individuare la probabilità con cui un'osservazione può valere uno dei due valori, inoltre date alcune caratteristiche permette di classificare le osservazioni nei due gruppi. In particolare, il modello restituisce la probabilità che tale valore sia pari a 1.

Per stimare i valori dei coefficienti delle variabili indipendenti il modello non utilizza il metodo OLS come nel caso delle precedenti regressioni lineari, ma la funzione di verosimiglianza che cerca di ottenere il risultato migliore attraverso un procedimento iterativo.

Dal momento che Y può assumere solo questi due valori, la funzione deve avere codominio  $C=[0,1]$ , quindi non può essere di tipo lineare infatti la derivata prima non deve essere costante. Per soddisfare questa caratteristica il modello utilizza la funzione di ripartizione normale  $\Phi(z)$ . La rappresentazione del modello di regressione Probit è la seguente:

*Equazione 7.6*

$$\Pr(Y = 1|X_1, \dots, X_k) = \Phi(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k) = \Phi(X^T \beta)$$

Di conseguenza la probabilità che la variabile dipendente assuma valore 0 sarà pari a

*Equazione 7.7*

$$\Pr(Y = 0|X_1, \dots, X_k) = 1 - \Pr(Y = 1|X_1, \dots, X_k)$$

Utilizzando come input su Stata “*probit y x1 x2*”, i risultati che si ottengono come output sono analoghi ai modelli di regressione lineare multipla, in quanto lo pseudo R2 ha un significato simile a R2.

Come variabili indipendenti si è scelto di utilizzare le sole variabili “statiche” che non cambiano nel tempo, in quanto l'obiettivo di questa analisi è capire come e quali fattori influenzano la scelta di un utente, quindi è necessario considerare i parametri presenti nel

momento in cui esso prende la decisione, purtroppo i dati che si hanno a disposizione da InsideAirbnb sono limitati per questo tipo di analisi perché sono scaricabili delle situazioni istantanee, che nel caso in questione è quello che si osserva il 7 luglio 2021, quindi le variabili come le recensioni, il prezzo, i titoli ottenuti e il numero di listing possono variare nel tempo e, di conseguenza, per non svolgere un'analisi sfalsata non possono essere considerate. Gli unici dati che possono essere quindi utilizzati oltre al genere dell'host, sono quelle relative alla struttura dell'alloggio (in particolare la tipologia d'alloggio e il bagno) e alla posizione, considerando sia la distanza dal centro che i quartieri in quanto determinati quartieri possono essere meno sicuri e quindi le donne che in genere prestano maggiore attenzione a questo fenomeno possono preferite alcune zone a discapito di altre; invece relativi all'host di sono considerati "host\_id\_verified" e "host\_has\_pic" in quanto sebbene possano variare nel tempo è molto probabile che essi siano abbastanza statici, inoltre si è creata una dummy "esperienza" che assume valore 1 se l'host è sulla piattaforma da più di 3 anni, questo perché anche considerando le prime recensioni del 2019 gli host erano già presenti sulla piattaforma da almeno 1 anno e quindi possono essere considerati non completamente inesperti e più affidabili.

La prima regressione in cui è solo presente una dummy che rappresenta il genere femminile dell'host, il valore del coefficiente permette di affermare che le ospiti donne prenotano maggiormente da padrone di casa dello stesso sesso, infatti, si ottiene un coefficiente positivo e altamente significativo. Il risultato è confermato in tutti i modelli indipendentemente dalle variabili considerate e che essi siano per alloggi privati o condivisi o per entrambi.

Per individuare la probabilità che il recensore sia femmina in caso di host femmina, ossia la variabile "fem\_h" assuma valore 1, considerando il primo modello si sostituiscono i valori ottenuti nell'equazione 7.6 nel seguente modo:  $\phi(0,0112 + 0,0442 * 1) = \phi(0,0554) \sim 52\%$ .

Successivamente sono state inserite tutte le variabili descritte in precedenza, si è così ottenuto che le donne preferiscono gli appartamenti condivisi con il bagno privato, in particolare, analizzando il secondo modello, posizionati nei quartieri "gracia", "horta\_guinardo" e "sarria\_sant\_gervasi" rispetto che nei distretti di "les-corts",

“sant\_marti” e “sants-montjiuc”, mentre le zone “ciutat-vella”, “nou\_barris” e “sant\_andreu” non risultano significative.

Inoltre, le ospiti donna si affidano maggiormente agli host con “esperienza” e “host\_id\_verified”, probabilmente perché li ritengono più affidabili, mentre sembrano non prestare attenzione alla variabile “host\_has\_pic” che non è significativa.

