

# POLITECNICO DI TORINO

Corso di laurea magistrale in Ingegneria Gestionale

Tesi di Laurea Magistrale

## L'importanza delle parole nel turismo digitale: analisi semantica sui titoli degli annunci e relative prestazioni economiche



**Politecnico  
di Torino**

**Relatore**

prof.ssa Elisabetta Raguseo

Co-relatore

ing. Francesco Milone

**Candidato**

Carlo CASTRONOVO

matricola: 277558

ANNO ACCADEMICO 2021-2022



*Un giorno, un uomo non vedente stava seduto sui gradini di un edificio con un cappello ai suoi piedi ed un cartello recante la scritta: "Sono cieco, aiutatemi per favore".*

*Un pubblicitario che passeggiava lì vicino si fermò e notò che aveva solo pochi centesimi nel suo cappello.*

*Si chinò e versò altre monete. Poi, senza chiedere il permesso dell'uomo, prese il cartello, lo girò e scrisse un'altra frase.*

*Quello stesso pomeriggio il pubblicitario tornò dal non vedente e notò che il suo cappello era pieno di monete e banconote.*

*Il non vedente riconobbe il passo dell'uomo: chiese se fosse stato lui ad aver riscritto il suo cartello e cosa avesse scritto. Il pubblicitario rispose "Niente che non fosse vero. Ho solo riscritto il tuo in maniera diversa", sorrise e andò via.*

*Il non vedente non seppe mai che ora sul suo cartello c'è scritto: "Oggi è primavera... ed io non la posso vedere."*

- dal web



# Abstract

La sharing economy e gli scambi commerciali peer-to-peer sono al centro della articolata scena del turismo digitale, nella quale si rende protagonista la famosa piattaforma turistica di AirBnb. Il seguente lavoro di tesi vuole esplorare in dettaglio, attraverso strumenti analitici quali l'analisi semantica e le regressioni statistiche, il processo di costruzione della fiducia tra pari negli scambi su piattaforma digitale, con un dettaglio sistematico sulle componenti semantiche degli annunci che comportano un marcato e tangibile effetto sulle performance economiche delle singole accommodations. In tale contesto di scambio tra pari l'abilità di produrre un sentimento di fiducia nel guest da parte dell'host gioca un ruolo cruciale, al fine di generare il massimo valore raggiungibile per entrambe le parti. Diviene, quindi, fondamentale sviluppare la capacità di trovare le "parole giuste" per attrarre il cliente e alla fine dei conti "vendere" il proprio appartamento con un vantaggio competitivo nel mercato della piattaforma online. Tali parole possono essere messe in evidenza dall'analisi semantica e studiate attraverso l'econometria per validare o falsificare l'ipotesi secondo la quale il titolo di un annuncio possa significativamente influire sulle prestazioni dello scambio su piattaforma, in base all'utilizzo di parole chiave specifiche che determinano tali performance, sia buone che cattive, in quanto esistono parole positive che hanno un effetto catalizzante e parole che presentano un'accezione negativa e che compromettono o alterano le prestazioni stesse. Nel dettaglio si sono svolte analisi di regressione con approccio top-down su tutti i dataset e ulteriori analisi con approccio bottom-up, ove necessario, al fine di isolare l'effetto della singola parola sui ricavi. I principali ritrovamenti necessitano una lettura consapevole alla luce di specifiche premesse fondamentali. Il presente lavoro si focalizza sui risultati relativi al livello semantico delle analisi svolte. In tal senso, è necessario sottolineare che le analisi osservano degli aspetti di dimensione estremamente ridotta rispetto all'intera cornice del fenomeno della sharing economy e degli scambi su piattaforma digitale di AirBnb. La trattazione riguarda fenomeni sociali complessi, difficili da modellare ma è possibile rappresentarne delle parti ed ottenere dei risultati di dettaglio significativi. L'ottica è proprio questa e i ritrovamenti che si riportano a seguire, possono essere trascurabili se si guarda agli aspetti che maggiormente influenzano le prestazioni

economiche degli annunci, ma rappresentano delle evidenze rilevanti se calate nell'ambito semantico, guardando ad aspetti secondari, ma importanti, degli scambi economici su piattaforma. Nel dettaglio, si può affermare con sicurezza che le prestazioni economiche delle abitazioni posizionate sulla piattaforma di AirBnb, in ambito hospitality, sono maggiormente influenzate da attributi esterni al campo semantico. Tuttavia, non è possibile negare che l'utilizzo di una parola piuttosto che un'altra nell'annuncio dell'abitazione abbia un effetto sulle prestazioni. L'aspetto semantico non è il primo su cui focalizzare i propri sforzi ma certamente da un contributo alla buona riuscita degli scambi su piattaforma.

# Indice

Abstract	5
Indice	7
<b>PARTE 1 - FASE TEORICA</b>	<b>9</b>
Introduzione	10
Contesto e obiettivi	10
AirBnb	12
Sharing Economy	13
Effetti di AirBnb sul business hospitality	14
Revisione della letteratura	16
<b>PARTE 2 - FASE EMPIRICA</b>	<b>34</b>
Domande di ricerca	35
Metodologia	35
Analisi econometrica	36
Analisi semantica	37
Collezione dei dati	38
Strumenti utilizzati	39
Prove sui dati	39
Pulizia del dataset	39
Analisi	42
Tipologia dei dati	42
Analisi descrittive	44
Analisi semantiche	47
Analisi semantica Torino	47
Analisi semantica Gallipoli	52
Analisi semantica Erice	56
Prime considerazioni semantiche	59
Analisi semantiche ulteriori	60
Analisi econometriche	62
Analisi econometriche Torino	67
Analisi econometriche Gallipoli	74
Analisi econometriche Erice	80
Analisi econometriche – confronto dei dataset	86
Analisi econometriche ulteriori	89

<b>PARTE 3 - FASE FINALE</b>	90
Conclusioni	91
Ritrovamenti	91
Opportunità di ricerca futura e punti aperti	94
Bibliografia, sitografia e citazioni	95

## **PARTE 1 - FASE TEORICA**

## Introduzione

AirBnb. Poche lettere e tantissimo significato. Tali parole non sono scelte a caso. Già il nome dell'azienda mostra quanto sia importante la scelta delle parole nel trasmettere sensazioni, stati d'animo, oltre che un messaggio chiaro. "Air", connesso alla natura digitale e dematerializzata di una piattaforma online che lega in remoto persone lontane, e allo stesso tempo al principale mezzo di viaggio, l'aria, che insieme ai velivoli che in essa transitano rappresenta l'emblema degli spostamenti turistici. Il tutto si lega al volo, concetto che da sempre stimola una sensazione di leggerezza e leggiadria, di semplicità con la quale ci si può spostare e catapultarsi da un luogo all'altro. Semplicità che rappresenta un altro caposaldo della piattaforma, che vuole proporsi come un semplificatore degli scambi tra pari, delle attività di ricerca di una sistemazione in viaggio. Tale semplicità è mostrata nella seconda parte del nome "Bnb", nella quale praticità e semplicità si legano in quella "n" che vuole richiamare fonicamente la tipica "E" commerciale ("&"), dando più leggerezza al nome e allo stesso tempo una certa modernità e voglia di innovare e proporsi in modo diverso dal tradizionale. Per non parlare del concetto di Bed & Breakfast, che rappresenta l'essenza della semplicità e praticità in viaggio. Insomma, sei sole lettere consentono la trasmissione di tutti i valori fondanti dell'azienda. Figuriamoci cosa si può generare con un intero titolo di un annuncio da creare.

## Contesto e obiettivi

Oggigiorno l'economia circolare e gli scambi commerciali peer-to-peer sono al centro della articolata scena del turismo digitale, nella quale si rende protagonista la ormai più che famosa piattaforma turistica di AirBnb. Molto è stato ricercato ed asserito relativamente alla stessa, ma a livello di analisi scientifica restano delle carenze in relazione alla formazione fattiva di un sentimento legato all'esperienza e ai beni offerti. Il seguente lavoro di tesi vuole esplorare in dettaglio, attraverso strumenti analitici quali l'analisi semantica e le regressioni statistiche, il processo di costruzione della fiducia tra pari negli scambi su piattaforma digitale, con un dettaglio sistematico sulle

componenti degli annunci che comportano un marcato e tangibile effetto sulle performance economiche delle singole accommodations.

In tale contesto di scambio tra pari, che di fatto non si conoscono l'un l'altro, l'abilità di produrre un sentimento di fiducia nel guest da parte dell'host gioca un ruolo cruciale, al fine di generare il massimo valore raggiungibile per entrambe le parti. Diviene, quindi, fondamentale sviluppare la capacità di trovare le "parole giuste" per attrarre il cliente e alla fine dei conti "vendere" il proprio appartamento con un vantaggio competitivo nel mercato della piattaforma online. Tali parole possono essere messe in evidenza dall'analisi semantica, al fine di comprendere se la visibilità dello specifico annuncio possa essere modificata in base alla presenza, o mancanza, di parole chiave che abbiano una verificata connessione con le migliori prestazioni, per lo meno a livello empirico di analisi dei dati.

La tesi è articolata come segue. Inizialmente, si riporta una dettagliata revisione della letteratura esistente trattante l'analisi semantica nel settore turistico e lo studio delle performance degli scambi economici in ambito hospitality effettuati mediante l'utilizzo della piattaforma digitale Airbnb. Nella seconda parte del lavoro, dopo aver evidenziato le domande di ricerca che spingono l'intero progetto di tesi, si passa alla fase empirica di analisi semantica e analisi delle prestazioni, utilizzando nello specifico big data provenienti dal database di Airbnb, relativi a una grande città, un borgo di mare e un borgo di montagna, al fine di differenziare il più possibile la tipologia di accommodations e di contesti nei quali le stesse si posizionano. A seguire si riportano i ritrovamenti empirici e le analisi connesse, con uno sguardo attento alle metriche econometriche di rilievo, costruite specificatamente in seguito alle analisi. In conclusione, vi è una sintesi dei risultati ottenuti con il lavoro svolto, mostrando spunti per prossimi lavori di ricerca e gli elementi di maggiore rilievo a supporto dell'attuale teoria nel campo di applicazione dell'hospitality.

## AirBnb

L'idea fondante di Airbnb costituisce un concetto che entra a gamba tesa sull'assetto tradizionale del settore turistico-ricettivo. Infatti, fino a circa una sola quindicina di anni fa, si distinguevano nettamente la figura di produttore/fornitori di beni/servizi e quella di fruitore/consumatore e i produttori/fornitori erano quasi esclusivamente operatori economici ed esperti del settore strutturati logisticamente ed economicamente, in genere locali fatta eccezione delle grosse catene dei circuiti internazionali, per fornire servizi ricettivi alla massa turistica a livello territoriale.

Il principio agente della piattaforma non ha semplicemente cambiato le modalità operative di un settore ma ne ha stravolto radicalmente la natura e la quantità rivolgendosi direttamente ai privati, potenziali produttori/fornitori e contemporaneamente fruitori/consumatori. Chiunque disponga di uno spazio, magari inutilizzato o non utilizzato continuativamente, che può ritenersi idoneo (sotto il profilo dell'abitabilità e dei servizi) ad ospitare persone, può attraverso la piattaforma renderlo disponibile in affitto a chi è in cerca di un alloggio o di una camera per brevi periodi. Cosicché nel giro di pochi anni (dal 2007 al 2012) il sistema Airbnb ha potuto contare alloggi in oltre 26.000 città per 10 milioni di notti prenotate in tutto il mondo.

Di fatto Airbnb ha un'offerta potenziale atta a contenere l'intera domanda mondiale quando a livello locale la disponibilità di camere negli hotel si rivelava spesso esaurita in tutto il mondo civile organizzato.

Nonostante alcuni aspetti controversi (episodi di razzismo, aumento dei prezzi delle case, evasione fiscale, informazioni false, esperienze negative degli ospiti) che comunque riguardano un numero di casi assai risibile rispetto ai 260 milioni di check-in nel quinquennio iniziale e ai quali la compagnia cerca di porre rimedio con fattivo interesse alla qualità e all'affidabilità del prodotto/servizio, è innegabile la potenzialità economica dell'innovativo approccio che non si pone come una mera soluzione alternativa alle risorse esistenti ma come una modalità fluida e moderna che consente una maggiore agilità negli spostamenti territoriali di qualunque natura (turistica, d'affari, culturale, eccetera eccetera) in linea e come diretta conseguenza dello sviluppo del web e delle app che l'odierna tecnologia offre. Da allora, la community di

Airbnb è cresciuta e oggi conta 4 milioni di host, che a loro volta hanno ospitato più di 1 miliardo di persone in quasi tutti i paesi del mondo. Ogni giorno, gli host offrono attività e alloggi unici, che consentono agli ospiti di scoprire il mondo in un modo più autentico e genuino, con l'obiettivo di consentire ai viaggiatori di tutto il mondo di sentirsi a casa, ovunque si trovino.

[fonti delle informazioni Wikipedia e AirBnb.com]

## Sharing Economy

La sharing economy, o economia della condivisione, è un concetto che si è andato affermando negli ultimi decenni, declinato in vari modi e applicato a diversi settori economici e sociali. Specialmente all'inizio il termine stesso è stato fonte di dibattito a livello internazionale, proprio perché il fenomeno è recente e l'area concettuale al quale fa riferimento è vasta e variegata. Così si sono sviluppate una serie di definizioni contigue, analoghe o parallele: da peer-to-peer economy a economia collaborativa, da gig economy a economia on-demand fino a consumo collaborativo. Termini a volte usati in modo intercambiabile, ma che, secondo gli esperti, indicano attività lievemente (o, a volte, sostanzialmente) diverse.

Tra gli esempi più noti di sharing economy si possono richiamare, oltre a Airbnb, eBay il sito di vendita e aste online fondato il 3 settembre 1995 in California; Uber servizio di taxi (ride sharing) che utilizza le auto private guidate dal loro proprietario, mediante un'app che consente agli utenti di richiedere l'auto e di pagare la corsa, nata nel 2009 a San Francisco e oggi è presente in tutto il mondo; Blablacar esempio perfetto di Sharing Economy, è una società francese che consente a chi viaggia in macchina da una località a un'altra di offrire passaggi a pagamento per riempire i posti non utilizzati, come alternativa a treni, bus o aerei, l'autista inserisce sul sito orario e luogo di partenza e di destinazione e i posti disponibili nell'auto, che possono essere riservati online. L'idea di sharing economy gode di consensi molto ampi, e non solo da parte dei suoi sempre più numerosi fruitori. Piace agli ambientalisti, perché la condivisione riduce il consumo delle risorse non rinnovabili e l'inquinamento. Piace anche a chi non

ama l'idea di proprietà, o considera le piattaforme digitali come uno strumento in grado di favorire l'inclusione sociale.

Ci si trova dunque di fronte a nuove forme di business e il quadro è molto complesso che certamente richiede nuove leggi atte a regolamentare questo "nuovo mondo" la cui affermazione è ormai un fatto ineluttabile che si inquadra anche nell'inarrestabile processo di cambiamento avviato con lo sviluppo della tecnologia: senza Internet e smartphone Airbnb sarebbe ancora una normale agenzia immobiliare e gli Uber e le altre ride sharing company non sarebbero nemmeno nate.

[fonti delle informazioni Wikipedia e web]

## Effetti di AirBnb sul business hospitality

Lo sviluppo della tecnologia digitale e di internet in generale ha fortemente stravolto il settore del turismo e dell'ospitalità (Litvin et al., 2008). Infatti, la piattaforma di AirBnb in questione ha consentito l'affermarsi della sharing economy nel campo hospitality, e di fatto tale tipologia di economia viene identificata comunemente proprio con i casi dei settori della hospitality e dei trasporti, nei quali effettivamente ha avuto il maggiore impatto e attraverso i quali si sta imponendo sulle dinamiche dell'economia moderna.

La sharing economy ha infatti portato davanti agli utenti un innumerevole quantità di alloggi turistici, impattando in maniera straordinaria il mercato e tutti gli elementi relativi. Il contesto collaborativo si presta molto per le applicazioni nel settore turistico, permeato da continui scambi, economici, culturali e sociali. In una cornice di condivisione le piattaforme online diventano un intermediario fondamentale, che catalizza le dinamiche condivisive grazie al bacino di utenze che permette di raggiungere e alle forti garanzie che riesce a fornire. In questo modo le persone possono offrire le proprie abitazioni libere in affitto (Sundararajan, 2014) a basso costo, sia dal punto di vista di ricerca che pratico (Henten & Windekilde, 2016).

Alla base del modello chiaramente si staglia un fattore che assume grande rilevanza nel panorama odierno di efficienza e sostenibilità: sfruttare della capacità inutilizzata,

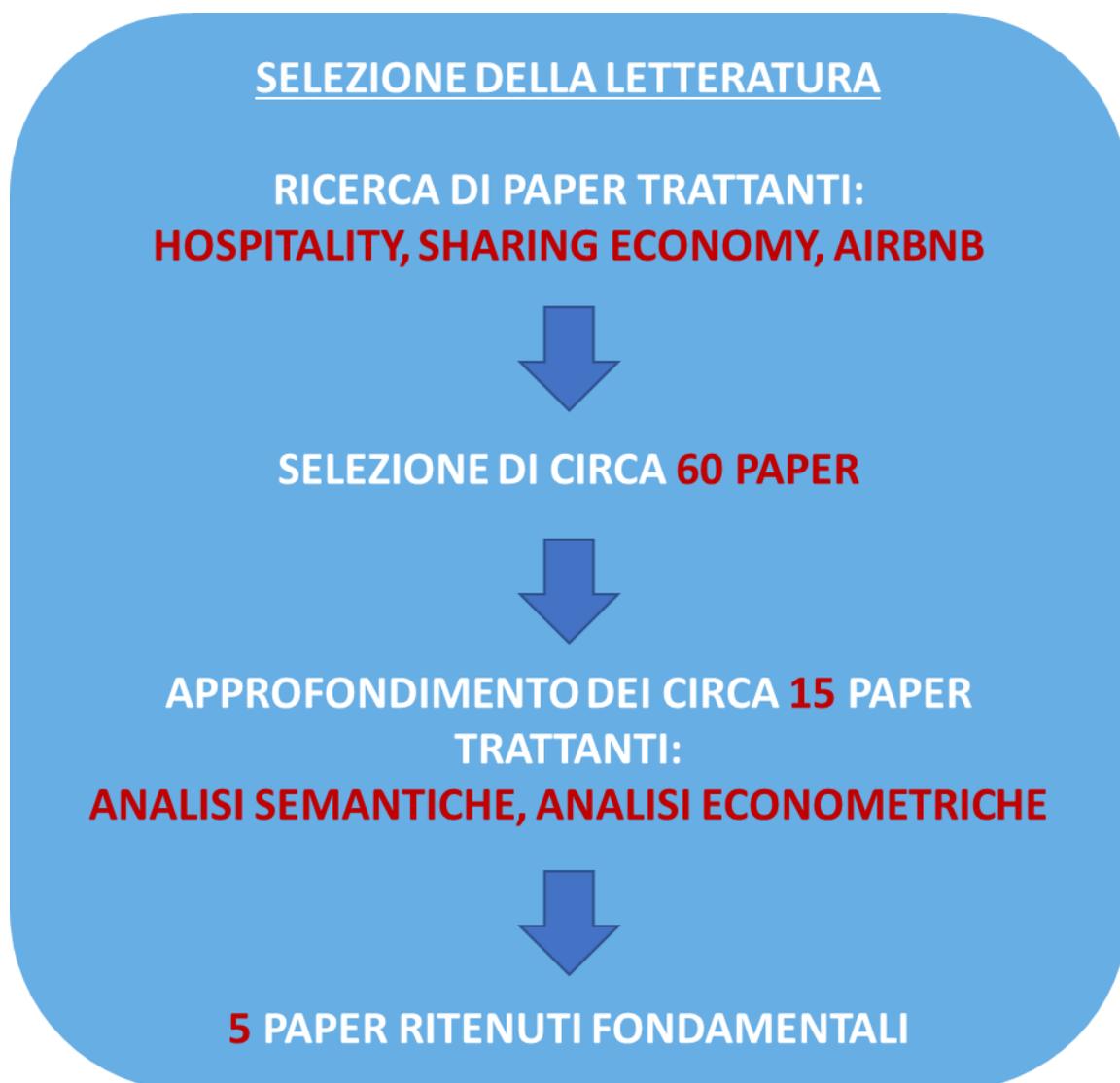
che rappresenterebbe un mero costo, e dare concreto valore a terzi interessati. E in questo fenomeno, la componente sociale diviene significativa (Lessig, 2008) a differenza dell'economia tradizionale nella quale il prezzo predomina gli altri attributi (Cohen & Kietzmann, 2014; Zervas et al., 2017).

