

POLITECNICO DI TORINO



Dipartimento di Ingegneria Gestionale e della Produzione

Corso di Laurea Magistrale
in Ingegneria Gestionale

Tesi di Laurea Magistrale

IL RUOLO DELLA GENTILEZZA NEL SISTEMA DI RECENSIONI DELLA PIATTAFORMA AIRBNB

Relatrice

Prof.ssa Laura Abrardi

Candidato

Stefano Rocco

Anno Accademico 2019/2020

Alla mia famiglia, vero motivo per essere la versione migliore di me stesso.

A che c'è stato, a chi c'è ora, a chi ci sarà

Indice

Sommario

1. Introduzione	5
2. Analisi della letteratura	7
2.1 Distribuzione distorta delle valutazioni	8
2.2 Importanza della relazione Host – Guest	11
2.3 Dai feedback al valore aggiunto	17
3. Descrizione dei Database di partenza	20
4. Analisi dati	30
4.1 Analisi Delle Variabili Di Interesse: L’host.....	30
4.1.1 Ubicazione dell’host	31
4.1.2 Tempo di risposta	34
4.1.3 Tasso di risposta e tasso di accettazione	35
4.1.4 Superhost	38
4.1.5 Presentazione dell’host	39
4.1.6 Numero di Annunci per host	41
4.1.7 Tipologia di annuncio	43
4.1.8 Anno di iscrizione	45
4.2 Analisi delle variabili del modello di regressione e processo	47
4.2.1 Foglio Review.....	50
4.2.2 Pulizia del Database.....	50
4.2.2.1 Pulizia delle recensioni in lingua non inglese	51
4.2.2.2 Pulizia di tipo temporale.....	52
4.2.3 Foglio overview	53

4.2.4 Creazione indice Kindness	56
4.2.4.1 Tipologia di indice Kindness.....	61
4.2.5 Scelte metodologiche	63
4.2.6 Altre Variabili di input.....	64
4.2.6.1 Variabili presenti nel dataset.....	65
4.2.6.2 Nuove Variabili introdotte.....	66
5. Modello Statistico.....	68
5.1 I modelli di regressione	70
5.1.1 I modelli di regressione semplici	73
5.1.1.1 Correlazione tra Rating complessivo e Tempo di risposta.....	73
5.1.1.2 Correlazione tra Rating Complessivo e Numero Delle Proprietà di un host.....	74
5.1.1.3 Correlazione tra Rating Complessivo e Numero dei bagni della proprietà	75
5.1.1.4 Correlazione tra Rating Complessivo e Numero dei letti della proprietà	76
5.1.1.5 Correlazione tra Rating Complessivo e Numero dei servizi della proprietà	78
5.1.1.6 Correlazione tra Rating Complessivo e numero delle parole di presentazione.....	79
5.1.1.7 Correlazione tra Rating e posizione dell'host	80
5.1.1.8 Correlazione tra Rating e numero di recensioni per mese	81
5.1.1.9 Correlazione tra Rating e Indice Kindness	82
5.1.1.10 Correlazione tra Rating e Indice Kindness * Superhost	84
5.1.2 Modelli divisi per tipologia di variabili.....	85
5.1.2.1 Correlazione tra Rating e variabili relative alla proprietà.....	85
5.1.2.2 Correlazione tra Rating e variabili relative alle performance dell'host ..	87

5.1.2.3 Correlazione tra Rating e variabili relative alla gentilezza dell'host	88
5.1.3 Modello finale: tutte le variabili	91
5.2 Limitazioni e indicazioni per studi futuri	97
6. Conclusioni	99
7. Bibliografia.....	102

1. Introduzione

Negli ultimi anni il fenomeno della Sharing Economy è diventato senza dubbio uno dei protagonisti dello scenario economico mondiale. Il successo dell'Economia collaborativa deriva principalmente dall'utilizzo sempre più frequente di piattaforme network in grado di connettere e soddisfare i bisogni di più categorie di utenti contemporaneamente. Detto per l'appunto two-sided market, il mondo delle piattaforme sharing sfrutta il desiderio di guadagno economico rendendo produttive quelle risorse in precedenza sottoutilizzate.

Questo lavoro di tesi si propone di indagare alcune caratteristiche proprie di questo nuovo modello economico, concentrandosi su uno dei modelli di maggior successo: Airbnb. Nata nel 2008 a San Francisco, la piattaforma internet peer-to-peer per eccellenza spopola in tutto il mondo, rivoluzionando il settore degli short-term rentals e imponendosi come principale competitor dell'intero sistema alberghiero.

Airbnb è diventata rapidamente oggetto di interesse da parte di ricercatori, imprenditori e non solo: la sua popolarità deriva in gran parte dalla possibilità, da parte degli utenti, di sfruttare ambienti domestici inutilizzati, ottenendo un guadagno "doppio", sia dal punto di vista economico, sia dell'esperienza personale.

Questo lavoro di tesi si propone di indagare i principali fattori che influenzano il sistema di valutazione utilizzato dalla piattaforma di Airbnb, in particolare andando a valutare l'effetto della gentilezza prodotto dal rating rispetto ad altre variabili collegate alla proprietà o all'host. Lo scopo dell'elaborato è infatti riuscire a capire se l'host possa migliorare la propria reputazione comportandosi con il proprio ospite in maniera gentile e disponibile, valutare l'impatto che avrebbe questo tipo di comportamento sul proprio rating e infine effettuare una comparazione tra differenti variabili per capire se effettivamente la gentilezza sia significativa oppure di minore importanza.

Per riuscire ad estrapolare l'effettivo impatto di questa caratteristica nella relazione host-guest è stata scelta come città campione Barcellona, località che si presta ad un'analisi dettagliata grazie al suo turismo diversificato, che comprende

un'ampia gamma di visitatori interni, esterni e il cosiddetto business tourism, in modo tale da avere una visione a 360 gradi.

È stato deciso di strutturare il lavoro partendo dall'analisi della letteratura dove è stato possibile rilevare alcune lacune nel sistema di rating di Airbnb, dovute alla presenza di alcuni bias che rendono poco veritiere le valutazioni finali degli alloggi. Il principale effetto di questi bias si riscontra nella distribuzione delle reviews, la quale si presenta eccessivamente sbilanciata agli estremi, in particolar modo verso l'estremo positivo, creando un grafico con una rappresentazione a forma di J, che verrà analizzata successivamente in dettaglio.

Inoltre, si procederà alla descrizione del dataset e del processo di creazione delle variabili in grado di poter esprimere il livello di gentilezza di un host nei confronti dei propri ospiti e alla predisposizione delle altre variabili da cui sarà valutato l'impatto sul rating complessivo e intermedio dell'host.

Infine, si procederà a illustrare la creazione dei modelli statistici per valutare l'impatto della gentilezza sul rating finale, sui rating intermedi e valutarne il livello di impatto rispetto ad altre variabili.

2. Analisi della letteratura

In una società sempre più globalizzata in cui i social media si sono fatti spazio nella vita quotidiana degli individui, irrompendo in ambiti inimmaginabili fino a pochi anni prima, la Sharing Economy sembra essere il decorso “naturale” delle tradizionali forme di economia sviluppatesi fino ad ora.

Nato nel 2008 negli Stati Uniti, questo nuovo modello commerciale esplose definitivamente nel 2013, grazie soprattutto allo sviluppo e alla diffusione di piattaforme network sempre più efficienti, in grado di essere usate da un numero sempre crescente di utenti.

Come sottolineato da Bernardi (Bernardi, 2015) ad oggi manca una definizione univoca di questo fenomeno, rendendo quindi complicato il tentativo di evidenziare nel dettaglio tutte le varie sfaccettature e caratteristiche che vengono toccate ed influenzate. Analizzando diverse definizioni di Sharing Economy emergono, tuttavia, alcune caratteristiche comuni, che permettono di identificare questa forma di mercato come un modello basato sulla condivisione e collaborazione tra individui, senza la necessità di intermediazione da parte di terzi. In un contesto commerciale di questo tipo, dove il rapporto tra le parti si svolge prettamente online, la fiducia e la reputazione diventano colonne portanti dell'intero meccanismo, rendendo necessario un sistema di rating efficiente che permetta ai consumatori di entrare in possesso di informazioni più o meno dettagliate relative al fornitore a cui si sta rivolgendo.

Luise e Chiappini (Luise e Chiappini, 2017), definiscono la reputazione come “una proprietà di un singolo individuo che può rappresentare la misura del valore individuale per come viene percepito all'interno della rete sociale”.

Hearn (Hearn, 2010) ancora sostiene che la reputazione sia “un attributo personale estremamente fluido, contingente e precario, generato interamente dalla percezione, dall'attenzione e dall'approvazione degli altri”.

Da queste definizioni è evidente come il sistema di rating utilizzato dalle piattaforme che sfruttano il concetto base di Sharing Economy, sia il punto cruciale per competere in un mercato in cui la concorrenza sembra non avere confini.

Per comprendere appieno l'importanza della reputazione, è necessario definire una delle principali forme di scambio contemporanee: l'approccio peer-to-peer. Questa categoria si presenta come un modello decentralizzato in cui gli individui coinvolti interagiscono, per l'appunto, in maniera paritaria, senza l'intermediazione di terze parti.

Diffusosi soprattutto grazie alle piattaforme network, il modello P2P viene introdotto negli anni '90 da eBay e Craigslist, piattaforme che consentivano a privati la condivisione o la commercializzazione di beni o servizi. L'efficienza di questo modello commerciale risiede nella possibilità da parte dei consumatori di recensire i fornitori in base al livello di gradimento raggiunto, sia attraverso sistemi di valutazione predefiniti (stelle), sia attraverso l'opzione di rilasciare un feedback più dettagliato, commenti e descrizioni che possono liberamente toccare vari aspetti del prodotto o servizio a cui si fa riferimento.

Questo sistema di reviews diventa quindi fondamentale sia dal lato della domanda, sia per quanto riguarda l'offerta; i consumatori hanno la possibilità di accedere ad informazioni rilasciate a seguito di un'interazione customer-supplier precedente, valutando così i possibili aspetti positivi e negativi della transazione, mentre i fornitori sono in grado di estrapolare eventuali punti di forza o debolezza. Quello che in linea teorica parrebbe un sistema efficiente, nella pratica riporta alcune criticità dovute non solo alla soggettività delle recensioni, ma anche ad un possibile uso strategico delle stesse; questi fattori lasciano spazio a possibili interpretazioni e generano un'esternalità di reputazione tra i venditori (C. Nosko e Tadelis, 2015).

2.1 Distribuzione distorta delle valutazioni

Airbnb, colosso delle transazioni peer-to-peer per eccellenza, non è esente da queste problematiche, seppure, nel panorama generale, abbia messo in atto qualche correttivo, per esempio rendendo le recensioni "blind". L'indagine proposta da Zervas et al (Zervas et al, 2015), attraverso un confronto tra TripAdvisor e Airbnb, mette in luce alcune differenze significative nella

distribuzione delle reviews, evidenziando come il sistema proposto da Airbnb presenti valutazioni mediamente molto alte rispetto all'altra piattaforma.

Questa distorsione stranamente positiva è stata oggetto di interesse di vari autori: in particolare Bridges e Vasquez (Bridges e Vasquez, 2016) hanno evidenziato come la distribuzione a J proposta per Amazon da Hu (Hu, 2009) si riscontri anche in Airbnb.

La distribuzione a "J" rappresenta il sistema di reviews in un grafico dove si formano due picchi, uno in corrispondenza della valutazione massima (5 punti nel caso di Airbnb) e uno lieve per quanto riguarda il livello minimo, da qui, appunto, la forma simile alla lettera J. Risulta quindi evidente che il sistema di rating della piattaforma di house-sharing presenti alcuni bias che rispecchiano l'inefficienza di quel meccanismo che in partenza pareva essere molto affidabile.

Prima di procedere all'analisi vera e propria, è opportuno indagare l'origine di questi bias, dimostrando come alcune caratteristiche proprie degli hosts di Airbnb possano influenzare più o meno positivamente le reviews rilasciate dai guests.

In primo luogo, Hu (Hu, 2009) identifica due bias tipici dei sistemi di reviews pubblici: il purchasing bias e l'under-reporting bias.

Il primo spiega l'eccesso di valutazioni estremamente positive come conseguenza del fatto che il consumatore, una volta presa la decisione di acquistare un prodotto o servizio, abbia la propensione a voler confermare la bontà della propria decisione valutando quindi la scelta effettuata come la migliore.

Come studiato anche da Lind et Al (Lind et al, 2017), successivamente all'acquisto di un prodotto vi è una naturale tendenza a rilevare e ricordare gli aspetti positivi minimizzando gli aspetti negativi, in modo da non sperimentare rimpianto verso un'eventuale scelta alternativa.

Per quanto riguarda invece l'under-reporting bias, è stato dimostrato come i consumatori siano più propensi a rilasciare un feedback in caso di estrema soddisfazione o estrema insoddisfazione, trascurando le esperienze considerate nella media: secondo questo principio si spiegherebbe quindi la scarsità di valutazioni intermedie, a vantaggio di quelle più estreme.

Come evidenziato da Dellarocas e Wood (Dellarocas e Wood, 2008), le cause dei reporting bias sono molteplici e complesse, spesso derivanti da fondamenti del

comportamento umano che inducono il consumatore ad essere restio, quando si tratta di “dare cattive notizie”, ad esprimersi, tesi confermata anche da Fradkin et al (Fradkin et al, 2020).

Una caratteristica specifica del sistema di rating di Airbnb, inoltre, è la cosiddetta rivelazione simultanea, introdotta dalla piattaforma a partire dal 2014. La scelta di questo nuovo meccanismo di rivelazione deriva dal tentativo di aumentare l’affidabilità del sito, eliminando le cosiddette “revenge reviews”, valutazioni effettuate al solo scopo di danneggiare il soggetto a cui ci si sta riferendo.

La rivelazione simultanea consiste nella pubblicazione delle rispettive valutazioni delle parti in causa, solo quando entrambe hanno effettuato la recensione. Questo meccanismo, come esposto da Fradkin et al (Fradkin et al, 2018), ha aumentato il numero di recensioni rilasciate, introducendo però un altro bias: la reciprocità. A sostegno di questa tesi, Fradkin et al (Fradkin et al, 2015) hanno dimostrato che la tendenza stranamente molto positiva della curva di reviews deriva anche da una tacita aspettativa di ricevere valutazioni reciprocamente positive, omettendo quindi informazioni che possano risultare sgradevoli.

I fattori distorsivi esposti fino ad ora non sono però le uniche fonti di pregiudizi: Bridges e Vasquez (Bridges e Vasquez, 2016) propongono come ulteriori cause alcune caratteristiche proprie del funzionamento di Airbnb. Nello specifico suggeriscono come le aspettative di partenza degli ospiti che si rivolgono a questo particolare sistema di house-sharing, siano mediamente inferiori rispetto a quelle dei consumatori che preferiscono grandi alberghi. Questo comporta inevitabilmente che sia molto più probabile che i guests rimangano “piacevolmente sorpresi”, implicando quindi una tendenza ad aumentare le valutazioni sulla base di aspettative spesso ampiamente superate.

Un altro fattore rilevante per la comprensione di questa distribuzione a J riportato dallo stesso studio è la mancanza di anonimato, in quanto tutte le recensioni presenti sono collegate al profilo dell’utente che le ha ricevute ed effettuate, sia per quanto riguarda gli hosts, sia per i guests. Questa policy di trasparenza di Airbnb è stata pensata per evitare reviews anonime, ritenute poco affidabili dai consumatori, ma viene considerata anche una delle cause della scarsità di recensioni negative, evidenziando la tendenza degli utenti a rilasciare valutazioni

positive quando possibile, e a mitigare quelle che potrebbero essere percepite come negative.

2.2 Importanza della relazione Host – Guest

È necessario ora ricercare le motivazioni che spingono gli utenti a scegliere di rivolgersi ai servizi offerti da Airbnb, analizzando nel dettaglio i fattori che influenzano il livello di soddisfazione post-transazione da loro raggiunto.

Liang e Joppe (Liang e Joppe, 2018) propongono una distinzione tra la soddisfazione dovuta alla transazione e la soddisfazione data dall'esperienza; la prima si riferisce al giudizio complessivo del consumatore per gli aspetti prettamente economici, mentre la seconda si concentra principalmente sul livello di gradimento raggiunto dal guest, tenendo in considerazione il confronto con le esperienze passate, il soggiorno e l'interazione con l'host.

A partire da questa distinzione sono state condotte diverse analisi sui fattori che spingono i consumatori a rivolgersi a questo tipo di servizio. Tra le motivazioni più evidenti come prezzo e posizione, Guttentag et al (Guttentag et al, 2017) identificano una particolare categoria di utenti che preferisce una sistemazione proposta da Airbnb, piuttosto che un albergo, per il valore aggiunto dato dall'interazione sociale con il proprio padrone di casa, definito un fattore pull. Questa categoria di persone, definita Collaborative Consumers, è attratta dalla possibilità di vivere un'esperienza locale autentica, grazie soprattutto al rapporto diretto con l'host, elemento importante per la soddisfazione data dall'esperienza riportata in precedenza.

Erose Sthapit e Jano Jiménez-Barreto (Erose Sthapit e Jano Jiménez-Barreto, 2018) hanno effettuato un'indagine prendendo campioni di popolazione diversi per età, sesso, provenienza e tipo di viaggio. La maggior parte degli intervistati ha menzionato la relazione con il proprio host come elemento fondamentale per la valutazione positiva dell'esperienza, sottolineando come la sua comunicazione e il suo atteggiamento abbiano influito sulla loro percezione del soggiorno. In questo studio viene esplicitamente chiesto agli intervistati se la loro ultima esperienza con Airbnb avesse coinvolto una forma di condivisione con l'host che li aveva

ospitati e le risposte sono state nella maggior parte dei casi positive. In particolare, gli intervistati hanno risposto con frasi del tipo “La nostra esperienza a Copenhagen è stata molto confortevole grazie al padrone di casa”, “Il nostro host Ivan ha reso la nostra esperienza molto memorabile” e ancora “In realtà, l'esperienza complessiva è stata davvero positiva, soprattutto grazie a Maria nostra padrona di casa”. Tutte le risposte che citavano il ruolo del padrone di casa sottolineavano come quest'ultimo avesse contribuito positivamente ai loro viaggi, offrendo consigli sulle attività turistiche locali, ristoranti tipici e talvolta mettendo a disposizione gratuitamente mezzi di trasporto come biciclette.

Sempre in questo studio viene riportato come spesso i guests tendano a rivolgersi ai padroni di casa direttamente con nomi propri e, quando questo avviene, il riferimento è sempre associato ad un commento positivo.

Infine, lo studio conclude distinguendo due tipi di relazione di condivisione tra host e guest: comunione e distribuzione. La dimensione della comunione si concentra sulla sfera dell'intimità, incorporando elementi come lo scambio di una storia o di un'emozione come un gesto di apertura o reciprocità. All'interno delle interviste, ogni riferimento alla disponibilità, gentilezza e ad atteggiamenti amichevoli, sono indicatori che sottintendono gesti di reciprocità propri di questa specifica categoria di relazione.

Per quanto riguarda invece la relazione distributiva, questa è stata definita dagli autori come la distribuzione o divisione di una risorsa tra individui con interessi economici comuni, intesa come la condivisione di un intero (“ad esempio una torta o una stanza”), oppure di oggetti materiali, emozioni o informazioni. Negli esempi riportati in precedenza è facilmente riscontrabile questa dimensione, evidenziata dal riferimento ad informazioni turistiche, luoghi di ristoro e condivisione di mezzi di trasporto.

Una ricerca effettuata da Arcidiacono et al (Arcidiacono et al, 2016) nell'ambito delle Sharing economy, ha riportato come “La dimensione di “informalità” che caratterizza il servizio sembra rendere gli utenti molto più tolleranti e comprensivi rispetto agli standard dello stesso”. Lo studio, inoltre, si spinge in una analisi dei lemmi ricorrenti all'interno dei feedback di Blabla Car, molto simile a quella che verrà affrontata nei capitoli successivi su Airbnb: termini come “puntuale”,

“ottimo”, “simpatico”, “piacevole”, “disponibile” e “gentile” si sono rivelati i più utilizzati dagli utenti. Procedendo con la lettura si ha un riscontro con quanto affermato in precedenza, ovvero i commenti associati a criticità come il ritardo sono legati ad apprezzamenti sull’atteggiamento del conducente o del passeggero a seconda dei casi, confermando quindi il contenimento nella percezione di possibili aspetti negativi.

Arcidiacono et al (Arcidiacono et al, 2016) concludono il loro lavoro sostenendo come il servizio offerto da BlaBlaCar, analogamente a quello offerto da Airbnb nel settore di house-sharing, fornisca una funzione definita “risocializzante”, fondamento dell’economia collaborativa. Viene inoltre precisato come l’utilizzo del servizio offerto dalla piattaforma vada al di là della semplice dimensione funzionale e dell’interesse individuale: la soddisfazione di bisogni che fanno riferimento ad aspetti pratici si lega a quella di fattori più impalpabili, di tipo più sociale, in questo modo si crea non solo un network digitale di transazioni tra sconosciuti, ma una vera e propria comunità, definita “Support Based”.

In altre parole, quello che inizialmente nasce come gruppo di persone che si aggrega a partire da un’esigenza utilitaristica e personale, sviluppa caratteristiche tipiche di una “community”: una “coscienza di specie”, visibile nei rapporti di fiducia e fidelizzazione alla piattaforma, una “condivisione di tradizioni e rituali”, fattore riscontrato nell’interazione tra gli utenti, e un “senso di responsabilità morale”, visibile soprattutto grazie alla pratica dei feedback.

La caratteristica di socialità è riscontrabile anche in Airbnb, dal momento che non esiste intermediario tra le parti in causa, ma, al contrario, host e guest si trovano spesso a condividere lo stesso appartamento. Questa condivisione degli spazi, che si traduce in un rapporto diretto, può quindi contenere la percezione di alcuni aspetti negativi che in situazioni più formali probabilmente non sarebbero tollerati.

L’analisi testuale delle recensioni di Airbnb proposta da Bridges e Vasquez (Bridges e Vasquez, 2016) ha dimostrato che nella maggior parte dei casi quando il riferimento all’host è fatto attraverso il nome proprio, il commento associato è

sempre positivo, anche quando la valutazione complessiva potrebbe risultare non pienamente soddisfacente.

Lo studio ha suddiviso 400 recensioni, metà scritte da ospiti e metà da hosts, secondo una logica semantica che colloca come positive tutte quelle recensioni che contengono aggettivi positivi come, ad esempio, meraviglioso, fantastico e ottimo, ponendo anche particolare attenzione laddove ci fosse una intensificazione degli avverbi (ad esempio molto, seriamente, decisamente) e un uso della punteggiatura che trasmettesse enfasi. Questa ricerca ha rivelato che l'aspetto menzionato più frequentemente riguarda caratteristiche ed atteggiamenti dell'host, seguito poi dalla pulizia e dal comfort dell'alloggio. Delle 200 reviews rilasciate dagli ospiti, 179 contenevano un diretto riferimento all'host, di queste 159 citavano il padrone di casa usando il suo nome di battesimo. Una ricerca diretta della parola "host" ha mostrato che gli hosts sono tipicamente descritti con un piccolo insieme di parole e frasi positive, spesso con riferimenti ai loro comportamenti e al loro stile di comunicazione (ad esempio accoglienti, amichevoli, premurosi, accomodanti, comprensivi, pazienti).