In seguito, si è diviso il database nei due gruppi alloggio privato e alloggio condiviso, dal momento che nelle regressioni svolte in precedenza era risultato che le caratteristiche variavano a seconda della tipologia considerata. Tale differenza emerge anche da questa analisi, infatti la distanza dal centro ha una correlazione significativa positiva nel caso di alloggio privato e negativa per l'alloggio condiviso, come ad indicare che in caso di sistemazione condivisa risulti molto importante la vicinanza al centro, mentre per gli alloggi privati esse preferiscono che siano posizionate in zone più lontane e quindi meno frequentate dai turisti, inoltre i quartieri “horta\_guinardo” e “sant\_andreu” sono positivi e significativi solo per gli alloggi condivisi, mentre “ciutat\_vella” lo è solo per gli alloggi privati.

In conclusione, le donne quando scelgono un soggiorno sembra che preferiscano affidarsi agli host dello stesso sesso e presenti sulla piattaforma da più tempo, probabilmente perché si sentono più sicure, anche i quartieri sono percepiti in maniera differente, mentre il fatto che abbiano l'identità verificata non sembra generare la stessa sensazione di sicurezza, infatti, il coefficiente di questi risulta significativo solo nel modello complessivo.

In particolare, considerando i modelli 3 e 5 si può affermare che la probabilità che un recensore sia femmina se l'host è femmina è più elevata negli alloggi condivisi in quanto  $\phi(-0,0183 + 0,0440 * 1) < \phi(0,0521 + 0,0382 * 1)$  e quindi l'area sottesa della curva è maggiore.

Anche gli ospiti maschi preferiscono scegliere un host dello stesso sesso e utilizzano maggiormente gli alloggi privati.

L'output delle regressioni svolte è mostrato nella seguente tabella.

Tabella 7.9 Risultati regressioni probit relative al genere degli ospiti

variabili	completa		alloggio_privato		alloggio_condiviso	
	genere	completo	genere	completo	genere	completo
fem_h	0.0442*** (0.00705)	0.0422*** (0.00709)	0.0440*** (0.00942)	0.0449*** (0.00948)	0.0382*** (0.0106)	0.0429*** (0.0108)
alloggio_privato		-0.127*** (0.0111)				
bathroom_private		0.0625*** (0.0116)				0.0598*** (0.0118)
distanza_centro		-0.00819 (0.00538)		0.0158** (0.00739)		-0.0304*** (0.00797)
ciutat_vella		-0.00724 (0.0106)		-0.0396** (0.0157)		0.000195 (0.0149)
gracia		0.0817*** (0.0134)		0.0820*** (0.0161)		0.0556** (0.0245)
horta_guinardo		0.0656*** (0.0243)		0.0559 (0.0383)		0.0927*** (0.0329)
les_corts		-0.0800** (0.0326)		-0.0728* (0.0423)		-0.135*** (0.0515)
nou_barris		0.00525 (0.0430)		0.211 (0.130)		0.0596 (0.0509)
sant_andreu		0.0370 (0.0396)		-0.0912 (0.0918)		0.106** (0.0467)
sant_marti		-0.105*** (0.0152)		-0.140*** (0.0202)		-0.0738*** (0.0235)
sants_montjiuc		-0.0619*** (0.0124)		-0.0533*** (0.0157)		-0.0780*** (0.0205)
sarria_sant_gervasi		0.0960*** (0.0225)		0.0574* (0.0294)		0.133*** (0.0353)
host_id_verified		0.0254** (0.0113)		0.0299 (0.0233)		0.0194 (0.0129)
host_has_pic		-0.0430 (0.155)				-0.0548 (0.157)
esperienza		0.0805*** (0.0124)		0.0962*** (0.0209)		0.0679*** (0.0155)
Constant	0.0112** (0.00491)	0.0114 (0.156)	-0.0183*** (0.00644)	-0.152*** (0.0312)	0.0521*** (0.00759)	0.0773 (0.159)
Observations	126,679	126,547	71,152	71,044	55,527	55,502
Pseudo R2	0,0002	0,0026	0,0002	00021	0,0002	0,0029

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

## 8 CONCLUSIONI

La parte empirica del lavoro ha permesso di confermare la presenza di un forte bias positivo sia nei commenti che nei punteggi di valutazione come descritto nella letteratura esistente. Inoltre, analizzando il contenuto testuale delle recensioni è emerso che in generale essi trattano argomenti abbastanza fissi quali l'esperienza con l'host, la posizione, i servizi offerti, la pulizia della casa, la raccomandazione del soggiorno e il ringraziamento; in particolare, sembra che il lessico del recensore non sia influenzato né dal proprio genere, né dal periodo in cui avviene il soggiorno (se precedente e contemporaneo alla pandemia), né dal tipo di proprietario, mentre come atteso le parole variano, in parte, in base al sentimento che esprimono.