La piattaforma assume un ruolo cruciale nel gestire la fiducia in degli scambi tra pari in cui effettivamente si vogliono delle garanzie da ambo le parti, che consistono principalmente nella tutela dei beni fisici e del business in genere per l'host e nella sicurezza delle informazioni per i guest. Inoltre, altro attore fondamentale che si affida alla piattaforma è rappresentato dalle istituzioni, che vedono nella stessa un garante delle corrette procedure legali relative ai soggiorni e al commercio nel settore hospitality, considerato che si tratti di uno scambio tra pari privati, con soggetti offerenti non strutturati come vere e proprie aziende del settore, meno informati e meno osservabili. Tutto ciò permette di strutturare i servizi, assicurando un'adeguata qualità e delle procedure di controllo capaci di garantire i requisiti minimi di un servizio nel settore hospitality.

Vi sono quindi dei marcati temi socioculturali e legali in caso di sharing economy, proprio per via del fatto che uno scambio tra pari non è regolarizzato allo stesso modo degli scambi commerciali tradizionali. La piattaforma, come evidenziano Zervas, Proserpio & Byers (2017), definisce esattamente le regole di partecipazione nel settore, e non rappresenta una mera vetrina: questo è solo uno degli svariati apporti che fornisce la piattaforma. Per esempio, nel 2015 gli stessi studiosi (Zervas, Proserpio & Byers, 2015) avevano già fatto notare come la piattaforma stabilisse concretamente gli attributi caratterizzanti i soggiorni e di conseguenza le valutazioni. E in un settore come quello dell'hospitality nel quale è possibile differenziare per qualità in maniera molto ampia, la piattaforma diventa una grande risorsa per posizionarsi al meglio nel mercato.

## Revisione della letteratura

Il procedimento in questione si è sviluppato seguendo un processo di selezione chiaro e definito che viene sintetizzato nella mappa sottostante:



In primo luogo, appunto, si è attuata una ricerca ad ampio spettro degli articoli trattanti tematiche relative alla sharing economy e alla piattaforma di AirBnb, chiaramente calate nel settore hospitality. Attraverso tale procedimento si sono selezionati circa 60 articoli, la maggior parte dei quali compare in bibliografia al fondo del presente lavoro. Tra questi si sono approfonditi in dettaglio una dozzina abbondante di paper trattanti nello specifico le analisi semantiche e le analisi econometriche. Da questo passaggio ulteriore si sono estratti gli articoli ritenuti fondamentali per le presenti ricerche.

Quindi, revisionando i più rilevanti articoli di ricerca, trattanti i temi in esame e pubblicati tra i recenti anni 2019 e 2021, si evidenziano lavori di analisi in paesi differenti, aventi al centro della metodologia l'analisi semantica, che mettono in luce le parole più ricorrenti e il loro valore. Tra di essi si trovano anche studi che analizzano le performance attraverso modelli econometrici. Si riportano in formato tabellare un riepilogo dei principali articoli e dei Journal in cui sono stati pubblicati.

#### Principali articoli di ricerca analizzati in dettaglio

<b>Autori</b>	<b>Articolo - Journal</b>	<b>Principali Apporti</b>
Cheng, Jin, 2019	What do Airbnb users care about? An analysis of online review comments - International Journal of Hospitality Management 76 (2019) 58-70	<ul style="list-style-type: none"> <li>- coinvolgimento emozionale nella relazione guest-host</li> <li>- bias della positività nei guest</li> <li>- predisposizione dei guest alla comunicazione testuale</li> <li>- sentiment analysis</li> <li>- trattamento dei big data</li> </ul>
Zhu, Cheng et al., 2019	The construction of home feeling by Airbnb guests in the sharing economy: A semantics perspective - Annals of Tourism Research 75 (2019) 308-321	<ul style="list-style-type: none"> <li>- principali dimensioni nel turismo legate all'home feeling</li> <li>- sentiment analysis</li> <li>- trattamento dei big data</li> </ul>
K. Ding, W. C. Choo et al., 2021	Exploring Sources of Satisfaction and Dissatisfaction in Airbnb Accommodation Using Unsupervised and Supervised Topic Modeling - Front Psychol	<ul style="list-style-type: none"> <li>- attributi chiave che determinano soddisfazione e insoddisfazione del guest</li> <li>- "conflitti nei guest" nella valutazione di stessi servizi</li> <li>- sentiment analysis</li> </ul>
E. J. Lee, Y. Cho, 2021	Investigating Citizen Perceptions and Business Performance of Airbnb in Korea - Journal of Asian Finance, Economics and Business Vol 8 No 8 (2021) 0167-0180	<ul style="list-style-type: none"> <li>- contesto della sharing economy</li> <li>- analisi econometriche sulle prestazioni delle abitazioni</li> <li>- attributi chiave che influenzano le prestazioni</li> <li>- percezioni e preferenze dei guest</li> </ul>
Dogru, Majid et al., 2021	Communicating quality while evoking loss – How consumers assess extra charges in the Airbnb marketplace - Tourism Management	<ul style="list-style-type: none"> <li>- impatto del prezzo sulla domanda</li> <li>- analisi econometriche in ambito hospitality</li> </ul>

Il primo degli articoli più rilevanti è quello di Cheng e Jin (2019), i cui ritrovamenti, relativi a un grande dataset di recensioni e commenti su accommodations a Sydney, mostrano che chi utilizza la piattaforma digitale in esame è portato a valutare la propria esperienza sulla base dei propri ricordi degli hotel in cui si è soggiornato. Il lavoro di ricerca pone l'attenzione su tre attributi chiave in ambito di hospitality, quali "location", "amenities" e "host", dimostrando che il prezzo non sia significativo in termini di effetti sulla percezione delle persone. Inoltre, le sensazioni sgradevoli non sembrano essere sistematiche e gli utenti mostrano una predisposizione favorevole a porsi positivamente nelle recensioni.

Attraverso le fonti (che si riportano a seguire) citate nell'articolo, avente un'accurata revisione della letteratura che consente di avere una visione d'insieme dell'ambito di ricerca, viene messo in luce anche la natura sociale che caratterizza lo scambio su piattaforma, come era già stato fatto notare da affermati lavori di ricerca (Festila and Müller, 2017; Lampinen and Cheshire, 2016; Tussyadiah and Pesonen, 2016; Yannopoulou, 2013). In aggiunta, viene posta in evidenza la creazione concreta di una comunità caratterizzata da vera e propria fiducia reciproca nelle esperienze di viaggio mediate dalla piattaforma (AirBnb, 2017).

Tra i precedenti lavori citati ed altri (Guttentag, 2016) non mancano però i dubbi relativi a questa sfera sociale e comunitaria, che per molti utenti si riduce a una poco significativa parte del valore creato dalla piattaforma digitale in questione, che rappresenta una fattiva alternativa alle usuali attività commerciali del settore turistico a minor costo. In questa ottica gli studi tecnici effettuati da Tussyadiah (2016) valutano statisticamente significativo per i clienti della piattaforma il risparmio in termini di costi, insieme alle amenità offerte dal contesto circostante le abitazioni e la fruibilità delle amenità stesse. Si evidenzia quindi un contrasto nei ritrovamenti dei vari studiosi precedenti, senza una ragione dimostratasi valida al fine di giustificare l'una o l'altra linea di pensiero.

Alcuni (Tussyadiah and Zach, 2016) pensano sia dovuto alla varietà di accommodations in giro per il mondo, senza che ci sia una tipologia fissa generalmente affermata; altri seguono lo stesso ragionamento ma in termini di varietà delle personalità dei vari utenti (Festila and Müller, 2017). Ad ogni modo, sulla base della revisione della letteratura l'articolo mostra come in precedenza le principali dimensioni influenti nello scambio su piattaforma siano la "location" (Tussyadiah and Zach, 2016), le "amenities" (Guttentag, 2015), la "pulizia" (Bridges and Vàsquez, 2016), la "interazione tra host e guest" (Festila and Müller, 2017) e ultima, non per importanza, la "convenienza economica" (Guttentag and Smith, 2017).

L'articolo, avente al centro l'analisi semantica, rappresenta anche un ottimo riferimento in termini di approccio ai big data. Nello stesso, infatti, è riportata un'analisi di cosa siano i big data analytics e viene messo in luce il fatto che nell'ambito in questione ci sono pochi lavori di ricerca che utilizzano questa metodologia (Fuchs et al., 2014; Xiang et al., 2015). In particolare, come sottolineato da Edwards et al. (2017) quest'ultima necessita di "un'iniziale comprensione della struttura dei dati e dei contenuti, un accurato processo di pulizia dei dati, un approccio multidisciplinare dietro al semplice testo e l'analisi statistica, oltre che una rappresentazione visuale chiara dei dati". Queste linee guide saranno alla base della parte di analisi empirica che caratterizza il presente lavoro di tesi.

Per quanto riguarda i risultati di text mining (che è una tecnica che mira alla elaborazione del linguaggio naturale per trasformare il semplice testo in dati strutturati e utilizzabili a livello statistico in quanto normalizzati) delle analisi di Cheng e Jin (2019), non sorprendentemente in ordine di importanza gli argomenti di maggiore peso si sono rivelati la "location", le "amenità" e "l'host", il quale rappresenta il vero soggetto mediatore con il quale il guest si interfaccia fisicamente a livello personale. Inoltre, un altro tema rilevante riguarda l'evidenza secondo la quale gli utenti della piattaforma

tendono a valutare le proprie esperienze sulla base di criteri solitamente utilizzati per i tradizionali alberghi, facendo spesso dei paragoni o dei riferimenti (come, ad esempio, il livello di qualità in termini di stelle su scala da 1 a 5).

Relativamente invece ai risultati della sentiment analysis, gli studiosi hanno trovato una marcata positività delle persone che hanno utilizzato la piattaforma per soggiornare a Sydney, andando quindi a supportare fortemente la precedente ipotesi giunta in fase di revisione della letteratura. Tutto ciò, però, è stato anche positivamente correlato (per mezzo delle loro analisi) all'ambiente cittadino di Sydney, estremamente servito e carismatico: la predisposizione degli utenti ad esprimere commenti positivi sembrerebbe fortemente influenzato al contesto generale del luogo che si va a visitare piuttosto che alle accommodations in sé.

Tirando le somme, ai fini del presente trattato di ricerca, l'evidenza più importante di questo articolo di Cheng e Jin (2019) è il maggiore coinvolgimento emozionale, rispetto ai tradizionali hotel, nella relazione interpersonale tra host e turista. Il maggiore tocco personale è evidente nei piccoli dettagli, come mostra il lavoro di Belarmino et al. (2017) nel quale si confronta il modo in cui i clienti riportino i nomi propri degli host di un b&b al contrario di ciò che accade nelle revisioni degli alberghi, nelle quali il personale è raramente nominato in prima persona e comunque senza un rapporto davvero diretto e informale. Inoltre, questo interfacciarsi direttamente con una singola persona, alla quale commenti e recensioni sono riferiti, piuttosto che a una entità impersonale, una personalità giuridica come il brand di un albergo, rende marcato il bias della positività nelle recensioni e nelle interazioni tra guest e host in generale. Sentirsi in una relazione faccia a faccia con una persona, rende chiaramente più sgradevole far pesare o mostrare qualcosa di negativo, soprattutto se minoritario rispetto al resto.

Tutto ciò rende necessario un lavoro di approfondimento relativamente a come può essere fortificato e reso intenso questo rapporto, a partire dalla primissima interazione che guest e host affrontano: le parole che l'host decide di immettere nella descrizione del proprio annuncio, al fine di rendere attrattivo l'appartamento e di trasmettere la "fiducia" che questo lavoro di tesi mette al centro della ricerca. Per di più sembrerebbe proprio che rispetto ad alberghi e strutture turistiche tradizionali nelle quali le linee guida sono standardizzate e lasciate al confronto orale, gli utenti della piattaforma siano predisposti a una comunicazione testuale, attraverso la piattaforma stessa e non solo, avvalorando la necessità dell'approfondimento di ricerca appena citato. Quest'ultimo va a contribuire, nei limiti del contesto di cui tratta, alla necessità, mostrata da Cheng ed Edwards (2017), relativa all'introduzione dell'analisi dei big data in ambito turistico e hospitality.

Ulteriore ricerca di fondamentale importanza è rappresentata dall'articolo di Zhu, Cheng et al. (2019), nel quale viene analizzata in dettaglio, con una prospettiva semantica, la costruzione della sensazione di sentirsi a casa durante gli scambi su piattaforma e le esperienze create attraverso la sharing economy, utilizzando come campione tre grandi metropoli urbane americane quali New York, Los Angeles e Chicago, che rappresentano le più grandi città aventi un diffuso utilizzo della piattaforma.

L'articolo presenta una copiosa e utile revisione della letteratura, che ricalca i temi trattati nell'articolo precedentemente citato. In particolare, anche questo lavoro di ricerca riporta la capacità della piattaforma di creare "una comunità costruita sulla condivisione" (AirBnb, 2018) e di permettere lo stabilirsi di un legame più genuino tra ospiti e gestori. Scintilla innescatrice dell'analisi semantica effettuata durante le ricerche è rappresentata nuovamente dall'importanza dei fattori sentimentali-emozionali che caratterizzano lo scambio in questione, che ha portato i ricercatori ad interessarsi a fondo del tema del "home feeling", sensazione di sentirsi a casa in un

accommodation in tutt'altro luogo e lontano dalla propria zona di comfort abituale, che può influenzare fortemente i comportamenti e le scelte dei clienti.

Alcuni studiosi asseriscono persino che il manifestarsi di questa sensazione appena descritta non possa avvenire nel settore del turismo tradizionale (Sigala, 2015), mentre molti altri studiosi non hanno posto la dovuta attenzione sul tema negli anni precedenti all'ultimo triennio, nonostante avessero nominato il fenomeno nei loro lavori. Di conseguenza tutt'oggi resta poco chiaro come effettivamente avvenga la costruzione di tale sensazione e diventa necessaria un'espansione delle ricerche in nuove aree geografiche e da varie prospettive, considerata l'importanza del fenomeno in questione e la poca analisi empirica effettuata a riguardo, alla luce del breve periodo intercorso tra il sorgere della piattaforma e l'affermarsi dello strumento nella vita di tutti i turisti del globo.

Ad ogni modo, l'articolo di Zhu, Cheng et al. (2019) può considerarsi una base concettuale di riferimento relativamente alla creazione di questa sensazione di "home feeling" e dell'influenza della stessa nei comportamenti futuri degli ospiti in ambito di scelte turistiche ed eventuali prenotazioni ripetute per la stessa accommodation.

Ai fini del presente lavoro di tesi diviene quindi d'interesse lo studio effettuato dagli studiosi sulle caratteristiche semantiche della parola casa, e sulle sensazioni che crea come descritto in precedenza: si palesa marcatamente una connessione tra la sensazione di sentirsi a casa, a proprio agio, e la fiducia che si vuole generare in un ospite al momento dello scambio su piattaforma, tema fondante della tesi stessa. I concetti che si possono trovare dietro singole parole sono molteplici e vanno ben oltre il mero significato della parola in questione. Diviene evidente proprio con la parola casa, che ha visto evolvere il proprio significato di luogo fisico in qualcosa di più profondo come le relazioni economiche, sociali e sessuali (Bowlby, Gregory, Mckie, 1997).

E non mancano anche in questo articolo fondamentale ([Zhu, Cheng et al., 2019](#)) le evidenze trovate da [Cheng e Jin \(2019\)](#) secondo le quali in fase di valutazione dei soggiorni gli ospiti che utilizzano la piattaforma di Airbnb tendono a comparare le proprie esperienze agli hotel in cui hanno soggiornato in passato durante viaggi precedenti. Inoltre, si mettono in evidenza anche in questo caso i ricorrenti attributi principali quali “location” ([Tussyadiah and Zach, 2016](#)), “amenities” ([Guttentag, 2015](#)) e “pulizia” ([Bridges and Vàsquez, 2016](#)).

Anche relativamente alla prospettiva semantica si ottengono ottimi spunti. Tale prospettiva viene descritta con le puntuali parole di [Saeed \(2015\)](#) che afferma che l’analisi semantica è lo studio del significato comunicato attraverso il linguaggio o le espressioni usate in un dato contesto. In tal senso il presente lavoro di tesi vuole analizzare ciò che comunicano le parole nel contesto della piattaforma online e delle conversazioni testuali tramite le descrizioni. L’articolo di [Zhu, Cheng et al., 2019](#) pone al centro della propria metodologia la prospettiva semantica e la utilizza sia nella selezione dei dati che nel processo di analisi dei dati stessi.

In questo caso il dataset è stato prelevato sempre da AirDNA, attraverso i database di Airbnb, attingendo dai commenti degli utenti, come era già stato fatto da studiosi che hanno effettuato ricerche precedenti nel campo dell’hospitality ([Kozinets, 2010](#); [Abbie-Gayle & Barbara, 2017](#); [Tussyadiah & Zach, 2017](#); [Brochado et al., 2017](#)).

Ai fini dell’analisi contenute nell’articolo, come di consueto, il periodo preso in esame è stato l’anno ed effettuando uno scrupoloso processo di pulizia dei dati si sono utilizzati, per le analisi di dettaglio, solamente i dati più significativi, processati con Python e ponendo alla base della ricerca il concetto di “casa”. Infatti, il primo step di analisi ha trattato una selezione delle parole più rilevanti a livello di insieme, che avessero un collegamento con il concetto di casa. In quest’ottica è stata creata una

rete di parole al fine di mettere in luce tale legame. Le varie parole sconnesse e la punteggiatura sono state rimosse attraverso Python. Al fine di analizzare la frequenza delle parole e delle associazioni tra le stesse sono state costruite nuvole di parole con programmi specifici, oltre che rappresentazioni delle parole in formato di network. Considerata la scarsa presenza di analisi simili nel dato contesto di applicazione, i valori stabiliti per valutare una parola come frequente sono stati dettati dall'intuito, dall'esperienza e dallo studio empirico che i ricercatori hanno effettuato: essi sono giunti a fissare un valore di 50 per le coppie di parole e di 1000 per la frequenza delle parole su un campione di 42085 recensioni.

Ma i ritrovamenti più importanti di questo articolo di [Zhu, Cheng et al., 2019](#) riguardano la concreta presentazione delle principali dimensioni che ruotano intorno al concetto di casa e che quindi impattano in modo determinante il senso di fiducia e l'esperienza in toto degli ospiti. Tutto ciò rappresenta una base imprescindibile per un lavoro di ricerca nell'ambito dell'hospitality e per tale motivo si riporta una dettagliata trattazione dei ritrovamenti in esame. Gli studiosi ([Zhu, Cheng et al., 2019](#)) effettuano una classificazione in "dimensioni generiche" basate sulla letteratura esistente e "nuove dimensioni" identificate attraverso la loro ricerca. Tra le prime si distingue ancora tra "dimensioni fisico-spaziali" e "dimensioni sociali".

Nella prima categoria si trovano elementi quali "location" (che non sorprendentemente ritorna continuamente nell'ambito in questione e rappresenta un concetto fondamentale); "space", in termini di spaziosità vera e propria ma anche di intimità e privacy, oltre che di veridicità di quanto presentato, relativamente allo spazio in questione, nella descrizione e sulla piattaforma in generale; "safe", parola connessa alla sicurezza percepita dagli ospiti durante il soggiorno per loro stessi e i propri averi. Nella seconda si trovano invece "Host", "Vicinato" e "Cane", che rispettivamente mostrano quanto a livello sociale, di comunità in cui ci si sente davvero bene e a casa, siano importante il rapporto diretto con il gestore del bene in affitto, che rappresenta

anche la figura di riferimento per le più svariate necessità; un vicinato autentico, amichevole e ricco in termini di conoscenza del luogo e delle culture locali; gli animali domestici che possono giocare un ruolo non banale per moltissimi ospiti.

Per quanto concerne, invece, le nuove dimensioni, gli autori suddividono in “dimensioni di ospitalità”, “dimensioni affettive” (distinte in due tipi diversi, 1 e 2, rispettivamente di affezione e di intrattenimento) e “dimensioni comparative con gli hotel” (che tornano nuovamente, dopo esser state messe in grande rilievo dal primo articolo fondamentale affrontato nella presente revisione della letteratura). Le prime racchiudono la collaboratività dell’host in termini sia di disponibilità e flessibilità ma anche di proattività nell’offrire servizi inaspettati e graditi; la “comunicazione” puntuale, gradevole e coerente dell’host; l’apprezzamento degli ospiti per l’ospitalità, i modi, la disponibilità e gli eventuali servizi inaspettati dell’host. Le seconde del primo tipo contengono “l’amore” degli ospiti per tutte le componenti dell’esperienza di soggiorno, “l’amabilità” di tali componenti oltre che dell’host stesso e “l’accoglienza” delle componenti della accommodation, in termini di spazi fisici, mobilio ma anche di atmosfera.