Lo studio prosegue poi con l'analisi delle recensioni negative: non sorprendentemente solo 27 feedback sono stati classificati come negativi, di cui 20 derivanti dagli ospiti. È importante sottolineare come la maggior parte delle recensioni classificate come negative, in realtà fossero principalmente positive, con l'aggiunta di qualche lamentela accanto a commenti positivi, lasciando solo otto recensioni su 400 (corrispondente al 2%) categoricamente negative.

Gli autori hanno mostrato che la maggior parte delle recensioni negative tende ad iniziare e finire con commenti positivi, inserendo la lamentela sotto forma di consiglio o comunque in forma indiretta.

Dai dati riportati fino a questo momento parrebbe che circa il 98% degli utenti di questa indagine abbia reputato la propria esperienza complessivamente soddisfacente, tranne qualche caso isolato che ha arricchito il proprio feedback con consigli per un miglioramento del servizio o della proprietà. Tuttavia, procedendo ad una scomposizione semantica di alcune reviews, gli autori hanno scoperto che alcune recensioni inizialmente classificate come positive, in realtà lasciavano trapelare solo una parziale soddisfazione del recensore. Queste

recensioni, definite “tiepide”, si caratterizzano per una lunghezza inferiore rispetto alle altre, nonché una scelta linguistica che non comprende l’uso degli aggettivi e degli avverbi sopracitati, nonostante l’utilizzo di termini come “buono” o “ok”. La ricerca conclude affermando che, spesso, il modo in cui un'esperienza non positiva viene comunicata in questo particolare contesto di recensione può avere più a che fare con ciò che non viene dichiarato, piuttosto che con ciò che viene dichiarato in un particolare messaggio. Naturalmente, una tale strategia di comunicazione è interpretabile solo da un pubblico di lettori che è in grado di leggere "tra le righe", di cogliere spunti sottili (come, ad esempio, la mancanza di alcuni segnali linguistici) e, in ultima analisi, di distinguere tra una recensione genuinamente positiva e una recensione apparentemente positiva che in realtà descrive un'esperienza non del tutto positiva.

Da questo studio si evince come il rapporto host-guest sia determinante per il raggiungimento di una totale soddisfazione da parte del consumatore, e come questa relazione porti ad un bias di positività laddove l’atteggiamento del proprietario sia stato molto gradito, anche se le caratteristiche oggettive dell'alloggio, avrebbero magari portato valutazioni più severe. In ogni caso i commenti negativi nella maggior parte dei casi sono riferiti alla proprietà o alla posizione, e molto raramente si riferiscono all’host.

Attraverso un’indagine statistica Fradkin et al (Fradkin et al, 2018) hanno dimostrato come la probabilità di un host di ricevere valutazioni estremamente positive sia funzione della quantità e della qualità di interazione sociale con i propri ospiti, tesi confermata anche da Alsudais (Alsudais, 2017). A partire da questa ipotesi gli autori hanno inoltre riscontrato che, a condizione che non ci siano raccomandazioni, il tasso di valutazioni massime si riduce per l’affitto di intere proprietà e per quegli host che gestiscono un numero di annunci superiore a tre. Questo accade perché, nei due casi sopra citati, la probabilità di interazione tra ospiti e padrone di casa si riduce notevolmente, confermando ancora una volta l’importanza della relazione host-guest. In particolare, in questo studio, sono state analizzate oltre mille reviews per trovare alcuni pattern ricorrenti che potessero fornire informazioni sul rapporto che si è instaurato tra guest e host in seguito all’esperienza. Per fare ciò, si è voluto analizzare nello specifico come il

guest si rivolgesse all'host nelle recensioni. Si evidenziano in particolare due tipi di pattern ricorrenti:

1. Riferimento all'host per il suo ruolo: vengono rilevate all'interno delle reviews parole come "proprietario", "signore", "donna" e tutte quelle parole che identificano l'host per il ruolo formale che ricopre;
2. Riferimento all'host per il suo nome di battesimo: qua vengono rilevati all'interno delle reviews il nome proprio dell'host oppure soprannomi o nomignoli amichevoli.

Alsudais conclude quindi il suo studio sottolineando che "le scoperte fatte suggeriscono che le interazioni offline tra guest e host sono estremamente importanti come comprovato dalla grande quantità di riferimenti espliciti al nome dell'host all'interno delle reviews" evidenziando il limite per cui "(...) è possibile che i guest più propensi a scrivere le reviews siano quelli più socievoli, e per quello potrebbero aver valutato le interazioni offline più importanti rispetto coloro che invece non hanno scritto recensioni"

Le considerazioni esposte fino a questo momento si concentrano sull'interazione personale tra individui, dimostrando come questa sia fondamentale nella fase di rating successiva al soggiorno.

Lee et al (Lee et al, 2015) hanno improntato la loro indagine sugli effetti delle caratteristiche sociali e sulla propensione all'acquisto degli utenti che si rivolgono a piattaforme Sharing, soffermandosi sugli elementi che influenzano i consumatori nella fase pre-acquisto. In primo luogo, hanno scoperto che, in un primo impatto, gli utenti sono condizionati dalla quantità di recensioni piuttosto che dalla media delle valutazioni. Questa considerazione apparentemente sorprendente potrebbe essere una conseguenza della tendenza straordinariamente positiva dei punteggi presenti sulla piattaforma, situazione che abbassa il potere distintivo di questi ultimi. In aggiunta a questa informazione, gli autori sostengono che la reattività dell'host, intesa come tasso e rapidità dei tempi di risposta, sia una componente importante nella fase di scelta dell'alloggio, percepita dal cliente come una dimostrazione di disponibilità da parte del padrone di casa.

Questo breve focus sulle interazioni online è utile per comprendere appieno tutti i fattori in grado di influenzare la percezione della qualità del servizio da parte dell'ospite, processo che ha origine nella fase precedente al soggiorno e che è in grado di generare pregiudizi positivi o negativi, che si riflettono nel rating post-consumo.

2.3 Dai feedback al valore aggiunto

Gli elementi analizzati nei paragrafi precedenti sono necessari per comprendere il contesto e le influenze che caratterizzano le piattaforme basate sulla Sharing Economy, in cui il concetto di valore assume un significato più ampio rispetto ad un mercato tradizionale.

La potenza di questo sistema economico risiede nella possibilità da parte dei fornitori di dirottare il proprio capitale di reputazione accumulato, in prezzo d'acquisto (Lee et al, 2015).

Per giungere al cuore di questa indagine è doveroso definire alcuni brevi concetti economici utili per la piena comprensione del potere che le caratteristiche sociali assumono in questo specifico settore. In economia, il valore aggiunto è una misura dell'incremento di valore dell'output dato dalla manipolazione e dall'attività produttiva dell'impresa, in grado di influenzare il prezzo di vendita. La disponibilità a pagare, in un contesto macroeconomico, rappresenta l'importo massimo che l'insieme dei consumatori è disposto a versare per il consumo di un bene o un servizio. I due indici appena descritti sono rispettivamente incorporati nelle funzioni di offerta e domanda, la cui intersezione definisce, per lo meno a livello teorico, il prezzo di mercato.

Si può quindi dedurre che il prezzo finale di un bene o servizio è influenzato dal valore aggiunto dato dall'impresa che, a parità di quantità, ha una relazione direttamente proporzionale al prezzo; specularmente, la disponibilità a pagare dei consumatori si riscontra in maniera diretta nella curva di domanda, dando anche un'indicazione sulla percezione che essi hanno del valore aggiunto.

In uno scenario in cui la condivisione diventa l'elemento base delle transazioni, Zhang et al (Zhang et al, 2018) espongono come siano coinvolti vari aspetti di

“valore”, a partire da quello monetario, fino ad arrivare a quello esperienziale, aspetti che non solo coinvolgono proprietari e utenti, ma derivano proprio dalla loro collaborazione. In questo studio vengono riportate diverse prospettive utili a questa indagine riguardo la co-creazione di valore: vengono proposti modelli basati sull’interazione diretta tra cliente e fornitore che comprendono il dialogo, l’accesso, i benefici del rischio e la trasparenza, mettendo in evidenza cinque pratiche (informare, salutare, consegnare, addebitare e aiutare) fondamentali per la creazione di un valore interattivo. Tutte le categorizzazioni presenti in questo articolo propongono sfumature diverse dello stesso concetto comune: la co-creazione del valore è strettamente legata all’interazione diretta tra supplier e consumer. Questo studio definisce il valore come “un’esperienza di consumo interattiva”, discutendo la sua co-creazione in tre distinte fasi: pre-consumption, mid-consumption e post-consumption.

Nello stadio che precede il consumo, le interazioni la guest e host sono in grado di influenzare la percezione del valore dei consumatori sotto vari aspetti, partendo dalla comunicazione. Per quanto possa risultare banale, una buona comunicazione è in grado di trasmettere all’interlocutore gentilezza e ospitalità, fattori chiave per la costruzione di un valore emotivo e di una relazione interpersonale. Durante questa fase di interazione online il consumatore è anche in grado di estrapolare informazioni utili per garantire che i servizi disponibili rispecchino tutte le esigenze del cliente, migliorando in via indiretta il valore funzionale.

Infine, gli host possono rafforzare il valore sociale che i consumatori attribuiscono ai servizi, segnalando che le sistemazioni abitative sono state precedentemente condivise con persone con cui i clienti si possono identificare (vengono riportati esempi come amanti degli animali e attivisti ambientali, categorie molto comuni tra gli individui che scelgono di rivolgersi ad un’economia di condivisione).

Durante il consumo, nella fase intermedia, l’ospite tende a percepire i valori funzionali del servizio acquistato, senza dimenticare che la cosiddetta “hospitality industry” coinvolge aspetti che vanno al di là del semplice funzionamento: in particolare la disponibilità e l’atteggiamento positivo dell’host contribuiscono alla creazione di valore emotivo percepito dal customer durante l’esperienza.

L'ultimo step di questo processo, il post-consumption stage, racchiude il significato complessivo di valore, definito come un concetto dinamico che incorpora aspetti funzionali, emotivi e sociali. Secondo Sánchez et al (Sánchez et al, 2006) i clienti sono in grado di quantificare e valutare la percezione di valore ricevuto nel suo complesso solo a seguito del consumo. La valutazione rilasciata in seguito ad un soggiorno può essere considerata come espressione del valore funzionale finale percepito dal cliente, mentre le recensioni testuali rispecchiano i valori emotivi e sociali accumulati prima, durante e dopo l'esperienza. In particolare, si riscontra come l'interazione e la percezione dell'approvazione sociale dopo l'uso di Airbnb siano associati al valore sociale costruito durante il soggiorno.

La distinzione di queste tre categorie di valore ha permesso di indagare quale di esse avesse la maggior influenza sulla disponibilità a pagare un Premium Price da parte dei consumatori. I risultati hanno confermato che il condizionamento predominante in tutte le fasi del consumo deriva dal valore sociale, che a sua volta ha origine dalle interazioni e dal rapporto che si instaura tra ospite e proprietario di casa.

A questo punto risulta chiaro come alcune caratteristiche personali degli host possano influire sulla percezione finale dell'esperienza del consumatore che, in caso positivo, si può tradurre nella disponibilità di quest'ultimo a pagare un prezzo superiore.

L'analisi che segue andrà a rilevare come la gentilezza sia un fattore determinante nella creazione di valore concreto e quantificabile in termini di valore.

3. Descrizione dei Database di partenza

L'analisi che seguirà nei capitoli successivi si sviluppa a partire da un set di dati di partenza: in questo capitolo si andrà ad investigare come sono stati ottenuti i dati, le variabili a disposizione (anche se non necessariamente utilizzate per costruire il modello statistico, per poter semplificare le analisi future e dare una visione generale di più ampio respiro) e si farà una panoramica dei fogli di lavoro che rappresentano la base dell'analisi.

Si è scelto di utilizzare come strumento di sorgente dati, per la costruzione del modello statistico la piattaforma Inside Airbnb (www.insideairbnb.com).

Accedendo alla piattaforma Inside Airbnb si ha la possibilità di raggiungere svariati set di dati aggregati in formato tabellare Microsoft Excel Comma Separated Values File (ossia tipologia di file ".csv"). Per selezionare la tabella utile, si è partito dalla selezione della città: la prima spaccatura di insideairbnb.com è infatti quella del luogo geografico di ubicazioni delle proprietà divisa per città in ordine alfabetico. La piattaforma fornisce inoltre la possibilità di visualizzare all'interno di una mappa ogni singolo annuncio.

È stato scelto di basare l'analisi sui dati degli annunci di Barcellona; questa decisione è stata dettata dal fatto che la città in questione si presta ad avere utenti della piattaforma Airbnb che sono sia business travellers internazionali, sia turisti provenienti da ogni parte del mondo: la vastità e diversità dei guests di cui saranno analizzate le recensioni, eliminerà (o quantomeno renderà poco rilevante) eventuali bias culturali tipici di un turismo interno alla nazione.

Inoltre, per ogni città, è possibile accedere all'archivio storico dei dati: è stato scelto di lavorare con i dati più recenti a disposizione a partire dal 2015 fino a febbraio 2020.

Una volta che è stata definita la località dove effettuare l'indagine e si è definito il periodo di riferimento per effettuare l'analisi dei dati, si hanno a disposizione i seguenti pacchetti di informazioni:

— ***"Detailed Listings"***:

in questo campo è possibile rilevare tutte le informazioni disponibili suddivise per proprietà. Questo set di dati, essendo il punto di partenza per l'analisi dei capitoli successivi, sarà descritto nei paragrafi che seguono.

- ***“Detailed Calendar Data”:***
 - Qui è possibile reperire l'informazione riguardo la disponibilità di ogni appartamento, il prezzo e il numero di notti massimo e minimo consentiti dall'host
- ***“Detailed Review Data”:***
 - questo sarà il file al centro dei calcoli statistici. In questa tabella sono infatti contenute tutte le recensioni lasciate dai guest suddivise per ogni proprietà e per ogni host
- ***“Summary information and metrics”:***
 - qua sono state raccolte per ogni post le informazioni relative alla posizione geografica, ed altre metriche utili come il numero di recensioni totali ricevute e il numero di recensioni al mese medie lasciate dai guests
- ***“Summary Review data and Listing ID”:***
 - per semplificare le analisi temporali viene riportato per ogni ID di ogni casa, la data di iscrizione alla piattaforma
- ***“Neighbourhood list”:***
 - per semplificare le analisi basate sulle posizioni delle case, viene presentato il quartiere di appartenenza e ogni sotto-quartiere
- ***“GeoJSON file of neighbourhood”:***
 - qui viene riportato lo stesso file precedente, ma in formato GeoJSON per effettuare analisi di tipo “Geo-based” più avanzate

Il foglio di lavoro iniziale è composto inizialmente da due fogli contenenti due tabelle di quelle sopra menzionate:

- il primo, che sarà il cuore pulsante del modello statistico, è il Data Base “Detailed Review Data”

- il secondo, che sarà utilizzato come Data Base accessorio al primo, è il file “Detailed Listings”

Ora si andrà ad analizzare in dettaglio la composizione e la granularità dei dati di partenza, evidenziando le peculiarità più importanti delle variabili più rilevanti.

All'interno del primo foglio del file excell, ogni riga contiene una recensione fatta da un guest: essendo ordinate per host le prime recensioni faranno riferimento al primo annuncio e così via.

Si può notare che il Database di partenza presenta numerose recensioni in lingue diverse dall'inglese e numerosi caratteri sconosciuti non riconosciuti da Excel.

Il primo Database utilizzato comprende una serie di informazioni utili per l'analisi che verrà svolta nel capitolo successivo, dove l'obiettivo sarà quello di ricercare quanti più segnali diretti ed indiretti della gentilezza dell'host.

Da questo set di dati è possibile identificare un ID associato alla proprietà del padrone di casa ospitante e al suo rispettivo nome di battesimo.

Per quanto riguarda la recensione, oggetto di interesse di questo lavoro, in questo file Excel sono esplicitati il testo completo della review, l'ID ad essa associato, il nome del guest che l'ha rilasciata con rispettivo ID cliente e infine la data in cui è stata scritta.

Queste informazioni sono il punto di partenza di questa indagine, la quale si propone di rielaborare i dati a disposizione per comprendere il ruolo della gentilezza all'interno dei servizi offerti dalla piattaforma, e la sua possibile monetizzazione da parte dell'host.

Per quanto riguarda il secondo foglio di lavoro, “Detailed Listings”, è necessaria un'analisi più approfondita, in quanto contiene le principali variabili rilevanti per il fine di questa indagine; a tal scopo si è deciso di descrivere i set di dati disponibili in questo file nel loro complesso, seppur non tutte le colonne saranno riprese nel capitolo successivo. Questa scelta è stata adottata per permettere una visione a 360 gradi delle informazioni disponibili riguardanti la piattaforma di Airbnb, in modo tale da illustrare come le discriminanti della gentilezza siano state estrapolate da un Database completo.

Mentre il database “Review Data” era principalmente incentrato sui dettagli relativi alla recensione e al suo autore, la base di dati “Detailed Listings” si propone di indagare le caratteristiche dell’host e del listing che Airbnb immagazzina, cataloga e rende disponibili a chiunque abbia la necessità di prenderne visione.

Procedendo con ordine a partire dalla prima colonna a sinistra del file, e proseguendo la lettura verso destra, si riscontrano tutta una serie di informazioni relative ad un padrone di casa, alla sua proprietà inserita nella piattaforma di house sharing e ai servizi aggiuntivi e non che esso offre.

Le prime due colonne sono relative rispettivamente al codice identificativo della proprietà e al nome dell’host, dettagli già riscontrati nel precedente foglio di lavoro analizzato; in aggiunta viene proposta una categoria che esplicita la data a partire dalla quale gli utenti hanno deciso di mettere a disposizione una o più proprietà, vale a dire la data in cui gli individui in analisi sono diventati host di Airbnb.

Il database prosegue poi con l’identificazione della locazione del proprietario di casa: è interessante notare che, contrariamente a quanto ci si aspetterebbe vista la scelta di focalizzarsi solo sulla città di Barcellona, non tutti gli host risultano residenti nei pressi o nei confini del centro abitato in questione. Scorrendo le diverse locazioni infatti, si incontrano diverse città della Spagna, più o meno vicine a Barcellona, ma saltano all’occhio anche residenze che si differenziano addirittura per lo stato, come ad esempio Leeds in Inghilterra (riga 18, host Cristina), Zurigo in Svizzera (riga 58, host Reinhard), Amsterdam in Olanda (riga 113, host Kitty) e molte altre. Per questa specifica categoria di ospitanti è evidente come il contatto diretto con i guest sia limitato alle sole interazioni online, non essendo possibile nella maggior parte dei casi un incontro personale delle due parti. Questo fattore di fatto limita la possibilità degli ospiti di accrescere la propria soddisfazione data dall’esperienza, in quanto, come è stato esposto nel capitolo precedente, uno degli elementi principali per il raggiungimento del massimo valore di soddisfazione complessiva è proprio la condivisione di alcuni aspetti della vita giornaliera e la relazione interpersonale che ospitato e ospitante

sviluppano, attraverso scambi di informazioni, consigli su come fruire al meglio il territorio, scambi di visioni e rappresentazioni della vita stessa.

Procedendo con la lettura si giunge poi ad una delle classificazioni più interessanti del file: la presentazione dell'host.

Questa categoria di dati permette di svolgere un'analisi preliminare più superficiale, soffermandosi quindi solo sulle scelte linguistiche e sulle caratteristiche che l'host ritiene rilevanti nel dare una prima impressione di sé stesso.

Per comprendere nel dettaglio come queste presentazioni siano utili per una prima scrematura di dati, è necessario riportare alcuni esempi, anche contrapposti, per identificare alcuni elementi che esprimano una propensione del padrone di casa alla condivisione. Data l'enorme disponibilità del database, si è scelto di riportare un campione di descrizioni casuali, senza applicare un particolare metodo discriminatorio, in modo tale da dimostrare l'evidenza dei diversi approcci personali degli individui, fattore risultato fondamentale durante la fase decisionale di una sistemazione offerta da Airbnb.

Scorrendo le varie presentazioni degli host è già evidente come queste si differenzino sia per la lunghezza, sia per le informazioni in esse presenti, e sia per la scelta dell'autore di scrivere in lingua inglese piuttosto che nella propria lingua madre (nella maggior parte dei casi spagnolo).

La policy della piattaforma lascia ai singoli host la decisione di lasciare o meno una propria descrizione, rendendola quindi volontaria e non obbligatoria. La caratteristica "opzionale" adottata da Airbnb di per sé può essere ritenuta un primo segnale di disponibilità ed interesse da parte del padrone di casa. Non è infatti scontata la scelta da parte dell'host di rilasciare un'auto descrizione, come si può riscontrare nelle righe 19, 23, 31, etc., che risultano prive di contenuto.

È opportuno sottolineare come la mancanza di testo in alcune celle non implichi necessariamente una scarsa propensione alla disponibilità da parte dell'host a cui sono associate, ma sicuramente è un'informazione (mancante in questo caso) di cui il consumatore tiene conto durante la scelta di un appartamento piuttosto che un altro.

Un discorso analogo può essere trasposto per quegli host che decidono di non effettuare una propria descrizione completa, ma preferiscono lasciare un semplice saluto (ed esempio “hi” riga 20, “Bienvenida” riga 733) o in alternativa specificare solo il loro tipo di occupazione (“trabajador” riga 718).

Sono poi presenti alcuni utenti che si possono definire come una via di mezzo tra coloro che preferiscono lasciare la propria presentazione per un momento futuro, e quelli che invece scelgono di esporsi maggiormente fornendo una descrizione più dettagliata.

Alcuni esempi di questa categoria sono riscontrabili alla riga 43, in cui l’host si descrive con tre aggettivi (servicial, alegre, amable), alla riga 734, la quale riporta il seguente testo “persona responsable, seria, agradable y sociable”, oppure ancora nella riga 744, in cui l’utente decide di descrivere sé stesso con il termine “equilibrio”.

Analizzando il Database senza applicare un metodo particolarmente critico, si riscontra che la maggior parte delle descrizioni dei padroni di casa siano mediamente superiori a quelle riportate in precedenza; è stato evidenziato infatti come gli host spesso si dilunghino in racconti di esperienze passate, interessi personali e talvolta spiegazioni delle motivazioni che li hanno spinti ad approcciarsi al modo dell’house sharing, piuttosto che soffermarsi su una visione superficiale dell’individuo, permettendo agli utenti interessati di entrare in possesso di possibili curiosità comuni, oltre al fatto di poter conoscere il sistema di valori e più in generale un atteggiamento verso la vita del padrone di casa.