Svolgendo dei test di ipotesi si è potuto confermare che maschi e femmine si comportano in molti casi in modo differente ed offrono servizi differenti, così come gli host amatoriali e quelli professionisti agiscono diversamente, infine anche le caratteristiche degli annunci variano significativamente a seconda della tipologia privata o condivisa dell'alloggio.

Dopo aver svolto numerose regressioni per individuare le principali variabili che possono influenzare il prezzo prestando attenzione al problema della multicollinearità, si sono presentati i risultati ottenuti suddividendo i parametri in macrogruppi: struttura dell'appartamento, notti, posizione, host, disponibilità e recensioni. Inizialmente si è effettuata l'analisi con la variabile dipendente "price", successivamente si è provato a trasformare la variabile in "lnprice". In generale è emerso in entrambi i modelli che le caratteristiche relative al numero di persone, alla tipologia di bagno e di alloggio influenzano significativamente il prezzo, anche i servizi come la tv, l'aria condizionata, la piscina e l'essere di fronte alla spiaggia contribuiscono ad incrementare questo valore, mentre le "commodity" come la lavatrice e il wifi risultano correlate negativamente. Gli alloggi che hanno un numero minimo di notti più elevato o che sono più distanti dal centro hanno dei prezzi più bassi, in particolare osservando entrambi i modelli i quartieri di Sarrià Sant Gervasi e Sant Martí risultano i più costosi mentre Ciutat Vella e Sants Montjuic i più economici. Inoltre, la maggiore disponibilità dell'alloggio, la prenotazione istantanea e l'host professionista influiscono positivamente sul prezzo, quest'ultimo quindi non necessita di sfruttare eventuali economie di scala che lo renderebbero più competitivo in

quanto sembra essere percepito dagli ospiti come persona più affidabile. Infine, il modello con la trasformata logaritmica permette di considerare le variabili relative alle recensioni positive e altamente significative, dando quindi la possibilità a coloro che hanno i listing con valori più elevati di questi di poter presentare un prezzo più elevato in quanto sono premiati da un maggior sentimento di fiducia nei loro confronti che spinge quindi gli utenti a spendere anche una percentuale in più per sentirsi più sicuri.

Successivamente si è valutato se il genere dell'host influenzasse il prezzo dell'annuncio, per farlo si è utilizzato un database ridotto ai soli listing a cui era possibile attribuire un genere agli host. È emerso che questo nuovo parametro fosse in effetti significativo e positivo, ciò a significare che le padrone di casa donne potevano avere un premio sul prezzo in quanto probabilmente sono ritenute più affidabili e precise.

Infine, dopo aver notato che gli alloggi privati e le stanze condivise presentavano numerose differenze significative tra le variabili utilizzate nelle regressioni, si è diviso il database in due parti a seconda della tipologia dell'alloggio e si sono svolte due regressioni differenti, ciò che è emerso è stato che effettivamente i fattori che influenzano il prezzo, sia considerando il price che lnprice, variano nei due sottogruppi. Da quest'ultima analisi è stata però confermata l'ipotesi che coloro che affittano un letto in una casa condivisa lo fanno per il minor costo; infatti, il coefficiente del numero di recensioni che può essere approssimato con la domanda è negativamente significativo, mentre nel caso di alloggi privati esso risulta positivo quindi non sembra emergere una particolare attenzione al prezzo per i soggetti che scelgono questo secondo tipo di soggiorno.

In generale è stato riscontrato che la trasformata logaritmica ha portato a dei modelli con  $R^2$  e  $R^2_{adj}$  più elevati; quindi, sembrano migliori in quanto riescono a spiegare maggiormente la varianza della variabile dipendente. Inoltre, sempre secondo  $R^2$  i modelli generali hanno percentuali più elevate.

L'introduzione delle variabili relative alle recensioni rilasciate dagli ospiti hanno confermato che le host donne hanno un premio di prezzo, mentre non hanno un'influenza sul prezzo se non in rari casi, ma i valori ottenuti non sono sempre conformi a quelli attesi; quindi, la significatività di tali modelli è limitata.

I risultati ottenuti dal modello Probit permettono di confermare la propensione degli ospiti a scegliere alloggi con host dello stesso sesso (Choi & Horvát, 2019) e in determinate aree, mentre le viaggiatrici femminili prediligono gli alloggi condivisi forse perché permettono di conoscere meglio la cultura locale come presentato da Sánchez-Franco & Alonso-Dos-Santos (2020), i maschi gli appartamenti privati probabilmente perché più simili agli hotel. La propensione a scegliere utenti dello stesso sesso è stata confermata anche nell'esperimento svolto sui viaggiatori singoli, ma in quel caso tale preferenza era emersa solo nel caso femminile.