Le altre del secondo tipo riguardano il “divertimento” in sé e la godibilità delle componenti fisiche, e non, dell’accomodation e del luogo. Nella terza ed ultima rientrano gli “hotel” e tutte le caratteristiche che possono essere prese a confronto tra l’industria alberghiera tradizionale e i b&b, quali, a titolo di esempio, l’economicità, la pulizia, l’ordine e le amenità. Attraverso queste dimensioni si può comprendere a fondo la totalità degli elementi che impattano le scelte e le sensazioni dei clienti, e di conseguenza vanno tenute ben a mente nel corso di una trattazione di ricerca nel campo dell’hospitality. In particolare, è necessario attenzionare le parole ricorrenti nelle descrizioni che siano legate a tali dimensioni, rendendo la ricerca più indirizzata e avente un obiettivo preciso.

Altro articolo fondamentale da osservare per le analisi del presente lavoro di tesi è quello denominato “Investigating Citizen Perceptions and Business Performance of Airbnb in Korea”, scritto da E. J. Lee e Y. Cho e pubblicato nel 2021 sul Journal of Asian Finance, Economics and Business, che rappresenta una delle fonti più attendibili nel panorama asiatico. Tale lavoro è posizionato nel contesto della sharing economy e cerca di comprendere quali siano le pratiche migliori da attuare al fine di diffondere nella società una cultura adeguata all’economia della condivisione. A tal proposito gli autori guardano alle recensioni e agli aspetti negativi che si possono trovare nelle procedure di condivisione degli appartamenti attraverso la piattaforma, attenzionando le percezioni che gli utenti ricevono durante il servizio. Di conseguenza vengono messi in luce i fattori che si riflettono in modo marcato sulle decisioni dei consumatori e sulle prestazioni del business in sé. Le analisi condotte, sul territorio coreano, hanno usufruito delle regole econometriche applicate alle determinanti delle prestazioni delle abitazioni e i risultati più significativi riguardano proprio le percezioni e preferenze dei guest prima citate.

La metodologia utilizzata nell’opera di Lee e Cho (2021) risulta di interesse, in quanto viene utilizzata la “content analysis” che si utilizza per interpretare materiale testuale. Di conseguenza l’articolo mostra un’altra tecnica utile alla comprensione dei dati qualitativi testuali per giungere a dati più quantitativi e scientificamente validi. In tale ottica anche l’analisi semantica trova spazio nell’opera, per comprendere l’accezione positiva, neutra o negativa del testo in analisi. Anche in questo caso i database di Airbnb rappresentano la fonte di informazioni cruciali per svolgere le analisi.

Anche tale articolo si focalizza sulle parole, cercando di formare dei raggruppamenti in base al significato delle parole stesse e ai loro legami reciproci. In tal senso gli autori hanno classificato su più livelli, definendo 6 temi e 73 concetti. Come evidenziato in tutti i lavori della letteratura precedenti, la rappresentazione grafica consente una maggiore chiarezza di visualizzazione, aumentando la capacità di comprendere

l'intero contesto in esame. In questo caso si è utilizzato un network di parole, con netta distinzione dei temi e le aree di intersezione degli stessi. Tale lavoro ha consentito di definire le macrocategorie principali che si affrontano nell'ambito hospitality: host, experience, sanitation, location, property e appreciation. Si nota che ritornano temi ricorsi anche nella precedente letteratura analizzata, tra cui i principali proprio "location" e "host".

Le evidenze principali del loro metodo riguardano la natura delle percezioni di host e guest: per il primo prevale la componente emotiva e comportamentale, per il secondo, a tale sfera, si affiancano le caratteristiche fisiche e materiali. Nello specifico gli autori focalizzano un'importante parte delle proprie analisi sulle recensioni negative, per comprendere la correlazione tra il sentimento negativo e le caratteristiche delle accommodation. A tal fine sviluppano varie ipotesi testate con la sentiment analysis e l'econometria. Per fornire un'idea delle scale di grandezza utilizzate, gli studiosi hanno utilizzato un campione di 629 recensioni negative. Esso è stato utile a validare il fatto che ci sia una correlazione significativa tra il tipo di abitazione e il tipo di reclamo, tra l'affidabilità dell'host e il tipo di reclamo e tra la zona dell'abitazione e il tipo di reclamo. Allo stesso tempo non è significativa la relazione tra i reclami e il tipo di proprietà (intera casa, porzione, ecc). Tutto ciò è servito a mostrare quali fattori si riflettano in sensazioni negative.

Similmente sono state analizzate le prestazioni economiche, utilizzando un campione di 17 mila osservazioni circa. Anche in questo caso sono state sviluppate delle ipotesi, che nello specifico riguardano il legame tra la performance economica e i vari attributi dell'abitazione o dell'esperienza ritrovabili nel database, che spaziano dalla tipologia di abitazione, agli stati dell'host (per esempio se in grado di fornire la prenotazione immediata e se superhost o meno) che delineano la capacità comunicativa dell'host o la sua reputazione.

Le principali evidenze delle analisi riguardano il fatto che i clienti preferiscono marcatamente la possibilità di ottenere l'intera abitazione per il soggiorno. Inoltre, si mette in luce che il numero di recensioni e il contenuto delle stesse impatti le scelte dei consumatori, per i quali la reputazione dell'host è importante e di conseguenza i parametri indicati poco prima relativi allo stato dell'host. In tal senso anche l'overall rating diventa un parametro di interesse che influenza fattivamente le decisioni degli utenti.

Dalle analisi econometriche si nota una significatività di tali fattori, ma bisogna evidenziare che il settore è particolarmente caratterizzato da numerosi aspetti non rappresentabili attraverso i dati del database della piattaforma. Infatti, le analisi di regressione hanno portato a degli R-squared ritenuti soddisfacenti già ai valori di 0.29 e 0.35, con casi d'eccezione che sono riusciti a raggiungere una rappresentazione pari a 0.41 e un massimo di 0.487, sempre inferiore al 50%. Nonostante ciò, analisi di questo genere restano decisamente utili per dettare delle tematiche manageriali specifiche, che se affrontate nel modo corretto possono determinare un concreto miglioramento delle prestazioni economiche degli appartamenti. Bisogna inoltre puntualizzare che anche in questo caso è presente una marcata maggioranza di recensioni positive e di conseguenza i campioni relativi ai sentimenti negativi non sono copiosi. Tuttavia, le analisi non sono state del tutto intaccate e si ritengono assolutamente valide, in quanto questi sono fattori che caratterizzano il settore.

Altro articolo di fondamentale importanza è stato scritto da Dogru, Majid et al., e pubblicato nel 2021 su "Tourism Management". L'articolo "Communicating quality while evoking loss – How consumers assess extra charges in the Airbnb marketplace" affronta, nel contesto della sharing economy, l'impatto del prezzo delle abitazioni sulla domanda e sfrutta analisi econometriche applicate al campo hospitality.

A differenza di quanto accadeva in precedenza con i tradizionali operatori di settore, di fatto gli utenti si trovano dinanzi a una sconfinata offerta di abitazioni differenti e questo causa chiare ripercussioni sul mercato, rendendo interessante studiare i prezzi e il loro partizionamento. Infatti, gli autori dimostrano che lavorare in tal senso sul prezzo può avere un provato impatto favorevole sulle prestazioni economiche dell'abitazione. Le analisi questa volta sono state svolte su territorio americano, utilizzando sempre gli utili database di AirBnb. Un apporto di significativa importanza è rappresentata dal periodo temporale in cui sono stati prelevati i campioni: a differenza di numerose ricerche gli esperti si sono spinti nell'osservazione di un periodo temporale ben più lungo di uno o due anni, prendendo in considerazione un dataset di annunci creati tra il 2008 e il 2018 e relativo agli interi Stati Uniti d'America: è chiaro che l'estensione delle ricerche è tra le più vaste del settore, con una numerosità del campione di accommodations superiore alle 840 mila occorrenze, selezionate e ripulite attraverso parametri di attività dell'annuncio ben definiti.

Tale articolo rappresenta un ottimo esempio di analisi econometriche svolte nell'ambito di interesse e consente di avere un'idea precisa degli ordini di grandezza in gioco. Nel dettaglio, ai fini delle analisi sono state sviluppate delle ipotesi relative ai vari aspetti che possano influenzare i comportamenti del cliente in base a diverse politiche di prezzo, o meglio in base alla tipologia di partizionamento dello stesso. Chiaramente il primo aspetto ha riguardato la creazione di una variabile dipendente utile a indicare l'andamento della grandezza di interesse sulla base di variabili indipendenti (legate agli attributi dell'annuncio) sia binarie, sia quantitative espresse in dollari.

Dopo avere introdotto tali variabili gli studiosi hanno creato 4 modelli di regressione, aventi formulazione lineare, ognuno con finalità di analisi specifiche, così da mettere in luce i fattori di interesse per le loro ipotesi prima formulate. Sulla base dei coefficienti e dei valori statistici legati ad ogni variabile, è stato possibile validare o meno le ipotesi,

esprimendo dei chiarimenti sugli andamenti riscontrati. Come evidenziato, diventa fondamentale avere un esempio dei valori caratterizzanti il settore, in quanto per vari indicatori non esistono dei valori fissati a priori dalla teoria, e pertanto le assunzioni e le evidenze degli esperti precedenti diventano un fondamentale punto di riferimento.

Nelle analisi in questione numerose variabili indipendenti sono risultate correlate in maniera significativa (a un livello del 10%, del 5% e dell'1%) alla variabile dipendente, ma bisogna porre l'attenzione sul fatto che le variabili utilizzate nel modello rappresentano una piccolissima parte della totalità degli attributi influenzanti la variabile dipendente, avendo ogni modello un R-squared sempre inferiore al 0.15 in ogni caso, pur utilizzando numerosi addendi nella regressione. Si rende evidente, quindi, che nel settore in esame è complesso creare dei modelli che siano fortemente rappresentativi, in quanto moltissimi fattori, che determinano i valori delle grandezze di cui si osserva l'andamento, non sono modellabili con dati precisi e quantitativi. Tuttavia, resta interessante vedere il tipo di correlazione, risultata significativa a livello statistico, che le variabili indipendenti mostrano, considerato che spesso questo è bastevole per smentire delle ipotesi o comunque descrivere un andamento dei fenomeni osservati.

L'ultimo articolo fondamentale al quale si è fatto riferimento è nominato "Exploring sources of satisfaction and dissatisfaction in Airbnb accommodation using unsupervised and supervised topic modeling" pubblicato nel 2021 su "frontiers in Psychology" e scritto dai rinomati autori Kai Ding, Wei Chong Choo, Keng Yap Ng, Siew Imm Ng e Pu Song (2021).

In tale articolo, calato sempre nell'ambito della sharing economy e dell'utilizzo della piattaforma digitale, i ricercatori mettono in chiaro gli attributi chiave che determinano soddisfazione e insoddisfazione del guest, determinando, per la prima volta in modo dettagliato, i "conflitti" ai quali differenti tipologie di guest sono soggetti nella

valutazione di stessi servizi: date la moltitudine di persone che utilizzano il servizio, la varietà di sfaccettature che compongono il singolo individuo e la componente emotiva che caratterizza questo tipo di scambi, spesso si verifica che utenti diversi abbiano reazioni totalmente in contrasto relativamente allo stesso identico servizio.

A tal fine la sentiment analysis ha rappresentato un grande alleato tecnico per rendere metodologicamente rigorose le analisi svolte. Quest'ultime hanno riguardato le recensioni dei clienti che hanno utilizzato i servizi della piattaforma e degli host, utilizzando un cospicuo numero di occorrenze che hanno raggiunto quasi le 60 mila unità. Anche in questo caso le analisi sono state effettuate distinguendo accuratamente tra recensioni positive e negative, costruendo due insiemi ben distinti.

I differenti utenti presentano concretamente svariati comportamenti nei confronti di stessi attributi del servizio, e questi fenomeni sono verificati anche in diverse tipologie di proprietà pubblicate su piattaforma. Gli autori, con un'ottica manageriale, hanno determinato i principali fattori che influenzano la previsione di un eventuale ritorno dei clienti all'utilizzo dei servizi, sulla base della loro soddisfazione o insoddisfazione, fortemente legata alla disponibilità dell'host e alla esperienza di soggiorno accostabile al normale concetto di casa.

Estrema importanza è riposta anche nello spunto relativo a come utilizzare al meglio i dati provenienti dai social al fine di sviluppare nuove strategie o tattiche, per generare maggiore soddisfazione nei clienti, e quindi valore, nello stimolante contesto di condivisione sotto studio.

In aggiunta, l'articolo presenta una ricca fonte di spunti per quanto riguarda la letteratura antistante, mostrando numerosi apporti precedenti di studiosi in tutto il mondo. In questo modo è stata osservata la moltitudine di tipologie di dati analizzati e di strumenti per svolgere le analisi in questione. Tra i tanti vengono messi in risalto

due filoni principali, che rispettivamente sfruttano metodologie tradizionali di interviste e questionari e metodologie più moderne basate sui dati provenienti dalla piattaforma digitale di AirBnb, come le recensioni degli utenti: è chiaro che il presente lavoro di tesi punta a posizionarsi in maniera innovativa all'interno del secondo filone di ricerca.

Per quanto concerne l'analisi del sentimento dei clienti vengono approfonditi i fondamenti teorici alla base di questa tipologia di ricerca, facendo riferimento anche in questo caso ai precedenti lavori di ricerca, e specificando i vari strumenti. Alcuni come Kumar e Zymbler (2019) si sono dedicati all'utilizzo di una sentiment analysis svolta attraverso sistemi di machine learning, altri, come i precedentemente citati in dettaglio lavori di Zhu et altri (2020), hanno preferito strumenti dedicati alla classificazione delle parole in base alla loro accezione positiva e negativa: tale orientamento è quello che si ritiene più opportuno per le analisi che si svolgono nel presente lavoro di tesi.

Al centro della metodologia di Kai Ding, Wei Chong Choo et altri (2021) viene posto anche il topic modeling, che rappresenta proprio un approccio di analisi basato sul machine learning, al fine di analizzare appunto documenti testuali e modellare i fenomeni connessi, come definito da Blei nel 2012. Nel dettaglio, anche gli autori in questione utilizzano un approccio basato su una classificazione delle parole tra sentimento positivo e negativo, poiché tale metodo consente di affrontare le limitazioni secondo le quali in estese documentazioni testuali o ridotte ma in cospicuo numero, numerose parole appaiono in entrambi i contesti, positivo e negativo, sulla base del contesto stesso o di loro specifiche accezioni, rischiando di rendere poco affidabili le fondamenta delle analisi: la divisione preventiva in due insiemi consente una maggiore sicurezza relativamente alla concreta attribuzione del sentiment alla parola. Resta comunque complesso ottenere una chiara e netta distinzione, a causa proprio della variabile natura delle parole e dei contesti in cui le stesse si posizionano. Tale

evidenza viene attenzionata durante lo svolgimento delle analisi che seguono, dando consapevolezza ai successivi ritrovamenti e dei limiti delle metodologie utilizzate.

Infatti, i maggiori spunti ottenuti attraverso l'intero procedimento di revisione della letteratura riguardano principalmente gli ordini di grandezza che caratterizzano il settore sotto esame, le metodologie di ricerca nello stesso, nonché le mancanze in tali lavori e le linee generali di contesto.

Si giunge così a rendere chiaro il posizionamento che si vuole dare, al presente lavoro di tesi: si nota, che nel contesto generale, e soprattutto in quello italiano, mancano analisi semantiche e analisi di regressione focalizzate sulle singole parole con confronti tra diversi anni e diversi luoghi. Inoltre, si evidenzia che in letteratura si è soliti studiare le recensioni, tralasciano i titoli, sui quali invece, il presente lavoro basa le proprie analisi.

## **PARTE 2 - FASE EMPIRICA**

## Domande di ricerca

Di conseguenza il presente lavoro di tesi si pone nel dibattito finora riportato in una posizione diversa, cercando di rispondere a specifiche domande, al fine di falsificare, o meno, la teoria sviluppata e le ipotesi ad essa relative. In particolare, si vuole validare il pensiero secondo il quale il titolo di un annuncio possa significativamente influire sulle prestazioni dello scambio su piattaforma, in base all'utilizzo di parole chiave specifiche che determinano tali performance, sia buone che cattive, in quanto esistono parole positive che hanno un effetto catalizzante e parole che hanno un'accezione negativa e che compromettono o alterano le prestazioni stesse. Si tratta, quindi, di validare anche se il ruolo dell'host, che è comparso tra i fattori fondamentali del settore durante il precedente lavoro di revisione della letteratura, è così centrale e tale da indirizzare le prestazioni della specifica accommodation attraverso le proprie doti personali di comunicazione, che mette in mostra per la primissima volta proprio con la scrittura del titolo dell'annuncio.

Quindi si può riassumere tali concetti nei due seguenti obiettivi:

- Comprendere se l'utilizzo di determinate parole in risposta a specifiche condizioni esterne comporti un effetto concreto sulle prestazioni economiche
- Studiare la risposta manageriale degli host alle condizioni esterne

## Metodologia

A tale scopo si effettua sistematicamente un'analisi semantica di grandi dataset di titoli degli annunci al fine di costruire delle metriche specifiche per il caso di applicazione. Fatto ciò, ci si prefigge di testare una relazione econometrica del tipo:

$$Performance = c + b \times Semantic + a \times X + e$$

Dove:

- "*Performance*" rappresenta la variabile dipendente da analizzare, la grandezza al centro del presente studio, della quale si vuole comprendere l'andamento alla luce delle variazioni di variabili indipendenti considerate di interesse e influenti in modo significativo;

- “c” intercetta della popolazione ignota;
- “a” e “b” stime dei parametri del modello, rispettivamente intercetta e pendenza;
- “*Semantic*”, variabile indipendente relativa al fattore semantico;
- “X” eventuale ulteriore variabile indipendente del modello;
- “e” termine di errore o residuo.

## Analisi econometrica

Di fatto si cerca di stimare con tale relazione degli effetti causali tra le variabili in gioco, utilizzando dati osservazionali e quindi non sperimentali. Infatti, in questo caso non si può attuare un esperimento controllato e casualizzato e di conseguenza si analizzano i dati che si hanno a consuntivo. Pertanto, la creazione di un modello econometrico diviene essenziale in tale scenario per giungere ai fini prefissati. Questa attività richiede dei passaggi specifici ben definiti, seguendo un rigore scientifico che porti ad un’analisi strutturata ed affidabile.

In prima istanza si definiscono le ipotesi da verificare e le grandezze in gioco. Il passaggio successivo è costituito dalla specificazione del modello, che prevede la scelta delle variabili indipendenti da inserire nel modello. Inoltre, in questa fase è necessario stabilire la forma funzionale che mette in relazione le variabili e fissare le assunzioni relative al termine di errore, che sarà una variabile casuale della quale bisogna ipotizzare i parametri caratteristici. Allo stesso modo, si rende necessario determinare delle assunzioni a riguardo della natura delle variabili indipendenti e della relazione delle stesse con il termine rappresentante il residuo.

A questo punto bisogna procurarsi i dati, poiché non sono noti i parametri caratterizzanti il modello. Essi devono essere stimati attraverso i dati campionari, che nella fattispecie in particolare sono stati utilizzati del tipo cross-sezionale: si è scelto di prendere dati di un periodo di riferimento limitato, lungo 2 anni (2019-2020), e relativi a più accommodations presenti in determinate città, senza concentrarsi sullo storico di una specifica abitazione. Questa tipologia di dati consente più facilmente di evitare campionamenti non identicamente e indipendentemente distribuiti, e di rispettare di

conseguenza una delle assunzioni fondamentali dei minimi quadrati. Infatti, si sceglie il meccanismo degli stimatori OLS che rappresenta un passe-partout per le analisi econometriche, essendo di semplice utilizzo e chiaramente leggibile. Il passo successivo prevede proprio la stima del modello econometrico, per il quale tale metodo di stima appena descritto è stato valutato il più indicato. Esso serve proprio a giungere alle stime dei parametri del modello che non sono noti, utilizzando come detto i dati osservazionali del campione scelto.

Chiaramente la costruzione del modello richiede una fase di controllo della corretta specificazione, che consta della verifica delle assunzioni, definite nei passaggi precedenti, sulla base dei dati campionari, i quali in riferimento alla loro natura specifica possono richiedere aggiustamenti al fine di rendere il modello effettivamente rappresentativo. Una volta effettuati tali controlli è possibile utilizzare il modello, verificando concretamente le ipotesi alla base del modello, l'effetto causale delle variabili indipendenti sulla variabile dipendente e l'attendibilità della previsione dell'andamento di tale variabile al variare dei regressori.