Alla riga 25, ad esempio, Nacharosa inizia con una brevissima descrizione della propria occupazione, “Soy economista, vivo en Barcelona, y me dedico al mundo de las finanzas (...)”, per poi proseguire con “(...) El hecho de recibir estudiantes y trabajadores en mis apartamentos, me hace sentir muy feliz, ya que me encanta intercambiar vivencias con personas de otros países”, sottolineando quindi la sua predisposizione ad instaurare un rapporto con i propri guest che vada oltre la mera transazione economica. In un’epoca in cui l’utilizzo dei social media ci consente continue incursioni nelle vite degli altri, nelle loro opinioni, oltre a proporre modelli di vita e di pensiero attraverso gli esempi dei cosiddetti “influencer”, gli utenti sono abituati, anche utilizzando una piattaforma di questo

tipo, a ricercare informazioni più personali, che fanno riferimento alla sfera valoriale del padrone di casa.

Tornando alle informazioni che si trovano nelle presentazioni, si trova, ad esempio, Lisette, riga 35, che scrive: “Hi! I'm a friendly and tranquil person, and Bimba is a sociable kitten. We will both do our best to make you feel welcome during your stay.

Passionate about food, art and architecture, I'm knowledgeable about this wonderful city, and able to provide you with helpful tips as well as direct you to the museums, restaurants, bars, nightlife, etc. that might be of interest to you.

You should have no problem communicating with me, for I speak fluent English, Spanish and Dutch, as well as some German, French, Catalan and a smattering of Italian. Should these prove insufficient, there's always hands and feet!”.

Oppure Albert, riga 424, si racconta tramite il seguente messaggio: “I am a 37 years old guy from Barcelona. I love cooking, nature, photography, traveling and meeting new people.”

Marta e Paul alla riga 757 dichiarano “Hello to everyone :) As owners we love being able to offer our apartments during your experience in the wonderful city of Barcelona. We welcome all cultures, religions and tastes. For us there is no discrimination but respect. Perfection does not exist but we strive to offer good conditions for your stay to be very positive. The only thing we like to ask is for our guests to be happy and positive. See you soon!”

Questi tre esempi mostrano i diversi approcci alla descrizione degli host, evidenziando come ogni individuo reputi utile specificare caratteristiche diverse rispetto ai concorrenti, focalizzandosi chi più sui possibili interessi comuni, chi sull'opportunità di comunicare in diverse lingue per facilitare l'interazione, e chi invece preferisce sottolineare la propria apertura mentale.

In generale, e non solo negli esempi sopracitati, si nota la volontà dei padroni di casa di fondare i rapporti con gli ospiti secondo quei criteri di condivisione e cooperazione tipici della Sharing Economy, che si innesta, come già detto, in un utilizzo più generale della rete per tanti aspetti della vita relazionale ed emotiva degli individui.

Tornando ora all'analisi del database si incontrano due colonne relative ai tempi e ai tassi di risposta degli host, seguiti poi dalla percentuale di accettazione dei guest da parte degli ospitanti. Il sistema propone successivamente la classificazione dei superhost, definiti dallo stesso Airbnb come "host esperti che rappresentano un eccellente modello da seguire per altri host".

In altre parole, i superhost sono quei padroni di casa che rispecchiano le seguenti caratteristiche:

- Hanno completato almeno dieci viaggi oppure tre prenotazioni per una durata complessiva di almeno cento notti.
- Hanno mantenuto un tasso di risposta pari almeno al 90%.
- Hanno mantenuto un tasso di cancellazione non superiore all'1% (1 cancellazione ogni 100 prenotazioni), fatta eccezione per i casi che rientrano nei Termini delle circostanze attenuanti di Airbnb.
- Hanno mantenuto una valutazione complessiva di 4,8 (facendo riferimento alle recensioni dei 365 giorni precedenti, in base alla data in cui ciascun ospite ha scritto la recensione e non a quella in cui ha effettuato il check-out).

Prima di arrivare alla catalogazione degli appartamenti, il sistema specifica tramite una colonna il numero di annunci disponibili dello stesso host presenti sulla piattaforma.

A questo punto il database inserisce una serie di informazioni sulla proprietà, partendo dalla definizione per tipo (appartamento, cottage, loft, casa indipendente, etc.), specificando la tipologia di camera laddove la sistemazione prevista si trovi all'interno di una proprietà condivisa, il numero di ospiti ammessi, il numero di bagni, camere da letto, posti letto con relativo modello.

Successivamente vengono elencati i benefit disponibili nella proprietà, come ad esempio l'accesso al Wi-Fi, la presenza di una o più televisioni, l'aria condizionata, l'ascensore, la lavatrice, il ferro da stiro e così via. In aggiunta l'host può decidere di specificare le dimensioni dell'alloggio che sta affittando, essendoci una sezione opzionale riguardante la metratura.

Il database pone poi in rilievo il prezzo, in particolare il prezzo per una notte, per una settimana e per un mese, laddove siano previsti degli sconti in base alla durata del soggiorno. Gli host hanno inoltre la possibilità di richiedere una cauzione, definita nella colonna "Security Deposit"; come si nota dal database, la definizione dell'ammontare del deposito cauzionario viene lasciata agli utenti ospitanti, i quali possono anche decidere di non richiedere nessun importo di precauzione.

Oltre a queste informazioni viene anche segnalata la presenza di spese aggiuntive per la pulizia al termine del soggiorno, che non sempre è inclusa nel prezzo finale della sistemazione; viene poi specificato il numero di persone extra che possono essere accolte nell'alloggio, il prezzo cadauno, il minimo e il massimo di notti richieste per la prenotazione.

A questo punto insideairbnb.com segnala, per ogni proprietà, la disponibilità immediata (in data 16 Marzo 2020) e in aggiunta offre una panoramica dell'ultimo anno, specificando anche i giorni in cui l'host ha reso accessibile la stanza o l'abitazione in questione, isolando rispettivamente i dati relativi agli ultimi 30, 60 e 90 giorni.

Le informazioni riportate fino a questo momento riguardano principalmente l'host, la sua proprietà e le tariffe, elementi sicuramente attrattivi a primo impatto, ma non sono sufficienti per effettuare una scelta consapevole dell'alloggio da affittare.

Le dodici colonne che seguono si riferiscono all'oggetto maggiormente discusso e citato nel capitolo precedente: le reviews.

Nel primo foglio di lavoro, "Detailed Review Data", sono stati riportati i testi delle recensioni e il loro ID; in questo secondo foglio invece, vengono specificati il numero totale di revisioni legate all'alloggio, definendo quante di esse siano state rilasciate negli ultimi dodici mesi, la data della prima e dell'ultima recensione. Successivamente viene esposto quello che viene chiamato "review score rating", vale a dire un punteggio complessivo della proprietà e dell'esperienza di soggiorno, estrapolato dalla manipolazione dei dati riportati in seguito, basati sul

sistema di valutazione scelto da Airbnb, il quale chiede ai suoi utenti di valutare separatamente pulizia, precisione e accuratezza nella descrizione dell'alloggio nell'annuncio, valore, inteso come rapporto qualità-prezzo, comunicazione con l'host prima e durante il soggiorno, check-in, posizione, la quale include anche la vicinanza e l'accesso ai trasporti pubblici o al centro città, e la sicurezza per la famiglia, e infine i servizi offerti.

Dopo aver analizzato questa serie di dati prettamente riguardanti il rating della proprietà da parte degli ospiti, il sistema specifica per ogni annuncio della città di Barcellona la policy di cancellazione scelta dal padrone di casa. Airbnb offre agli individui che decidono di usufruire della piattaforma mettendo a disposizione i propri immobili, la possibilità di scegliere tra sei metodi di cancellazione che differiscono per tolleranza, partendo da una policy flessibile, fino ad arrivare ai termini più rigidi e vincolanti.

Gli ospiti intenzionati a rivolgersi ad un host piuttosto che un altro hanno le risorse necessarie per conoscere e comprendere le condizioni di cancellazione della prenotazione direttamente dall'annuncio, entrando quindi in possesso di queste informazioni prima di una eventuale comunicazione con il proprietario di casa.

Una volta definito l'immobile prescelto, il guest entra in contatto con l'host, il quale, a sua discrezione, può richiedere una serie di informazioni aggiuntive riportate nelle colonne seguenti, come una foto del soggetto intenzionato ad effettuare la prenotazione, ed una verifica telefonica.

Le ultime informazioni riportate in questo foglio Excel specificano, per ogni host relativo a una proprietà, eventuali altri immobili posseduti dallo stesso, suddivisi in intere case indipendenti, stanze private all'interno di una proprietà condivisa, e infine gli eventuali locali affittabili esclusivamente in condivisione con altri.

4. Analisi dati

In questo capitolo verrà proposta un'analisi dettagliata delle informazioni presenti nei database descritti precedentemente, secondo la seguente struttura:

- Analisi delle variabili di interesse: l'host
 - Ubicazione
 - Tempo di risposta
 - Tasso di risposta e tasso di accettazione
 - Superhost
 - Presentazione dell'host
 - Numero di annunci per host
 - Tipologia di annuncio
 - Anno di iscrizione
- Variabili del modello di regressione

4.1 Analisi Delle Variabili Di Interesse: L'host

Per cercare di capire il principale attore del sistema Airbnb, di seguito verranno fatte alcune analisi sulle variabili a nostra disposizione riguardanti l'host. Infatti, poiché si andrà successivamente ad investigare il ruolo della gentilezza nel rapporto tra guest e host, è importante studiare ed approfondire le caratteristiche del soggetto più incentivato ad offrire ai propri ospiti un'esperienza sociale, oltre che al semplice servizio di affitto della stanza.

È importante sottolineare come non tutte le variabili analizzate in questo capitolo saranno utilizzate all'interno del modello statistico, tuttavia è estremamente importante avere una panoramica completa degli elementi presenti, per meglio comprendere tutte le sfumature della piattaforma e per poter permettere e facilitare analisi future.

I dati che verranno presentati di seguito sono stati raccolti dal foglio Overview del foglio di calcolo in allegato, che sarà poi il punto di partenza per l'analisi statistica. Poiché all'interno del Database iniziale, ogni riga contiene un annuncio (o *listing*) e poiché molti host offrono più di una proprietà, è stata necessaria un'attività di

pulizia dei dati, rimuovendo gli annunci multipli in modo tale che ogni host fosse presente solo una volta (solo per questo tipo di analisi, successivamente saranno utilizzate tutte le righe).

Per la città di Barcellona sono 10365 gli host attualmente attivi, per un totale di 21116 annunci. In questo paragrafo si analizzeranno i dati relativi al proprietario: residenza, tempo di risposta medio e tasso di accettazione/risposta, numero e tipologia di annunci, descrizione, se è stato identificato come Superhost da Airbnb e anno di iscrizione.

4.1.1 Ubicazione dell'host

Si riportano di seguito le tabelle contenenti la divisione degli host per luogo geografico: prima seguendo una divisione per stato, poi seguendo una divisione per regione spagnola.

Stato	Numero	Percentuale
Spain	9325	90,11%
France	142	1,37%
United Kingdom	141	1,36%
United States	121	1,17%
Italy	77	0,74%
Germany	72	0,70%
Russia	65	0,63%
Argentina	53	0,51%
Netherlands	29	0,28%
Switzerland	23	0,22%
Brazil	21	0,20%
Sweden	20	0,19%
Mexico	19	0,18%

Stato	Numero	Percentuale
Belarus	3	0,03%
Ukraine	3	0,03%
Panama	2	0,02%
New Zealand	2	0,02%
Uruguay	2	0,02%
Latvia	2	0,02%
Bulgaria	2	0,02%
Egypt	2	0,02%
Dominican Republic	2	0,02%
Morocco	2	0,02%
Georgia	2	0,02%
Slovenia	2	0,02%
Guatemala	2	0,02%

China	17	0,16%
Portugal	15	0,14%
Venezuela	12	0,12%
Chile	12	0,12%
Australia	11	0,11%
Belgium	10	0,10%
Turkey	9	0,09%
United Arab Emirates	7	0,07%
Canada	7	0,07%
Peru	6	0,06%
Norway	5	0,05%
Israel	5	0,05%
Austria	5	0,05%
Andorra	5	0,05%
Poland	4	0,04%
Czech Republic	4	0,04%
Romania	4	0,04%
Hong Kong	4	0,04%
Ecuador	4	0,04%
Ireland	4	0,04%
Finland	4	0,04%
Singapore	3	0,03%
Denmark	3	0,03%
South Korea	3	0,03%
Luxembourg	3	0,03%
Thailand	3	0,03%

Kenya	2	0,02%
Cyprus	2	0,02%
Philippines	2	0,02%
Costa Rica	1	0,01%
Bolivia	1	0,01%
Qatar	1	0,01%
Montenegro	1	0,01%
Lebanon	1	0,01%
Serbia	1	0,01%
Federal District	1	0,01%
Montserrat	1	0,01%
Taiwan	1	0,01%
Vietnam	1	0,01%
Tunisia	1	0,01%
South Africa	1	0,01%
India	1	0,01%
Bahamas	1	0,01%
Estonia	1	0,01%
Brittany	1	0,01%
Puerto Rico	1	0,01%
Bogotá	1	0,01%
Cayman Islands	1	0,01%
Hungary	1	0,01%
Jordan	1	0,01%
Totale	10347	100,00%

Regione	Conteggio	Percentuale
Catalonia	6980	74,85%
Non Specificato	2237	23,99%
Community of Madrid	39	0,42%
Andalusia	16	0,17%
Balearic Islands	15	0,16%
Valencian Community	14	0,15%
Galicia	5	0,05%
Aragon	4	0,04%
Basque Country	3	0,03%
Castile and León	3	0,03%
Asturias	2	0,02%
Region of Murcia	2	0,02%
Castile-La Mancha	2	0,02%
Navarre	1	0,01%
Canary Islands	1	0,01%
Córdoba	1	0,01%
Totale complessivo	9325	100,00%

La prima variabile di interesse è l'ubicazione dell'host: si tratta del luogo fisico in cui risiede l'host, che spesso non coincide con la posizione della sua proprietà in affitto su Airbnb. Il luogo fisico dell'host è estremamente rilevante nella nostra analisi, in quanto il rapporto personale tra guest e host sarà con estrema probabilità più forte se i due soggetti hanno la possibilità di interagire di persona. Qualora l'host non dovesse risiedere nella regione di Barcellona, le possibilità di una interazione personale sono infatti molto ridotte e quindi la possibilità che la gentilezza giochi un ruolo predominante nella soddisfazione dell'esperienza del guest sono molto ridotte. L'ipotesi di lavoro utilizzata, infatti, è che nonostante sia

vero che host e guest possano interagire tramite mezzi di comunicazione telematici, sia estremamente difficoltoso costruire una relazione di tipo empatico senza la possibilità di avere contatti personali.

Osservando i dati, si può evincere che più del 90% degli host è ubicato all'interno dei confini di stato, permettendo così di evitare alcune barriere comunicative tra cui la differenza di fuso orario, maggiori costi di comunicazione telefonici o eventuali barriere linguistiche.

Il dato più rilevante di tutti è proprio il numero di persone che non solo risiede all'interno dello stato spagnolo, ma che sia ubicato anche all'interno della regione di Barcellona. Si tratta di circa il 75% degli host presenti sul territorio nazionale, in valore assoluto quasi 7000 host. Per avere un confronto con la totalità degli host di Barcellona, si tratta di circa il 67% del totale. Questi dati sono estremamente confortanti, in quanto ci indicano che una significativa percentuale degli host è presente fisicamente sul territorio e avrà quindi la possibilità di interagire direttamente con il guest e di sviluppare eventualmente una vera e propria relazione con quest'ultimo.

4.1.2 Tempo di risposta

Tempo di Risposta	Conteggio	Percentuale
within an hour	4673	45,08%
N/A	2998	28,92%
within a few hours	1304	12,58%
within a day	989	9,54%
a few days or more	401	3,87%
Totale complessivo	10365	100,00%

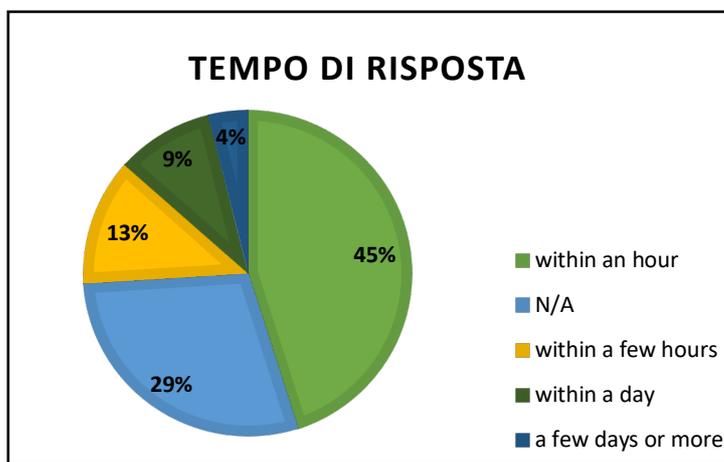


Figura 1: dati percentuali del tempo di risposta degli host

Come riportato dai grafici precedenti, circa il 45% degli host risponde alle informazioni richieste dai guests in un tempo inferiore ad un'ora e che gli host rispondono in un tempo superiore a un giorno sono solamente il 4%. È necessario sottolineare come questo dato non abbia una diretta correlazione con la gentilezza, bensì rappresenta un parametro di efficienza. Tuttavia, qualora in futuro si volesse approcciare alla gentilezza in un senso più ampio, questo parametro potrebbe risultare interessante in quanto possibile indicatore di disponibilità dell'host nei confronti del guest.

Inoltre, è importante segnalare che questo parametro sarà direttamente correlato con il riconoscimento di Superhost da parte della piattaforma Airbnb, che sarà analizzata in maniera approfondita successivamente.

4.1.3 Tasso di risposta e tasso di accettazione

Si riporta di seguito la tabella

Tasso risposta	Conteggio	Percentuale
0%	265	3,60%
50%	132	1,79%
67%	82	1,11%
70%	75	1,02%
80%	151	2,05%
90%	276	3,75%
100%	5543	75,24%

N/A	2998	28,92%
Totale complessivo	7367	100,00%

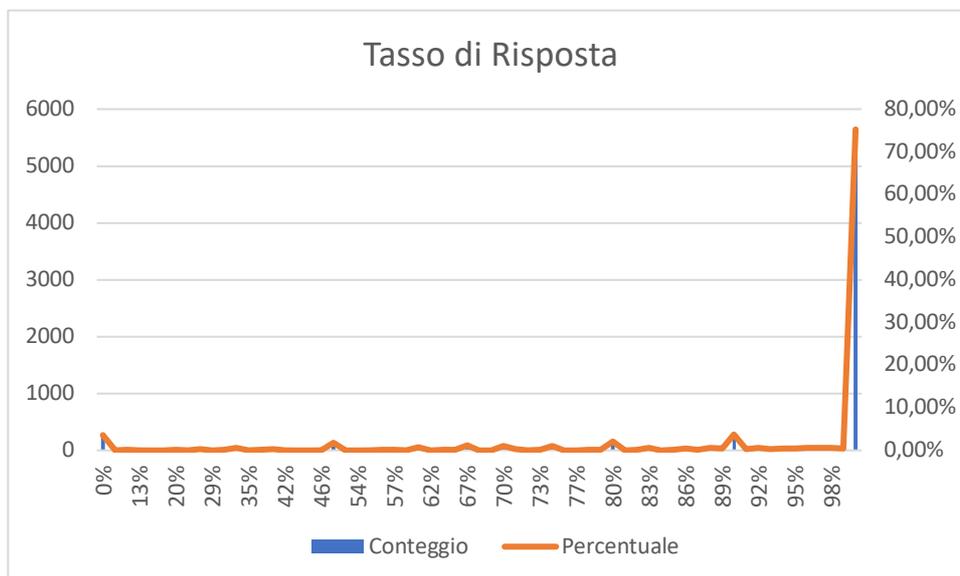


Figura 2: tasso di risposta in funzione della percentuale degli host

I dati riportano un tasso di risposta del 100% per oltre il 75% dei proprietari di casa. Il tasso di risposta essendo un elemento logistico fondamentale per far sì che possa avvenire la transazione, difficilmente può essere correlato alla gentilezza anche solo nel senso ampio del termine.

Tuttavia, non si può dire la stessa cosa per la variabile riportata di seguito, ossia il tasso di accettazione.

Tasso di Accettazione	Conteggio	Percentuale
Non disponibile	2099	20,25%
0%	230	2,22%
92%	138	1,33%
93%	142	1,37%
94%	150	1,45%

95%	162	1,56%
96%	228	2,20%
97%	269	2,60%
98%	347	3,35%
99%	443	4,27%
100%	3796	36,62%
Totale	8004	77,22%

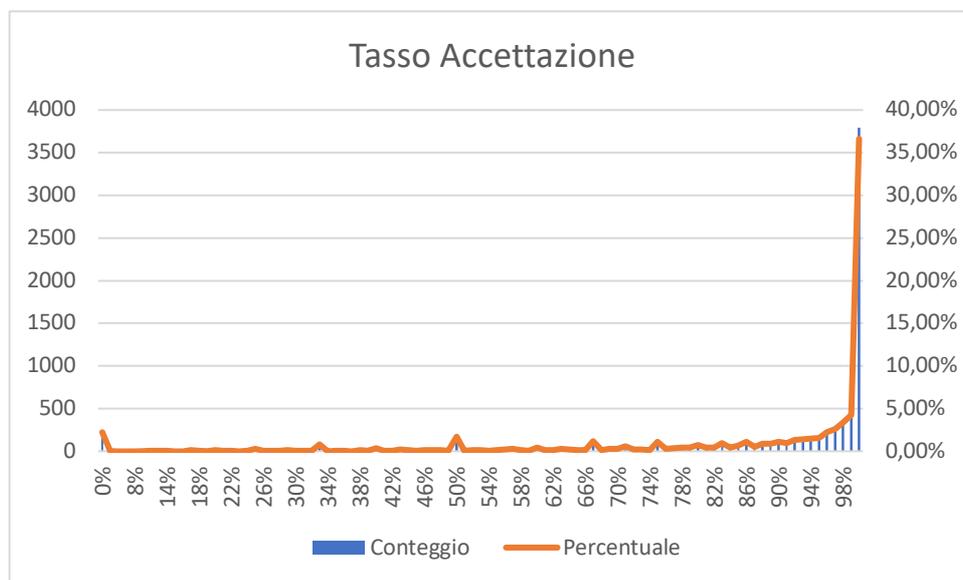


Figura 3: tasso di accettazione in funzione della percentuale degli host

La curva del tasso di accettazione presenta una maggiore distribuzione a partire dal 92%, fino a raggiungere il picco corrispondente al 100%. Questo dato, sporcato dalla mancanza dei dati di circa il 20% degli host, fornisce due tipologie di informazioni: la prima è che proprio per il senso di comunità che crea Airbnb, gli host sono propensi ad accettare quasi tutte le richieste; la seconda è che grazie alla funzionalità di “instant booking” attivata dalla piattaforma, alla quale hanno aderito in massa gli host, il guest non deve più superare l’approvazione di quest’ultimo portando quindi un picco nel tasso di accettazione proprio in corrispondenza del 100% .