## 9 LIMITI DELL'ANALISI

Lo studio si è limitato ad analizzare la città di Barcellona in un preciso periodo temporale quindi i risultati potrebbero variare modificando questi parametri. Inoltre, lavorando con un numero di dati molto elevato è stato necessario automatizzare le operazioni e utilizzare alcuni pacchetti di Python che possono generare alcuni errori. In particolare, per classificare il genere degli utenti si è ricorso a un algoritmo del pacchetto gender che classifica i nomi in femminili, maschili, sconosciuti o “andy” (che corrisponde a un nome che può essere sia maschile che femminile), ma come si è potuto notare il nome che è disponibile scaricando i dati di InsideAirbnb può non essere un nome di persona o al contempo questo può essere maschile o femminile a seconda del paese di origine, ma tra le informazioni disponibili quest'ultimo dato non emerge; quindi, si è considerato quello inglese di default. Per queste motivazioni la scelta di classificare il genere in base al nome può non essere ottimale; infatti, in molti studi in cui si analizzano le caratteristiche dell'host come razza, sesso e bellezza personale si risolve il problema di non avere questi dati direttamente scaricabili incaricando alcune persone ad osservare le foto profilo dei singoli utenti per procedere ad una classificazione manuale. Analogo discorso per individuare il sentimento che emerge dalle recensioni.

Un altro limite è stato quello di analizzare solo una piccola percentuale ritenuta più significativa dei servizi elencati in amenities, mentre probabilmente si fossero inseriti tutti i parametri ne sarebbero emersi altri di significativi.

Infine, i database scaricati non sempre hanno i dati che corrispondono a quanto indicato direttamente sulla piattaforma e talvolta anche quelli presenti sul sito se osservati attentamente non risultano avere senso.

## **INDICE FIGURE**

Figura 2.1 Esempio annuncio1 (screenshot della pagina) .....	45
Figura 2.2 Esempio annuncio2(screenshot della pagina) .....	45
Figura 4.1 Distribuzione appartamenti Barcellona.....	72
Figura 5.1 Word cloud generale .....	84
Figura 7.1 Matrice di correlazione delle variabili .....	109

## **INDICE GRAFICI**

Grafico 1.1 Predicibilità utilizzo verso chunkiness di alcuni beni .....	17
Grafico 4.1 Distribuzione delle recensioni di prenotazioni cancellate .....	70
Grafico 4.2 Distribuzione temporale delle recensioni .....	71
Grafico 4.3 Tipologia di camere.....	72
Grafico 4.4 Andamento prezzo listing a notte .....	73
Grafico 4.5 Distribuzione delle lingue dei commenti.....	74
Grafico 4.6 Percentuali genere dei recensori e degli host .....	75
Grafico 5.1 Distribuzione punteggi complessivi .....	77
Grafico 5.2 Distribuzione variabili rilasciate dai feedback .....	77
Grafico 5.3 Distribuzione numero parole nei commenti .....	79
Grafico 5.4 Relazione tra soggettività e polarità .....	80
Grafico 5.5 Sentimento dei commenti.....	81

Grafico 5.6 Sentimento delle singole frasi delle recensioni .....	82
Grafico 5.7 Correlazioni del numero di parole con la polarità e la soggettività.....	82
Grafico 6.1 Distribuzione numero listing .....	97

## **INDICE TABELLE**

Tabella 1.1 Risparmio risorsa a notte rispetto hotel .....	22
Tabella 1.2 Comportamenti host sharing economy .....	22
Tabella 4.1 Esempio tabella reviews .....	71
Tabella 5.1 Statistiche punteggi recensioni .....	78
Tabella 5.2 Statistiche relative al numero di parole nei commenti .....	78
Tabella 5.3 Ripartizione del sentimento dei commenti .....	81
Tabella 5.4 Ripartizione del sentimento delle singole frasi nei commenti.....	82
Tabella 5.5 Incidenza nome proprio e termine host sulle recensioni .....	83
Tabella 5.6 Percentuale delle recensioni in cui compaiono le parole più citate confrontando quelle generali, con il sentimento e il genere .....	85
Tabella 5.7 Parole più utilizzate nelle recensioni confrontando quelle complessive, con quelle pre-durante covid e professionisti e amatoriali.....	86
Tabella 6.1 Differenza nelle recensioni rilasciate .....	92
Tabella 6.2 Differenza nei ricavi .....	93
Tabella 6.3 Differenza tra gli host.....	94
Tabella 6.4 Differenza nei feedback degli annunci .....	94
Tabella 6.5 Distribuzione recensioni genere recensori e host .....	95