## Analisi semantica

Ai fini dell'analisi semantica si ricercano e analizzano le parole più frequenti in ogni dataset, per anno e per città, costruendo delle nuvole di parole che rendono di facile individuazione le parole in questione. Svolta un'analisi statistica descrittiva sulle parole, si studiano nel dettaglio le frequenze e le frequenze percentuali delle singole parole in ogni dataset, giungendo alla selezione di quelle ritenute adeguate al lavoro di analisi. Vengono tralasciate nelle analisi le parole direttamente connesse a luoghi specifici o legate ad aspetti strettamente non semantici. Inoltre, si analizzano in dettaglio tutte le casistiche, per ogni singolo dataset (per anno e per città) in cui il titolo ha subito una variazione di anno in anno, in cerca delle parole aggiunte o rimosse che sono legate a determinati cambiamenti nell'indice delle prestazioni definito, giungendo a una selezione di ulteriori parole ritenute da analizzare.

Per costruire le nuvole di parole e analizzare i dati si utilizza un procedimento di analisi basato sugli spunti indicati da [Natassha Selvaraj, 2020](#) nell'articolo "A Beginner's Guide to Sentiment Analysis with Python - An end to end guide on building word clouds, beautiful visualizations, and machine learning models using text data" presente sul sito [towardsdatascience.com](https://towardsdatascience.com), al quale si consiglia caldamente di fare riferimento per analisi semantiche, per via della efficacia ed esaustività dei passaggi che l'autrice presenta, i quali al tempo stesso risultano snelli e poco articolati, consentendo di effettuare analisi semantiche seppur avendo poca esperienza con le stesse.

Nel dettaglio, per le analisi principali del presente lavoro di tesi si utilizza la parte relativa alla costruzione delle nuvole di parole, ma l'esperta di data science, riporta una tecnica che consente di analizzare un pezzo di testo e di mettere in luce la tipologia di sentimento che esso vuole esprimere. Tale procedimento, è stato utilizzato durante le ricerche per svolgere ulteriori analisi semantiche: si tratta fattivamente di eseguire una categorizzazione in base alla positività, neutralità o negatività espressa dal dato pezzo di testo. Diviene chiaro, dunque, il motivo per cui tale strumento sia molto utile ai fini della ricerca che si vuole effettuare: è possibile classificare le parole che effettivamente generano un sentimento positivo nel cliente, ed analizzarle per capire se presentano concretamente un legame con le prestazioni dei beni registrati sulla piattaforma. Per analizzare questo legame si crea una variabile binaria che indichi la presenza o meno di tale parola e si inserisce nel modello econometrico prima descritto, applicando i vari step della procedura.

## Collezione dei dati

I dati utilizzati per eseguire le analisi sono in formato csv e, come anticipato, provengono dal database di AirBnb, per mezzo di AirDNA, e riguardano un borgo di mare, uno di montagna e una grande città situati in Italia, rispettivamente Gallipoli, sita in Puglia, Erice, sul promontorio Siciliano che affianca la città di Trapani sita a nord-ovest dell'isola, e Torino, città tra le più grandi per estensione e fermento a livello hospitality ed utilizzo della piattaforma turistica di AirBnb.

## Strumenti utilizzati

Relativamente allo strumento informatico utilizzato per eseguire le analisi semantiche e le rappresentazioni grafiche si è deciso di sfruttare Jupyter Notebook (anaconda3), per via della sua fruibilità. Anche Microsoft Excel ha rappresentato uno strumento essenziale per operare direttamente sui dataset e per svolgere le analisi di dettaglio. Per quanto concerne le analisi econometriche, invece, il software statistico utilizzato è StataIC 12, che rappresenta uno degli strumenti econometrici più diffuso e di semplice utilizzo.

## Prove sui dati

Il primo campione analizzato è stato Erice, che avendo un ridotto numero campionario, 897 osservazioni, si è rivelato un ottimo punto di partenza per testare e mettere in uso gli algoritmi implementati su python, il modello di analisi semantica e per comprendere come gestire la tipologia di dati specifici del dataset.

## Pulizia del dataset

Al fine di svolgere le indagini si è effettuata una pulizia del dataset. Sono stati controllati tutti i dati in analisi e portati alla corretta unità di misura: per esempio i dati con formato “intero” riportavano una unità in più che è stata eliminata, riportando i valori alla quantità corretta. Per effettuare il controllo è stato utilizzato Microsoft Excel, attraverso le varie funzioni del software o strumenti di visualizzazione dei dati che agevolano il procedimento, come ad esempio le tabelle pivot.

Altra manipolazione importante ha riguardato l'eliminazione delle righe non idonee allo svolgimento delle analisi. A titolo di esempio, sono state rimosse dal dataset relativo a Gallipoli le righe non riportanti il “listing title” e aventi numerosi campi senza valori (ogni qualvolta si è verificata l'assenza del “listing title” si è ritrovata anche tale mancanza, a dimostrazione che tali casi fossero riportanti errori di registrazione o

pochi dati, e quindi poco significativi ai fini della ricerca o addirittura dannosi, in quanto outliers non attendibili, dovuti ad errori, che avrebbero distorto le inferenze dall'effettivo andamento della distribuzione del campione). Allo stesso modo è stato ripulito il dataset delle altre città in esame. Allo scopo di pulire i dataset si sono eseguiti vari step su Excel, al fine di eliminare prima gli annunci non aventi ricavi nell'anno in questione, e risultanti pertanto mai attivati concretamente seppur presenti tra i dati. In seguito, si passa alla rimozione dei duplicati in termini di codici univoci, presenti per via di ripetizioni di stessi annunci nel dataset. Per analisi ulteriori sono state utilizzate anche specifiche funzioni implementate direttamente in python, delle quali si riporteranno in seguito degli esempi durante la presentazione delle analisi svolte.

#### Resoconto pulizia dataset per città e anno

Città	Torino	
Anno	2019	2020
N. Annunci	12110	7125
N. Annunci aventi ricavi	6086	5592
N. Annunci al netto dei duplicati	6086	4829

Città	Gallipoli	
Anno	2019	2020
N. Annunci	4425	3448
N. Annunci aventi ricavi	2278	2378
N. Annunci al netto dei duplicati	2278	1567

Città	Erice	
Anno	2019	2020
N. Annunci	529	368
N. Annunci aventi ricavi	296	283
N. Annunci al netto dei duplicati	296	229

# Analisi

## Tipologia dei dati

Prima di presentare le analisi, si riporta una tabella riepilogativa per mostrare la tipologia dei dati e la relativa descrizione.

<u>Campo</u>	<u>Tipologia</u>	<u>Descrizione</u>
Property ID	string	Codice identificativo della abitazione
Listing Title	string	Titolo dell'annuncio
Property Type	enum	Tipologia di abitazione
Listing Type	enum	Tipologia di soggiorno (Intera casa, stanza, ecc)
Created Date	date	Data di creazione dell'annuncio
Country	enum	Paese in cui è situata l'abitazione
Latitude	float	Latitudine in cui è situata l'abitazione
Longitude	float	Longitudine in cui è situata l'abitazione
City	string	Città in cui è situata l'abitazione
Neighborhood	enum	Quartiere in cui è situata l'abitazione
Annual Revenue LTM (USD)	integer	Ricavo annuo degli ultimi 12 mesi in dollari
Count Reservation Days LTM	integer	Conteggio dei giorni in cui l'abitazione è riservato negli ultimi 12 mesi
Count Available Days LTM	integer	Conteggio dei giorni in cui l'abitazione è disponibile negli ultimi 12 mesi

Count Blocked Days LTM	integer	Conteggio dei giorni in cui l'abitazione è bloccato negli ultimi 12 mesi
Bedrooms	integer	Numero di stanze da letto presenti nell'abitazione
Bathrooms	integer	Numero di bagni presenti nell'abitazione
Max Guests	integer	Massimo numero di ospiti consentito
Calendar Last Updated	date	Data di ultimo aggiornamento dell'annuncio
Response Rate	float	Tasso di risposta dell'host alle eventuali richieste dei clienti
Airbnb Superhost	boolean	Indicatore della qualifica di Superhost assegnata all'host e verificato dalla piattaforma
HomeAway Premier Partner	boolean	Indicatore della qualifica di HomeAway Premier Partner assegnata all'host e verificato dalla piattaforma
Cancellation Policy	enum	Regolamentazione cancellazione della prenotazione effettuata
Security Deposit (USD)	integer	Ammontare in dollari del deposito di sicurezza richiesto dall'host
Cleaning Fee (USD)	integer	Ammontare in dollari della tariffa del servizio di pulizia
Extra People Fee (USD)	integer	Ammontare in dollari della tariffa per persona aggiuntiva
Published Monthly Rate (USD)	integer	Tariffa mensile media pubblicata per l'abitazione in dollari
Published Weekly Rate (USD)	integer	Tariffa settimanale media pubblicata per l'abitazione in dollari
Check-in Time	string	Orario o intervallo di tempo in cui è possibile introdursi nell'appartamento per il soggiorno
Checkout Time	string	Orario o intervallo di tempo in cui è necessario lasciare l'appartamento alla fine del soggiorno
Minimum Stay	integer	Durata minima del soggiorno in giorni per effettuare la prenotazione
Number of Reviews	integer	Numero di recensioni pubblicate sull'annuncio

Number of Photos	integer	Numero di foto pubblicate sull'annuncio
Instantbook Enabled	boolean	Indicatore della possibilità di prenotazione istantanea su piattaforma
Overall Rating	float	Valutazione complessiva dell'esperienza di soggiorno da parte degli utenti su scala da 1 a 5
Airbnb Host ID	string	Codice identificativo dell'host
HomeAway Property ID	string	Codice identificativo delle abitazioni HomeAway
Last Year	integer	Ultimo anno di rilevazione dei dati

## Analisi descrittive

Il primo passaggio fondamentale, successivo alla pulizia dei dati, riguarda un ulteriore approfondimento delle frequenze in relazione ai casi di annunci che sono presenti in entrambi gli anni e ai casi di quelli che, tra questi, hanno subito variazioni al titolo.

Di seguito il dettaglio delle analisi descrittive dei dataset per città e anno:

Città	Torino	
Anno	2019	2020
N. Annunci	12110	7125
N. Annunci aventi ricavi	6086	5592
N. Annunci al netto dei duplicati	6086	4829
N. Annunci presenti anche l'anno precedente		3955
N. Annunci con titolo cambiato		335

Città	Gallipoli	
Anno	2019	2020
N. Annunci	4425	3448
N. Annunci aventi ricavi	2278	2378
N. Annunci al netto dei duplicati	2278	1567
N. Annunci presenti anche l'anno precedente		1180
N. Annunci con titolo cambiato		147

Città	Erice	
Anno	2019	2020
N. Annunci	529	368
N. Annunci aventi ricavi	296	283
N. Annunci al netto dei duplicati	296	229
N. Annunci presenti anche l'anno precedente		184
N. Annunci con titolo cambiato		11

Dalle analisi si evidenzia che per Torino nel 2020 gli annunci si sono ridotti del 21%. L'82% degli annunci è già presente nell'anno precedente e soltanto l'8% di questi ha cambiato titolo. Relativamente a Gallipoli si evidenzia un calo del numero di annunci di addirittura 31 punti percentuali, con il 75% di annunci che è presente già nel 2019 e il 12% di questi che ha subito una effettiva variazione nel titolo. Per Erice si evidenzia un calo meno del numero degli annunci meno marcato, pari al 23%. Tra tutti gli annunci del 2020, l'80% risulta già pubblicato e attivo su piattaforma l'anno precedente con soltanto il 6% di tali annunci che presenta un cambiamento del titolo.

Quindi, nel complesso è possibile affermare che dal 2019 al 2020 si è andati incontro a una riduzione, tra i 20 e i 30 punti percentuali, degli annunci attivi sulla piattaforma

(in linea con le stringenti condizioni esterne dovute all'avvento della pandemia). Circa l'80% degli annunci in ogni dataset si è mantenuto negli anni, e di questi soltanto il 10% circa ha subito una variazione nel titolo: è evidente che pochissimi annunci sul totale presentano cambiamenti nel titolo di anno in anno.

Svolte le analisi descrittive complete è possibile passare alle analisi semantiche sulle singole parole. Si riportano a seguire i vari passaggi di analisi, classificati per tipologia di analisi e città.

## Analisi semantiche

### Analisi semantica Torino

Il primo passo di analisi semantica prevede per ciascuna annualità di riferimento la costruzione di una nuvola di parole nella quale sono evidenziate le singole parole in base alla loro frequenza nel dataset. Per questo passaggio, come specificato in precedenza, si utilizzano specifici comandi implementati in python. Nella costruzione delle nuvole, vengono bloccate le parole di disturbo, molto frequenti ma non significative, come preposizioni, articoli e simili. Il risultato per le due annualità è il seguente:



2019



Si rende evidente che il nome della città in esame si afferma in entrambi gli anni come la parola più frequente. A seguire la parola “casa”, con le sue versioni in lingua inglese, e le parole legate alla centralità dell’abitazione.

Il risultato più interessante è il fatto che le parole più frequenti sono le stesse in entrambi gli anni, con frequenze simili e ordine tra le parole di poco variato: la maggior parte degli host non ha intrapreso decisioni evidenti a livello semantico atte a rispondere allo shock esterno dettato dalla pandemia e le conseguenti variazioni dei fattori esterni di contorno.

La lettura di una scarsa risposta è lecita, considerato che senza azioni atte a intervenire è intuitivo pensare, come confermano i dati, che in uno scenario in cui la maggior parte degli annunci (82%) è stata confermata dal 2019 al 2020 senza alcun cambiamento, se non in solo l’8% dei casi, gli ulteriori annunci creati (restante 20%) si allineino nelle modalità di creazione ed espressione a quelli già presenti su piattaforma. Ecco che le parole più frequenti rimangono grosso modo le stesse nelle due annualità in maniera non stupefacente.

Inoltre, si nota che dalla selezione per mezzo delle frequenze non si sono evidenziate parole strettamente e direttamente connesse al covid, come per esempio “sanificazione” o “igienizzazione” e le parole correlate. In tale direzione si decide di selezionare ulteriori parole attraverso un’altra tipologia di analisi semantica: si osservano nel dettaglio tutti i casi relativi ad annunci in cui il titolo è cambiato dal 2019 al 2020, confrontando il nuovo titolo con quello precedente e selezionando tutte le parole che sono considerate interessanti dal punto di vista semantico o che siano chiaramente comparse o scomparse da un anno all’altro.

In tale confronto viene utilizzato l’indice di performance dell’annuncio (definito in dettaglio nella sezione delle analisi econometriche a seguire) al fine di mettere in evidenza singoli casi in cui l’aggiunta o la rimozione di una specifica parola avviene in corrispondenza di una variazione netta, positiva o negativa, delle performance, così da selezionare la data parola per le analisi econometriche atte a verificare se tale caso sia effettivamente legato alla specifica parola in esame e se tale risultato sia

generalizzato all'intero dataset in analisi. Di seguito un estratto dei casi descritti e un riepilogo delle parole selezionate con il dettaglio dello studio di frequenza attuato sulle stesse per entrambe le annualità (in ordine secondo le frequenze del 2020):

Titolo2019	Titolo2020	P2019	P2020	Delta%
COSY STUDIO ORMEA 164	STUDIO ORMEA 164 WITH COSY TERRACE	1.45	10.99	655%
RIVERSIDE DA PO 9 - APT 1	RIVERSIDE DA PO 9 APT. 1 - Ai piedi della collina	12.13	41.02	238%
Atelier Relax Petit Turin - Centralissimo	Atelier Relax Petit Turin - Centrale e Sanificato	17.15	34.51	101%
Walking to the moon-Il Monviso b&b	Turin green House	8.95	17.47	95%
Casa Verazzano	Maison Verazzano by Wonderful Italy	32.95	56.63	72%
Monocale luminoso con soppalco e balcone	[Metro Dante] Suite con soppalco e balcone	21.45	30.62	43%
Una stanza comoda vicino ai mezzi	FLAT: automatic entrance with code.Super SANITIZED	12.18	3.04	-75%
pop home!! pieno centro di torino!	Pop home nel cuore di Torino	43.34	6.71	-85%
Vintage is the center of Turin	B&B TURIN HISTORICAL CENTER	13.60	2.09	-85%
Sweet House Corso Francia	Corso Francia di fronte alla metro	39.29	3.10	-92%

Città	Torino			
	2019		2020	
Parole	2019	% su 6086	2020	% su 4829
maison	46	0.76%	40	0.83%
sweet	48	0.79%	36	0.75%
space	39	0.64%	33	0.68%
wonderful	13	0.21%	27	0.56%
giardino	33	0.54%	27	0.56%
comfortable	35	0.58%	26	0.54%
confortevole	43	0.71%	25	0.52%
collina	36	0.59%	25	0.52%
vintage	26	0.43%	23	0.48%
super	28	0.46%	23	0.48%
delizioso	24	0.39%	21	0.43%
independent	20	0.33%	17	0.35%
indipendente	29	0.48%	17	0.35%
relax	24	0.39%	16	0.33%
historical	19	0.31%	16	0.33%
terrazza	21	0.35%	16	0.33%
green	31	0.51%	14	0.29%
riverside	14	0.23%	11	0.23%
offer	7	0.12%	5	0.10%
best	8	0.13%	5	0.10%
price	1	0.02%	2	0.04%
sanificato	0	0.00%	2	0.04%
sanitized	0	0.00%	1	0.02%

Si nota subito che le parole covid-related ricercate, ovviamente non presenti nel 2019, sono mancanti anche nel 2020, se non per un paio di casi (le parole con frequenza nulla in entrambi gli anni non sono riportate). Diviene evidente che non è possibile attuare analisi econometriche su dei casi con frequenze talmente ristrette. Probabilmente il 2020 è un anno iniziale nel quale è ancora presto per trovare i termini in questione, che verosimilmente saranno presenti maggiormente in dataset relativi ad anni successivi. Bisogna comunque sottolineare che anche per le altre parole ricercate le frequenze, seppur più alte, non siano elevate: per le parole che non presentano frequenze inutilizzabili (minori del 0.3%) si effettua comunque la costruzione del modello attenționando i risultati alla luce di tali considerazioni sulle frequenze. Inoltre, anche in questo caso, seppur con eccezioni più marcate le parole assumono delle frequenze simili nei due diversi anni.

Selezionate attraverso questo processo le parole da testare con i procedimenti econometrici è possibile passare allo step successivo, ma prima si riporta la parte di analisi semantica relativa alle altre due città sotto esame.





Si rende evidente ancora una volta che il nome della città si conferma di gran lunga come la parola più frequente, che nel caso di Gallipoli risulta in quasi la metà dei titoli di entrambi gli anni. In maniera poco stupefacente per una località balneare, a seguire si trova la parola “mare” insieme alla sua traduzione in lingua inglese, e nuovamente “casa” e le parole legate alla centralità dell’abitazione.

Salta all’occhio il fatto che anche per questa città le parole più frequenti sono le stesse in entrambi gli anni, con frequenze simili e ordine tra le parole di poco diverso: la maggioranza degli host non ha intrapreso decisioni evidenti a livello semantico atte a rispondere allo shock esterno dettato dalla pandemia e le conseguenti variazioni dei fattori esterni di contorno. La scarsa risposta è confermata dai dati, posto che senza azioni atte a intervenire è intuitivo pensare che in uno scenario in cui tre quarti degli annunci (75%) è stata confermata dal 2019 al 2020 senza alcun cambiamento (se non nel 12% dei casi) gli ulteriori annunci creati (restante 25%) si allineino nelle modalità di creazione ed espressione a quelli già presenti su piattaforma. Ecco che le parole più frequenti rimangono per lo più simili nei due anni in maniera poco sorprendente e anche in questo caso si ottiene un risultato simile al precedente: dalla selezione tramite le frequenze non si sono evidenziate parole covid-related.

Come effettuato per Torino si selezionano ulteriori parole: si confrontano nel dettaglio tutti i casi relativi ad annunci in cui il titolo è variato negli anni, selezionando tutte le parole che sono considerate interessanti dal punto di vista semantico o che siano chiaramente apparse o sparite dal 2019 al 2020. Il confronto viene attuato sempre osservando l’indice di performance dell’annuncio così da mettere in evidenza singoli casi in cui l’aggiunta o la rimozione di una specifica parola si verifica in corrispondenza di una variazione netta, positiva o negativa, delle performance, così da selezionare la data parola per le analisi econometriche atte a testare se tale caso sia effettivamente legato alla specifica parola e se tale risultato sia generalizzato all’intero dataset in analisi. Di seguito si riporta un dettaglio dei casi descritti e un riepilogo delle parole selezionate con lo studio di frequenza attuato sulle stesse per entrambe le annualità (in ordine secondo le frequenze del 2020).