È estremamente interessante segnalare come l’host tenda a fare poca selezione sul guest che verrà ospite: questo può essere da un lato segno di apertura

mentale e di volontà di non creare discriminazioni, dall'altro lato indica che l'host non ricerca un guest con particolari caratteristiche in quanto direttamente interessato a costruire un legame (ad esempio passioni in comune, tipologie di comportamento ecc.) ma piuttosto si comporta come una tradizionale struttura alberghiera prediligendo l'aspetto economico.

4.1.4 Superhost

Si riporta di seguito la tabella contenente l'informazione sul conteggio degli host con il flag di Superhost e la relativa percentuale.

Superhost	Conteggio	Percentuale
No	8270	79,79%
Si	2095	20,21%
Totale complessivo	10365	100,00%

In questo paragrafo, si va ad analizzare quanti host presenti a Barcellona siano effettivamente Superhost e quanti non lo siano. Si ricorda che per essere segnalati sul sito di Airbnb come Superhost, bisogna superare alcuni requisiti imposti dalla piattaforma, che per comodità riportiamo di seguito:

- La media della valutazione complessiva (frutto della media di valutazioni intermedie come: pulizia, comunicazione, location, rapporto qualità/prezzo ecc.) di tutte le review dell'anno precedente deve essere superiore a 4,8 su 5
- Devono essere stati completati nell'ultimo anno 10 soggiorni, oppure 100 notti suddivise in 3 soggiorni; questo requisito è utile per garantire agli ospiti un host con esperienza
- Tasso di cancellazione inferiore all'1%, per garantire agli ospiti tranquillità sulla prenotazione effettuata
- Tasso di risposta superiore al 90% entro le 24 ore, per garantire agli ospiti un host presente e disponibile

È importante sottolineare come sia molto utile per un host essere segnalato come Superhost da Airbnb. I benefici principali per un Superhost sono svariati, tra cui maggiori guadagni dati dalla maggiore visibilità e fiducia dei guest, l'aumento del numero di ospiti per via della maggiore visibilità che offre la piattaforma ad annunci Superhost e infine alcuni premi riservati dalla piattaforma tra cui bonus superiori a quelli canonici e crediti di viaggio.

Analizzando i dati del paragrafo precedente, una lettura combinata delle tabelle relative a tempo e tasso di risposta e tasso di accettazione, suggerirebbe una presenza elevata di Superhost nel territorio della città di Barcellona. Tuttavia, come evidenziato dalla tabella sovrastante, i proprietari di casa delle abitazioni in analisi che rispecchiano i requisiti necessari per la qualifica di Superhost sono poco più del 20%. Questo risultato, sebbene non direttamente collegato alla gentilezza, potrebbe sottolineare una forte influenza del rating dell'abitazione, il quale tiene conto di diversi fattori, tra cui alcuni indici indiretti della stessa.

4.1.5 Presentazione dell'host

La maggior parte degli host decide di non descriversi e non raccontare nulla di sé. Infatti, su 10365 host analizzati, solo 5044 hanno deciso di spendere qualche parola su loro stessi, la loro vita e passioni. 4847 utenti, ovvero il 47%, ha deciso di non compilare questo campo, nonostante in realtà possa costituire un punto critico nel momento della scelta degli ospiti.

Numero parole	Conteggio Host	Percentuale
1-30	2185	43,32%
31-60	1350	26,76%
61-90	686	13,60%
91-120	337	6,68%
121-150	173	3,43%
151-180	107	2,12%
181-210	68	1,35%

211-240	40	0,79%
241-270	26	0,52%
271-300	18	0,36%
301-330	18	0,36%
331-360	6	0,12%
361-390	10	0,20%
391-420	6	0,12%
421-450	1	0,02%
451-480	6	0,12%
481-510	2	0,04%
>511	5	0,10%
Totale complessivo	5044	100,00%

Chi decide di descriversi normalmente non utilizza più di 100 parole, nella maggior parte dei casi (43,32%) sono sufficienti meno di 30 parole. In pochissimi casi l'host si dilunga nella descrizione di sé stesso e in ancora meno, mettendosi nei panni dei clienti, offre questa descrizione in più lingue.

Un esempio è Alex, che si racconta in 3 lingue per un totale di 900 caratteri:

“Dearest traveller, thank you very much for visiting my page (...)

Apreciado viajero, Muchas gracias por visitar mi página (...)

Cher voyageur, Merci beaucoup pour visiter ma page (...)”

O Donato, che ha reso disponibile la sua descrizione in 4 lingue per un totale di 626 caratteri:

“I am friendly and sociable, but I also like privacy. I am a little crazy and artist (...)

Soy persona sociable pero me gusta mi intimidad, soy un poco loco y sensible al arte (...)

Sono una persona generalmente socievole ma allo stesso tempo mi piace la mia intimità (...)

Je suis amical et sociable, mais j'aime aussi la vie privée. Je suis un peu fou et artiste (...)”

Questa variabile è molto significativa ed indica la volontà di un host di raccontare di sé, quindi indirettamente anche una eventuale predisposizione a creare un legame con il guest. La sua significatività sarà studiata nel capitolo successivo come variabile di controllo nella correlazione tra rating e gentilezza dell'host.

4.1.6 Numero di Annunci per host

Si riporta di seguito la tabella con il numero di proprietà collegate a un host, il conteggio degli host e le proprietà complessive. Prima verranno presentate a fasce di 10 proprietà alla volta e poi verrà fatto il focus sulla prima fascia.

Numero annunci per host	Conteggio host	Percentuale	Numero proprietà complessive	Percentuale
<10	10144	97,82%	15510	73,45%
10-19	139	1,34%	1869	8,85%
20-29	39	0,38%	901	4,27%
30-39	19	0,18%	642	3,04%
40-49	12	0,12%	515	2,44%
50-59	3	0,03%	164	0,78%
60-69	1	0,01%	68	0,32%
70-79	1	0,01%	70	0,33%
80-89	2	0,02%	163	0,77%
90-100	4	0,04%	377	1,79%
>100	6	0,06%	837	3,96%
Totale complessivo	10365	100,00%	21116	100,00%

In questo paragrafo si andrà ad analizzare il numero di annunci collegati ad ogni host. Questa variabile è interessante da analizzare in questo lavoro, in quanto un host con un numero importante di annunci tenderà a comportarsi come "amministratore" delle sue proprietà a differenza di un host con pochi annunci (o addirittura nel caso si tratti di un annuncio solo, in cui la proprietà è la stessa

abitazione dell'host) che tenderà a sviluppare in media un rapporto più profondo e vicino con l'host.

Analizzando i dati, si può vedere come la maggior parte degli host (97,82%) ha meno di 10 annunci disponibili ed il restante 1,34% degli host ha meno di 20 annunci attivi sulla piattaforma. Solo pochi proprietari hanno a disposizione un numero maggiore di annunci, fino ad arrivare ad un massimo di 159 proprietà. Sono comunque rilevanti gli annunci degli host con oltre 100 proprietà, che con 837 unità costituiscono quasi il 4% degli annunci totali. Il 73,45% degli annunci nella città di Barcellona è però in mano a host di piccole dimensioni.

Numero annunci per host	Conteggio host	Percentuale	Numero proprietà complessive	Percentuale
1	7378	72,60%	7378	34,94%
2	1599	15,74%	3198	15,14%
3	564	5,55%	1692	8,01%
4	236	2,32%	944	4,47%
5	142	1,40%	710	3,36%
6	90	0,89%	540	2,56%
7	62	0,61%	434	2,06%
8	43	0,42%	344	1,63%
9	30	0,30%	270	1,28%
10	18	0,18%	180	0,85%
Totale complessivo	10162	100,00%	15690	100,00%

In questa seconda tabella focalizziamo l'attenzione sugli host di piccola dimensione, con un numero di annunci inferiore o pari a 10 unità.

Questo specchio sugli host di piccole dimensioni offre la possibilità di notare come la maggior parte di essi non superino le due proprietà (88,34%). Il numero di

host ha un tasso decrescente al crescere del numero di annunci per host, come si evidenzia dal grafico seguente.

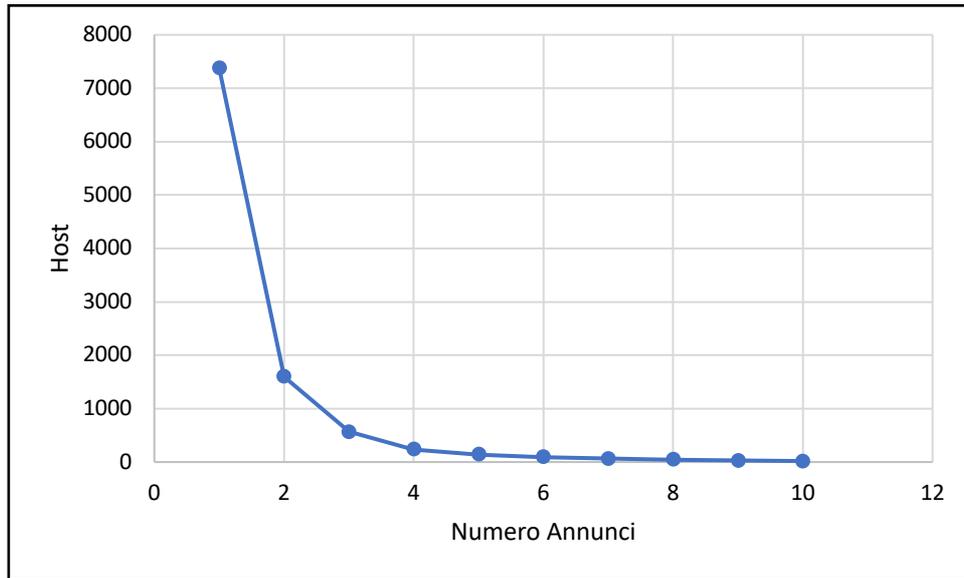


Figura 4: numero annunci con il rispettivo conteggio host

Questo dato è estremamente incoraggiante, in quanto ci mostra come gli host che amministrano un numero esiguo di proprietà siano in netta maggioranza rispetto agli altri. Questo dato rende l'interazione umana host – guest (e di conseguenza la gentilezza) una variabile molto rilevante all'interno di un soggiorno con la piattaforma Airbnb.

4.1.7 Tipologia di annuncio

Di seguito viene riportata la tabella contenente i dati con il tipo di annuncio, il conteggio degli host e la rispettiva percentuale.

Tipo di Annuncio	Conteggio	Percentuale
Annunci Alloggio Intero	9626	46,51%
Annunci Camera Privata	10840	52,37%
Annunci Camera Condivisa	231	1,12%
Annunci Totali	21116	100,00%

Gli host di Airbnb offrono spazi di ogni tipo, da camere condivise fino a isole private e gli annunci si suddividono per tipologia, in una delle 3 categorie seguenti:

- Alloggio intero
- Stanza privata
- Camera condivisa

Alloggio intero

La scelta migliore se ciò che cerchi è una seconda casa per il tuo viaggio. Questo tipo di alloggio è l'ideale per chi desidera avere un'intera casa a disposizione. Di solito include una camera da letto, un bagno, una cucina e un ingresso separato dedicato. Nella descrizione, gli host dovrebbero indicare se si troveranno o meno nell'alloggio (ad esempio: "L'host abita al primo piano") e fornire ulteriori dettagli sull'alloggio.

Stanze private

Sono perfette se vuoi fare conoscenza con le persone del luogo senza rinunciare alla tua privacy. In questo modo puoi dormire in una camera privata, facendo conoscenza con altri viaggiatori negli spazi comuni (cucina,sala..). Per entrare in camera, potresti dover attraversare spazi interni occupati da un altro ospite o dall'host.

Camere condivise

Se scegli una camera condivisa, sicuramente non hai problemi a condividere lo stesso spazio con altri sia di giorno che di notte. Le camere condivise sono gettonate tra i viaggiatori più flessibili, che vogliono spendere meno e conoscere nuovi amici.

Dalla tabella è evidente come gli Host Airbnb di Barcellona prediligano le prime tipologie di case, mentre la camera condivisa sia praticamente inesistente. Infatti, sono disponibili solo 231 annunci che offrono una camera condivisa, su un totale di 21116 annunci, ovvero l'1,12% degli annunci.

Una spiegazione per la scarsa disponibilità di annunci con camere condivise può essere la competizione che si andrebbe a creare con gli ostelli. Poiché negli ostelli la tipologia di camera principalmente disponibile è quella condivisa e dato l'alto numero di ostelli presenti in Barcellona, pari a 9731 unità (dato aggiornato al 2018), questa tipologia di annuncio non promette alti margini di guadagno agli Host Airbnb.

4.1.8 Anno di iscrizione

Anno Iscrizione	Conteggio	Percentuale
2008	1	0,01%
2009	5	0,05%
2010	54	0,52%
2011	296	2,86%
2012	761	7,34%
2013	1130	10,90%
2014	1139	10,99%
2015	1273	12,28%
2016	1276	12,31%
2017	1058	10,21%
2018	1076	10,38%
2019	1692	16,32%
2020	604	5,83%
Totale complessivo	10365	100,00%



Figura 5: numero dei nuovi iscritti suddivisi per anno

Si è scelto di analizzare in ultima istanza il numero di host presenti sulla piattaforma, non tanto per la sua correlazione con la gentilezza o con il rapporto host – guest, bensì poiché è interessante vedere come dal 2011 i nuovi iscritti per anno siano aumentati sensibilmente per due anni di fila, fino a raggiungere un tasso di crescita più stabile. Il numero di nuovi iscritti, oltre a segnalarci la diffusione della piattaforma nella città di Barcellona a partire dal 2012 in ritardo con il più maturo mercato statunitense, ci segnala che con tutta probabilità una crescita così importante di tutta la comunità avrà portato a una significativa diversificazione degli host iscritti. In particolare, come già descritto nel capitolo 1 di letteratura, se i primi utenti iscritti davano priorità all’esperienza sociale rispetto al puro compenso economico, con il passare del tempo si sono aggiunti anche quegli utenti con ottica più imprenditoriale ed orientata al profitto invece che all’esperienza.

4.2 Analisi delle variabili del modello di regressione e processo

In questo capitolo, si andrà a presentare il lavoro svolto sul modello statistico, in particolare modo soffermandosi sui seguenti punti:

- Descrizione delle variabili delle regressioni
- Descrizione e riassunto del processo

Il modello che si andrà ad analizzare in questo ultimo capitolo è basato sulla regressione lineare: ci servirà per verificare l'ipotesi che la gentilezza e la buona comunicazione dell'host costituisca un fattore determinante nel rating finale che l'utente assegna nella recensione al termine dell'esperienza di soggiorno.

Questa analisi è stata possibile grazie al database pubblico di Airbnb, dove sono presenti diversi documenti, suddivisi per le città principali, in cui sono raccolti gli elenchi degli host, le proprietà e le recensioni finali degli utenti presentati in maniera molto approfondita nei capitoli precedenti. Prima di iniziare il processo di analisi, inoltre, è stato necessario effettuare una pulizia del dataset, andando ad eliminare dagli elenchi tutti quei dati che non erano utilizzabili o di interesse per dimostrare la tesi del modello di regressione. Al termine del lavoro di pulizia, è stato possibile ottenere un database finale in cui le informazioni, inizialmente ripartite su diversi documenti, fossero concentrate in pochi fogli di calcolo e presentate in maniera comprensibile.

Si andrà a spiegare quale tra le variabili utilizzate per la costruzione del modello di regressione lineare fossero già presenti all'interno del database di Airbnb e quali invece sono state create in quanto necessarie per valutare l'ipotesi di partenza.

Esempi di questi tipi di variabili create, che saranno analizzate nei capitoli successivi, sono l'indice di gentilezza e la variabile interagita, frutto del prodotto tra l'indice di gentilezza e la caratteristica "SuperHost" automaticamente assegnata da Airbnb.

Il percorso di approfondimento del processo di costruzione del modello inizia descrivendo brevemente i database utilizzati (non l'intero dataset come descritto nel capitolo 2) per capire poi come sono stati puliti e processati i dati.

Prima di entrare nel vivo dell'analisi, al fine di rendere più chiara possibile la stessa, è importante andare a riprendere gli attori principali, i loro principali attributi e il glossario della piattaforma:

- **Host:** è l'attore principale, ed è colui che gestisce le proprietà (o *listing*). Ha la possibilità di presentarsi attraverso una descrizione in cui presenta sé stesso e i suoi principali interessi e passioni. Ha il compito di rispondere ai messaggi dei guest (dalla quale si ricava il *response rate*), di accettare o rifiutare le richieste dei guest che vogliono pernottare nelle proprietà dell'host (da cui si ricava il *tasso di accettazione*). Non necessariamente l'host è presente nello stesso luogo della proprietà, per questo è disponibile anche la posizione dell'host (o *location host*)
- **Guest:** è colui che, se la sua richiesta di pernottamento viene accettata dall'host, pernotta in una proprietà di quest'ultimo. Al termine della sua esperienza potrà rilasciare una recensione (o *review*) associata a una valutazione in termini di stelle (da 0 a 5) su più parametri, che sommati insieme andrà a costituire un *rating* complessivo. A sua volta sarà recensito dall'host con una review e un rating: queste recensioni non saranno analizzate in questo elaborato
- **Rating e Review:** si hanno 14 giorni di tempo dopo il check-out per scrivere la recensione e valutare con un rating il soggiorno (nel caso dell'host) e l'host (nel caso del guest). Inoltre, come descritto in precedenza, per evitare commenti vendicativi o distorti, le recensioni vengono rese pubbliche solo quando entrambe le parti, Host e Guest, hanno completato il processo di recensione, oppure allo scadere dei 14 giorni. Dal punto di vista degli ospiti, nelle recensioni si possono sottolineare gli aspetti critici o particolarmente positivi del soggiorno, la relazione che si è creata con l'host ed altre peculiarità. Ogni recensione, dopo essere stata pubblicata, potrà essere letta pubblicamente da tutti gli utenti e potrà quindi influenzare la reputazione dell'host. Come detto in precedenza, gli ospiti, in aggiunta alla recensione, possono anche rilasciare dei rating in stelle (da 0 a 5) sull'esperienza complessiva o su singole categorie. Essendo questo punto particolarmente importante per l'analisi

svolta nei capitoli successivi, andiamo ad analizzare una categoria di rating alla volta

- Esperienza complessiva: il gradimento del soggiorno complessivo
- Pulizia: il livello di pulizia e ordine della proprietà al momento del check-in
- Precisione: valutazione della precisione dell'annuncio rispetto allo stato corrente della proprietà. È importante che l'ospite possa disporre di fotografie e informazioni aggiornate rispetto lo stato corrente
- Valore: valutazione del rapporto qualità prezzo secondo l'ospite
- Comunicazione: valutazione della qualità di comunicazione con l'host sia in fase di organizzazione del soggiorno, sia durante il soggiorno stesso. Per gli ospiti è infatti importante rispondere in maniera celere, precisa e affidabile ai messaggi
- Check-in: valutazione della qualità in termini di fluidità o intoppi durante il processo di check-in
- Posizione: valutazione del quartiere che può comprendere vicinanza e accesso ai trasporti, ad attività commerciali, vicinanza al centro città, livello di sicurezza o segnalazione di eventuali rumori molesti
- Servizi: valutazione dei servizi disponibili all'interno dell'annuncio. È importante che i servizi descritti all'interno dell'annuncio siano tutti funzionanti e disponibili

È inoltre importante sottolineare, come la valutazione complessiva del soggiorno sia direttamente ricavata dai rating intermedi appena elencati. Si segnala anche che sono necessarie almeno 3 recensioni con le rispettive valutazioni per ogni proprietà, prima che siano resi pubblici a tutti gli utenti: ciò con l'unico scopo che non sia un'unica esperienza particolarmente buona o particolarmente sfortunata a determinare la reputazione di un host.

Ora che sono stati analizzati i principali attori, gli attributi e le variabili di riferimento della piattaforma, si passa alla descrizione delle tabelle e degli elenchi

di partenza, per poi passare ad analizzare con dettaglio il processo di costruzione del modello di regressione lineare.

4.2.1 Foglio Review

Questo foglio di calcolo è quello principale per l'analisi statistica svolta. In questo database, erano inizialmente disponibili oltre 115.000 recensioni, tutte riferite a proprietà presenti nella città di Barcellona, in un arco temporale che inizia nel 2008 ed arriva fino ai giorni nostri.

Per ogni recensione sono presenti queste informazioni:

- **Listing ID:** rappresenta il codice della proprietà a cui la recensione è associata. È importante quindi comprendere come la recensione non sia correlata direttamente all'host, ma alla proprietà dell'host. In questo modo l'host ottiene un rating diverso in ciascuna delle proprietà da lui gestite sulla piattaforma. Questo fatto è fisiologico, in quanto secondo quanto già detto in precedenza, il rating integra alcune variabili peculiari della singola proprietà come ad esempio la posizione, il rapporto qualità prezzo, la precisione dell'annuncio e i servizi disponibili all'interno della proprietà
- **ID:** Rappresenta il codice identificativo dell'ospite che ha effettuato la recensione.
- **Date:** Data in cui è stata effettuata la recensione da parte dell'utente.
- **Reviewer_name:** Nome dell'utente che ha effettuato la recensione.
- **Comments:** Recensione scritta dall'utente. È stata l'informazione di maggior interesse per il tipo di analisi svolta, poiché costituisce la testa da cui è stato poi calcolato l'indice di gentilezza.

4.2.2 Pulizia del Database

Essendo i dati a disposizione mal organizzati e troppo ricchi rispetto l'analisi che si intendeva svolgere, è stato necessario un processo di pulizia di quelle recensioni

che non erano utilizzabili per la creazione dell'indice di gentilezza. Questo processo di pulizia si è svolto in diversi passaggi.

4.2.2.1 Pulizia delle recensioni in lingua non inglese

Il primo passaggio obbligatorio è stato andare a pulire il database da tutte le recensioni scritte in lingua diversa dall'inglese (molte di esse erano appunto in spagnolo) o erroneamente tradotte da Airbnb in simboli non supportati da Excel e quindi anch'esse da scartare.

Per ottenere un elenco di recensioni adatte all'analisi sono state ricercate nel testo presente nella colonna "Comments", che è la colonna che contiene la recensione (o *review*) vera e propria, le dieci parole più utilizzate nella lingua inglese.

Si riportano di seguito le parole ricercate all'interno della recensione:

- "the"
- "he"
- "is"
- "have"
- "Of"
- "location"
- "And"
- "My"
- "In"
- "that"
- "This"

Sono state create 11 nuove colonne, dove si andava a segnalare la presenza o meno della parola in oggetto sulla recensione della riga di riferimento. La ricerca della parola restituiva come valore 1 nel momento in cui la parola era presente e 0 se non fosse stata trovata.