Tabella 6.6 Risultati esperimento viaggiatori solitari.....	96
Tabella 6.7 Riassunto statistiche host amatoriali e professionisti .....	98
Tabella 6.8 Genere recensore e tipologia di host.....	100
Tabella 6.9 Riassunto statistiche appartamenti privati e condivisi.....	100
Tabella 7.1 Statistiche descrittive delle variabili (N=numero, D=dummy, %=percentuale, €=valuta).....	106
Tabella 7.2 Correlazione variabili struttura alloggio.....	110
Tabella 7.3 Risultati regressione con price.....	114
Tabella 7.4 Risultati regressione con lnprice.....	116
Tabella 7.5 Risultati regressione con genere host con variabili dipendenti price e lnprice .....	118
Tabella 7.6 Risultati regressioni separando gli alloggi condivisi da quelli privati .....	121
Tabella 7.7 Risultati regressioni con inserimento variabili relative alle recensioni rilasciate .....	127
Tabella 7.8 Regressione polarità recensioni relative al genere degli host.....	130
Tabella 7.9 Risultati regressioni probit relative al genere degli ospiti .....	134

## BIBLIOGRAFIA

- Abrate, G., & Viglia, G. (2019). Personal or Product Reputation? Optimizing Revenues in the Sharing. *Economy. Journal of Travel Research*, 58(1), 136-148. doi:<https://doi.org/10.1177/0047287517741998>
- Aloisi, A. (2016). Commoditized Workers. Case Study Research on Labour Law Issues Arising from a Set of 'On-Demand/Gig Economy' Platforms. *Comparative Labor Law&Policy Journal*, 37(3), 38.
- Alsudais, A., & Teubner, T. (2019). Large-Scale Sentiment Analysis on Airbnb Reviews from 15 Cities. *Twenty-fifth Americas Conference on Information Systems (amcis)*. Cancun.
- Ayouba, K., Breuillè, M., Grivault, C., & Le Gallo, J. (2019). Does Airbnb Disrupt the Private Rental Market? An Empirical Analysis for French Cities. *International Regional Science Review*, 43, 76-104. doi:<https://doi.org/10.1177/0160017618821428>
- Barzilay, A.R., & Ben-David, A. (2017). Platform Inequality: Gender in the Gig-Economy. *Seton Hall Law Review*, 47(393).
- Belk, R. (2014). You are what you can access: Sharing and collaborative consumption online. *Journal of Business Research*, 67(8), 1595-1600.
- Boto-García, D., Mayor, M., & De la Vega, P. (2021). Spatial price mimicking on Airbnb: Multi-host vs single-host. *Tourism Management*, 87.
- Bridges, J., & Vásquez, C. (2018). If nearly all Airbnb reviews are positive, does that make them meaningless? *Current Issues in Tourism*, 21(18), 2057-2075.
- Celata, F., Hendrickson, C. Y., & Sanna, V. S. (2017). The sharing economy as community marketplace? Trust, reciprocity and belonging in peer-to-peer accommodation platforms. *Cambridge Journal of Regions, Economy and Society*, 10(2), 349-363. doi:<https://doi.org/10.1093/cjres/rsw044>

- Chattopadhyay, M., & Mitra, S. (2019). Do airbnb host listing attributes influence room pricing homogenously? *International Journal of Hospitality Management*, *81*, 54-64. doi:<https://doi.org/10.1016/j.ijhm.2019.03.008>
- Cheng, M., & Jin, X. (2019). What do Airbnb users care about? An analysis of online review comments. *International Journal of Hospitality Management*, *76*, 58-70.
- Cheng, X., Fu, S., Sun, J., Bilgihan, A., & Okumus, F. (2019). An investigation on online reviews in sharing economy driven hospitality platforms: A viewpoint of trust. *Tourism Management*, *71*, 3, 71(366-377). doi:<https://doi.org/10.1016/j.tourman.2018.10.020>
- Cheung, R. (2014). The Influence of Electronic Word-of-Mouth on Information Adoption in Online Customer Communities. *Global Economic Review: Perspectives on East Asian Economies and Industries*, *43*(1), 42-57. doi:10.1080/1226508X.2014.884048
- Choi, E., & Horvát, E.-Á. (2019). Airbnb's Reputation System and Gender Differences Among Guests: Evidence from Large-Scale Data Analysis and a Controlled Experiment. *Social Informatics*, *11864*, 3-17. doi:[https://doi.org/10.1007/978-3-030-34971-4\\_1](https://doi.org/10.1007/978-3-030-34971-4_1)
- Cui, R., Li, J., & Zhang, D. J. (2019). Reducing Discrimination with Reviews in the Sharing Economy: Evidence from Field Experiments on Airbnb. *Management Science*, *66*(3), 1071-1094. doi:<https://doi.org/10.1287/mnsc.2018.3273>
- Deboosere, R., Kerrigan, D. J., Wachsmuth, D., & El-Geneidy, A. (2019). Location, location and professionalization: a multilevel hedonic analysis of Airbnb listing prices and revenue. *Regional Studies, Regional Science*, 143-156.
- Dellarocas, C. (2010). Online Reputation Systems: How to Design One That Does What You Need. *MIT Sloan Management Review*, *5*, 51(3), 33-38.
- Edelman, B. G., & Geradin, D. (2015). Efficiencies and Regulatory Shortcuts: How Should We Regulate Companies like Airbnb and Uber? *Stanford Technology Law Journal*, *19*(2), 293-328. doi:<https://doi.org/10.2139/ssrn.2658603>