Titolo2019	Titolo2020	P2019	P2020	Delta%
BAIA VERDE 400 MT DAL MARE	GALLIPOLI BAIA VERDE BEACH	10.20	35.64	249%
La Casa di Bianca	Le suite del dormiglione	7.96	22.75	186%
Nuovissimo monolocale in centro a Gallipoli	Suite monolocale in centro a Gallipoli	29.34	73.90	152%
Grande casa con giardino 50m spiaggia a Gallipoli	Casa con giardino su spiaggia con distanza sociale	85.34	109.29	28%
La Dimora Del Cinque	Case Vacanza "La Dimora Del Cinque"	57.80	62.13	7%
Baia Verde Appartamento a 300m dal mare.	Baia Verde Appartamento vicino mare con posto auto	27.86	25.43	-9%
Gallipoli Corso Roma	Gallipoli relax e divertimento	132.61	93.76	-29%
Villetta di fronte alla spiaggia	Bilocale indipendente sulla spiaggia a Gallipoli	83.06	58.06	-30%
Nuovissimo appartamento nel centro di Gallipoli	Gallipoli apartment a few meters from the sea	146.37	59.74	-59%

Città	Gallipoli			
Anno	2019		2020	
Parola	Frequenza	% su 2278	Frequenza	% su 1567
green	17	0.75%	13	0.83%
super	8	0.35%	12	0.77%
center	14	0.61%	11	0.70%
central	27	1.19%	11	0.70%
relax	21	0.92%	8	0.51%
centrale	10	0.44%	6	0.38%
sweet	3	0.13%	2	0.13%
distanza	0	0.00%	1	0.06%
sanitized	0	0.00%	0	0.00%
sanificato	0	0.00%	0	0.00%

Si presenta una situazione simile a quella di Torino, anche in questo caso le parole ricercate strettamente connesse alla pandemia, chiaramente non presenti nel 2019, risultano assenti pure nell'anno successivo, se non per singoli casi. Risulta evidente che non è possibile attuare analisi econometriche su dei casi con frequenze talmente ristrette. Si rafforza l'ipotesi secondo cui il 2020 è un anno iniziale nel quale è ancora presto per trovare i termini in questione, che verosimilmente saranno presenti maggiormente in dataset relativi ad anni successivi. Bisogna comunque sottolineare che anche per le altre parole ricercate le frequenze non siano elevate: per le parole che presentano frequenze accettabili si effettua la costruzione del modello attenționando i risultati alla luce di tali considerazioni sulle frequenze.



Di queste parole evidenziate, sono state studiate nel dettaglio le frequenze in ciascun dataset, che si riportano nelle liste seguenti:

Città	Erice	
Anno	2019	
Parole	Frequenze	% su 295
Erice	65	22.03%
Casa	56	18.98%
Mare	35	11.86%
vacanze	19	6.44%
House	13	4.41%
beach	12	4.07%
vista	12	4.07%
B&B	12	4.07%
Sea	11	3.73%
residence	10	3.39%
Home	10	3.39%
garden	7	2.37%
antico	6	2.03%

Città	Erice	
Anno	2020	
Parole	Frequenze	% su 229
Erice	47	20.52%
Casa	36	15.72%
Mare	28	12.23%
beach	15	6.55%
vacanze	14	6.11%
residence	13	5.68%
Sea	11	4.80%
vista	11	4.80%
B&B	11	4.80%
House	9	3.93%
antico	9	3.93%
garden	7	3.06%
Home	6	2.62%

Si rende chiaro nuovamente che il nome della città si piazza in cima alla lista delle parole più frequenti in entrambi gli anni. Come per Torino, a seguire si ripresenta “casa”, insieme alle parole correlate e come per Gallipoli, località balneare, si ritrova molte volte la parola “mare” insieme alla sua traduzione in lingua inglese. Questa volta non sono presenti le parole legate alla centralità dell’abitazione.

Come per i 4 dataset precedenti, riferiti alle altre città, pure per Erice le parole più frequenti sono le stesse nel 2019 e nel 2020, con frequenze molto simili e ordine tra le parole di poco cambiato: anche qui gli host non ha intrapreso azioni nette a livello semantico per rispondere allo shock esterno del covid e le connesse variazioni dei fattori esterni di contorno. La scarsa risposta è evidente, soprattutto se assunto che senza tali azioni è intuitivo pensare, come confermano i dati, che in uno scenario in cui l’80% degli annunci è stato confermato negli anni in questione senza alcun cambiamento, considerando irrisorio il valore del 6% di titoli che hanno subito un cambiamento, gli ulteriori annunci creati (rimanente 20%) si allineino nelle modalità di

creazione ed espressione a quelli già presenti sulla piattaforma. Le parole più presenti risultano simili tra 2019 e 2020 in maniera non casuale e anche in questo caso si ottiene un risultato simile ai precedenti: dalla selezione tramite le frequenze non si sono evidenziate parole covid-related.

Anche per Erice si passa allo step successivo di analisi semantica e si selezionano ulteriori parole. Di seguito si riporta un dettaglio dei casi studiati e il riepilogo delle parole selezionate con annesso lo studio di frequenza attuato sulle stesse per entrambe le annualità (in ordine secondo le frequenze del 2020).

Titolo2019	Titolo2020	P2019	P2020	Delta%
Casa vacanze Le Terrazze	Casa "Le Terrazze"	19.16	36.43	90%
Erice Pietre Antiche Medieval house and rooms	Pietre Antiche Residence "Begonia upper floor"	21.63	33.59	55%
Bed & Breakfast ad Erice gialla	Antico Borgo Erice gialla	5.71	6.98	22%
APPARTAMENTO TULIPANO ROSSO	APPARTAMENTO TULIPANO ROSSO BICICLETTE WI-FI FREE	14.93	12.39	-17%
Bed & Breakfast ad Erice rossa	Antico Borgo Erice rossa	7.46	5.65	-24%
Casa a 30 metri dal mare	Finestra sul mare	55.29	41.07	-26%
Sea View Apartment Trapani	Sea Side Apartment Trapani	65.68	45.62	-31%

Città	Erice			
	2019		2020	
Parole	Frequenze	% su 295	Frequenze	% su 229
suite	1	0.34%	5	2.18%
spiaggia	3	1.02%	4	1.75%
relax	3	1.02%	4	1.75%
giardino	5	1.69%	4	1.75%
vetta	5	1.69%	3	1.31%
beachside	0	0.00%	3	1.31%
dimora	4	1.36%	3	1.31%
terrazza	2	0.68%	3	1.31%
borgo	1	0.34%	3	1.31%
cuore	2	0.68%	2	0.87%
verde	2	0.68%	2	0.87%
super	2	0.68%	2	0.87%
finestra	0	0.00%	1	0.44%
privata	2	0.68%	1	1.31%
terrace	2	0.68%	1	0.44%
sweet	1	0.34%	0	0.00%
sanitized	0	0.00%	0	0.00%
sanificato	0	0.00%	0	0.00%

Si ottiene un risultato in linea con le precedenti analisi semantiche delle altre città: anche in questo caso le parole covid-related ricercate, chiaramente non presenti nel 2019, risultano non esserci nel dataset del 2020. Risulta evidente che non è possibile attuare analisi econometriche su dei casi con frequenze talmente ristrette. Si rafforza ulteriormente l'ipotesi secondo cui il 2020 è un anno iniziale nel quale è ancora presto per trovare le parole in questione, e che verosimilmente è necessario analizzare dataset relativi ad anni successivi. Questa volta anche per le altre parole ricercate le frequenze non sono adeguate alle analisi econometriche: si effettua la costruzione dei modelli attenendosi ai risultati alla luce di tali considerazioni sulle frequenze e con riferimento ai dataset di Torino e Gallipoli, per avere un confronto, senza potere di fatto asserire nettamente posizioni relative a queste ultime parole sulla base del solo dataset di Erice.

#### Prime considerazioni semantiche

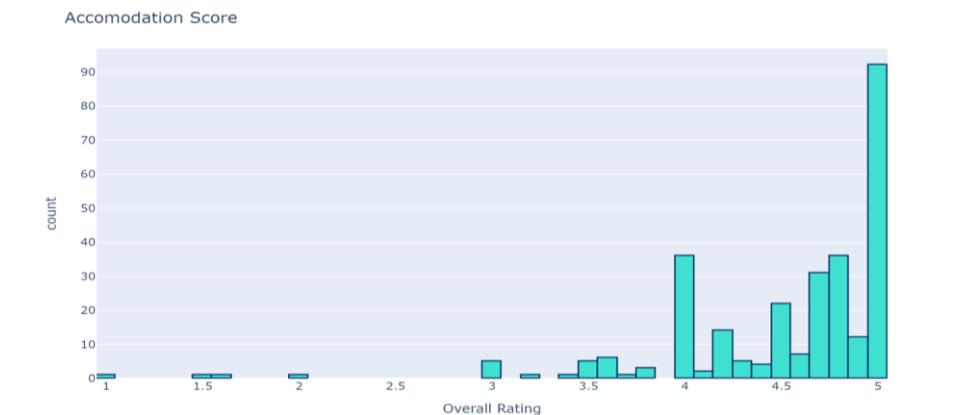
Svolte tutte le analisi semantiche su ogni dataset, per città e annualità, si evidenziano dei primi risultati di valore. In prima istanza, bisogna sottolineare che le parole utilizzate di anno in anno si mantengono, andando incontro a poche variazioni. In aggiunta, tale fenomeno non si presenta soltanto a livello di singola città, ma le parole più frequenti ricorrono in tutti i dataset presi in esame: le stesse parole compaiono con simile frequenza percentuale nelle diverse città, evidenziando un lessico e un vocabolario comune che caratterizza la pubblicazione degli annunci su piattaforma digitale.

In seconda istanza, si riscontra una inerzia degli host diffusa, per lo meno a livello semantico, nella risposta al cambiamento delle condizioni esterne in seguito a un forte shock come il diffondersi del virus e delle misure di contenimento correlate.

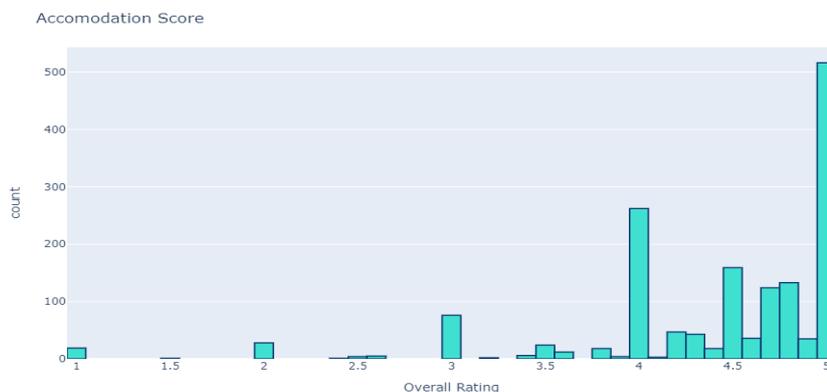
## Analisi semantiche ulteriori

Nel periodo di ricerca durante il quale si sono svolte le analisi, sono state testate più metodologie, portando a determinati procedimenti che ritenuti validi in partenza si sono dimostrati inadeguati per specifici fattori emersi durante il proseguo delle analisi stesse, ma di grande valore per mettere in risalto le criticità che le metodologie finora presentate riescono a superare con successo e altre evidenze relative ai dataset considerati. In questa sezione si riportano i principali riscontri delle ulteriori analisi che si sono svolte a livello semantico con un focus sul sentimento, per mezzo del procedimento della sentiment analysis.

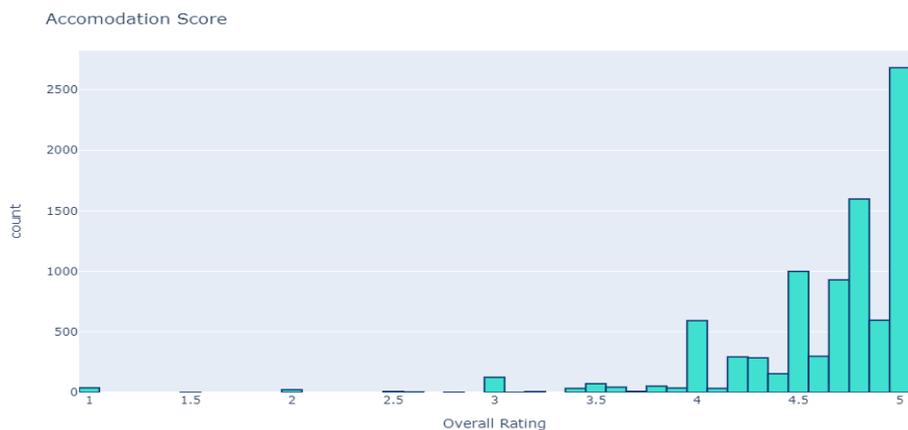
Durante le analisi si è rappresentata la distribuzione delle valutazioni degli annunci per ogni singola città:



*Erice*



*Gallipoli*



### *Torino*

Si evidenzia immediatamente, attraverso queste rappresentazioni, il bias di positività degli utenti che rilasciano una valutazione sull'abitazione di cui hanno usufruito, proprio come mostrato dai precedenti lavori di ricerca analizzati durante la revisione della letteratura. La distribuzione è completamente sbilanciata verso destra con moda pari al valore massimo della scala. In quest'ottica è chiaro che il campione relativo alle parole con accezione positiva è di gran lunga maggiore in numerosità campionaria.

Questo aspetto fortemente influente sui dati, porta a rendere i modelli previsionali di regressione logistica testati poco performanti e di scarsa affidabilità. La classificazione del sentiment effettuata sulle città risultava verosimile ma non valida, a causa delle distorsioni introdotte dal bias di positività. Questa tipologia di analisi ha permesso di mettere in luce tale caratteristica del settore, mostrando che analisi di questo genere non sono del tutto valide a causa di sproporzioni tra casi positivi e negativi. La metodologia precedente, invece, ritenuta solida e valida, permette di effettuare le analisi senza influenze del bias di positività e conseguenti distorsioni dei dati.

## Analisi econometriche

Come descritto in precedenza, tali analisi richiedono dei passaggi ben definiti e certo rigore scientifico. A tal fine, una volta identificate le parole chiave attraverso l'analisi semantica precedente, si definiscono delle ipotesi della tipologia seguente:

- La presenza della parola "xxxxx" nel titolo dell'annuncio di una data abitazione ha un effetto positivo significativo sulle prestazioni economiche dell'abitazione stessa;
- La presenza della parola "yyyyy" nel titolo dell'annuncio di una data abitazione ha un effetto negativo significativo sulle prestazioni economiche dell'abitazione stessa.

Il passaggio successivo è costituito dalla specificazione del modello, che prevede la scelta delle variabili indipendenti da inserire nello stesso. Nel caso specifico, per verificare le ipotesi precedenti relative alla semantica, si inserisce una variabile dummy, quindi booleana, legata alla presenza o meno della data parola in esame: tale variabile indipendente assume valore pari a 1 quando la parola stessa è contenuta nel titolo, diversamente ha valore pari a zero nel caso contrario. Allo stesso modo si costruisce una variabile relativa al cambio di titolo, per ogni caso in cui nel 2020 si è verificato un cambiamento nel titolo, così da analizzare l'ulteriore aspetto semantico del cambiamento di titolo in generale.

Chiaramente, al di là delle ipotesi relative ai fattori semantici, si vogliono osservare anche gli andamenti delle grandezze associate agli attributi degli annunci. Infatti, per consentire al modello di rappresentare in modo efficace la variazione delle performance si inseriscono ulteriori variabili di controllo, descrittive fenomeni al di fuori del fattore semantico, come per esempio la dimensione dell'appartamento (osservata attraverso il massimo numero di ospiti consentiti), il numero di recensioni presenti a sistema e il numero di foto pubblicate sull'annuncio. Il parametro per misurare la bontà di tale rappresentazione è l'R-squared, osservando attentamente se valori elevati siano solamente causati da un eccessivo utilizzo di variabili indipendenti o se il modello sia effettivamente verosimile. Si anticipa già in questa sede che i valori

connessi a tale parametro non sono elevati, indicando, quindi, che i fattori di regressione spiegano soltanto una piccola parte delle variazioni nella variabile obiettivo, ma si riconosce che nel settore in esame, caratterizzato da R-squared molto bassi, come dimostrato dallo studio della letteratura, tali valori siano ritenuti soddisfacenti.

Inoltre, in questa fase si stabilisce che la forma funzionale più adeguata a legare le variabili è una relazione lineare tra le stesse e si fissano come assunzioni relative al termine di errore, che esso è una variabile casuale che si distribuisce come una normale per la legge dei grandi numeri, in quanto i campioni utilizzati sono caratterizzati da numerosità tali da giustificare ampiamente i requisiti minimi per tale considerazione. Relativamente alla forma funzionale si è attenzionato l'andamento delle variabili selezionate al fine di scegliere quella che meglio lega e rappresenta le variabili stesse. Alcune delle variabili di controllo hanno chiaramente un effetto incrementale minore al crescere vertiginoso dei valori della variabile stessa, e questo condurrebbe alla scelta della forma logaritmica. Tuttavia, i valori in analisi non assumono valori estremamente elevati e risulta quindi assumibile che per gli ordini di grandezza in gioco ci si trovi in una porzione ancora lineare della curva logaritmica. Per esempio, considerando il massimo numero di guest nell'abitazione si assume che ad ogni aumento unitario di tale parametro l'effetto incrementale sia più o meno costante, per lo meno sino a numerosità inferiori ai 7 ospiti. Considerato che i casi superiori a questa soglia sono estremamente esigui si opta per la forma lineare. Considerazioni di questo genere sono applicabili anche alle altre variabili di controllo, mentre le variabili semantiche, tutte booleane hanno un andamento a gradini che si può approssimare a un andamento lineare, in favore della costruzione della regressione (lo stesso vale per i regressori di controllo binari). Ecco, quindi, che alla luce di queste assunzioni si è preferito utilizzare una forma di più semplice lettura e formulazione come la relazione lineare tra le variabili.

A riguardo della natura delle variabili indipendenti e della relazione delle stesse con il termine rappresentante il residuo si evidenzia che la presenza o meno della parola nel titolo dell'annuncio è certa, mentre il numero massimo di ospiti, seppur un buon indice, non rappresenta esattamente le dimensioni dell'abitazione; il numero di foto e di

recensioni non indica la qualità delle stesse, che potrebbero presentare ripetizioni o casi di foto non adeguatamente visibili e di recensioni fittizie o non chiare.

Le analisi in questione si sono svolte con un sottoinsieme dei campi presenti nel dataset: sono stati considerati quelli più idonei a una rappresentazione in forma lineare e aventi un maggiore significato a livello di ipotesi. Si riporta di seguito un dettaglio dei campi utilizzati durante le analisi, con il nome e il tipo della variabile ad essi associati.

	<b>Campo</b>	<b>Descrizione</b>	<b>Nome Variabile</b>	<b>Tipologia</b>
<b>Variabile Dipendente</b>	Annual Revenue LTM (USD)	Ricavo annuo degli ultimi 12 mesi in dollari	ann_rev	integer
<b>Variabili Indipendenti</b>	Number of Photos	Numero di foto	num_photos	integer
	Number of Reviews	Numero di recensioni	num_reviews	integer
	Response Rate	Tasso di risposta dell'host	response	integer
	AirBnb Superhost	Qualifica di superhost dell'host	superhost	boolean
	Instantbook Enabled	Possibilità di effettuare una prenotazione istantanea	instantbook	boolean
	Max Guests	Numero massimo di ospiti ammessi nell'abitazione	max_guests	integer
	Minimum Stay	Numero minimo di giorni di soggiorno	min_stay	integer

Si evidenzia che sono stati presi in considerazione maggiormente dei campi direttamente influenzabili dalle scelte dell'host, ossia scelti e gestiti dall'host: per esempio, l'host decide quale numero di foto inserire nell'annuncio, o quanti posti letto possiede l'abitazione, ma non può gestire direttamente il valore dell'overall rating, che dipende dalla valutazione dei clienti. Tale approccio si attua in vista dell'obiettivo del presente lavoro di studiare l'apporto dei comportamenti manageriali degli host.

Per le variabili indipendenti binarie (definite in precedenza) relative alle parole trovate attraverso la ricerca semantica si introducono nel modello delle dummies nominate come la parola stessa e della tipologia “boolean”. Esse sono create attraverso Microsoft Excel come rappresentato nei dettagli che si riportano direttamente in seguito per le singole città.