Una volta svolta la ricerca delle parole, per riuscire a separare le recensioni in lingua inglese da quelle in altre lingue è stata successivamente creata una

variabile unica, che restituisse la somma delle singole ricerche per ciascuna parola. È importante segnalare per precisione, che per essere più agevolati nell'utilizzare i valori, sono state create altre colonne che contenessero solo i valori delle celle e non l'intera formula. La variabile creata assume valori compresi tra 0 (nessuna delle parole ricercate è stata riscontrata nel testo) e 10 (tutte le parole sono presenti nel testo della recensione).

Per mantenere un margine di sicurezza nella selezione dei commenti solo in lingua inglese sono stati considerati solo quei commenti che presentavano un valore in questa variabile maggiore o uguale a 6. Questa scelta è stata adottata, in quanto andando a ricercare fino a 4 parole, nonostante la maggior parte dei commenti fosse interamente in lingua inglese, non venivano scartati alcuni commenti in lingue straniere che contenevano all'interno alcuni anglicismi. Andando a eliminare tutte le recensioni con un numero minore o uguale a 6 delle 10 parole cercate, è stato empiricamente possibile scoprire che le recensioni fossero tutte in lingua inglese (dato confermato anche da un controllo manuale riga per riga).

Delle 115mila recensioni inizialmente presenti nel database, solo 15mila hanno superato questo test, mentre le altre sono state scartate in quanto non utilizzabili.

4.2.2.2 Pulizia di tipo temporale

Il secondo processo di pulizia del Database è stato effettuato considerando la scala temporale: le recensioni considerate nel modello sono state solo quelle effettuate nell'intro 2019 e i primi mesi del 2020, mentre quelle scritte in anni precedenti sono state escluse, in modo da ottenere un indice di gentilezza dell'host basato sulle esperienze più recenti, non complessive di tutte le recensioni effettuate negli anni. Questa decisione è stata presa per svariate ragioni: in prima istanza perché alcuni dati aggregati presenti nel secondo foglio di calcolo che si andrà ad analizzare successivamente, sono su base annua sull'ultimo anno intero disponibile (ad esempio il prezzo medio per notte); in secondo luogo per permettere di avere un numero di campioni rappresentativo, ma allo stesso tempo gestibile in termini di possibili controlli manuali sul database.

In questa seconda selezione sono stati ottenuti 7836 campioni finali

4.2.3 Foglio overview

Il secondo documento utilizzato per la creazione del modello descrive quelle che sono le caratteristiche di ciascuna proprietà presente sulla piattaforma per la città di Barcellona.

In questo documento sono presenti molte informazioni aggiuntive rispetto a quello delle recensioni, molte della quali sono risultate essere superflue ai fini della nostra ricerca. È stato possibile collegare le informazioni presenti nei due documenti grazie al codice (ID) identificativo della proprietà: mentre nel primo documento ogni riga corrisponde ad una recensione, ed ogni recensione è collegata a una proprietà (per dovere di chiarezza è importante specificare che nel primo database, più righe potevano quindi rispondere alla stessa proprietà), nel secondo rappresenta invece una proprietà. Poiché ogni recensione è però associata univocamente ad una proprietà, è stato possibile unire le informazioni dei due database inizialmente separati.

Si riportano di seguito le informazioni presenti per ogni proprietà:

- **Caratteristiche della proprietà:**
 - **Listing ID:** Identifica il codice della proprietà.
 - **Listing URL:** Link dell'annuncio sulla piattaforma
 - **Name:** Nome dell'annuncio sulla piattaforma
 - **Summary:** Breve riassunto introduttivo della proprietà
 - **Space:** Descrizione della zona in cui la proprietà è ubicata
 - **Description:** Descrizione più dettagliata della proprietà e degli spazi comuni
 - **Notes:** Aggiunta opzionale da parte dell'host nel caso in cui voglia sottolineare qualcosa di importante per gli utenti
 - **Transit:** Aggiunta opzionale da parte dell'host nel caso in cui voglia indicare agli utenti come raggiungere la proprietà

- **Access:** Descrizione di quali spazi comuni sono inclusi nella prenotazione e quali invece sono riservati al proprietario/altri utenti presenti nella casa
- **Interaction:** Descrive come verrà effettuato il check-in, a quali persone rivolgersi in caso di necessità, numeri di emergenza ecc.
- **House_rules:** Definisce quelle che sono le regole da rispettare riguardo ad eventuali orari di ingresso e check-in nel momento in cui anche l'host vive nella casa, oppure il numero di persone massimo che la casa può ospitare.
- **Caratteristiche dell'host**
 - **Host ID:** Identifica il codice dell'host a cui la proprietà è associata. E' possibile che ad un host siano associati più annunci presenti nella piattaforma.
 - **Host name:** Nome dell'host dell'annuncio.
 - **Host since:** Data in cui l'host si è iscritto alla piattaforma.
 - **Host location:** Definisce l'indirizzo di residenza del proprietario di casa. Nel nostro caso la maggior parte di essi vive in Barcellona, ma un numero considerevole è suddiviso nelle diverse regioni spagnole piuttosto che in paesi europei e non.
 - **Host about:** Uno spazio dedicato alla vita personale, passioni, attitudini, punti di interesse ma anche modi di vivere del proprietario di casa. Per il cliente in cerca di una sistemazione temporanea può essere molto utile questa informazione, soprattutto considerando che in molti casi l'host stesso vive nella proprietà e condivide gli spazi comuni. In questo modo l'utente ha la possibilità di scegliere in base a passioni comuni.
 - **Host response time:** Definisce l'arco temporale in cui mediamente l'host risponde alle domande da parte degli utenti.
 - **Host response rate:** Tasso di risposta dell'host
 - **Host acceptance rate:** Tasso di accettazione delle richieste di soggiorno degli utenti da parte dell'host
 - **Host is superhost:** Flag che identifica se l'host è un Superhost.

- **City:** Definisce in che città vive l'host
- **Host listing count:** Definisce il numero di annunci presenti sulla piattaforma pubblicati dallo stesso host
- **Specifiche della proprietà**
 - **Property type:** Definisce il tipo di proprietà. Esse possono essere: Appartamento, Castello, Chalet, Condominio, Fattoria, Ostello, Bed&Brekfast, Loft, Villa ecc.
 - **Room type:** Definisce il grado di libertà all'interno della proprietà. Esso può essere: Intera casa, camera privata o camera condivisa.
 - **Accomodate:** Numero di spazi
 - **Bathroom:** Numero di bagni
 - **Bedroom:** Numero di camere da letto
 - **Bed:** Posti letti disponibili e tipo di letto
 - **Amenities:** Confort disponibili all'interno della casa. Incluse la TV, l'aria condizionata, il wi-fi, ecc
 - **Square_feet:** Superficie della proprietà misurata in piedi quadrati
 - **Price:** Esistono tre tipologie di prezzo indicate dagli host: per una singola notte, settimanale e mensile.
 - **Extra fee:** esistono tre tipologie di costi aggiuntivi a discrezione dell'host:
 - Acconto
 - Tassa di pulizia (giornaliera/fine soggiorno)
 - Persone extra (nel caso in cui si aggiungano persone)
 - **Instant bookable:** Definisce se la proprietà è direttamente prenotabile o è necessaria l'accettazione da parte dell'host.
 - **Cancellation Policy:** Definisce quelle che sono le politiche di cancellazione. Essa può essere: Flessibile, Moderata, Rigida, Lungo Termine, Super Rigida 30 o 60 giorni. L'utente può ottenere un rimborso in caso di cancellazione la cui

percentuale dipende dalla politica di cancellazione scelta dall'host. La tipologia scelta inoltre influisce sul numero di giorni massimo anteriori al check-in in cui è possibile ottenere questa percentuale in caso di cancellazione. Una volta superata questa soglia, non è possibile ottenere il rimborso.

- **Valutazione annuncio**
 - **Rating:** Costituisce il punteggio assegnato alla proprietà. Esso è composto da un punteggio sull'esperienza complessiva, su scala da 0 a 100, come già descritto in precedenza frutto della media di cinque punteggi specifici in:
 - Precisione
 - Pulizia
 - Check-in
 - Comunicazione
 - Posizione
 - Valore
 - **Number of review:** Definisce il numero di recensioni presenti per l'annuncio in questione. Costituisce un dato importante, poiché spesso è decisivo per la scelta da parte dell'utente, poiché a maggior numero di recensioni corrisponde una maggiore fiducia e sicurezza da parte di chi sta valutando.

4.2.4 Creazione indice Kindness

L'indice di gentilezza è una variabile creata per verificare se esiste una correlazione tra la gentilezza dell'host e la valutazione complessiva assegnata dall'utente.

Uno sforzo particolare di creatività è stato infatti richiesto per cercare di misurare il grado di gentilezza: non esistono nel database iniziale variabile numeriche che

possano dare una buona approssimazione del grado di gentilezza che sia stato tenuto dall'host nei confronti del suo ospite.

Per ovviare a questo problema, si è deciso di utilizzare le recensioni scritte dall'ospite dopo aver concluso la sua esperienza con l'host. Infatti, all'interno delle recensioni esistono un numero significativo di elementi che possano misurare non solo il grado di gentilezza tenuto dall'host, ma possono anche dare una serie di indicazioni sulla relazione host – guest che si è andata a creare.

Per completezza di analisi, si riporta un esempio di una recensione, per capirne meglio il contenuto:

“Staying at Pedro’s house was the best Airbnb experience I’ve ever had yet. If you are looking for more than just a simple place to stay, then this is it. Very kind and friendly host. Very open minded and willing to sit and chat with you if that is what you like. I was lucky enough to enjoy a trip up the coast thanks to him and I’ve learnt a lot. I also would not hesitate to inquire about one of his massage sessions. Top class host”

Moath

In questo esempio, è facile notare come l'ospite nella sua recensione nomini l'host per nome e in maniera molto informale, come ad indicare non solo che si siano effettivamente conosciuti di persona, ma anche che abbiano creato un vero e proprio legame. Inoltre, sono presenti alcuni aggettivi che indicano come indubbiamente l'host abbia scelto di essere molto gentile con il suo ospite. Questo si può vedere nei passaggi come “very kind and friendly host” oppure “Very open minded and willing to sit and chat with you if that is what you like”.

Per questa ragione, la base di partenza per creare un indice di gentilezza dell'host sono state proprio le recensioni, nelle quali si è deciso di andare a ricercare alcune particolari formule o aggettivi che sembravano ricorrenti all'interno di molte recensioni presenti nel dataset.

Per valutare la gentilezza dell'host sono state ricercate 29 parole all'interno del testo presente nelle recensioni e per ciascuna delle parole presenti è stato incrementato il valore della variabile di uno. In particolare, sono state create 29

nuove colonne, dove in ogni colonna si va a ricercare all'interno della recensione nella riga di riferimento la parola di interesse. Il valore della cella può essere:

- 0: la parola non è stata trovata all'interno della recensione
- 1: la parola è stata trovata all'interno della recensione

Le parole, che vengono riportate nell'immagine di seguito, sono:

- Good service
- Sweet
- Helpful
- Excellent
- Nice
- Thoughtful
- Lovely
- Friendly
- Genuine
- Flexible
- Positive
- Energy
- Responsive
- Spirit
- Gentle
- Intresting
- Funny
- Peaceful
- Warm
- Special
- Great
- Generous
- Open mind
- Presenza del nome dell'host

L'ultimo punto, fa riferimento alla ricerca del nome dell'host all'interno del commento, in quanto si è visto che in numerose recensioni, un buon indicatore del fatto che si fosse venuto a creare un legame tra host e guest, era proprio il fatto che nella recensione l'ospite si rivolgesse a lui in maniera molto informale e chiamandolo direttamente con il suo nome.

Una volta fatta questa operazione, è stato necessario sommare tutti i record ottenuti per ogni riga, andando a creare l'indice di gentilezza della singola recensione.

Quindi l'indice di gentilezza ha un intervallo di valori compresi tra 0 e 29. In questo modo, un alto valore di questa variabile associato alla recensione indica un maggior livello di gentilezza, disponibilità e comunicazione, poiché l'utente ha descritto molto positivamente l'esperienza nel commento finale. Viceversa, per valori nulli o bassi della variabile, si suppone che l'utente non abbia avuto un'esperienza particolarmente positiva dal punto di vista del trattamento da parte dell'host. Questo non significa che non sia stata nel complesso una esperienza positiva, ma che l'interazione e la comunicazione con l'host non è stata degna di particolari attenzioni nel commento finale.

Parole	Conteggi o	Percentuale		Parole	Conteggi o	Percentuale
Good Service				Sweet		
0	7832	99,95%		0	7712	98,42%
1	4	0,05%		1	124	1,58%
Totale complessivo	7836	100,00%		Totale complessivo	7836	100,00%
Helpful				Excellent		
0	6695	85,44%		0	7302	93,19%
1	1141	14,56%		1	534	6,81%
Totale complessivo	7836	100,00%		Totale complessivo	7836	100,00%
Nice				Thoughtful		
0	6078	77,57%		0	7650	97,63%
1	1758	22,43%		1	186	2,37%
Totale complessivo	7836	100,00%		Totale complessivo	7836	100,00%

Lovely				Friendly		
0	7064	90,15%		0	7139	91,11%
1	772	9,85%		1	697	8,89%
Totale complessivo	7836	100,00%		Totale complessivo	7836	100,00%
Kind				Genuine		
0	7241	92,41%		0	7800	99,54%
1	595	7,59%		1	36	0,46%
Totale complessivo	7836	100,00%		Totale complessivo	7836	100,00%
Flexible				Positive		
0	7714	98,44%		0	7781	99,30%
1	122	1,56%		1	55	0,70%
Totale complessivo	7836	100,00%		Totale complessivo	7836	100,00%
Pleasure				Energy		
0	7754	98,95%		0	7817	99,76%
1	82	1,05%		1	19	0,24%
Totale complessivo	7836	100,00%		Totale complessivo	7836	100,00%
Responsive				Spirit		
0	7474	95,38%		0	7828	99,90%
1	362	4,62%		1	8	0,10%
Totale complessivo	7836	100,00%		Totale complessivo	7836	100,00%
Gentle				Interesting		
0	7810	99,67%		0	7726	98,60%
1	26	0,33%		1	110	1,40%
Totale complessivo	7836	100,00%		Totale complessivo	7836	100,00%
Generous				Open mind		
0	7781	99,30%		0	7836	100,00%
1	55	0,70%		(vuoto)		0,00%
Totale complessivo	7836	100,00%		Totale complessivo	7836	100,00%
Funny				Peaceful		

0	7817	99,76%		0	7771	99,17%
1	19	0,24%		1	65	0,83%
Totale complessivo	7836	100,00%		Totale complessivo	7836	100,00%
Super				Presenza del nome		
0	6759	86,26%		0	3983	50,84%
1	1077	13,74%		1	3853	49,16%
Totale complessivo	7836	100,00%		Totale complessivo	7836	100,00%
Warm				Special		
0	7498	95,69%		0	7448	95,05%
1	338	4,31%		1	388	4,95%
Totale complessivo	7836	100,00%		Totale complessivo	7836	100,00%
Tips				Great		
0	7235	92,33%		0	4493	57,34%
1	601	7,67%		1	3343	42,66%
Totale complessivo	7836	100,00%		Totale complessivo	7836	100,00%

4.2.4.1 Tipologia di indice Kindness

A questo punto, il nostro indice kindness misura il livello di gentilezza che c'è stato nell'esperienza singola di un singolo guest. Se ipoteticamente lo stesso guest fosse tornato più volte durante l'anno, avrebbe avuto indici kindness differenti, anche a fronte di uno stesso atteggiamento da parte dell'host.

Per ovviare a questo problema, è stato necessario raggruppare tutte le recensioni della stessa proprietà con i rispettivi indici kindness associati, e procedere a fare una media di questi indici: la variabile creata è quindi una media degli indici kindness di ogni proprietà ed è stata nominata "Indice Kindness Proprietà".

Nel caso in cui un host dovesse avere più di una proprietà, potrebbe rientrare nel caso in cui nonostante si stia parlando della stessa persona (e quindi con la

tendenza a rigor di logica di trattare tutti gli ospiti allo stesso modo, o almeno con lo stesso grado di gentilezza) possa avere valori di Indice Kindness Proprietà diversi per le sue diverse proprietà. Se ad esempio un host dovesse avere una sola proprietà in una zona particolarmente turistica e non fosse in grado di parlare lingue straniere, nonostante il livello di gentilezza verso i propri ospiti sia il medesimo, potrebbe avere recensioni basse solo su quella particolare proprietà. Per ovviare a questo effetto distorsivo dell'indice proprietà, sono state raggruppate tutte le proprietà appartenenti agli host, ed è stata calcolata una media degli indici kindness della proprietà. Questo nuovo indice è ora espressione della gentilezza dell'host nei confronti del suo guest medio, in una delle sue proprietà.

Questo indice, insieme a tutte le altre variabili riguardanti unicamente gli host e non le proprietà nel suo complesso, sono state raggruppate in un foglio a parte all'interno del file in allegato a questo elaborato chiamato "host".

Per riassumere il processo di scambio di informazioni tra i diversi database è stata creata questa tabella, in cui si vuole evidenziare quali interazione sono presenti e quali variabili sono state utilizzate e create:

Foglio Recensione		Foglio Proprietà		Foglio Host	
Recensione1	Proprietà1	Proprietà1	Host1	Host1	
Recensione2	Proprietà1	Proprietà2	Host1	Host2	
Recensione3	Proprietà2	Proprietà3	Host2	Host3	
Recensione4	Proprietà2	Proprietà4	Host2	Host4	
Recensione5	Proprietà2	Proprietà5	Host2	Host5	
Recensione6	Proprietà2	Proprietà6	Host2	Host6	
Recensione7	Proprietà3	Proprietà7	Host3	Host7	
Recensione8	Proprietà3	Proprietà8	Host4	Host8	
Recensione9	Proprietà3	Proprietà9	Host4	Host8	

Recensione10	Proprietà4	Proprietà19	Host4	Host10	
Numero Dati					
7836 Recensioni Analizzate		665 Proprietà		472 host	
Variabili					
Indice Gentilezza recensione		Indice Gentilezza proprietà, Rating proprietà		Indice Gentilezza Host, Rating Host	

Nel foglio recensioni per ogni recensione è associato un *indice di gentilezza recensione*, creato attraverso la ricerca delle parole precedentemente elencate nel testo della recensione. Ogni recensione è associata univocamente ad una ed una sola proprietà. Le recensioni vengono quindi accorpate nelle rispettive proprietà e si determina una nuova variabile, ovvero *l'indice di gentilezza proprietà*, che è frutto della media dell'indice di gentilezza associato alle recensioni per quella proprietà. In questo modo notiamo come 7836 recensioni analizzate inizialmente si raggruppano in 665 proprietà. Questo vuole dire che in media ogni proprietà analizzata è stata recensita circa 11 volte. Le proprietà a loro volta sono associate univocamente ad uno ed un solo host, che nella maggior parte dei casi dispone di più annunci nella piattaforma di Airbnb. Quindi, seguendo la stessa logica utilizzata nel passaggio da recensioni a proprietà, così si compone un indice di gentilezza relativo all'host, che è il risultato della media aritmetica dell'indice di gentilezza di ciascuna delle proprietà a lui associate, e un Rating Host, che è il frutto della media aritmetica dei Rating ottenuti in ciascuna delle proprietà a lui associate. Queste due ultime variabili verranno poi utilizzate nel modello di regressione successivamente spiegato.

4.2.5 Scelte metodologiche

In questo paragrafo si andranno ad analizzare le scelte metodologiche messe in atto durante la creazione dell'indice di gentilezza.

È stato scelto di creare un indice senza andare alcuni aggettivi di carattere negativo che abbassassero il punteggio per diverse ragioni. La prima ragione è

sicuramente poiché come spiegato all'interno del capitolo di Review della letteratura, è stato studiato come all'interno della piattaforma Airbnb, i commenti negativi fossero sempre estremamente tiepidi e mai diretti. Inoltre, è stato riscontrato come gli aggettivi negativi si possano rivolgere in prevalenza alla struttura, alla posizione o ai servizi: è infatti estremamente raro che l'host venga definito "scortese" o "maleducato", ma piuttosto sarebbe più possibile trovare come "il rapporto con l'host non è stato pienamente soddisfacente".

Creando un indice di gentilezza che integri anche le parole negative, c'è il rischio di assegnare punteggi estremamente bassi a host che nonostante nel loro piccolo fossero stati estremamente cordiali e cortesi, offrissero una proprietà non all'altezza delle aspettative del cliente (in termini magari di struttura, posizione e servizi) che per questo inseriva una recensione negativa.

Per i lavori futuri, qualora fosse possibile disporre di un tool che possa contestualizzare le parole ricercate, potrebbe essere interessante integrare all'interno dell'indice di gentilezza anche alcune formule negative, ma sincerandosi però che siano effettivamente riferite all'host e non all'esperienza in sé o alla struttura.

Inoltre, sempre in ottica di nuovi lavori futuri, sarebbe interessante andare a scoprire grazie a tools più potenti e flessibili di analisi del testo, laddove il nome dell'host è presente con nomignoli oppure nomi abbreviati.

4.2.6 Altre Variabili di input

In questo paragrafo, si andrà ad analizzare altre variabili, oltre all'indice di gentilezza, da poter correlare al rating.

Studiare la unica correlazione tra rating e gentilezza, non sarebbe infatti stato sufficiente, in quanto al fine di ottenere un modello completo e con valori di R quadro significativi le variabili devono necessariamente essere molteplici.

Sarebbe estremamente limitante provare a ricercare la correlazione tra una variabile come il rating complessivo di una esperienza su Airbnb solamente con l'indice di gentilezza creato.

Per questa ragione, al fine di svolgere un'analisi più puntuale e panoramica, è stato necessario introdurre nuove variabili, che per ipotesi potessero essere in qualche modo correlate con il rating finale di una proprietà.

Sono state introdotte due tipologie di variabili all'interno del modello: quelle che erano già presenti all'interno del dataset iniziale e quelle che invece sono state create manipolando i dati a disposizione.

Si partirà ad esplorare la prima categoria di variabili per passare successivamente alla seconda tipologia

4.2.6.1 Variabili presenti nel dataset

In questo paragrafo si andranno ad analizzare le variabili già presenti nel dataset di partenza e che sono state utilizzate per la costruzione del modello statistico.