- Edelman, B., Luca, M., & Svirsky, D. (2017). Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment. *American Economic Journal: Applied Economics*, 9(2), 1-22. doi:<https://doi.org/10.1257/app.20160213>
- Ert, E., & Fleischer, A. (2019). The evolution of trust in Airbnb: A case of home rental. *Annals of Tourism*, 75, 279-287. doi: <https://doi.org/10.1016/j.annals.2019.01.004>
- Ert, E., Fleischer, A., & Magen, N. (2016). Trust and reputation in the sharing economy: The role of personal photos in Airbnb. *Tourism Management*, 55, 62-73. doi:<https://doi.org/10.1016/j.tourman.2016.01.013>
- Falk, M., Larpin, B., & Scaglione, M. (2019). The role of specific attributes in determining prices of Airbnb listings in rural and urban locations. *International Journal of Hospitality Management*, 83, 132-140. doi:<https://doi.org/10.1016/j.ijhm.2019.04.023>
- Filippas, A., Horton, J., & Zeckhauser, R. J. (2019). Owning, Using and Renting: Some Simple Economics of the “Sharing Economy”. *Management Science*, 66(9). doi:<https://doi.org/10.1287/mnsc.2019.3396>
- Fradkin, A., Grewal, E., Holtz, D., & Pearson, M. (2015). Bias and Reciprocity in Online Reviews: Evidence From Field Experiments on Airbnb. *EC '15: Proceedings of the Sixteenth ACM Conference on Economics and Computation*, (p. 641). doi:<https://doi.org/10.1145/2764468.2764528>
- Frenken, K. (2017). Political economies and environmental futures for the sharing economy. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2095(20160367), 375. doi:<https://doi.org/10.1098/rsta.2016.0367>
- Frenken, K., Meelen, T., Arets, M., & Van de Glind, P. (2015). Smarter Regulation for the Sharing Economy. *The Guardian* .
- Garbarino, E., & Strahilevitz, M. (2004). Gender differences in the perceived risk of buying online and the effects of receiving a site recommendation. *Journal of Business Research*, 57(7), 768-775.

- Garcia-López, M., Jofre-Monseny, J., Martínez-Mazza, R., & Segú, M. (2020). Do short-term rental platforms affect housing markets? Evidence from Airbnb in Barcelona. *Journal of Urban Economics*, *119*. doi:<https://doi.org/10.1016/j.jue.2020.103278>
- Granitz, N., & Ward, J. (1996). Virtual community: A sociocognitive analysis. *Advances in Consumer Research*, *23*, 161-166.
- Gunter, U. (2018). What makes an Airbnb host a superhost? Empirical evidence from San Francisco and the Bay Area. *Tourism Management*, *66*, 26-37. doi:<https://doi.org/10.1016/j.tourman.2017.11.003>
- Guttentag, D., Smith, S., Potwarka, L., & Havitz, M. (2017). *Journal of Travel Research*, *57*(3), 342-359. doi:<https://doi.org/10.1177/0047287517696980>
- Hamari, J., Sjöklint, M., & Ukkonen, A. (2016). The sharing economy: Why people participate in collaborative consumption. *Journal of the Association for Information*, *67*(9), 2047-2059.
- Hennig-Thurau, T., & Walsh, G. (2003). Electronic Word-of-Mouth: Motives for and Consequences of Reading Customer Articulations on the Internet. *International Journal of Electronic Commerce*, *8*(2), 51-74. doi:<http://dx.doi.org/10.1080/10864415.2003.11044293>
- Heo, C. Y., Blal, I., & Choi, M. (2019). What is happening in Paris? Airbnb, hotels, and the Parisian market: A case study. *Tourism Management*, *70*(78-88). doi:<https://doi.org/10.1016/j.tourman.2018.04.003>
- Hong, J., Kim, B., & Park, K. (2019). Optimal risk management for the sharing economy with stranger danger and service quality. *European Journal of Operational Research*, *279*(3), 1024-1035.
- Hu, N., Pavlou, P., & Zhang, J. (2009). Overcoming the J-shaped distribution of product reviews. *Communications of the ACM*, *52*(10), 144-147.
- Huete-Alcocer, N. (2017). A Literature Review of Word of Mouth and Electronic Word of Mouth: Implications for Consumer Behavior. *Frontiers in psychology*, *8*(1256).