Successivamente, si utilizzano i dati, per stimare attraverso il software statistico i parametri caratterizzanti il modello, che non sono noti. Essi devono essere stimati attraverso i dati campionari, che nella fattispecie in particolare, come già evidenziato, sono stati utilizzati del tipo cross-sezionale. Il passo successivo prevede proprio la stima del modello econometrico, processo per il quale il meccanismo degli stimatori OLS è stato valutato il più indicato. Si giunge così alle stime dei parametri del modello che non sono noti, utilizzando i dati osservazionali del campione scelto.

In generale, la metodologia scelta, ritenuta la più solida e valida, consiste nella formulazione di numerosi modelli contenenti tutti i regressori di controllo e una sola variabile binaria semantica: l’obiettivo è mettere in luce l’apporto della singola parola e la sua specifica correlazione con le prestazioni economiche. Tale modello ha una formulazione del tipo seguente:

$$performance = c + (b_i \times word_i) + \sum(a_i \times X_i) + e$$

L’indice delle prestazioni economiche, variabile dipendente del modello, viene costruito utilizzando i ricavi annui e i giorni di effettiva attività dell’annuncio. Nello specifico si ottiene dividendo il valore in dollari dei ricavi annui per il numero dei giorni di disponibilità dell’appartamento dell’anno, intesi come la somma dei giorni in cui l’appartamento è “affittato” e i giorni in cui lo stesso è libero e disponibile per il soggiorno. Questo indice permette di standardizzare le prestazioni tra annunci che hanno diversa durata di permanenza sulla piattaforma: un annuncio presente 300 giorni, pur avendo prestazioni scarse, potrebbe avere un maggiore ammontare di ricavi di uno molto performante ma presente per soli 60 giorni. Quindi, osservando i

ricavi al giorno in dollari, come definito, si riesce a comprendere l'effettiva prestanza economica del singolo annuncio. Le  $X_i$  rappresentano i vari regressori di controllo definiti in precedenza, mentre le  $word_i$  rappresentano i regressori semantici denominati come la parola in questione.

Il modello viene costruito per ogni parola selezionata (circa una ventina per ogni anno), per ciascuna città, distinguendo i dataset del 2019 e del 2020. Tale processo porta alla formulazione di più di 150 modelli che vengono descritti in dettaglio (per ogni anno e città) a seguire.

Una volta analizzati tali modelli, l'obiettivo è di confrontare i risultati per comprendere se ci sono delle relazioni causa-effetto generalizzabili tra le parole e le performance. Risulta interessante sia il confronto di anno in anno sia quello tra città distinte: tale apporto rappresenta una mancanza in letteratura, come evidenziato dagli studi degli articoli di settore eseguiti all'inizio del presente lavoro di tesi. Inoltre, la metodologia selezionata consente di superare una forte criticità eliminando l'effetto distorsione dovuto alla multicollinearità delle variabili semantiche in esame.

## Analisi econometriche Torino

Alla luce delle parole evidenziate per Torino si definiscono le seguenti variabili semantiche: casa, centro, center, house, central, cozy, home, accogliente, cuore, piazza, suite, B&B, terrace, maison, sweet, space, wonderful, giardino, vintage, super, indipendente, relax, historical, terrazza. Di seguito un estratto del foglio di calcolo relativo alla costruzione delle variabili:

Property ID	Listing Title	Casa	House	Home	cuore	center	centro	central	...	suite
ab-150626	Studio in Vintage house downtown	0	1	0	0	0	0	0	...	0
ab-65383	IN THE HEART OF TURIN STYLISH HOME	0	0	1	0	0	0	0	...	0
ab-150573	Elegant apartment close to city centre	0	0	0	0	0	0	0	...	0
ab-99242	Private room, WiFi, near JuventusStadium	0	0	0	0	0	0	0	...	0
ab-116123	Panoramic Flat, parking , WI -FI	0	0	0	0	0	0	0	...	0
ab-73870	Suite with private bathroom	0	0	0	0	0	0	0	...	1
ab-74163	Wide apartment in the city center	0	0	0	0	1	0	0	...	0
ab-137951	B&B Borgo Dora Balòn	0	0	0	0	0	0	0	...	0
ab-36144	Garden House - Tube-Internet-A/C	0	1	0	0	0	0	0	...	0
ab-7870	Central Penthouse Historical Turin	0	1	0	0	0	0	1	...	0
...	...	...	...	...	...	...	...	...	...	...
ab-41448155	Rilassati nel cuore di Torino	0	0	0	1	0	0	0	...	0
ab-41426744	Monolocale Matrimoniale Soppalcato	0	0	0	0	0	0	0	...	0
ab-41305078	Charming apartment in Palazzo Martini di Cigala	0	0	0	0	0	0	0	...	0
ab-41304691	CASA TORRICELLI NEL CENTRALE QUARTIERE CROCETTA	1	0	0	0	0	0	1	...	0
ab-41291326	TERRAZZA MARTINI	0	0	0	0	0	0	0	...	0
ab-41265141	RESIDENZA STADIO	0	0	0	0	0	0	0	...	0
ab-41317003	Posto letto su divano - Studio regina margherita	0	0	0	0	0	0	0	...	0
ab-41222738	Appartamento bello e comodo in zona Crocetta	0	0	0	0	0	0	0	...	0
6086	<b>Somma</b>	712	295	222	144	319	654	250	...	113
	<b>Frequenza percentuale sul totale</b>	11.70%	4.85%	3.65%	2.37%	5.24%	10.75%	4.11%	...	1.86%

2019

Property ID	Listing Title	Casa	House	Home	cuore	center	centro	central	...	suite
ab-65383	IN THE HEART OF TURIN STYLISH HOME	0	0	1	0	0	0	0	...	1
ab-73633	At home,in the heart of Turin	0	0	1	0	0	0	0	...	1
ab-137951	B&B Borgo Dora Balòn	0	0	0	0	0	0	0	...	1
ab-150573	Elegant apartment close to city centre	0	0	0	0	0	0	0	...	1
ab-150626	Studio in Vintage house downtown	0	1	0	0	0	0	0	...	1
ab-178935	Holiday studio, near Metro station	0	0	0	0	0	0	0	...	1
ab-183925	Palazzo Piossasco di Rivalba 1	0	0	0	0	0	0	0	...	1
ab-222774	study of central and comfortable	0	0	0	0	0	0	1	...	1
ab-1302519	Comfortable apt. center of Turin	0	0	0	0	1	0	0	...	1
ab-1312476	Turin green House	0	1	0	0	0	0	0	...	1
...	...	...	...	...	...	...	...	...	...	...
ab-46945531	Otherside House in Torino	0	1	0	0	0	0	0	...	0
ab-47016812	Cozy and modern house	0	1	0	0	0	0	0	...	0
ab-47118900	"Giardini di sabbia": deliziosa camera con bagno	0	0	0	0	0	0	0	...	0
ab-47202154	COZY ROOM CROCETTA ELEGANT DISTRICT	0	0	0	0	0	0	0	...	0
ab-47250843	Vista su Torino	0	0	0	0	0	0	0	...	0
4829	<b>Somma</b>	585	232	171	122	252	532	195	...	98
	<b>Frequenza percentuale sul totale</b>	12.11%	4.80%	3.54%	2.53%	5.22%	11.02%	4.04%	...	2.03%

2020

Una volta costruite le variabili semantiche si genera per ciascun anno il dataset da analizzare attraverso il software statistico STATA. Tali dataset contengono sia il codice identificativo dell'appartamento che le variabili, semantiche e di controllo, ad esso connesse:

id	performance	max_guests	response_rate	superhost	minimum_stay	num_reviews	number_photos	instantbook	cambio_titolo	casa	house	...	home
ab-150626	43.13	4	100	1	2	383	23	1	0	0	1	...	0
ab-65383	23.00	4	100	1	3	36	28	0	0	0	0	...	1
ab-150573	44.18	7	100	1	2	161	27	1	0	0	0	...	0
ab-99242	5.80	2	89	1	1	105	15	0	0	0	0	...	0
ab-74163	45.06	5	100	0	2	54	23	1	0	0	0	...	0
ab-73855	146.17	12	90	0	1	8	35	0	0	0	0	...	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
ab-137951	13.36	5	100	0	1	9	23	0	0	0	0	...	0
ab-36144	12.26	2	100	0	13	13	20	0	0	0	1	...	0
ab-7870	12.00	2	100	0	1	6	14	0	0	0	1	...	0
ab-37844	28.29	3	100	0	3	2	9	0	0	0	0	...	0
ab-73907	5.37	2	90	0	1	7	12	0	0	0	0	...	0
ab-111974	9.20	2	100	0	3	1	6	0	0	0	0	...	0
ab-73633	18.98	4	100	1	3	20	23	0	0	0	0	...	1

2019

id	performance	max_guests	response_rate	superhost	minimum_stay	num_reviews	number_photos	instantbook	cambio_titolo	casa	house	...	home
ab-37844	32.45	3	100	1	3	5	9	0	1	0	0	...	0
ab-65383	83.31	4	100	1	2	36	28	0	0	0	0	...	1
ab-73633	41.14	4	100	1	2	31	23	0	0	0	0	...	1
ab-73855	24.55	12	90	0	1	8	35	0	0	0	0	...	0
ab-73907	18.97	2	90	0	1	7	12	0	0	0	0	...	0
ab-78815	3.19	2	100	0	5	2	4	0	0	0	0	...	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
ab-99242	1.45	2	100	1	1	108	14	0	0	0	0	...	0
ab-130791	49.50	4	100	1	2	103	28	0	1	0	0	...	0
ab-137951	6.88	5	100	0	1	9	23	0	0	0	0	...	0
ab-150573	36.25	7	100	0	2	170	28	1	0	0	0	...	0
ab-150626	26.77	4	100	0	2	419	23	1	0	0	1	...	0
ab-178935	4.85	2	83	0	1	4	14	0	0	0	0	...	0
ab-193996	8	2	100	1	20	23	12	1	0	0	0	...	0

2020

Ottenuti i dataset idonei alle analisi econometriche si sono sviluppati tutti i modelli seguendo la metodologia descritta in precedenza. Per ognuno di questi si riportano a seguire i risultati ottenuti (con tutti i parametri di interesse) e le considerazioni annesse.

	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020
casa	-1.207 (-1.43)	-0.578 (-0.68)										
centro			1.57 (1.71)	1.503 (1.59)								
center					-0.0697 (-0.06)	0.168 (0.13)						
house							-2.223 (-1.96)	-0.688 (-0.61)				
central									2.702* (1.98)	2.205 (1.64)		
cozy											-0.603 (-0.59)	-0.00314 (-0.00)
max_guests	6.392*** (14.96)	4.706*** (9.04)	6.380*** (15.01)	4.701*** (9.06)	6.378*** (14.99)	4.699*** (9.05)	6.382*** (15.00)	4.701*** (9.06)	6.372*** (14.97)	4.698*** (9.05)	6.374*** (14.97)	4.699*** (9.04)
response_rate	0.0568*** (3.49)	-0.00143 (-0.09)	0.0562*** (3.46)	-0.00159 (-0.10)	0.0566*** (3.48)	-0.00162 (-0.10)	0.0567*** (3.49)	-0.00152 (-0.10)	0.0562*** (3.46)	-0.00172 (-0.11)	0.0567*** (3.49)	-0.00161 (-0.10)
superhost	3.171*** (4.56)	4.242*** (6.28)	3.117*** (4.49)	4.177*** (6.23)	3.138*** (4.51)	4.228*** (6.25)	3.160*** (4.55)	4.228*** (6.28)	3.151*** (4.54)	4.235*** (6.29)	3.134*** (4.51)	4.231*** (6.28)
minimum_stay	0.0193 (0.74)	-0.0494* (-2.21)	0.0199 (0.76)	-0.0491* (-2.17)	0.0195 (0.74)	-0.0496* (-2.20)	0.0189 (0.72)	-0.0499* (-2.21)	0.0193 (0.74)	-0.0493* (-2.19)	0.0195 (0.75)	-0.0496* (-2.20)
num_reviews	0.00987 (1.96)	0.00458 (1.00)	0.0105* (2.09)	0.00493 (1.09)	0.0103* (2.08)	0.00468 (1.04)	0.0105* (2.09)	0.00481 (1.06)	0.00937 (1.87)	0.00419 (0.92)	0.0103* (2.06)	0.00475 (1.05)
number_photos	0.210*** (6.00)	0.168*** (5.47)	0.210*** (6.00)	0.169*** (5.49)	0.210*** (5.98)	0.168*** (5.46)	0.211*** (6.01)	0.168*** (5.46)	0.210*** (5.98)	0.167*** (5.45)	0.210*** (5.97)	0.168*** (5.46)
instantbook	5.543*** (9.05)	5.524*** (9.47)	5.540*** (9.06)	5.536*** (9.49)	5.527*** (9.03)	5.528*** (9.48)	5.528*** (9.03)	5.536*** (9.50)	5.564*** (9.08)	5.563*** (9.51)	5.522*** (9.02)	5.528*** (9.48)
_cons	0.647 (0.37)	2.012 (1.52)	0.428 (0.24)	1.797 (1.34)	0.581 (0.33)	1.976 (1.49)	0.633 (0.36)	1.995 (1.51)	0.529 (0.30)	1.912 (1.45)	0.610 (0.35)	1.980 (1.49)
N	6085	4829	6085	4829	6085	4829	6085	4829	6085	4829	6085	4829
R2	0.2357	0.194	0.2359	0.1943	0.2355	0.1939	0.2358	0.1939	0.2359	0.1943	0.2355	0.1943

I modelli appena riportati rappresentano quelli relativi alle parole più frequenti nei dataset di Torino. Si nota immediatamente come tutte le parole, a meno di “central”, risultino non significative: evidentemente per queste parole molto frequenti non si sono distinti andamenti ben delineati tali da risultare dalle analisi econometriche. Ma in questa ottica assume ancora più riguardo l’interessante ritrovamento relativo alla parola “central” che sembrerebbe essere andata incontro a una riduzione di attrattività tra il 2019 e il 2020: la parola positivamente influente nel primo anno, risulta non significativa nell’anno successivo.

Tale risultato è in linea con quanto accaduto durante il covid, considerato che le persone si siano dimostrate disinteressate ai luoghi centrali e affollati. Per la restante parte dei dati non si ritrovano, come detto, casi di particolare interesse. Relativamente al parametro di rappresentatività si evidenziano valori più bassi nel 2020, ma in generale tali valori risultano in linea con gli ordini di grandezza del settore. Il coefficiente di intercetta risulta sempre non significativamente diverso da zero come intuibile: senza l’apporto degli altri termini, che rappresentano tutte componenti fondamentali degli annunci, è ragionevole che le prestazioni tendano a zero e che rese costanti tali altre variabili non si evidenzino variazioni nelle prestazioni.

Per quanto riguarda i regressori di controllo si nota che quattro variabili risultano sempre fortemente significative: quelle relative al massimo numero di ospiti, alla qualifica di superhost, al numero di foto pubblicate e alla possibilità di prenotazione istantanea. Esse sono sempre positivamente correlate alle prestazioni economiche degli annunci pubblicati e rappresentano i fattori di maggiore interesse su cui gli host dovrebbero concentrarsi. Al contrario, il minimo numero di giorni di soggiorno è l'unica variabile che ha impatto significativamente negativo, seppur solamente nel 2020. Tale risultato si traduce in una preferenza degli ospiti per soggiorni più flessibili, cosicché rendere questo parametro più stringente rappresenta da parte dell'host un'azione che impatta negativamente le performance.

Il tasso di risposta dell'host risulta positivamente impattante in modo saltuario mentre poco si può asserire relativamente al numero di recensioni, generalmente non significativo, se non per sparuti casi. Tale fenomeno è interessante: l'unico aspetto non direttamente controllabile dall'host risulta essere non importante in termini di prestazioni dell'abitazione sulla piattaforma.

Bisogna sottolineare che i coefficienti di regressione, associati alle variabili indipendenti, presentano dei valori verosimili in relazione al tipo di variabile ai quali sono connessi: le variabili binarie, come superhost e instant booking, presentano dei coefficienti più elevati, in quanto in base al valore che assumono, "0" o "1", si ha un vero e proprio salto discontinuo, mentre variabili discrete, come il numero di foto o il tasso di risposta, hanno un'influenza proporzionale al valore della variabile. Non si deve erroneamente pensare che la caratteristica di superhost o di prenotazione istantanea abbiano un effetto decisamente migliore sulle prestazioni rispetto agli altri fattori inseriti nella regressione: è chiaro che essi abbiano un effetto positivo, ma la determinazione del fatto che sono più o meno influenti rispetto alle altre componenti, dipende proprio dai valori di quest'ultime. Si può chiaramente affermare che l'aggiunta di una foto abbia un impatto sicuramente inferiore alla possibilità di prenotare istantaneamente, ma un numero superiore di foto potrebbe portare a un maggiore apporto, positivo, sulle prestazioni economiche.

	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020
home	-0.0696 (-0.04)	-0.716 (-0.48)										
accogliente			-1.254 (-0.75)	-0.951 (-0.72)								
cuore					3.605* (2.24)	-1.599 (-1.15)						
piazza							0.456 (0.24)	1.448 (0.84)				
suite									10.33*** (3.55)	7.130** (3.24)		
B&B											-2.544 (-1.22)	-5.441** (-2.66)
max_guests	6.378*** (15.01)	4.701*** (9.06)	6.378*** (14.99)	4.699*** (9.06)	6.380*** (15.00)	4.700*** (9.06)	6.378*** (14.98)	4.698*** (9.05)	6.395*** (15.03)	4.711*** (9.08)	6.377*** (14.99)	4.700*** (9.06)
response_rate	0.0566*** (3.48)	-0.00172 (-0.11)	0.0567*** (3.48)	-0.00163 (-0.10)	0.0570*** (3.50)	-0.00193 (-0.12)	0.0566*** (3.48)	-0.00169 (-0.11)	0.0580*** (3.64)	-0.000782 (-0.05)	0.0568*** (3.49)	-0.00124 (-0.08)
superhost	3.137*** (4.52)	4.237*** (6.29)	3.130*** (4.51)	4.231*** (6.28)	3.149*** (4.54)	4.223*** (6.27)	3.137*** (4.52)	4.238*** (6.30)	3.067*** (4.41)	4.158*** (6.18)	3.131*** (4.51)	4.213*** (6.26)
minimum_stay	0.0194 (0.74)	-0.0497* (-2.20)	0.0198 (0.74)	-0.0494* (-2.19)	0.0198 (0.76)	-0.0498* (-2.20)	0.0195 (0.74)	-0.0493* (-2.19)	0.0206 (0.79)	-0.0482* (-2.14)	0.0190 (0.72)	-0.0508* (-2.23)
num_reviews	0.0103* (2.05)	0.00476 (1.05)	0.0105* (2.09)	0.00487 (1.07)	0.00999* (1.99)	0.00486 (1.07)	0.0103* (2.05)	0.00480 (1.06)	0.0113* (2.25)	0.00542 (1.20)	0.0101* (2.01)	0.00453 (1.00)
number_photos	0.210*** (5.95)	0.168*** (5.47)	0.209*** (5.93)	0.167*** (5.43)	0.210*** (5.97)	0.168*** (5.47)	0.210*** (5.98)	0.168*** (5.45)	0.206*** (5.89)	0.165*** (5.39)	0.211*** (6.01)	0.170*** (5.52)
instantbook	5.528*** (9.02)	5.522*** (9.47)	5.514*** (9.00)	5.519*** (9.46)	5.531*** (9.04)	5.529*** (9.48)	5.525*** (9.02)	5.525*** (9.47)	5.433*** (8.96)	5.423*** (9.31)	5.516*** (9.01)	5.527*** (9.48)
_cons	0.581 (0.33)	2.006 (1.51)	0.622 (0.36)	2.013 (1.52)	0.445 (0.26)	2.041 (1.54)	0.573 (0.33)	1.963 (1.48)	0.304 (0.18)	1.813 (1.37)	0.612 (0.35)	1.993 (1.51)
N	6085	4829	6085	4829	6085	4829	6085	4829	6085	4829	6085	4829
R2	0.2355	0.1939	0.2356	0.1939	0.2359	0.1940	0.2355	0.1940	0.2383	0.1959	0.2357	0.1947

L'andamento si mantiene simile anche per il secondo gruppo di modelli. La maggior parte delle parole risulta non significativamente impattante sulle prestazioni, a meno di "cuore", "suite" e "B&B": i ritrovamenti su quest'ultime parole sono interessanti.