Queste altre variabili verranno presentate e illustrate in dettaglio di seguito:

- **Location host:** questa variabile dà l'informazione della presenza o meno dell'host all'interno della regione di Barcellona. La variabile può assumere due valori: 0 se l'host non è presente all'interno della regione, 1 se l'host è presente all'interno della regione
- **Numero proprietà degli host:** questa variabile dà l'informazione del numero di proprietà attive per ogni host. Si ricorda infatti che ogni host può avere svariate proprietà, con un numero minimo di 1 proprietà. Questa variabile assume quindi valori da 1 fino a svariate centinaia
- **Bathrooms:** questa variabile indica il numero di bagni presenti all'interno di ogni proprietà
- **Numero di Servizi:** questa variabile indica il numero di servizi (o *amenities*) della proprietà. Tra questi servizi compaiono ad esempio la TV, l'aria condizionata, la piscina, la cucina ecc.
- **Beds:** questa variabile indica il numero di letti presenti all'interno della casa. È importante sottolineare come questa variabile faccia riferimento al numero di posti letti presenti all'interno della casa senza aggiungere ulteriori ospiti: non comprende quindi eventuali brandine o letti mobili

4.2.6.2 Nuove Variabili introdotte

Si vanno ad analizzare in questo paragrafo le variabili introdotte e costruite ad hoc, che sono state utilizzate per la costruzione del modello statistico.

Si riportano di seguito le variabili in oggetto:

- **Tempo di risposta:** per la costruzione di questa variabile, si è deciso di affidare un punteggio da 1 a 4 a seconda della velocità di risposta dell'host: più l'host risponde velocemente, più alto sarà il punteggio che riceve.

In particolare, i punteggi che sono stati sono stati:

- 1: se l'host tende a rispondere ai messaggi degli ospiti entro qualche giorno
- 2: se l'host tende a rispondere ai messaggi degli ospiti entro il primo giorno
- 3: se l'host tende a rispondere ai messaggi degli ospiti entro un numero esiguo di ore
- 4 se l'host tende a rispondere ai messaggi degli ospiti entro soltanto un'ora

Questa variabile è stata quindi creata con l'obiettivo di premiare la responsività ai messaggi da parte dell'host con un numero più alto

- **SuperHost * Indice Kindness Host:** questa variabile è una variabile interagita tra due ulteriori variabili: la prima è quella SuperHost, ossia se l'host è un Superhost (0 se non è un superhost, 1 se si tratta di un superhost) e la seconda è l'indice Kindness host descritta con estrema precisione nel paragrafo precedente. L'interazione di queste due variabili, ci mette nella condizione di poter analizzare l'effetto congiunto di essere dei SuperHost (ricordiamo essere una buona approssimazione del fatto di lavorare bene, puntuali ed essere gentili) e dell'indice kindness. Questa è probabilmente la migliore approssimazione della gentilezza, in quanto unisce l'analisi fatta esclusivamente sulle recensioni scritte, all'effetto Superhost che è estremamente più completo e variegato
- **Numero Parole Presentazione:** questa variabile fornisce l'informazione del numero di parole che l'host ha utilizzato per descriversi all'interno della

sua pagina personale. Questa variabile ha un set di valori estremamente variegato: esistono numerosi host che decidono di non inserire la propria descrizione (in questo caso il valore della variabile Numero Parole Presentazione è uguale a zero) e altri che invece si dilungano raccontando in dettaglio la loro storia, le loro passioni, gli hobby ecc.

Nel grafico di seguito è possibile vedere come il picco sia in corrispondenza delle 61 parole, ma che la distribuzione sia molto ripida all'inizio, e che vada a diminuire più lentamente: essendo un parametro molto importante per l'host (sono infatti non rilevanti gli host che presentano 0 parole all'interno della loro descrizione) sarà importante capire successivamente se sia una discriminante per la reputazione dell'host.

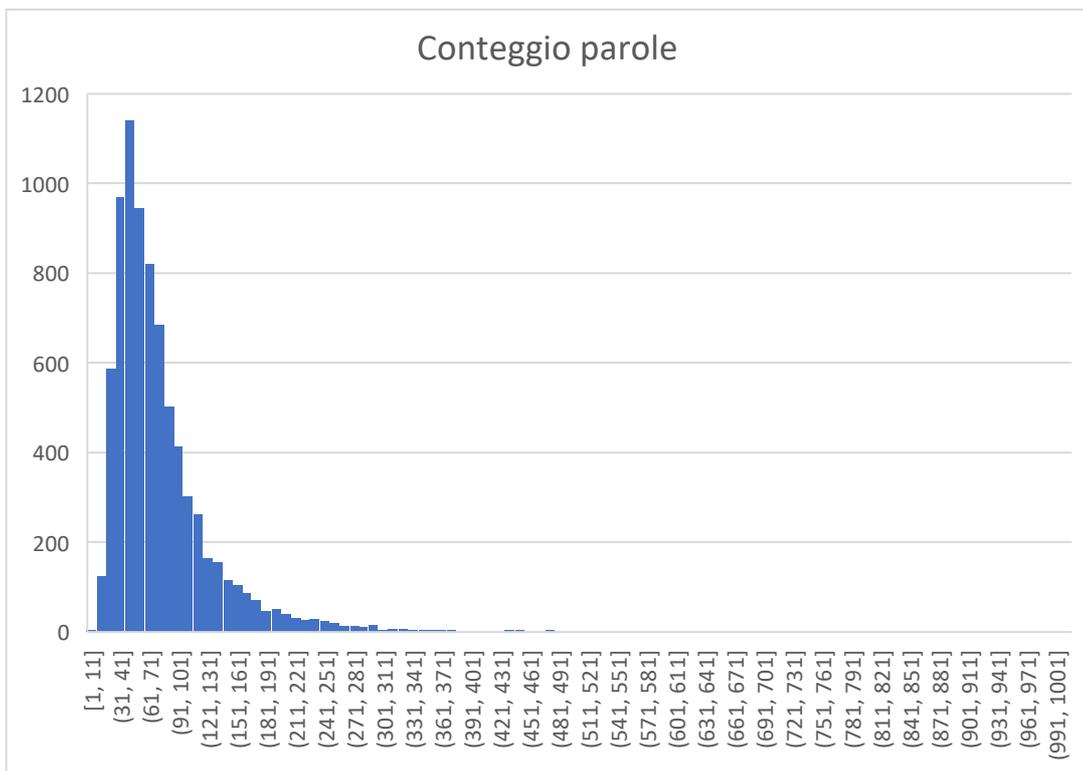


Figura 6: numero parole della descrizione dell'host con il rispettivo conteggio degli host

5. Modello Statistico

In questo capitolo, si andrà ad investigare come è stato costruito il modello di regressione lineare utilizzando le variabili viste precedentemente in dettaglio.

Il tool informatico utilizzato per lo studio della regressione lineare è Stata, molto rinomato e spesso utilizzato nella creazione di modelli statistici in ambito ingegneristico, economico e di scienze sociali.

Per agevolare l'utilizzo del software Stata, si è deciso di creare una baseline in Excel in formato ".xls". Si tratta di un unico foglio di calcolo chiamato "Baseline Modello" dove al proprio interno sono stati riportati unicamente le colonne di dati utili per l'analisi statistica. Per effettuare questa operazione, è stato necessario riportare come prime due colonne "ID Proprietà" e "ID host": grazie a queste due informazioni che fanno da chiave primaria per le altre tabelle, è stato possibile utilizzando la funzione Cerca Verticale di Excel riportare tutte le informazioni utili.

Una volta creato il modello, è stato sufficiente utilizzare Stata per importare i dati dal documento Excel, selezionando il foglio "Baseline Modello" come base di dati, indicando che la prima riga del foglio corrispondeva al nome della variabile. La maggior parte delle variabili del modello è di tipo discreto continuo, quindi possono assumere qualsiasi tipo di valore numerico in un determinato range, mentre solo due variabili sono di tipo dummy, ovvero possono assumere solo due valori [0;1].

In questo modello non sono presenti variabili di categoria, ovvero variabili che non derivano da operazioni di misura (non hanno unità di misura), ma da operazioni di classificazione e confronto. Un esempio di variabile categoria potrebbe essere il colore della macchina. Se ipotizziamo come scelte possibili il giallo, verde o rosso, la macchina può assumere un valore tra le tre scelte possibili, ma qualsiasi colore sia non si possono applicare funzioni matematiche come somma o prodotto.

Per le due variabili FlagBarcellona, che assume valore 1 se l'host vive a Barcellona e 0 in caso negativo, e **SuperHost**, che assume valore 1 nel momento in cui l'host è identificato come Superhost, è stato necessario eseguire il comando in STATA che permette la creazione della dummy:

tabulate VARIABILE1, generate (Dummy_var1)

Questo processo crea un numero di variabili pari al numero di valori che la variabile può assumere: in questo caso ha creato due variabili poiché i valori erano appunto solo 0 e 1. È importante ricordare che nel modello non sono state inserite entrambi le dummy create, ma solo una dummy per ciascuna variabile.

Nel successivo capitolo verranno ripercorse con precisione la generazione delle variabili dummy, esposti i risultati del modello di regressione.

Il modello è suddiviso in 4 sotto modelli, in cui la variabile dipendente è sempre la stessa, ovvero il Rating della proprietà. I sotto modelli si differenziano per le variabili dipendenti considerate: inizialmente si è svolta una analisi considerando le singole variabili indipendenti separatamente.

Infine, verrà presentato l'ultimo modello che considera tutte le variabili precedentemente analizzate contemporaneamente e dimostra quelle che maggiormente influenzano il rating finale assegnato dall'utente alla proprietà.

5.1 I modelli di regressione

In questo capitolo, si presenteranno e si commenteranno i risultati ottenuti dai diversi modelli statistici impostati ed elaborati con Stata.

Si inizierà presentando la creazione e i dati relativi alle variabili dummy, poi si passerà all'analisi dei modelli di regressione lineare semplici con singole variabili: qui sarà possibile valutare l'effetto di ogni singola variabile sul rating complessivo. Successivamente si passerà all'analisi dei modelli intermedi, ossia quei modelli che mettono in relazione le variabili raggruppate per tipologia con i rating complessivi delle proprietà.

Infine, si passerà ai modelli finali costruiti con più variabili e verranno commentati i risultati ottenuti: qui sarà possibile andare ad analizzare l'effetto di tutte le variabili presentate nel capitolo precedente con il rating complessivo di una proprietà, e con i rating intermedi più significativi.

Per completezza di analisi, si ricorda che il modello di regressione lineare è stato costruito per ricercare eventuali correlazioni tra il rating finale di una proprietà con altre variabili di varia natura, in particolare legate della gentilezza dell'host, al suo atteggiamento e alle caratteristiche della proprietà. Tuttavia, il rating complessivo è una variabile estremamente complessa, derivante da altri rating intermedi, quindi può essere che questo determini la presenza al suo interno di possibili bias di valutazione ed opinioni personali. Si è perciò deciso di cercare la correlazione da variabili oggettivamente misurabili e presentate in dettaglio nel capitolo precedente.

Iniziamo con l'analisi andando a vedere le due variabili dummy utilizzate all'interno del modello statistico: si riporta di seguito il prospetto creato con Stata.

```
. tabulate FlagBarcellona,generate (D1_Location_Barcellona)
```

Flag Barcellona	Freq.	Percent	Cum.
0	3,538	24.89	24.89
1	10,676	75.11	100.00
Total	14,214	100.00	

```
. tabulate Superhost,generate (D2_Superhost)
```

Superhost	Freq.	Percent	Cum.
0	10,852	76.36	76.36
1	3,359	23.64	100.00
Total	14,211	100.00	

Figura 7: tabella con frequenza e percentuale delle variabili dummy

Le variabili dummy sono quelle variabili che possono tipicamente assumere 0 o 1 come valore, e che solitamente rappresentano il verificarsi o meno di una certa condizione (solitamente la variabile vale 1 quando la condizione è verificata e vale invece 0 quando la variabile non è verificata). Queste variabili servono per catturare l'effetto sul modello statistico della condizione a cui fa riferimento la variabile stessa.

Le prima variabile dummy che è stata creata, è "D1_Location_Barcellona" calcolata in maniera diretta dalla colonna del file Excel chiamata "Flag_Barcellona". Possiamo vedere che su tutti i campioni su cui è disponibile l'informazione della location dell'host, ossia 14.214 misurazioni disponibili, in valore assoluto 3.538 sono al di fuori della regione di Barcellona mentre 10.676 sono al suo interno. Si ricorda per completezza di analisi, che gli host al di fuori dalla regione di Barcellona, possono essere sia all'interno del confine di stato

spagnolo, sia all'esterno. Per tenere la variabile snella ed efficace, è stato scelto come ipotesi di lavoro che l'impatto principale a livello di interazione host guest fosse la presenza fisica: non vengono quindi considerati altri fattori come le barriere linguistiche o le difficoltà di comunicazione dettate da differenti fusi orari.

I dati sono molto incoraggianti, in quanto il 75,11% degli host è presente all'interno della regione di Barcellona: in questo caso c'è la possibilità di interagire quindi di persona con il proprio ospite.

Il rimanente 24,89%, invece, è residente all'esterno della suddetta regione, e sarà quindi in questo caso impossibile il verificarsi di un incontro e quindi una interazione diretta con il proprio ospite.

La seconda variabile Dummy è stata creata per andare a lavorare sulla variabile interagita: si tratta della variabile "D2_SuperHost" ricavata a partire dai dati della colonna "SuperHost" del foglio Excel.

Questa variabile ci mostra come, tra tutti i 14.211 campioni per i quali l'informazione è disponibile, 10.852 campioni non presentano il flag di SuperHost, contro i 3.359 che invece lo possiedono.

In termini percentuali, si tratta del 76,36% dei campioni a non avere il riconoscimento offerto da Airbnb per prestazioni particolarmente elevate, contro il 23,64% di host con il riconoscimento.

5.1.1 I modelli di regressione semplici

In questo paragrafo si andranno ad analizzare tutti i modelli di regressione semplici, dove si andranno a studiare i singoli effetti di una variabile alla volta sul rating complessivo di una proprietà.

È importante ricordare l'approccio metodologico utilizzato per la creazione dei modelli: è stato scelto di utilizzare il rating complessivo di una proprietà come variabile dipendente (Y) in quanto il prodotto finale dell'esperienza dell'utente chiaramente dipende in maniera molto stretta da numerosi altri fattori (che saranno le variabili indipendenti).

Inizialmente però i modelli di regressione che saranno studiati sono quindi della tipologia:

- $Y = m X_1 + c_1$
- $Y = m X_2 + c_2$
- ...

5.1.1.1 Correlazione tra Rating complessivo e Tempo di risposta

. regress RATING Tempo_Risposta						
Source	SS	df	MS	Number of obs	=	14,061
Model	128.588288	1	128.588288	F(1, 14059)	=	1.26
Residual	1430315.84	14,059	101.736669	Prob > F	=	0.2609
Total	1430444.42	14,060	101.738579	R-squared	=	0.0001
				Adj R-squared	=	0.0000
				Root MSE	=	10.086

RATING	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Tempo_Risposta	-.0837929	.0745324	-1.12	0.261	-.2298863 .0623005
_cons	91.317	.2663085	342.90	0.000	90.795 91.839

Figura 8: modello Rating complessivo e Tempo di risposta

Il primo modello di regressione semplice che si va a studiare è la correlazione tra il rating complessivo di una proprietà e la variabile "Tempo_Risposta" per quella determinata proprietà.

Partendo da un numero di campioni pari a 14.061 per i quali è disponibile l'informazione riguardante il tempo di risposta, il primo risultato che salta

all'occhio è un coefficiente negativo: chiaramente il fatto di far aspettare l'ospite non può che avere un impatto negativo sul rating finale.

Tuttavia, il P-value del 26,1%, sottolinea la scarsa significatività della variabile Tempo di Risposta sul rating finale: nonostante il rating complessivo dell'appartamento sia ricavata direttamente dal rating intermedio di comunicazione, è importante segnalare che non subisce una significativa influenza dal tempo di attesa. Questo può portare a pensare che sia effettivamente il fatto di ricevere una risposta la vera discriminante, e non il tempo in cui l'host fornisce le informazioni.

Inoltre, è importante sottolineare come il valore del parametro R-Squared sia allo 0,01% per due ragioni: in prima battuta, poiché il modello non è sufficientemente spiegato da una variabile sola, in seconda battuta perché, come detto in precedenza, questa variabile non ha un impatto significativo sull'andamento del rating. Sarà tuttavia importante avere un valore di R-Squared significativo nel modello finale per avere la certezza di aver creato una regressione lineare a più variabili che sia effettivamente significativa.

5.1.1.2 Correlazione tra Rating Complessivo e Numero Delle Proprietà di un host

. regress RATING Numero_proprietà						
Source	SS	df	MS	Number of obs	=	14,061
Model	18837.6746	1	18837.6746	F(1, 14059)	=	187.62
Residual	1411606.75	14,059	100.405914	Prob > F	=	0.0000
				R-squared	=	0.0132
				Adj R-squared	=	0.0131
Total	1430444.42	14,060	101.738579	Root MSE	=	10.02

RATING	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Numero_proprietà	-.0201132	.0014684	-13.70	0.000	-.0229914	-.0172349
_cons	91.40822	.0888259	1029.07	0.000	91.23411	91.58233

Figura 9: modello Rating Complessivo e Numero Delle Proprietà di un host

Questo modello di regressione rappresenta la correlazione tra il rating complessivo di un appartamento (variabile dipendente) e il numero di proprietà di un host (variabile indipendente).

Anche in questo il numero di campioni disponibili sulla quale è stata fatta l'analisi è pari a 14.061. È molto interessante anche in questo caso riscontrare che ci sia un coefficiente di correlazione negativo: come da ipotesi, all'aumentare del numero di proprietà corrisponde una diminuzione del livello di reputazione dell'host. Questo effetto può essere spiegato per due principali ragioni: in primo luogo poiché dal momento che un host inizia a dover gestire un numero consistente di proprietà, ha un decremento della qualità del servizio offerto ai suoi ospiti in termini di tempo da dedicare, prontezza a rispondere e disponibilità. In secondo luogo, poiché come descritto nel capitolo di review della letteratura, l'utente medio di Airbnb sceglie la piattaforma poiché, diversamente dal servizio alberghiero classico, oltre al pernottamento semplice è alla ricerca di una esperienza sociale vera e propria. Questa esperienza viene però a mancare se invece che un host che ti accoglie a casa sua, si ha davanti un vero e proprio "gestore" di diverse proprietà.

Inoltre, è estremamente interessante osservare il valore del P-value prossimo allo zero: questo è indice di grande significatività della variabile in oggetto. Il valore R-squared all'1,32% indica che il modello non è molto rappresentato dalla variabile della regressione in oggetto.

5.1.1.3 Correlazione tra Rating Complessivo e Numero dei bagni della proprietà

. regress RATING Bagni						
Source	SS	df	MS	Number of obs	=	14,062
Model	2533.5914	1	2533.5914	F(1, 14060)	=	24.95
Residual	1428000.3	14,060	101.564744	Prob > F	=	0.0000
				R-squared	=	0.0018
				Adj R-squared	=	0.0017
Total	1430533.89	14,061	101.737706	Root MSE	=	10.078
RATING	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Bagni	.0652252	.0130593	4.99	0.000	.090823	-.0396273
_cons	91.90487	.194192	473.27	0.000	91.52423	92.28551

Figura 10: modello Rating Complessivo e Numero dei bagni della proprietà

In questo paragrafo si va a studiare ed analizzare il modello di correlazione tra il Rating complessivo della proprietà e il numero di bagni presenti all'interno della proprietà.

I dati validano le ipotesi fatte nei capitoli precedenti: essendo il coefficiente di correlazione positivo, all'aumentare del numero dei bagni in una proprietà (a parità di altre variabili) aumenterà anche il rating complessivo.

Questa correlazione esprime un concetto non particolarmente inaspettato: chiaramente, soprattutto quando si viaggia in gruppo, è un fatto molto gradito avere a disposizione più di un bagno: in particolar modo per le case con un numero significativo di ospiti, avere più bagni a disposizione può migliorare sensibilmente la qualità del soggiorno, la qualità della privacy e diminuire i tempi di sosta a casa.

Per quanto detto, non è sorprendente che il P-value prossimo allo 0 indichi la rilevanza statistica della variabile in oggetto.

Anche in questo caso però, è importante sottolineare come il valore di R-squared sia solamente dello 0,18%: anche in questo caso la variabile rating non è rappresentata sufficientemente dal modello creato.

5.1.1.4 Correlazione tra Rating Complessivo e Numero dei letti della proprietà

. regress RATING Letti						
Source	SS	df	MS	Number of obs	=	13,980
Model	7883.54449	1	7883.54449	F(1, 13978)	=	78.50
Residual	1403841.51	13,978	100.432215	Prob > F	=	0.0000
				R-squared	=	0.0056
				Adj R-squared	=	0.0055
Total	1411725.05	13,979	100.988987	Root MSE	=	10.022
RATING	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Letti	-.3895447	.0439676	-8.86	0.000	-.4757271	-.3033622
_cons	91.91831	.1324286	694.10	0.000	91.65873	92.17789

Figura 11: modello Rating Complessivo e Numero dei letti della proprietà

In questo paragrafo si va a investigare la relazione tra il numero di letti presenti all'interno della proprietà e il rating complessivo della proprietà.

Il numero di campioni per il quale è disponibile l'informazione del numero di letti è leggermente inferiore (ma non significativamente) rispetto ai modelli visti in precedenza: si tratta di 13.980 campioni.

I risultati del modello in oggetto, indicano una correlazione negativa tra il numero di letti e il livello di reputazione dell'host su una determinata proprietà. Questo dato è giustificato dal fatto che nessun ospite gradisce, a parità di numero di bagni, avere tanti letti: per scongiurare l'effetto "dormitorio" è infatti necessario che i servizi della proprietà, la superficie a disposizione e il numero di bagni siano sufficienti per tutti gli ospiti. Quindi a fronte di questa considerazione, un aumento del numero di letti a parità di tutto il resto porta minore comfort agli ospiti, che si traduce in una diminuzione del rating della proprietà. È interessante notare come in valore assoluto, il coefficiente è di un ordine di grandezza superiore rispetto a quelli visti fino ad ora: ciò può portare a pensare che inserire all'interno della proprietà un numero sproporzionato di posti letti per massimizzare il profitto, può portare a una perdita di reputazione in termini di rating.

Anche in questo caso un valore prossimo allo 0 del P-value ci indica estrema significatività della variabile in oggetto.

Tuttavia, anche in questo caso, il valore di R-squared di 0,55% ci indica che il modello non è spiegato dalla variabile che abbiamo scelto.

5.1.1.5 Correlazione tra Rating Complessivo e Numero dei servizi della proprietà

. regress RATING Numero_Servizi						
Source	SS	df	MS	Number of obs	=	14,064
Model	43979.525	1	43979.525	F(1, 14062)	=	446.02
Residual	1386571.03	14,062	98.6041128	Prob > F	=	0.0000
Total	1430550.56	14,063	101.724423	R-squared	=	0.0307
				Adj R-squared	=	0.0307
				Root MSE	=	9.93

RATING	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Numero_Servizi	.1732828	.008205	21.12	0.000	.1571999	.1893657
_cons	87.20716	.1995753	436.96	0.000	86.81597	87.59836

Figura 12: modello Rating Complessivo e Numero dei servizi della proprietà

In questo modello, si va a ricercare la correlazione tra il rating complessivo della proprietà e il numero di servizi che essa dispone (o *amenities*).