doi:10.3389/fpsyg.2017.01256

- Koopman, C., Mitchell, M. D., & Thierer, A. D. (2015). The Sharing Economy and Consumer Protection Regulation: The Case for Policy Change. *The Journal of Business, Entrepreneurship & the Law*, 8(2), 529-545. doi:<https://doi.org/10.2139/ssrn.2535345>
- Lawani, A., Reed, M., Mark, T., & Zheng, Y. (2018). Reviews and price on online platforms: Evidence from sentiment analysis of Airbnb reviews in Boston. *Regional Science and Urban Economics*, 22-34. doi:<https://doi.org/10.1016/j.regsciurbeco.2018.11.003>
- Lee, C., Tse, Y., Zhang, M., & Ma, J. (2020). Analysing online reviews to investigate customer behavior in the sharing economy: The case of Airbnb. *Information Technology & People*, 33(3), 945-961. doi: <https://doi.org/10.1108/ITP-10-2018-0475>
- Lee, S., & Kim, D. Y. (2018). Brand personality of Airbnb: Application of user involvement and gender differences. *Journal of Travel & Tourism Marketing*, 35(1), 32-45.
- Liang, S., Schuckert, M., Law, R., & Chen, C. (2017). Be a “Superhost”: The importance of badge systems for peer-to-peer rental accommodations. *Tourism Management*, 60, 454-465. doi:<https://doi.org/10.1016/j.tourman.2017.01.007>
- Litvin, S., Goldsmith, R., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, 29, 458-468.
- Marchenko, A. (2019). The impact of host race and gender on price on Airbnb. *Journal of Housing Economics*, 46.
- Miller, S. R. (2016). First Principles for regulating the sharing economy. *SSRN Electronic Journal*, 53(1), 147-202. doi:DOI:10.2139/ssrn.2568016
- Nieuwland, S., & van Melik, R. (2018). Regulating Airbnb: how cities deal with perceived negative externalities of short-term rentals. *Current Issues in Tourism*, 23(2), 1-15. doi:10.1080/13683500.2018.1504899

- Ordenes, F., Burton, J., Theodoulidis, B., Gruber, T., & Zaki, M. (2013). Analyzing customer experience feedback using text mining: A linguistics-based approach. *Journal of Service Research*, 17(3), 278-295.
- Ranchordás, S. (2015). Does Sharing Mean Caring? Regulating Innovation in the Sharing Economy. *Minnesota Journal of Law, Science & Technology*, 16(1), 414-475. Tratto da [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2492798](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2492798)
- Sánchez-Franco, M., & Alonso-Dos-Santos, M. (2020). Exploring gender-based influences on key features of Airbnb accommodation. *Ekonomiska Istraživanja / Economic Research*. doi:10.1080 / 1331677X.2020.1831943
- Sánchez-Franco, M., Navarro-García, A., & Rondán-Cataluña, F. (2016). Online Customer Service Reviews in Urban Hotels: A Data Mining Approach. *Psychology & marketing*, 33(12), 1174-1186. doi:<https://doi.org/10.1002/mar.20955>
- Santos, G., Mota, V., Benevenuto, F., & Silva, T. (2020). Neutrality may matter: sentiment analysis in reviews of Airbnb, Booking, and Couchsurfing in Brazil and USA. *Social Network Analysis and Mining*, 10(45). doi:<https://doi.org/10.1007/s13278-020-00656-5>
- Schoenbaum, N. (2016). Gender and the Sharing Economy. *Fordham Urban Law Journal*(53).
- So, K. K., Oh, H., & Min, S. (2018). Motivations and constraints of Airbnb consumers: Findings from a mixed-methods approach. *Tourism Management*, 67, 224-236. doi:<https://doi.org/10.1016/j.tourman.2018.01.009>
- Sthapit, E., & Jiménez-Barreto, J. (2018). Sharing in the host–guest relationship: perspectives on the Airbnb hospitality experience. *Anatolia*, 29(2), 282-284.
- Stocker, V., Kolleck, A., Brause, S., & Schawe, N. (2021). Navigating the Landscape of the Sharing and Gig Economy Literature: A Systematic and Interdisciplinary Review. doi:<http://dx.doi.org/10.2139/ssrn.3773458>
- Tadelis, S. (2016). Reputation and Feedback Systems in Online Platform Markets. *Annual*

*Review of Economics*, 8(1), 321-340.