Come per "central" nel primo gruppo, sembrerebbe che la parola "cuore" è andata incontro a una riduzione di attrattività nel passaggio di anno in anno, infatti la parola in questione è positivamente influente nel 2019 e risulta non significativa nel 2020. D'altronde le parole sono simili per valenza considerato che solitamente tale parola viene intesa con significato assimilabile a centrale o centro in esempi come "nel cuore della città" nei quali indica centralità e vicinanza ai punti di interesse. Tale risultato è in linea con quanto accaduto durante la pandemia, considerato che le persone si sono dimostrate disinteressate ai luoghi centrali e affollati.

Per la parola "suite" il discorso è diverso: essa è la prima che risulta significativamente influente in entrambe le annualità, sempre con apporto positivo. Sembrerebbe, dunque, che tale parola abbia un'accezione positiva, probabilmente connessa generalmente al campo del lusso, che comporta una certa attrattività della stessa nel titolo. Gli host dovrebbero ispirarsi a tale tipologia di impostazione dell'alloggio mettendo in risalto tale questione nell'annuncio. Similmente, il segno meno in "B&B"

potrebbe esser letto con un'accezione negativa se comparato all'immagine lussuosa legata a "suite": nonostante si stia parlando proprio di B&B, sembrerebbe meglio non sottolineare tale aspetto, in favore di un'ottica più focalizzata sulla bellezza dell'alloggio che sulla sua economicità. Le considerazioni sui regressori di controllo, il parametro di rappresentatività e il coefficiente di intercetta risultano analoghe ai modelli del primo gruppo analizzato.

	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020
terrace	8.266*	8.801*										
	(2.57)	(2.06)										
maison			5.801	1.129								
			(1.84)	(0.32)								
sweet					-6.817**	-4.322						
					(-3.24)	(-1.89)						
space							-3.495	-1.192				
							(-1.35)	(-0.40)				
wonderful									1.543	2.280		
									(0.29)	(0.59)		
giardino											0.449	3.142
											(0.08)	(0.41)
max_guests	6.382***	4.701***	6.366***	4.698***	6.373***	4.698***	6.381***	4.700***	6.378***	4.697***	6.378***	4.698***
	(14.99)	(9.06)	(14.93)	(9.04)	(14.98)	(9.06)	(14.99)	(9.06)	(14.99)	(9.04)	(15.00)	(9.07)
response_rate	0.0569***	-0.00176	0.0564***	-0.00170	0.0570***	-0.00186	0.0567***	-0.00166	0.0566***	-0.00169	0.0566***	-0.00150
	(3.50)	(-0.11)	(3.47)	(-0.11)	(3.50)	(-0.12)	(3.49)	(-0.10)	(3.48)	(-0.11)	(3.48)	(-0.09)
superhost	3.174***	4.264***	3.127***	4.226***	3.133***	4.240***	3.162***	4.237***	3.138***	4.242***	3.137***	4.238***
	(4.57)	(6.34)	(4.50)	(6.26)	(4.51)	(6.30)	(4.55)	(6.29)	(4.52)	(6.30)	(4.52)	(6.26)
minimum_stay	0.0195	-0.0500*	0.0197	-0.0495*	0.0194	-0.0500*	0.0193	-0.0497*	0.0195	-0.0495*	0.0194	-0.0496*
	(0.74)	(-2.20)	(0.75)	(-2.19)	(0.74)	(-2.21)	(0.74)	(-2.20)	(0.74)	(-2.19)	(0.74)	(-2.20)
num_reviews	0.00875	0.00339	0.0106*	0.00481	0.0106*	0.00485	0.0103*	0.00478	0.0103*	0.00481	0.0103*	0.00482
	(1.74)	(0.74)	(2.11)	(1.06)	(2.11)	(1.07)	(2.06)	(1.05)	(2.05)	(1.06)	(2.06)	(1.07)
number_photos	0.205***	0.163***	0.209***	0.168***	0.209***	0.167***	0.210***	0.168***	0.210***	0.168***	0.210***	0.168***
	(5.79)	(5.26)	(5.96)	(5.46)	(5.95)	(5.43)	(5.98)	(5.46)	(5.98)	(5.46)	(5.98)	(5.47)
instantbook	5.555***	5.534***	5.522***	5.523***	5.546***	5.542***	5.524***	5.527***	5.527***	5.518***	5.528***	5.521***
	(9.08)	(9.50)	(9.02)	(9.48)	(9.06)	(9.51)	(9.02)	(9.48)	(9.03)	(9.46)	(9.03)	(9.48)
_cons	0.551	2.008	0.608	1.989	0.623	2.045	0.578	1.990	0.576	1.986	0.574	1.957
	(0.32)	(1.52)	(0.35)	(1.50)	(0.36)	(1.54)	(0.33)	(1.50)	(0.33)	(1.50)	(0.33)	(1.47)
N	6085	4829	6085	4829	6085	4829	6085	4829	6085	4829	6085	4829
R2	0.2365	0.1957	0.2359	0.1939	0.2360	0.1942	0.2356	0.1939	0.2355	0.1940	0.2355	0.1940

Nel terzo gruppo di modelli si evidenzia per "terrace" un caso analogo al precedente relativo alla parola "suite": pubblicizzare nell'annuncio la presenza di una terrazza risulta positivo. Negli alloggi che possiedono questo tipo di caratteristica aggiuntiva gli host dovrebbero assolutamente mettere in risalto tale aspetto, che risulta positivamente correlato con le prestazioni economiche dell'alloggio. Risulta invece poco interpretabile il segno meno legato a sweet nel quinto modello, mentre tutte le altre parole non risultano di interesse, data la loro non significatività. Anche in questo gruppo le considerazioni sulle variabili di controllo, il valore di R-squared e il coefficiente di intercetta sono simili ai modelli dei gruppi precedenti.

	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020
vintage	-5.684 (-1.33)	-7.045* (-2.14)										
super			-8.197** (-3.16)	-3.460 (-0.72)								
independente					-4.075 (-1.21)	-3.467 (-1.30)						
relax							-9.304* (-2.32)	2.360 (0.38)				
historical									17.03 (1.96)	7.337 (0.98)		
terrazza											4.404 (1.13)	3.122 (1.04)
max_guests	6.380*** (15.00)	4.701*** (9.06)	6.373*** (14.98)	4.701*** (9.06)	6.375*** (14.98)	4.696*** (9.05)	6.386*** (15.01)	4.698*** (9.05)	6.383*** (15.01)	4.699*** (9.06)	6.379*** (14.99)	4.700*** (9.06)
response_rate	0.0566*** (3.48)	-0.00150 (-0.09)	0.0559*** (3.43)	-0.00162 (-0.10)	0.0564*** (3.47)	-0.00159 (-0.10)	0.0567*** (3.48)	-0.00167 (-0.11)	0.0567*** (3.48)	-0.00169 (-0.11)	0.0564*** (3.47)	-0.00164 (-0.10)
superhost	3.127*** (4.50)	4.207*** (6.25)	3.149*** (4.54)	4.246*** (6.30)	3.134*** (4.51)	4.229*** (6.28)	3.132*** (4.52)	4.231*** (6.28)	3.172*** (4.59)	4.235*** (6.29)	3.140*** (4.52)	4.235*** (6.29)
minimum_stay	0.0193 (0.74)	-0.0499* (-2.21)	0.0200 (0.76)	-0.0496* (-2.20)	0.0193 (0.74)	-0.0497* (-2.20)	0.0192 (0.73)	-0.0495* (-2.20)	0.0196 (0.75)	-0.0494* (-2.19)	0.0195 (0.74)	-0.0496* (-2.20)
num_reviews	0.0104* (2.07)	0.00495 (1.09)	0.0103* (2.05)	0.00465 (1.02)	0.0102* (2.03)	0.00469 (1.03)	0.0103* (2.05)	0.00475 (1.05)	0.00963* (2.00)	0.00456 (1.01)	0.0103* (2.05)	0.00476 (1.05)
number_photos	0.211*** (6.00)	0.169*** (5.49)	0.210*** (5.99)	0.168*** (5.46)	0.210*** (5.98)	0.168*** (5.46)	0.210*** (5.98)	0.168*** (5.46)	0.209*** (5.96)	0.167*** (5.46)	0.210*** (5.98)	0.168*** (5.46)
instantbook	5.531*** (9.03)	5.553*** (9.52)	5.531*** (9.04)	5.525*** (9.47)	5.526*** (9.03)	5.521*** (9.47)	5.511*** (9.00)	5.532*** (9.48)	5.525*** (9.03)	5.535*** (9.49)	5.514*** (9.00)	5.519*** (9.46)
_cons	0.581 (0.33)	1.965 (1.48)	0.689 (0.40)	1.996 (1.51)	0.634 (0.36)	2.010 (1.52)	0.597 (0.34)	1.983 (1.50)	0.531 (0.30)	1.973 (1.49)	0.582 (0.33)	1.970 (1.49)
N	6085	4829	6085	4829	6085	4829	6085	4829	6085	4829	6085	4829
R2	0.2357	0.1944	0.2360	0.1940	0.2356	0.1940	0.2360	0.1939	0.2368	0.1943	0.2356	0.1940

Nell'ultimo gruppo, invece, non si evidenziano casi di particolare interesse per il presente lavoro di ricerca, seppur, chiaramente, vengano confermati gli andamenti delle variabili di controllo.

## Analisi econometriche Gallipoli

Alla luce delle parole evidenziate per Gallipoli si definiscono le seguenti variabili semantiche: mare, casa, centro, vacanze, vista, sea, house, B&B, spiaggia, lido, beach, suite, residence, giardino, cuore, dimora, lungomare, home, terrazza, terrace, super, center, central, relax. Di seguito un estratto del foglio di calcolo relativo alla costruzione delle variabili:

Property ID	Listing Title	Casa	House	Home	Mare	Sea	cuore	center	...	centro
ab-153486	Ancient Gallipoli Exclusive holiday	0	0	0	0	0	0	0	...	0
ab-112767	VIVI NATURA AGRITURISMO CINQUE LETTI	0	0	0	0	0	0	0	...	0
ab-130827	Apartment with sea view - GALLIPOLI	0	0	0	0	1	0	0	...	0
ab-132601	Villetta a due passi dal mare	0	0	0	1	0	0	0	...	0
ab-87951	Gallipoli baia verde vicino spiaggia	0	0	0	0	0	0	0	...	0
ab-112793	VIVI NATURA CITTA' CAMERA DOPPIA	0	0	0	0	0	0	0	...	0
...	...	...	...	...	...	...	...	...	...	...
ab-41396628	b&b SUNSHINE	0	0	0	0	0	0	0	...	0
ab-41172287	VENTIDIMAREGALLIPOLI ROOM RAFFICA1	0	0	0	1	0	0	0	...	0
ab-41256039	Appartamento Violetta - Gallipoli	0	0	0	0	0	0	0	...	0
ab-41307927	Casa vacanze Gallipoli- vista mare con piscina	1	0	0	1	0	0	0	...	0
ab-41256176	Baia Verde - Gallipoli, Spazioso Appartamento	0	0	0	0	0	0	0	...	0
ab-41285184	Appartamento Gandhi - Gallipoli	0	0	0	0	0	0	0	...	0
2278	<b>Somma</b>	266	95	36	547	118	31	14	...	145
	<b>Frequenza percentuale sul totale</b>	12%	4%	2%	24%	5%	1%	1%	...	6%

2019

Property ID	Listing Title	Casa	House	Home	Mare	Sea	cuore	center	...	centro
ab-87951	Gallipoli baia verde vicino spiaggia	0	0	0	0	0	0	0	...	0
ab-130827	Apartment with sea view - GALLIPOLI	0	0	0	0	1	0	0	...	0
ab-153486	Ancient Gallipoli Exclusive holiday	0	0	0	0	0	0	0	...	0
ab-1300094	Gallipoli Baia verde apartment	0	0	0	0	0	0	0	...	0
ab-695728	Apartment 200 meters from the beaches 12 places	0	0	0	0	0	0	0	...	0
ab-1412424	Gallipoli. Loft immerso nel verde	0	0	0	0	0	0	0	...	0
ab-1200635	Gallipoli old town	0	0	0	0	0	0	0	...	0
ab-1224506	CENTRO GALLIPOLI CASA INDIPENDENTE	1	0	0	0	0	0	0	...	1
...	...	...	...	...	...	...	...	...	...	...
ab-46207303	Tenuta Cotriero	0	0	0	0	0	0	0	...	0
ab-46207966	Appartamento Mimosa	0	0	0	0	0	0	0	...	0
ab-46208042	Attic Iolanda	0	0	0	0	0	0	0	...	0
ab-46208221	Lido Giovani	0	0	0	0	0	0	0	...	0
ab-46208383	Pearl Apartment	0	0	0	0	0	0	0	...	0
ab-46592413	Monolocale vista mare	0	0	0	1	0	0	0	...	0
1567	<b>Somma</b>	196	61	20	373	83	28	11	...	97
	<b>Frequenza percentuale sul totale</b>	13%	4%	1%	24%	5%	2%	1%	...	6%

2020

Una volta costruite le variabili semantiche si genera per ciascun anno il dataset da analizzare attraverso il software statistico STATA. Tali dataset contengono sia il codice identificativo dell'appartamento che le variabili, semantiche e di controllo, ad esso connesse:

id	performance	max_guests	response_rate	superhost	minimum_stay	num_reviews	number_photos	instantbook	cambio_titolo	casa	house	...	home
ab-43992	51.08	5	100	1	3	44	13	0	0	0	0	...	0
ab-630376	23.17	2	0	0	2	0	20	0	0	1	0	...	0
ab-691949	69.59	10	70	0	3	2	22	1	0	0	0	...	1
ab-613492	18.63	2	100	0	1	141	33	1	0	0	0	...	1
ab-538365	32.34	10	70	0	3	6	36	0	0	0	0	...	1
ab-496758	22.59	2	100	0	2	46	20	0	0	0	0	...	0
ab-1068202	28.74	4	100	0	5	0	13	0	0	0	0	...	0
ab-454412	10.97	2	100	0	2	50	14	0	0	0	1	...	0
ab-1030353	17.39	6	100	0	3	2	37	0	0	1	0	...	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
ab-226324	42.68	5	100	0	5	16	27	0	0	0	0	...	0
ab-1266525	20.26	5	100	0	1	104	26	1	0	0	0	...	0
ab-1205176	18.18	4	100	0	3	12	10	0	0	0	0	...	0
ab-1508966	1.81	3	100	0	1	12	61	0	0	0	0	...	0
ab-1469427	9.53	4	0	0	1	1	18	0	0	0	0	...	1
ab-1669614	1.39	4	0	0	3	1	14	0	0	0	0	...	0
ab-1518285	9.31	2	63	0	1	0	10	1	0	0	0	...	0
ab-1532125	38.26	3	100	0	1	0	33	0	0	0	0	...	0
ab-2038021	65.68	5	100	1	2	68	39	1	0	0	0	...	0
ab-2041142	79.89	5	100	1	2	2	40	1	0	0	0	...	0

## 2019

id	performance	max_guests	response_rate	superhost	minimum_stay	num_reviews	number_photos	instantbook	cambio_titolo	casa	house	...	home
ab-43992	84.43	5	100	1	3	46	13	0	0	0	0	...	0
ab-226324	54.36	5	100	0	5	16	27	1	0	0	0	...	0
ab-691949	165.91	10	38	0	7	2	22	1	0	0	0	...	1
ab-1205176	17.06	4	100	0	3	18	10	0	0	0	0	...	0
ab-1266525	24.12	5	100	0	2	104	26	1	0	0	0	...	0
ab-630376	11.66	2	100	0	2	0	24	0	0	1	0	...	0
ab-454412	23.42	2	100	0	2	53	14	0	0	0	1	...	0
ab-613492	31.20	2	100	0	1	141	33	1	0	0	0	...	1
ab-1469427	4.60	4	13	0	1	1	18	0	0	0	0	...	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...
ab-1508966	2.05	3	100	0	1	12	61	0	0	0	0	...	0
ab-1532220	6.88	2	100	0	1	2	54	0	0	0	0	...	0
ab-1068202	10.05	4	0	0	5	0	13	0	0	0	0	...	0
ab-496758	23.08	2	100	0	2	46	20	0	0	0	0	...	0
ab-2038021	45.62	5	100	1	1	73	39	1	1	0	0	...	0
ab-2041142	48.07	5	100	1	2	8	42	1	0	0	0	...	0
ab-2041416	36.44	5	100	1	1	14	37	1	0	0	0	...	0
ab-2044683	28.12	5	100	1	2	19	42	1	0	0	0	...	0
ab-2235713	5.65	2	100	0	1	28	15	1	1	0	0	...	0

## 2020

Ottenuti i dataset idonei alle analisi econometriche si sono sviluppati tutti i modelli seguendo la metodologia descritta in precedenza. Per ognuno di questi si riportano a seguire i risultati ottenuti (con tutti i parametri di interesse) e le considerazioni annesse.

(se)	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020
mare	-2.841 (3.823)	-2.237 (2.689)										
casa			-12.79*** (3.544)	-4.599 (3.028)								
centro					-13.57*** (3.642)	-5.606 (3.468)						
vacanze							-10.41** (4.020)	-4.429 (4.853)				
vista									-10.67** (3.579)	-4.140 (4.512)		
sea											13.31* (6.628)	-0.353 (5.462)
max_guests	6.215*** (0.861)	7.592*** (0.768)	6.215*** (0.861)	7.565*** (0.764)	6.121*** (0.859)	7.532*** (0.766)	6.179*** (0.861)	7.567*** (0.763)	6.196*** (0.862)	7.570*** (0.765)	6.185*** (0.857)	7.560*** (0.764)
response_rate	0.163*** (0.0407)	0.185*** (0.0294)	0.160*** (0.0406)	0.183*** (0.0293)	0.162*** (0.0406)	0.185*** (0.0293)	0.163*** (0.0408)	0.185*** (0.0293)	0.161*** (0.0408)	0.184*** (0.0294)	0.162*** (0.0406)	0.184*** (0.0292)
superhost	20.43* (9.113)	16.94*** (4.440)	21.23* (9.054)	17.22*** (4.438)	20.65* (9.062)	16.89*** (4.447)	19.97* (9.090)	16.89*** (4.444)	20.11* (9.082)	17.00*** (4.458)	20.24* (9.126)	17.01*** (4.454)
minimum_stay	-0.151 (0.478)	-0.482 (0.430)	-0.154 (0.471)	-0.510 (0.430)	-0.242 (0.470)	-0.519 (0.428)	-0.148 (0.472)	-0.497 (0.428)	-0.148 (0.471)	-0.491 (0.428)	-0.171 (0.470)	-0.494 (0.428)
num_reviews	-0.430** (0.141)	-0.0177 (0.0781)	-0.443** (0.142)	-0.0219 (0.0779)	-0.404** (0.139)	-0.00857 (0.0788)	-0.435** (0.142)	-0.0205 (0.0779)	-0.427** (0.141)	-0.0173 (0.0781)	-0.460** (0.141)	-0.0173 (0.0795)
number_photos	0.988*** (0.230)	0.499*** (0.107)	0.994*** (0.231)	0.499*** (0.107)	0.982*** (0.231)	0.496*** (0.107)	0.993*** (0.231)	0.497*** (0.107)	0.991*** (0.231)	0.498*** (0.107)	0.965*** (0.234)	0.502*** (0.108)
instantbook	17.93*** (3.188)	9.184*** (2.351)	17.28*** (3.269)	9.024*** (2.356)	17.72*** (3.267)	9.050*** (2.368)	17.83*** (3.266)	9.105*** (2.355)	18.16*** (3.253)	9.270*** (2.358)	17.86*** (3.244)	9.270*** (2.363)
_cons	-12.42* (5.832)	-8.056 (4.563)	-11.20 (5.742)	-7.560 (4.567)	-11.24 (5.925)	-7.699 (4.612)	-12.58* (5.900)	-8.039 (4.593)	-12.44* (5.891)	-8.177 (4.596)	-12.87* (5.910)	-8.346 (4.582)
N	2278	1567	2278	1567	2278	1567	2278	1567	2278	1567	2278	1567
R2	0.108	0.187	0.111	0.187	0.110	0.187	0.109	0.187	0.109	0.187	0.110	0.186

I modelli appena riportati rappresentano quelli relativi alle parole più frequenti nei dataset di Gallipoli. Si nota immediatamente come tutte le parole risultino non significative nel 2020: evidentemente non si sono distinti andamenti ben delineati tali da risultare dalle analisi econometriche. Per la restante parte dei dati, comunque, non si evidenziano casi di particolare interesse, anzi alcuni risultano di difficile interpretazione e contro intuitivi come il segno negativo legato a casa, vista e vacanze. La parola “mare” è l'unica non significativa in entrambi gli anni, così da perdere di importanza per la ricerca: verosimilmente il fenomeno è dovuto alla presenza della parola in un numero di annunci tale che non si evidenzi una determinata correlazione con le prestazioni (d'altronde essa è la parola di gran lunga più frequente nel dataset). Mentre la versione in lingua inglese di tale parola (sea) risulta positivamente influente nel 2019 e non più nel 2020.