I risultati del modello confermano ciò che si poteva facilmente ipotizzare: il coefficiente di correlazione è positivo, quindi all'aumentare del numero di servizi a disposizione per gli ospiti aumenterà il rating complessivo dell'appartamento. Anche se il valore aggiunto di alcuni servizi non è particolarmente importante per alcuni ospiti, il solo fatto di poter avere un servizio aggiuntivo a disposizione fa sì che il valore percepito della casa a parità di tutti gli altri fattori aumenti.

La variabile del numero di servizi della proprietà è molto significativa come ci indica il P-value prossimo allo zero.

Il valore di R-squared poco superiore al 3% indica però che il modello non è sufficientemente spiegato dalla variabile scelta.

5.1.1.6 Correlazione tra Rating Complessivo e numero delle parole di presentazione

. regress RATING Numero_parole_presentazione						
Source	SS	df	MS	Number of obs	=	14,064
Model	634.439568	1	634.439568	F(1, 14062)	=	6.24
Residual	1429916.12	14,062	101.68654	Prob > F	=	0.0125
Total	1430550.56	14,063	101.724423	R-squared	=	0.0004
				Adj R-squared	=	0.0004
				Root MSE	=	10.084
RATING						
	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Numero_parole_presentazione	.0030577	.0012241	2.50	0.013	.0006582	.0054572
_cons	90.88673	.1032751	880.05	0.000	90.6843	91.08916

Figura 13: modello Rating Complessivo e numero delle parole di presentazione

In questo paragrafo si va ad analizzare l'impatto del numero di parole della presentazione dell'host sul rating finale. Ci si chiede se sia remunerato lo sforzo dell'host di descriversi al meglio rispetto a chi invece fa una descrizione molto breve e sintetica di sé.

I risultati ottenuti dopo aver fatto girare il modello sono estremamente interessanti: vi è correlazione positiva tra il numero di parole che l'host scrive nella sua presentazione e il rating che ottiene. Questo risultato ci segnala che agli ospiti piace sapere in anticipo chi è l'host, cosa fa e quali sono le sue passioni. Inoltre, avendo una descrizione dettagliata dell'host, l'ospite può più agevolmente farsi ospitare da persone con interessi in comune o che comunque condividano qualcosa.

Il coefficiente però è in valore assoluto poco elevato, quindi lo sforzo dell'host di pubblicare una descrizione adeguata potrebbe non essere adeguatamente ricompensato in termini di ritorno sulla sua reputazione. Tuttavia, per quanto minimo sia l'effetto, per massimizzare il proprio rating l'host dovrebbe procedere all'inserimento di una sostanziosa descrizione della propria persona.

Il valore del P-value è molto buono, in quanto è 1,3%: ci indica una grande significatività della variabile all'interno del modello impostato.

Tuttavia, un valore dello 0,04% di R-squared indica una totale mancanza di significatività del modello nel suo complesso.

5.1.1.7 Correlazione tra Rating e posizione dell'host

. regress RATING D1_Location_Barcellona2						
Source	SS	df	MS	Number of obs	=	14,064
Model	7113.31513	1	7113.31513	F(1, 14062)	=	70.27
Residual	1423437.24	14,062	101.225803	Prob > F	=	0.0000
Total	1430550.56	14,063	101.724423	R-squared	=	0.0050
				Adj R-squared	=	0.0049
				Root MSE	=	10.061
RATING						
	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
D1_Location_Barcellona2	1.649665	.196791	8.38	0.000	1.263928	2.035401
_cons	89.79049	.1707973	525.71	0.000	89.4557	90.12528

Figura 14: modello Rating e posizione dell'host

Questo modello studia la correlazione tra la variabile dummy legata alla presenza fisica dell'host all'intero dell'area di Barcellona e il rating complessivo della proprietà.

Dai risultati si scopre come l'ipotesi formulata nei capitoli precedenti, trova una conferma numerica: vi è correlazione con coefficiente positivo.

Infatti, il fatto di essere presenti fisicamente nell'area di Barcellona, produce un effetto positivo sul rating dell'host. Questo fenomeno è spiegabile dal fatto che poter interagire di persona con il proprio ospite, crea in lui un vero e proprio valore aggiunto che si traduce in maggiore soddisfazione che a sua volta porta a un aumento del rating complessivo della proprietà. Se ciò che infatti l'ospite ricerca è un'esperienza sociale oltre che un semplice pernottamento, la presenza fisica è una condizione necessaria (anche se assolutamente non sufficiente) per far sì che questo si verifichi.

È importante sottolineare come il valore in termini assoluti del coefficiente sia significativamente più alto di quelli visti fino a questo punto: essendo il coefficiente uguale a circa 1,65, è caldamente consigliabile all'host di rimanere in prossimità delle proprietà in affitto per avere il massimo riscontro in termini di rating.

Andando ad analizzare i parametri, vediamo che il valore uguale a 0 del P-value indica grande significatività della variabile nel modello creato.

Tuttavia, come i modelli precedenti (e come è fisiologico che sia) andando a inserire una sola variabile indipendente per spiegare una variabile complessa come il rating di una proprietà il valore di R-squared è significativamente basso: 0,5%. Il modello non è quindi sufficientemente spiegato.

5.1.1.8 Correlazione tra Rating e numero di recensioni per mese

. regress RATING Recensione_per_mese						
Source	SS	df	MS	Number of obs	=	14,064
Model	21500.3522	1	21500.3522	F(1, 14062)	=	214.57
Residual	1409050.21	14,062	100.202689	Prob > F	=	0.0000
Total	1430550.56	14,063	101.724423	R-squared	=	0.0150
				Adj R-squared	=	0.0150
				Root MSE	=	10.01

RATING	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Recensione_per_mese	.0073533	.000502	14.65	0.000	.0063693 .0083373
_cons	89.76421	.1209502	742.16	0.000	89.52713 90.00129

Figura 15: modello Rating e numero di recensioni per mese

In questo paragrafo si andrà ad analizzare la correlazione tra il Rating complessivo della proprietà e la media delle recensioni ricevute per mese della proprietà.

Possiamo vedere che, come da ipotesi, più una proprietà è recensita più aumenta il valore del rating della casa. È interessante osservare come, così come si è visto precedentemente per le recensioni eccessivamente positive rispetto quanto dovrebbero essere (riferimento al capitolo 1), anche in questo caso si può notare lo stesso effetto distorto: osservando un coefficiente positivo, possiamo trarre la conclusione che più una casa ha ricevuto recensioni, più il suo rating complessivo aumenta. Nel caso in cui la proprietà fosse particolarmente scadente, oppure l'host particolarmente sgradevole dovrebbe però essere il contrario: all'aumentare del numero di utenti insoddisfatti dovrebbe diminuire il rating complessivo. Tuttavia, il fatto che il coefficiente sia positivo, porta alla conclusione che l'effetto di positività supera quello di negatività in termini assoluti. Ossia

statisticamente sono più rilevanti gli host che vedono benefici nell'essere recensiti bene rispetto agli altri.

Il valore del P-value uguale a 0, indica significatività della variabile all'interno del modello creato, nonostante il valore del coefficiente in termini assoluti sia sostanzialmente poco impattante sul rating finale.

Il valore di R-squared allo 1,5% indica che il modello non è adeguatamente spiegato.

5.1.1.9 Correlazione tra Rating e Indice Kindness

. regress RATING Indice_Gentilezza_Host						
Source	SS	df	MS	Number of obs	=	1,615
Model	1047.27194	1	1047.27194	F(1, 1613)	=	15.67
Residual	107817.066	1,613	66.8425705	Prob > F	=	0.0001
				R-squared	=	0.0096
				Adj R-squared	=	0.0090
Total	108864.338	1,614	67.4500236	Root MSE	=	8.1757
RATING	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Indice_Gentilezza_Host	1.303316	.3292657	3.96	0.000	.6574823	1.949149
_cons	88.0167	.6143994	143.26	0.000	86.8116	89.22181

Figura 16: modello Rating e Indice Kindness

In questo paragrafo si analizza invece il modello più interessante e centrale dell'elaborato: l'effetto della gentilezza rappresentato in questo caso dall'indice kindness sul rating complessivo della proprietà.

Si ricorda che l'indice gentilezza è stato calcolato unicamente basandosi sulle recensioni degli ospiti, ed è basato su alcuni aggettivi o formule che associate all'host vanno a misurare la gentilezza in senso stretto del termine. Inoltre, è importante sottolineare come non essendoci una categoria di "disponibilità dell'host" nella quale l'ospite può fornire una valutazione diretta (seppur personale) della gentilezza, è stato necessario affidarsi ad una variabile indiretta, che sarà quindi inevitabilmente meno precisa.

Inoltre, il numero di osservazioni è sensibilmente inferiore rispetto agli altri modelli (si attesta a 1615 osservazioni), in quanto sono state prese in

considerazione solo le proprietà collegate agli host per le quali sia stato possibile costruire un indice di gentilezza, che avesse, quindi, le caratteristiche descritte nel capitolo precedente (recensioni scritte in inglese e almeno tre recensioni scritte negli anni 2019 e 2020).

I risultati del modello sono estremamente positivi: il valore del P-value uguale a zero, indica estrema significatività della variabile dell'indice kindness sul rating complessivo di una proprietà. Il coefficiente positivo indica che all'aumentare del livello di gentilezza di un host nei confronti del suo guest, l'effetto sul rating complessivo della sua proprietà è positivo: conviene quindi introdurre la gentilezza all'interno della propria strategia di host, se l'obiettivo è quello di massimizzare la propria reputazione e quindi il rating delle proprie proprietà.

È inoltre estremamente importante sottolineare, come il valore assoluto del coefficiente (pari a 1,3) sia molto più impattante sul rating finale rispetto a tutte le variabili legate alla proprietà viste fin ora (come il numero di bagni o il numero dei servizi): questa informazione sottolinea l'importanza della relazione interpersonale che si stabilisce tra host e guest, che rappresenta qualcosa di molto diverso dalla relazione che si forma con un albergatore tradizionale.

Il valore di R-Squared è prossimo all'1%, in continuità con i modelli precedentemente analizzati.

La prima conclusione è quindi la correlazione positiva tra la gentilezza e il rating, ma ancora manca l'informazione di quanto effettivamente sia rilevante nel quadro complessivo rispetto ad altre variabili.

5.1.1.10 Correlazione tra Rating e Indice Kindness * Superhost

. regress RATING Superhost_IndiceGentilezza							
Source	SS	df	MS	Number of obs	=	14,064	
Model	8260.11244	1	8260.11244	F(1, 14062)	=	81.67	
Residual	1422290.45	14,062	101.14425	Prob > F	=	0.0000	
				R-squared	=	0.0058	
				Adj R-squared	=	0.0057	
Total	1430550.56	14,063	101.724423	Root MSE	=	10.057	
RATING		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Superhost_IndiceGentilezza		2.208277	.2443606	9.04	0.000	1.729298	2.687256
_cons		90.91112	.085872	1058.68	0.000	90.7428	91.07944

Figura 17: modello Rating e Indice Kindness * Superhost

In questo modello, si va ad analizzare la correlazione che c'è tra la variabile interagita formata da Indice Kindness * Superhost e il rating complessivo associato alla proprietà.

È interessante ricercare questa correlazione, poiché questa variabile ci da una precisa indicazione dell'interazione della variabile di gentilezza pura espressa dall'indice kindness e la buona capacità di offrire un servizio di livello dell'host: è quindi la perfetta misura di gentilezza espressa in senso più ampio, intesa come l'unione della buona logistica, ottima disponibilità e appunto cortesia verso i propri ospiti.

Si può notare che il coefficiente associato a questa variabile è positivo: l'aumento della variabile interagita Indice Kindness * SuperHost si traduce nell'aumento del rating complessivo dell'abitazione.

Questo risultato conferma le ipotesi fatte in precedenza ed è allineato con quanto scoperto nei grafici precedenti: sarà tuttavia interessante scoprire nei paragrafi successivi il peso di questa variabile interagita rispetto alle altre e rispetto al puro e semplice indice kindness.

Si sottolinea inoltre che questa variabile possiede il coefficiente maggiore in termini assoluti rispetto a tutti quelli visti fino ad ora (valore positivo di circa 2,2). Si tratta quindi del parametro più significativo che un host dovrebbe monitorare per massimizzare il proprio rating.

Sarà fondamentale poter capire se il peso e l'importanza della gentilezza può superare altri parametri come quelli fisici della casa a livello di impatto sul rating complessivo (come i servizi, il numero di bagni, ecc.) o i servizi di logistica richiesti da un buon host (avere bassi tempi di risposta, non avere intoppi al check-in, essere disponibile durante il soggiorno, ecc.), o se semplicemente possa avere un effetto coadiuvante a parità di condizioni, oppure se sia meno rilevante.

5.1.2 Modelli divisi per tipologia di variabili

In questo paragrafo si andranno ad analizzare alcuni modelli intermedi, creati con la logica di raggruppare le variabili per le seguenti tipologie e cercare di capire quali effetti siano più marcati rispetto agli altri:

- Variabili relative alla proprietà: si andrà a studiare tra tutte le variabili relative alla proprietà, quali sono più significative delle altre
- Variabili legate all'host: si andrà a studiare quale delle variabili legate alle performance e caratteristiche dell'host siano più significative delle altre
- Variabili legate alla gentilezza: si andrà a studiare quale delle variabili create sia più significativa in termini di impatto di gentilezza

Nei prossimi paragrafi si andranno ad analizzare uno alla volta i modelli appena menzionati, analizzando per ognuno i risultati prodotti.

5.1.2.1 Correlazione tra Rating e variabili relative alla proprietà

In questo paragrafo si andrà ad analizzare il modello che mette in relazione il rating complessivo della proprietà con le variabili relative alla proprietà stessa: è infatti interessante vedere quali tra le variabili fisiche prese in considerazione nello studio possa influenzare maggiormente il rating e quali siano meno rilevanti. Si riportano di seguito i risultati ottenuti dal modello.

. regress RATING Bagni Letti Numero_Servizi						
Source	SS	df	MS		Number of obs	= 13,979
Model	65647.342	3	21882.4473		F(3, 13975)	= 227.18
Residual	1346076.74	13,975	96.3203392		Prob > F	= 0.0000
					R-squared	= 0.0465
					Adj R-squared	= 0.0463
Total	1411724.08	13,978	100.996143		Root MSE	= 9.8143

RATING	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Bagni	.0243945	.0144904	1.68	0.092	.0527977	.0040087
Letti	-.599425	.0498986	-12.01	0.000	-.6972329	-.5016171
Numero_Servizi	.2050156	.008378	24.47	0.000	.1885936	.2214377
_cons	88.18932	.2474174	356.44	0.000	87.70435	88.67429

Figura 18: modello Rating e variabili relative alla proprietà

Si può osservare dai dati che tra le variabili della casa, le due più rilevanti da tenere sotto controllo dall'host per cercare di massimizzare il proprio rating complessivo sono il numero di letti, correlato negativamente e il numero di servizi. Da questa analisi si evince che il guest preferisce una maggior attenzione alla privacy (determinata dal numero di letti), probabilmente associata anche ad un maggior comfort e relax, piuttosto che ai servizi offerti. Infatti, l'impatto dei servizi aggiuntivi sul rating complessivo è inferiore in valore assoluto a quello associato al numero di letti. Si può notare che il numero di bagni non abbia un impatto così significativo rispetto al resto, probabilmente perché il guest non valuta di significativo valore aggiunto avere un numero di bagni superiore ad un certo numero soglia (ad esempio, avere più di due bagni talvolta può essere percepito come di scarso valore aggiunto). Anche il valore del P-value della variabile collegata al numero di bagni, che vale 9,2%, indica minore significatività delle altre due variabili (numero dei servizi della proprietà e numero di letti) che invece hanno valore uguale a 0.

Tuttavia, il valore di R-squared al 4,65% indica che il modello non è sufficientemente spiegato con le sole variabili fisiche relative ai servizi e alla proprietà.

5.1.2.2 Correlazione tra Rating e variabili relative alle performance dell'host

In questo paragrafo si andrà ad analizzare la correlazione tra il rating complessivo di una proprietà e le variabili collegate alle caratteristiche dell'host e delle sue performance.

Si riportano di seguito i risultati ottenuti dal modello.

. regress RATING Tempo_Risposta Numero_proprietà Numero_parole_presentazione D1_Location_Barcellona2						
Source	SS	df	MS			
Model	28337.3983	4	7084.34957	Number of obs	=	14,061
Residual	1402107.03	14,056	99.7514958	F(4, 14056)	=	71.02
				Prob > F	=	0.0000
				R-squared	=	0.0198
				Adj R-squared	=	0.0195
Total	1430444.42	14,060	101.738579	Root MSE	=	9.9876

	RATING	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Tempo_Risposta		-.0034504	.0747542	-0.05	0.963	-.1499786	.1430778
Numero_proprietà		-.0214805	.0014845	-14.47	0.000	-.0243903	-.0185706
Numero_parole_presentazione		.0026274	.0012526	2.10	0.036	.0001721	.0050826
D1_Location_Barcellona2		1.779179	.2006074	8.87	0.000	1.385962	2.172396
_cons		89.97908	.2956176	304.38	0.000	89.39963	90.55853

Figura 19: modello Rating e variabili relative alle performance dell'host

Si può osservare dai dati ottenuti dal modello che tutte le variabili risultano significative, fatta eccezione del tempo di risposta, che possiede un P-value del 96,3% oltre che un coefficiente negativo ma in valore assoluto poco significativo. Altre 2 variabili, ossia il numero di proprietà dell'host e la presenza fisica dell'host nella regione di Barcellona presentano valori di P-value prossimi allo zero. La variabile collegata al numero di parole impiegate dall'host per presentarsi ha un valore di P-value del 3,6%, indice comunque di grande significatività della variabile.

In questo modello intermedio, è possibile vedere come la variabile che possiede in valore assoluto l'impatto superiore è la presenza fisica dell'host nella regione di Barcellona: questo poiché tutto il valore aggiunto collegato all'host, non può realmente attivarsi se l'host non è presente in persona.

Il modello riesce infatti a esprimere numericamente come per quanto ci si possa impegnare nel ridurre al massimo i tempi di risposta e presentarsi nel miglior

modo possibile, descrivendo le proprie passioni e attitudini, non si otterrà mai un impatto sul rating paragonabile a quello determinato dalla presenza fisica sul luogo.

È inoltre interessante sottolineare come l'effetto negativo dettato dalla rapidità della risposta da parte dell'host, non sia particolarmente elevato in termini di valore assoluto del coefficiente: questo fatto probabilmente è dovuto al fatto che questa fase sia a livello temporale molto distaccata dalla fase in cui il guest procede con la stesura della recensione e assegnazione dei rating. Di solito i maggiori contatti si hanno in fase di organizzazione, quindi precedentemente al soggiorno vero e proprio, mentre la recensione viene scritta al termine del soggiorno stesso: questo lasso di tempo può permettere al guest di valutare aspetti più legati all'esperienza vissuta e meno alla fase preliminare.

La variabile collegata al numero delle proprietà, in conclusione, presenta correlazione negativa come da ipotesi; tuttavia dato lo scarso valore in termini di valore assoluto del coefficiente, si è portati a pensare che effettivamente sia meno influente delle altre.

Il valore di R-squared è vicino al 2%, e indica che il modello non sia sufficientemente spiegato.

5.1.2.3 Correlazione tra Rating e variabili relative alla gentilezza dell'host

In questo paragrafo si andrà a valutare il modello che prende in considerazione la correlazione tra il rating complessivo delle proprietà con le variabili costruite ad hoc collegate alla gentilezza dell'host.

Si riportano di seguito i risultati ottenuti dal modello.

. regress RATING Indice_Gentilezza_Host Superhost_IndiceGentilezza						
Source	SS	df	MS	Number of obs	=	1,615
Model	13825.2669	2	6912.63343	F(2, 1612)	=	117.25
Residual	95039.0712	1,612	58.9572402	Prob > F	=	0.0000
				R-squared	=	0.1270
				Adj R-squared	=	0.1259
Total	108864.338	1,614	67.4500236	Root MSE	=	7.6784

RATING	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Indice_Gentilezza_Host	.6962548	.3377487	2.06	0.039	1.358728	.0337821
Superhost_IndiceGentilezza	3.34427	.2271635	14.72	0.000	2.898704	3.789837
_cons	89.92826	.5914514	152.05	0.000	88.76817	91.08836

Figura 20: modello Rating e variabili relative alla gentilezza dell'host

Come nel sotto modello dell'indice gentilezza (o indice *kindness*) è possibile notare come il numero di osservazioni sia sensibilmente inferiore rispetto agli altri modelli, in quanto sono state prese in considerazione solo le proprietà collegate agli host per la quale sia stato possibile costruire un indice di gentilezza, che avesse quindi le caratteristiche descritte nel capitolo precedente (recensioni scritte in inglese e almeno tre recensioni scritte negli anni 2019 e 2020); il numero di campioni è tuttavia molto significativo.

Guardando i risultati del modello, si può osservare come entrambe le variabili siano significative, seppur la variabile interagita tra il Superhost e l'indice di gentilezza abbia un valore di P-value uguale a 0 contro il 3,9% del semplice indice di gentilezza dell'host (un valore che comunque indica significatività della variabile in oggetto).

Come da aspettative, nonostante ci sia effettivamente un impatto della gentilezza dell'host sul rating complessivo della proprietà, si può osservare come l'impatto sia estremamente più significativo quando si parla della variabile interagita con il Superhost (informazione che ci viene fornita dal valore assoluto del coefficiente relativo alla variabile in oggetto). Si evince quindi che per massimizzare il rating, per un host è estremamente importante essere gentile nel senso più ampio del termine: quindi unire alla gentilezza nell'accezione più stretta del termine, quella più ampia come la disponibilità a rispondere sempre ai potenziali guest, a rispondere in fretta e rifiutare il minor numero possibile di richieste di prenotazione.

Tuttavia, è interessante osservare come anche se non si volesse eccellere nella logistica, il puro effetto della gentilezza è presente, seppur meno significativo: per un host conviene quindi sempre essere gentile nei confronti del proprio guest per avere un miglioramento del proprio rating.

Questo risultato è spiegato chiaramente dal fatto che l'indice gentilezza è presente sia singolarmente sia in forma interagita: nel primo caso possiamo vedere l'effetto della sola componente, nel secondo caso possiamo vedere l'effetto del primo sommato all'effetto (chiaramente percepito di grande valore aggiunto) dell'aver le qualità del SuperHost. Ciò spiega perché il coefficiente della variabile interagita sia quasi 5 volte superiore a quello della variabile associata all'indice kindness.

5.1.3 Modello finale: tutte le variabili

Fino ad ora sono state analizzate i singoli effetti delle singole variabili sul rating finale; poi sono stati studiati i loro effetti nei modelli intermedi per tipologia di variabili. Ora si andranno ad analizzare gli effetti di tutte le variabili sul rating complessivo, cercando di capire quali variabili siano più significative, quali meno e quali siano quelle maggiormente impattanti.