- Tong, B., & Gunter, U. (2020). Hedonic pricing and the sharing economy: how profile characteristics affect Airbnb accommodation prices in Barcelona, Madrid, and Seville. *Current Issues in Tourism*.
- Voltes-Dorta, A., & Sánchez-Medina, A. (2020). Drivers of Airbnb prices according to property/room type, season and location: A regression approach. *Journal of Hospitality and Tourism Management*, 45, 266-275.
- Xiang, Z., Schwartz, Z., Jr, G., H., J., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, 44, 120-130.
- Ye, Q., Law, R., Gu, B., & Chen, W. (2011). The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. *Computers in Human Behavior*, 27(2), 634-639.
- Zervas, G., Proserpio, D., & Byers, J. (2021). A first look at online reputation on Airbnb, where every stay is above average. *Mark Lett*, 1-16. doi:<https://doi.org/10.1007/s11002-020-09546-4>
- Zervas, G., Proserpio, D., & Byers, J. W. (2017). The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry. *Journal of Marketing Research*, 54(5), 687–705. doi:<https://doi.org/10.1509%2Fjmr.15.0204>
- Zhang, Z., Chen, R., Han, L., & Yang, L. (2017). Key Factors Affecting the Price of Airbnb Listings: A Geographically Weighted Approach. *Sustainability*, 9.
- Zhu, Y., Cheng, M., Wang, J., Ma, L., & Jiang, R. (2019). The construction of home feeling by Airbnb guests in the sharing economy: A semantics perspective. *Annals of Tourism Research*, 75, 308–321. doi:<https://doi.org/10.1016/j.annals.2018.12.013>
- Zloteanu, M., Harvey, N., Tuckett, D., & Livan, G. (2018). Digital Identity: The effect of trust and reputation information on user judgement in the Sharing Economy. *PLOS ONE*, 13(12). doi:<https://doi.org/10.1371/journal.pone.0209071>

## SITOGRAFIA

[www.airbnb.it](http://www.airbnb.it)

[Growth hacking | La strategia di Airbnb | MARKETING TORINO](#)

[Il growth hacking di AirBNB? Prova a copiarlo - StartupBusiness](#)

[www.datacamp.com/community/tutorials/text-analytics-beginners-nltk](http://www.datacamp.com/community/tutorials/text-analytics-beginners-nltk)

[www.python.org/](http://www.python.org/)

[www.nltk.org/](http://www.nltk.org/)

## RINGRAZIAMENTI

Arrivata alla fine di questo impegnativo ma bellissimo percorso voglio ringraziare tutti coloro che mi sono stati accanto.

Vorrei ringraziare la professoressa Laura Rondi per avermi seguito nella stesura di questa tesi.

Un enorme grazie va alla mia famiglia che mi ha permesso di svolgere questo cammino supportandomi e sopportandomi anche nei momenti più complicati.

Desidero ringraziare le mie due coinquiline Cava e Fede con cui ho condiviso questi 5 anni perché dall'inizio alla fine ci siamo sempre sostenute a vicenda e Leo su cui ho sempre potuto contare.

Ringrazio inoltre il Poli per tutte le fantastiche persone che mi ha fatto incontrare e che mi hanno permesso di rendere più leggere e divertenti le lezioni, i momenti di studio e quelli di svago. Un grazie sincero a Sara, Simo, Jamy, Marina, Cate, Ludo e Dani.

Grazie anche a tutti gli amici cuneesi e torinesi che soprattutto in questi ultimi anni di pandemia mi hanno fatto vivere dei bei momenti e passare dei sabati sera "in compagnia".

Ringrazio inoltre Antonio e la signora Rossi per avermi dato l'opportunità di entrare nel mondo lavorativo, tutto il team della Wejo e della Tpl per aver sempre risposto alle mie domande e al bel clima che mi hanno fatto vivere, e a Francesco per tutte le cose che con pazienza ed entusiasmo mi ha spiegato.

Grazie quindi a tutti voi per essermi stati accanto in questi anni e avermi aiutato a crescere.