Si riportano di seguito i dati ottenuti dal modello appena descritto.

```
. regress RATING Tempo_Risposta Numero_proprietà Bagni Letti Numero_Servizi Indice_Gentilezza_Ho
> ceGentilezza Numero_parole_presentazione Recensione_per_mese D1_location_barcellona2
```

Source	SS	df	MS	Number of obs	=	1,613
Model	26300.5066	10	2630.05066	F(10, 1602)	=	51.05
Residual	82538.9727	1,602	51.5224548	Prob > F	=	0.0000
				R-squared	=	0.2416
				Adj R-squared	=	0.2369
Total	108839.479	1,612	67.5182874	Root MSE	=	7.1779

	RATING	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Tempo_Risposta		-.266575	.2551772	-1.04	0.296	-.7670914 .2339413
Numero_proprietà		-.0618459	.0067586	-9.15	0.000	-.0751024 -.0485893
Bagni		.1091494	.032722	3.34	0.001	.044967 .1733319
Letti		-.3146038	.0944375	-3.33	0.001	-.4998379 -.1293697
Numero_Servizi		.1173532	.0195526	6.00	0.000	.0790019 .1557046
Indice_Gentilezza_Host		.4670804	.3229255	1.45	0.148	1.100481 .1663206
Superhost_IndiceGentilezza		2.202495	.2271877	9.69	0.000	1.756878 2.648111
Numero_parole_presentazione		.0088401	.0026298	3.36	0.001	.003682 .0139983
Recensione_per_mese		.0065991	.0014485	4.56	0.000	.003758 .0094402
D1_location_barcellona2		.2082881	.7369095	0.28	0.777	-1.23712 1.653696
_cons		86.96921	1.46718	59.28	0.000	84.09141 89.847

Figura 21: modello Rating complessivo e tutte le variabili di controllo

Il modello si basa sulle osservazioni fatte sulle proprietà per le quali tutte le informazioni sono disponibili: si tratta di 1613 proprietà.

Andando ad analizzare i risultati in maniera più approfondita, è subito importante sottolineare che si tratta di un modello ben spiegato dalle variabili: questa informazione si può osservare dal valore di R-squared che si attesta a 24,16%. Questo ottimo risultato è stato ottenuto proprio per la presenza di tutte le variabili che nei paragrafi precedenti sono state analizzate singolarmente: incorpora quindi le caratteristiche fisiche della proprietà, le caratteristiche e le performance dell'host e gli indici di gentilezza creati.

Questo modello può presentare alcune differenze rispetto ai modelli visti in precedenza in quanto l'effetto di alcune variabili può essere assorbito o stimolato dalla presenza di altre variabili. È possibile vedere ad esempio come la variabile collegata alla presenza fisica dell'host all'interno dell'area di Barcellona non presenti particolare significatività rispetto i modelli visti in precedenza: il P-value al 77,7% indica proprio il fatto che si tratta della variabile meno significativa del modello. In accordo con il modello visto in precedenza, si ha conferma della scarsa significatività della variabile collegata al tempo di risposta che presenta un P-value del 29,6%.

È importante sottolineare che le altre variabili mostrate nel modello presentano coefficienti di segno e impatto simili a quelle dei sotto modelli visti nei paragrafi precedenti.

La variabile con il coefficiente più impattante in termini di valore assoluto è quella collegata alla variabile interagita tra indice gentilezza e la dummy Superhost. Essendo questa la variabile chiave del modello e considerando le altre variabili più significative, se l'obiettivo dell'host è quello di massimizzare il proprio rating, dovrà agire sulle seguenti leve:

- Essere gentile con il proprio ospite: essere accogliente, disponibile, fornire indicazioni se richieste sulla città e dedicare del tempo al rapporto con il proprio ospite se richiesto
- Avere ottime performance per ottenere il flag di superhost: questo si può avere cercando di tenere la proprietà sempre prenotata, non respingendo le richieste di soggiorno (tasso di cancellazione basso) e rispondendo tempestivamente ai messaggi

Le variabili legate al numero di parole utilizzate per la presentazione e il numero di recensioni medio per mese, per quanto siano molto significative (valore del P-value uguale o prossimo allo zero) presentano coefficienti poco rilevanti in termini di valore assoluto. Per questa ragione, l'host per avere un comportamento perfettamente volto alla massimizzazione del rating, dovrà anche attuare i seguenti comportamenti seppur meno rilevanti:

- Produrre una descrizione molto dettagliata della propria persona, dei propri hobby e le proprie passioni

- Richiedere al proprio guest una recensione del soggiorno una volta terminato il soggiorno, sottolineandone l'importanza

Una volta terminata questa analisi, si è voluto approfondire ulteriormente lo studio chiedendosi, dal momento che il rating complessivo è prodotto da altri rating intermedi, su quali di questi avesse avuto maggior effetto. In particolare, si è scelto di studiare quei rating intermedi che non dipendono da variabili esogene (come il rating sulla pulizia, che dipende appunto solo dalla qualità della pulizia dell'appartamento): si tratta del rating che esprime la qualità/prezzo e il rating di comunicazione. È estremamente importante sottolineare che, poiché l'intenzione è quella di fare un confronto tra modelli, è stato necessario mantenere le stesse variabili indipendenti per i due modelli che verranno illustrati tra poco.

Si inizia lo studio riportando di seguito i risultati ottenuti correlando il rating intermedio comunicazione con le stesse variabili indipendenti.

regress COMMUNICATION Tempo_Risposta Numero_proprietà Bagni Letti Numero_Servizi Indice_Gentilezza_Host_IndiceGentilezza Numero_parole_presentazione Recensione_per_mese D1_location_barcellona2						
Source	SS	df	MS	Number of obs	=	1,612
Model	172.668309	10	17.2668309	F(10, 1601)	=	32.94
Residual	839.204519	1,601	.524175215	Prob > F	=	0.0000
				R-squared	=	0.1706
				Adj R-squared	=	0.1655
Total	1011.87283	1,611	.628102315	Root MSE	=	.724

COMMUNICATION	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Tempo_Risposta	.0613388	.0257396	2.38	0.017	.010852	.1118256
Numero_proprietà	-.0060327	.0006817	-8.85	0.000	-.0073698	-.0046955
Bagni	.0049229	.0033012	1.49	0.136	-.0015523	.011398
Letti	-.0269602	.0095294	-2.83	0.005	-.0456516	-.0082688
Numero_Servizi	.0079229	.0019731	4.02	0.000	.0040527	.011793
Indice_Gentilezza_Host	.0195124	.0325752	0.60	0.549	.083407	.0443822
Superhost_IndiceGentilezza	.1439961	.0229214	6.28	0.000	.099037	.1889552
Numero_parole_presentazione	.0003245	.0002653	1.22	0.221	-.0001958	.0008449
Recensione_per_mese	.0006279	.0001461	4.30	0.000	.0003414	.0009145
D1_location_barcellona2	.0344826	.0743295	0.46	0.643	-.1113107	.1802759
_cons	9.081848	.1480157	61.36	0.000	8.791523	9.372173

Figura 22: modello Rating intermedio Comunicazione e tutte le variabili di controllo

È possibile notare dai risultati del modello, al contrario di quanto ci si potesse immaginare, che non vi è significativo impatto delle variabili sul rating comunicazione. Questo è sottolineato dal valore di R-squared del 16,55%, che

indica un basso livello di spiegazione del modello. Inoltre, le variabili che dovrebbero impattare maggiormente il rating comunicazione, risultano avere coefficienti molto poco significativi: è il caso della variabile collegata al numero di proprietà che nonostante abbia un P-value uguale a 0 (che indica la significatività della variabile all'interno del modello) presenta un coefficiente negativo estremamente basso. È il caso anche della location dell'host nell'area di Barcellona, che intuitivamente risulterebbe essenziale per la buona comunicazione: tuttavia non sembra essere una variabile significativa all'interno del modello dato il valore del P-value del 64,3% che indica assoluta mancanza di significatività della variabile.

Questo risultato ambiguo è facilmente spiegato andando a vedere empiricamente la colonna dei valori dei rating collegati alla comunicazione: si tratta di un set di valori dove raramente ci si discosta dalla votazione di 10/10. Essendo quasi una costante, risulta difficile trovare correlazioni significative con tante variabili: infatti che l'host sia gentile o meno difficilmente gli impatti saranno registrati all'interno di questo rating intermedio.

Per questa ragione si è deciso di proseguire le analisi andando a comparare l'impatto delle medesime variabili indipendenti con il rating che valuta il rapporto qualità / prezzo dell'esperienza.

Si riportano di seguito i risultati ottenuti dal modello che studia la correlazione tra le variabili di riferimento e il rating qualità prezzo.

```
. regress Rapporto_QualitàPrezzo Tempo_Risposta Numero_proprietà Bagni Letti Numero_Servizi Indice_Gent
> t Superhost_IndiceGentilezza Numero_parole_presentazione Recensione_per_mese Dl_location_barcellona2
```

Source	SS	df	MS	Number of obs	=	1,612
Model	249.632263	10	24.9632263	F(10, 1601)	=	37.65
Residual	1061.38201	1,601	.66294941	Prob > F	=	0.0000
				R-squared	=	0.1904
				Adj R-squared	=	0.1854
Total	1311.01427	1,611	.813789117	Root MSE	=	.81422

Rapporto_QualitàPrezzo	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Tempo_Risposta	-.0172996	.028947	-0.60	0.550	-.0740775 .0394784
Numero_proprietà	-.0056058	.0007667	-7.31	0.000	-.0071096 -.004102
Bagni	.0136944	.0037126	3.69	0.000	.0064125 .0209764
Letti	-.0350139	.0107169	-3.27	0.001	-.0560345 -.0139934
Numero_Servizi	.0098458	.002219	4.44	0.000	.0054934 .0141982
Indice_Gentilezza_Host	.0310833	.0366345	0.85	0.396	.1029399 .0407732
Superhost_IndiceGentilezza	.2245363	.0257776	8.71	0.000	.1739748 .2750978
Numero_parole_presentazione	.0000908	.0002984	0.30	0.761	-.0004944 .000676
Recensione_per_mese	.0008006	.0001643	4.87	0.000	.0004783 .0011229
Dl_location_barcellona2	.0066712	.0835917	0.08	0.936	.1706318 .1572894
_cons	8.591931	.16646	51.62	0.000	8.265429 8.918434

Figura 23: modello Rating intermedio qualità prezzo e tutte le variabili di controllo

Anche in questo caso, è possibile vedere come la significatività del modello è meno alta di quanto sperato: il valore di R-squared è del 19,04% che indica discreta spiegazione del modello. È interessante osservare come ci sono variabili che non sembrano significare particolarmente all'interno del modello che sono il tempo di risposta (valore del P-value al 55%), l'indice gentilezza (con valore del P-value al 39,6%), il numero di parole della presentazione (con valore del P-value al 76,1%) e la location dell'host (con valore di P-value al 93,6%). Inoltre, si può osservare che l'effetto di tutte le variabili con valore di P-value considerabile sia un ordine di grandezza inferiore rispetto il valore del coefficiente della variabile interagita, che è quindi la variabile chiave del modello.

Si può concludere che la variabile interagita di Superhost * indice gentilezza sia risultata la variabile chiave su cui agire per massimizzare i rating intermedi collegati alla comunicazione e alla percezione dell'host di valore (rapporto qualità / prezzo). Tuttavia, è interessante osservare come l'effetto della variabile interagita sul rating complessivo sia estremamente più intenso che sulle sui rating intermedi: ciò significa che l'effetto della gentilezza produce delle distorsioni di valutazione anche su altri rating intermedi che invece non dovrebbero essere in

alcun modo correlati con la gentilezza o la disponibilità dell'host (ad esempio il rating sulla qualità dei servizi o della pulizia).

5.2 Limitazioni e indicazioni per studi futuri

In questo paragrafo verranno analizzate le limitazioni dello studio fatto e offerti spunti per eventuali studi futuri.

Iniziando ad analizzare i limiti dell'analisi svolta in questo elaborato, la prima limitazione che salta all'occhio è proprio l'utilizzo del tool Excel, che si presta a molte tipologie di analisi ma non è specifico per l'analisi detta "sentiment analysis". In Excel infatti, è semplicemente possibile cercare delle determinate parole e assegnare un punteggio: non vi è nessun tipo di supporto che permetta di trovare un contesto alle parole utilizzate. Infatti, ovunque ci sia un problema di contesto, vengono evidenziati i limiti del tool; in particolare nel calcolo dell'indice gentilezza e nei seguenti casi specifici:

- Quando la parola ricercata sia associata a un soggetto diverso che non sia l'host (ad esempio la parola "nice" può essere associata all'animale domestico)
- Quando vi è un "not" davanti a un complimento, per sottolineare in maniera tiepida che l'host non si è comportato come ci aspettava (ad esempio nel caso di "not very thoughtful")
- Quando al posto del nome dell'host viene utilizzato un nome abbreviato oppure un nomignolo, che non viene rilevato (come ad esempio "Myke was really nice")
- Quando viene espresso un complimento verso l'host con una locuzione non presente tra quelle ricercate

Si può quindi dire che l'indice gentilezza sia una buona approssimazione, ma che il metodo utilizzato per costruirlo non sia privo di limiti.

Qualora in futuro si volessero ampliare o perfezionare gli studi, sarebbe auspicabile utilizzare un tool apposito per svolgere la "sentiment analysis", in modo da classificare il grado di emozioni di un commento (positivo, negativo o neutrale). Questi strumenti sono utilizzati dalle aziende per identificare il sentimento dei propri clienti analizzando le recensioni online e i feedback. In altri studi (Cheng et al, 2019) è stato utilizzato Leximancer, uno strumento che

permette di svolgere analisi evolute di testi: in questo modo è sicuramente possibile avere approssimazioni più precise nell'ambito di questo genere di analisi.

6. Conclusioni

Di seguito si andranno a riassumere e presentare le conclusioni che si possono trarre dalla lettura, dall'analisi e dall'interpretazione dei dati dei modelli costruiti. In particolare, si andrà a comparare l'effetto della gentilezza manifestata dal padrone di casa verso il proprio ospite, con l'effetto prodotto dalle altre variabili del servizio offerto, infine si valuterà la strategia ottimale da tenere per un host per massimizzare il suo ritorno in termini di qualità della sua reputazione.

Si riportano di seguito le tre variabili più impattanti nello studio dei singoli effetti di ciascuna di esse sul rating:

- la variabile interagita tra l'indice di gentilezza e la dummy Superhost: il coefficiente è più alto sia nei modelli a variabili singole che nei modelli complessivi.
- la variabile che indica la presenza dell'host all'interno dell'area di Barcellona: il coefficiente è pari a circa 1,65
- la variabile collegata alla semplice gentilezza: il coefficiente è pari a circa 1,3.

È possibile vedere come due tra le tre variabili più influenti siano proprio quelle collegate alla gentilezza: è interessante osservare che sia quindi estremamente più significativo porsi in maniera cordiale con il proprio ospite, per avere un ritorno maggiore in termini di reputazione rispetto all'aumento del numero di servizi (che prevederebbe anche un investimento economico) o all'aumento del numero di bagni presenti nella casa. Inoltre, è importante sottolineare come nonostante l'effetto della semplice gentilezza produca effettivamente un ritorno importante in termini di aumento di reputazione, l'effetto maggiore si ha cercando di ottenere il flag di superhost.

Riguardo i modelli intermedi, è interessante andare a comparare i livelli di R-squared, per riuscire a capire quali aggregati di variabili spieghino meglio il rating complessivo: se le variabili relative alla proprietà, quelle relative all'host oppure quelle legate alla gentilezza. I risultati sono estremamente interessanti: il modello contenente le variabili di gentilezza ha un valore del 12,7%, ossia un ordine di grandezza in più rispetto agli altri due modelli. Questo risultato indica quindi che le variabili collegate alla cortesia spiegano il rating estremamente meglio delle

variabili collegate alla proprietà fisica e rispetto alle performance dell'host: controllare l'indice di gentilezza fornisce quindi maggior potere all'host di poter influenzare il rating complessivo del proprio listing.

Per quanto riguarda il modello finale, sono diverse le considerazioni doverose da fare. In primo luogo, si può verificare come la variabile che spiega il modello maggiormente e che ha l'impatto maggiore sul rating complessivo maggiore (valori sottolineati dal valore del coefficiente e del P-value) si conferma essere la variabile interagita tra l'indice gentilezza e la dummy Superhost: si conferma il fatto che la gentilezza produca un effetto maggiore quando vi sono altre caratteristiche di disponibilità e buone performance dell'host certificate appunto dal flag di superhost.

Oltre a questo, è interessante andare a paragonare il valore dell'R-squared tra il rating complessivo e i rating intermedi di comunicazione e qualità prezzo. Al contrario di quanto ci si potrebbe aspettare, il valore di R-squared è più alto nel modello con variabile dipendente il rating complessivo rispetto ai modelli con i rating intermedi: le variabili legate alla gentilezza hanno quindi significativamente più impatto sul rating complessivo che non sui rating intermedi.

Questo porta alla chiara conclusione che l'effetto della gentilezza porti alcuni bias di valutazione anche e soprattutto su quei rating intermedi che dovrebbero essere indipendenti ed andare a misurare caratteristiche specifiche della casa: sarà quindi probabile che il guest tenderà a dare voti più alti anche su alcuni parametri come il livello pulizia o la funzionalità dei servizi solo perché il proprio host è stato gentile e disponibile.

A questo punto appare evidente l'importanza, per il padrone di casa, di curare la relazione con il proprio ospite, facilitando uno scambio comunicativo più ampio e favorendo una vera e propria esperienza sociale: essere a disposizione per raccontargli la città, farlo sentire a casa, instaurare un rapporto di amicizia e ed essere quindi in generale gentile con lui.

Oltre a questo, il secondo parametro più importante sarà quello di ottenere il flag di superhost: per farlo sarà quindi necessario limitare il più possibile le disdette e rispondere a tutti nel più breve tempo possibile. Questi due aspetti consentono di

migliorare sensibilmente la reputazione del padrone di casa, tuttavia anche curare la propria descrizione personale sulla piattaforma di Airbnb, richiedere al proprio ospite, al termine dell'esperienza, di lasciare una recensione per incrementare il numero di recensioni e rispondere tempestivamente ai messaggi aiuta ad aumentare la valutazione generale.

In un mondo dove la competizione economica e interpersonale assume spesso carattere aggressivo, e la comunicazione, anche attraverso i cosiddetti social media, scopre sempre di più aspetti legati all'odio se non addirittura alla violenza, c'è, evidentemente, un bisogno di ricercare benessere attraverso rapporti basati sul rispetto, sulla condivisione di valori, di visioni del mondo. Quello che in questa analisi viene rappresentato dal termine "gentilezza", che appare particolarmente trainante in merito al gradimento degli utenti Airbnb, al punto da essere un aspetto distorsivo delle valutazioni, sembra quindi rappresentare una modalità di interazione che rende l'incontro tra cliente e fornitore qualcosa di più rilevante, più ampio, più profondo ed evidentemente più soddisfacente.

7. Bibliografia

Alsudais (2017): Alsudais A., "Quantifying the Offline Interactions between Hosts and Guests of Airbnb", *Twenty-third Americas Conference on Information Systems, Boston, 2017*

Arcidiacono et al (2016): Arcidiacono D., Mainieri M., Pais I., "Quando la sharing economy fa innovazione sociale. Il caso BlaBlaCar", Collaboriamo.org, 2016

Bernardi (2015): Bernardi, M., "Un'introduzione alla Sharing Economy", Fondazione GianGiacomo Feltrinelli, 2015

Bridges & Vásquez (2016): Bridges J., Vásquez C., "If Nearly all Airbnb Reviews are Positive, Does that Make them Meaningless?", *Current Issues in Tourism*, 2018, Volume 21, [2057-2075]

Dellarocas et al. (2008): Dellarocas, Chrysanthos, Charles A Wood, "The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias." *Management Science*, 2008, Volume 54 (3), [460–476].

Fradkin et al (2018): Fradkin A., Grewal E., Holtz D., "The Determinants of Online Review Informativeness: Evidence from Field Experiments on Airbnb", Working Paper, 2018

Fradkin et al (2019): Fradkin A., Grewal E., Holtz D., "Reciprocity in Two-sided Reputation Systems: Evidence from an Experiment on Airbnb", Working Paper, 2019

Guttentag D., Smith S., Potwarka L., Havitz M., (2017): "Why Tourists Choose Airbnb: A Motivation-Based Segmentation Study", *Journal of Travel Research*, 2017, Volume 57 (3), [342-359]

Hearn (2010): Hearn A., "Structuring feeling: Web 2.0, online ranking and rating, and the digital 'reputation' economy", *Ephemera*, 2010, Volume 10 (4/4), [412-438]

Hu et al. (2009): Hu N., Zhang J., Pavlou P.A., "Overcoming the J-shaped distribution of product reviews", *Communications of the ACM*, Ottobre 2009, Volume 52 (10), [144-147]

Lee et al (2015): Lee D., Hyun W., Ryu J., Lee W.J., Rhee W., Suh B., "An Analysis of Social Features Associated with Room Sales of Airbnb", *CSCW'15 Companion: Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*, 2015, [219-222]

Liang et al (2018): Liang L.J., Choi H.C., Joppe M., "Exploring the relationship between satisfaction, trust and switching intention, repurchase intention in the

context of Airbnb”, International Journal of Hospitality Management, 2018, Volume 69, [41-48]

Lind et al (2017): Lind M., Visentini M., Mäntylä, Del Missier F., “*Choice-Supportive Misremembering: A New Taxonomy and Review*”, Front Psychol, Dicembre 2017, Volume 8

Luise V., Chiappini L., (2017): “*Sharing Economy e nuove forme di precarietà: problematiche, resistenze e possibili soluzioni*”, Futuri, 2017, Volume 8, [85-97]

Mingming Cheng & Xin Jin (2019): “*What do Airbnb users care about? An analysis of online review comments*”, International Journal of Hospitality Management 76, [58-70]

Nosko & Tadelis (2014): Nosko C., Tadelis S., “*The limits of reputation in platform markets: An empirical analysis and field experiment.*”, NBER Paper No. 20830, Gennaio 2015

Sthapit & Jiménez-Barreto (2018): Sthapit E., Jiménez-Barreto J., “*Sharing in the host–guest relationship: perspectives on the Airbnb hospitality experience*”, Anatolia, 2018, Volume 29, [282-284]

Zervas et al (2015): Zervas G., Proserpio D., Byers J., “*A First Look at Online Reputation on Airbnb, Where Every Stay is Above Average*”, Working Paper, 2015

Zhang et al (2018): Zhang T.C., Jahromi M.F., Kizildag M., “*Value co-creation in a sharing economy: The end of price wars?*”, International Journal of Hospitality Management, Aprile 2018, Volume 71, [51-58]