Relativamente al parametro di rappresentatività si evidenziano valori marcatamente più bassi nel 2019, mentre i valori del 2020 risultano in linea con i precedenti. Il

coefficiente di intercetta risulta nella maggioranza dei casi non significativamente diverso da zero come intuibile: senza l'apporto degli altri termini, che rappresentano tutte componenti fondamentali degli annunci, è ragionevole che le prestazioni tendano a zero e che rese costanti tali altre variabili non si evidenzino variazioni nelle prestazioni importanti. Nei pochi casi in cui tale valore è significativamente diverso da zero, assume segno negativo, facendo pensare che potrebbero essere stati tralasciati dei fattori esterni che impattano negativamente le prestazioni.

In relazione alle variabili di controllo si ritrova che questa volta la maggior parte dei regressori risultano sempre fortemente significativi: anche il tasso di risposta è strettamente significativo e con un apporto importante. Diversamente dal caso di Torino, il minimo numero di giorni di soggiorno non ha impatto significativo e il numero di recensioni, in modo contro intuitivo, assume importanza nel dataset del 2019 con segno negativo.

(se)	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020
house	13.08 (6.961)	-10.08 (5.493)										
B&B			-7.056* (3.091)	-3.645 (4.282)								
spiaggia					-0.228 (4.986)	5.010 (6.234)						
lido							-9.689** (3.192)	-9.235 (5.578)				
beach									11.24 (9.810)	-3.944 (6.298)		
suite											10.62 (6.962)	17.03** (6.452)
max_guests	6.154*** (0.860)	7.593*** (0.765)	6.049*** (0.876)	7.510*** (0.779)	6.164*** (0.862)	7.531*** (0.769)	6.174*** (0.861)	7.582*** (0.765)	6.139*** (0.850)	7.567*** (0.765)	6.248*** (0.862)	7.712*** (0.768)
response_rate	0.166*** (0.0409)	0.183*** (0.0293)	0.163*** (0.0407)	0.181*** (0.0297)	0.163*** (0.0407)	0.183*** (0.0293)	0.161*** (0.0407)	0.184*** (0.0293)	0.163*** (0.0407)	0.185*** (0.0293)	0.165*** (0.0408)	0.188*** (0.0296)
superhost	20.57* (9.106)	16.78*** (4.449)	20.15* (9.106)	17.13*** (4.462)	20.28* (9.117)	16.62*** (4.481)	20.22* (9.086)	16.91*** (4.487)	19.88* (9.172)	17.04*** (4.465)	20.12* (9.103)	16.71*** (4.418)
minimum_stay	-0.154 (0.471)	-0.460 (0.429)	-0.241 (0.473)	-0.530 (0.430)	-0.173 (0.471)	-0.505 (0.426)	-0.140 (0.471)	-0.469 (0.430)	-0.175 (0.471)	-0.478 (0.431)	-0.101 (0.472)	-0.360 (0.432)
num_reviews	-0.439** (0.142)	-0.00916 (0.0790)	-0.436** (0.142)	-0.0205 (0.0778)	-0.432** (0.142)	-0.0162 (0.0784)	-0.436** (0.142)	-0.0189 (0.0780)	-0.425** (0.142)	-0.0191 (0.0782)	-0.425** (0.141)	-0.0102 (0.0774)
number_photos	0.985*** (0.231)	0.505*** (0.107)	0.994*** (0.231)	0.502*** (0.107)	0.991*** (0.231)	0.503*** (0.107)	0.987*** (0.231)	0.498*** (0.107)	0.979*** (0.233)	0.506*** (0.108)	0.984*** (0.232)	0.487*** (0.107)
instantbook	17.59*** (3.252)	9.507*** (2.377)	18.03*** (3.256)	9.339*** (2.365)	18.07*** (3.255)	9.321*** (2.358)	17.87*** (3.262)	9.282*** (2.357)	17.85*** (3.224)	9.326*** (2.362)	17.87*** (3.250)	8.700*** (2.366)
_cons	-13.28* (5.905)	-8.435 (4.593)	-11.69 (6.042)	-7.697 (4.822)	-12.85* (5.927)	-8.305 (4.601)	-12.31* (5.909)	-8.189 (4.585)	-12.74* (5.890)	-8.506 (4.615)	-13.75* (5.898)	-9.883* (4.648)
N	2278	1567	2278	1567	2278	1567	2278	1567	2278	1567	2278	1567
R2	0.109	0.188	0.108	0.187	0.108	0.187	0.109	0.187	0.109	0.187	0.109	0.189

Anche nel secondo gruppo di modelli i risultati non sono copiosi, però questa volta risulta interessante un caso: nuovamente, come per Torino seppur solamente nel 2020, la parola "suite" risulta positivamente correlata alle prestazioni economiche. Di conseguenza ne risulta una conferma relativa a quanto asserito su questa parola specifica: sembrerebbe ancora una volta che tale parola abbia un'accezione positiva,

verosimilmente legata al campo del lusso, che comporta una attrattività della stessa nel titolo. In questa ottica gli host dovrebbero ispirarsi a tale tipologia di impostazione dell'alloggio mettendo in risalto tale aspetto nell'annuncio.

Analogamente, i ragionamenti espressi per i modello della parola "B&B" durante le analisi su Torino, sembrano trovare conferma e possono ritenersi maggiormente valide: nonostante nel 2020 la parola non sembra influenzare le performance, essa appare nel 2019 con correlazione negativa con le performance.

Poco si può asserire sulle restanti parole, mentre, chiaramente, restano simili ai modelli appena precedenti le considerazioni relative alle variabili di controllo, la rappresentatività e il coefficiente di intercetta.

(se)	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020
residence	-2.354 (6.394)	-5.398 (9.075)										
giardino			22.53 (33.73)	15.74 (10.33)								
cuore					-12.01 (6.930)	-0.0879 (9.148)						
dimora							-1.691 (5.052)	11.52 (9.139)				
lungomare									3.179 (8.428)	2.075 (8.142)		
home											56.05 (29.28)	3.208 (8.247)
max_guests	6.166*** (0.860)	7.552*** (0.765)	6.110*** (0.871)	7.532*** (0.766)	6.158*** (0.859)	7.561*** (0.764)	6.158*** (0.860)	7.606*** (0.767)	6.151*** (0.866)	7.555*** (0.768)	6.182*** (0.858)	7.565*** (0.765)
response_rat	0.163*** (0.0408)	0.185*** (0.0294)	0.162*** (0.0406)	0.182*** (0.0292)	0.164*** (0.0407)	0.184*** (0.0294)	0.163*** (0.0407)	0.185*** (0.0294)	0.163*** (0.0407)	0.184*** (0.0293)	0.160*** (0.0404)	0.184*** (0.0293)
superhost	20.20* (9.102)	16.88*** (4.456)	19.88* (9.210)	16.77*** (4.452)	20.12* (9.091)	17.01*** (4.455)	20.24* (9.094)	17.04*** (4.462)	20.23* (9.081)	16.99*** (4.453)	19.59* (9.164)	16.96*** (4.453)
minimum_st	-0.171 (0.472)	-0.486 (0.429)	-0.220 (0.471)	-0.567 (0.428)	-0.189 (0.472)	-0.495 (0.429)	-0.179 (0.473)	-0.454 (0.431)	-0.179 (0.472)	-0.494 (0.428)	-0.0740 (0.474)	-0.492 (0.428)
num_review	-0.432** (0.141)	-0.0183 (0.0781)	-0.427** (0.140)	-0.0138 (0.0782)	-0.430** (0.141)	-0.0180 (0.0780)	-0.432** (0.142)	-0.0144 (0.0782)	-0.432** (0.141)	-0.0178 (0.0781)	-0.435** (0.143)	-0.0189 (0.0779)
number_pho	0.990*** (0.231)	0.501*** (0.107)	0.987*** (0.229)	0.497*** (0.107)	0.993*** (0.231)	0.501*** (0.107)	0.992*** (0.231)	0.497*** (0.107)	0.991*** (0.231)	0.501*** (0.107)	0.954*** (0.224)	0.500*** (0.107)
instantbook	18.09*** (3.257)	9.387*** (2.364)	18.17*** (3.215)	9.462*** (2.351)	18.14*** (3.253)	9.265*** (2.354)	18.06*** (3.257)	9.429*** (2.360)	18.07*** (3.255)	9.265*** (2.359)	17.58*** (3.159)	9.285*** (2.360)
_cons	-12.84* (5.911)	-8.336 (4.595)	-12.63* (5.899)	-8.120 (4.600)	-12.71* (5.907)	-8.353 (4.606)	-12.76* (5.902)	-8.989 (4.657)	-12.81* (5.935)	-8.357 (4.597)	-12.98* (5.906)	-8.380 (4.593)
N	2278	1567	2278	1567	2278	1567	2278	1567	2278	1567	2278	1567
R2	0.108	0.187	0.109	0.188	0.108	0.186	0.108	0.187	0.108	0.186	0.117	0.186

Nel terzo gruppo di modelli, purtroppo, si evince ancora meno: tutte le parole non mostrano significatività e risultano di scarso interesse per le analisi. Non si verificano casi da attenzionare anche per le variabili di controllo.

(se)	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020
terrazza	-13.53** (4.671)	1.965 (8.991)										
terrace			21.44 (12.69)	-2.898 (12.94)								
super					20.68 (16.31)	-3.271 (12.67)						
center							10.35 (22.68)	-15.35* (6.531)				
central									-5.522 (7.423)	-2.433 (7.828)		
relax											-7.963 (7.506)	8.436 (28.15)
max_guests	6.156*** (0.859)	7.561*** (0.765)	6.160*** (0.859)	7.567*** (0.765)	6.168*** (0.860)	7.558*** (0.765)	6.170*** (0.859)	7.554*** (0.765)	6.160*** (0.860)	7.561*** (0.765)	6.156*** (0.860)	7.564*** (0.765)
response_rate	0.163*** (0.0407)	0.184*** (0.0293)	0.164*** (0.0407)	0.184*** (0.0293)	0.163*** (0.0407)	0.184*** (0.0293)	0.163*** (0.0407)	0.185*** (0.0294)	0.163*** (0.0407)	0.184*** (0.0293)	0.163*** (0.0407)	0.183*** (0.0293)
superhost	20.10* (9.093)	17.01*** (4.453)	18.88* (9.214)	17.16*** (4.360)	20.12* (9.102)	16.98*** (4.457)	20.12* (9.125)	16.63*** (4.478)	20.19* (9.099)	16.99*** (4.455)	20.28* (9.092)	16.99*** (4.460)
minimum_stay	-0.176 (0.471)	-0.495 (0.428)	-0.151 (0.472)	-0.500 (0.427)	-0.166 (0.471)	-0.498 (0.427)	-0.172 (0.471)	-0.503 (0.428)	-0.181 (0.472)	-0.496 (0.428)	-0.170 (0.472)	-0.496 (0.428)
num_reviews	-0.431** (0.141)	-0.0178 (0.0781)	-0.444** (0.142)	-0.0170 (0.0789)	-0.432** (0.142)	-0.0179 (0.0780)	-0.437** (0.142)	-0.00763 (0.0791)	-0.432** (0.141)	-0.0180 (0.0780)	-0.432** (0.141)	-0.0182 (0.0781)
number_photos	0.995*** (0.232)	0.501*** (0.107)	0.982*** (0.232)	0.503*** (0.107)	0.992*** (0.231)	0.500*** (0.107)	0.989*** (0.231)	0.506*** (0.107)	0.990*** (0.231)	0.501*** (0.107)	0.990*** (0.231)	0.500*** (0.105)
instantbook	17.98*** (3.260)	9.290*** (2.365)	17.91*** (3.256)	9.293*** (2.365)	18.01*** (3.255)	9.282*** (2.358)	18.08*** (3.257)	9.238*** (2.358)	18.07*** (3.255)	9.251*** (2.359)	18.04*** (3.256)	9.288*** (2.365)
_cons	-12.68* (5.904)	-8.389 (4.596)	-12.88* (5.910)	-8.396 (4.600)	-12.96* (5.914)	-8.312 (4.597)	-12.90* (5.911)	-8.358 (4.595)	-12.73* (5.925)	-8.324 (4.596)	-12.72* (5.919)	-8.348 (4.589)
N	2278	1567	2278	1567	2278	1567	2278	1567	2278	1567	2278	1567
R2	0.108	0.186	0.109	0.186	0.108	0.186	0.108	0.187	0.108	0.186	0.108	0.187

Il quarto e ultimo gruppo di modelli, tuttavia, riporta uno dei ritrovamenti più interessanti delle ricerche su Gallipoli: la parola “center” sembrerebbe essere andata incontro a una marcata riduzione di attrattività tra il 2019 e il 2020: la parola non influente nel 2019 (risulta non significativa), risulta invece fortemente impattante in senso negativo nell’anno successivo.

Tale risultato è in linea con quanto accaduto durante il covid, considerato che le persone si siano dimostrate disinteressate ai luoghi centrali e affollati. Questa evidenza conferma quanto già visto nelle analisi relative al dataset della città di Torino e inizia a delineare uno scenario preciso: sembrerebbe che gli host che non hanno reagito in modo adeguato alle necessità dettate dal forte shock esterno dovuto al covid siano andati incontro a un peggioramento delle prestazioni economiche dei propri alloggi.

Per la restante parte dei dati non si evidenziano casi di particolare interesse, mentre si ritrovano gli stessi andamenti per tutti gli altri fattori non semantici della regressione.

## Analisi econometriche Erice

Alla luce delle parole evidenziate per Erice si definiscono le seguenti variabili semantiche: casa, mare, beach, vacanze, residence, sea, vista, B&B, house, antico, garden, home, suite, spiaggia, relax, giardino, vetta, beachside, dimora, terrazza, privata, cuore, super, terrace. Di seguito un estratto del foglio di calcolo relativo alla costruzione delle variabili:

Property ID	Listing Title	Casa	House	Home	Mare	Sea	...	beach
ab-43992	Flat in front of the beach!!!	0	0	0	0	0	...	1
ab-630376	Casa Vista Mare_ Spiaggia Trapani	1	0	0	1	0	...	0
ab-613492	Bruno Homes for Erice	0	0	1	0	0	...	0
ab-538365	Erice-Trapani charming delux home up to 10 people	0	0	1	0	0	...	0
ab-496758	Quiet apartment with sunny terrace.	0	0	0	0	0	...	0
ab-454412	Erice-Trapani lovely house	0	1	0	0	0	...	0
ab-1030353	La Casa della Danza e della Musica	1	0	0	0	0	...	0
ab-226324	Open Space San Cusumano in Villa ideal position	0	0	0	0	0	...	0
...	...	...	...	...	...	...	...	...
ab-37451236	Posizione strategica in zona residenziale.	0	0	0	0	0	...	0
ab-38428068	Casa Francisco on the Sea	1	0	0	0	1	...	0
ab-40876712	"Ginestra" medieval apartment on the upper floor	0	0	0	0	0	...	0
ab-40945925	LEO APPARTAMENTI con vista panoramica	0	0	0	0	0	...	0
ab-41711455	Casa vacanze Le Terrazze	1	0	0	0	0	...	0
295	<b>Somma</b>	56	13	10	35	11	...	12
	<b>Frequenza percentuale sul totale</b>	19%	4%	3%	12%	4%	...	4%

2019

Property ID	Listing Title	Casa	House	Home	Mare	Sea	...	beach
ab-43992	Flat in front of the beach!!!	0	0	0	0	0	...	1
ab-226324	Open Space San Cusumano in Villa ideal position	0	0	0	0	0	...	0
ab-1205176	apartment with garden	0	0	0	0	0	...	0
ab-1266525	Appartamentino Pomelia	0	0	0	0	0	...	0
ab-630376	Casa Vista Mare_ Spiaggia Trapani	1	0	0	1	0	...	0
ab-454412	Erice-Trapani lovely house	0	1	0	0	0	...	0
ab-613492	Bruno Homes for Erice	0	0	1	0	0	...	0
ab-1469427	TORREBIANCA HOMEHOLIDAYS. ELEGANCE!	0	0	1	0	0	...	0
ab-1508966	B&B Il Tipico!	0	0	0	0	0	...	0
...	...	...	...	...	...	...	...	...
ab-44591681	Popi & Silver Apartment, a due passi dal mare!	0	0	0	1	0	...	0
ab-44603786	La tua stanza che si affaccia sul mare	0	0	0	1	0	...	0
ab-44608046	Inviting Holiday Home in Erice with Garden	0	0	1	0	0	...	0
ab-45048904	Domus Sicily - Erice Suite	0	0	0	0	0	...	0
ab-45109404	LEO APPARTAMENTI 2 con vista panoramica	0	0	0	0	0	...	0
ab-45149348	Da Angela	0	0	0	0	0	...	0
ab-45398192	B&B Bella vista :Stanza privata per 2 persone	0	0	0	0	0	...	0
229	<b>Somma</b>	36	9	6	28	11	...	15
	<b>Frequenza percentuale sul totale</b>	16%	4%	3%	12%	5%	...	7%

2020

Una volta costruite le variabili semantiche si genera per ciascun anno il dataset da analizzare attraverso il software statistico STATA. Tali dataset contengono sia il codice identificativo dell'appartamento che le variabili, semantiche e di controllo, ad esso connesse:

id	performance	max_guests	response_rate	superhost	minimum_stay	num_reviews	number_photos	instantbook	cambio_titolo	casa	house	...	home
ab-153486	225.08	6	100	1	4	70	16	1	0	0	0	...	0
ab-112767	16.37	5	100	0	3	0	4	0	0	0	0	...	0
ab-879462	147.54	10	0	0	7	7	22	0	0	0	1	...	0
ab-1074207	55.27	4	100	1	2	86	35	0	0	0	0	...	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
ab-1083100	46.31	5	0	0	7	0	10	0	0	0	1	...	0
ab-394609	32.16	3	100	0	3	0	6	0	0	0	0	...	0
ab-1224506	38.37	6	100	0	2	55	27	1	0	1	0	...	0
ab-1267903	35.90	5	100	0	2	16	34	0	0	1	0	...	0
ab-1263739	75.31	6	78	0	7	0	7	1	0	0	0	...	0

### 2019

id	performance	max_guests	response_rate	superhost	minimum_stay	num_reviews	number_photos	instantbook	cambio_titolo	casa	house	...	home
ab-87951	28.46	4	100	0	2	5	8	0	0	0	0	...	0
ab-130827	215.35	6	0	0	5	0	15	0	0	0	0	...	0
ab-1224506	36.22	6	100	0	2	55	27	0	0	1	0	...	0
ab-1263098	58.94	6	0	0	4	15	20	0	0	0	0	...	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
ab-1267903	53.25	5	100	0	2	17	34	0	0	1	0	...	0
ab-1012895	74.61	5	100	0	5	7	23	0	0	0	0	...	0
ab-879462	33.81	10	0	0	7	7	22	0	0	0	1	...	0
ab-567872	21.80	4	60	0	6	21	15	0	0	0	0	...	0
ab-1674416	27.77	6	100	0	3	1	8	0	0	1	0	...	0

### 2020

Ottenuti i dataset idonei alle analisi econometriche si sono sviluppati tutti i modelli seguendo la metodologia descritta in precedenza. Per ognuno di questi si riportano a seguire i risultati ottenuti (con tutti i parametri di interesse) e le considerazioni annesse.

(se)	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020
casa	-11.60 (7.475)	-6.792 (4.706)										
mare			1.807 (6.359)	2.055 (4.657)								
beach					2.467 (5.367)	1.822 (6.241)						
vacanze							-13.23 (7.601)	0.247 (8.804)				
residence									25.72*** (7.635)	15.99* (6.176)		
sea											1.594 (9.872)	13.98 (10.79)
max_guests	5.155* (2.088)	2.210* (1.051)	4.951* (2.072)	2.080* (1.007)	4.938* (2.082)	2.068* (1.012)	5.088* (2.112)	2.072* (1.037)	4.991* (2.076)	2.078* (1.005)	4.949* (2.071)	2.081* (0.993)
response_rate	0.104 (0.0627)	0.0327 (0.0449)	0.125* (0.0617)	0.0454 (0.0444)	0.125* (0.0617)	0.0422 (0.0439)	0.117 (0.0596)	0.0436 (0.0443)	0.135* (0.0618)	0.0358 (0.0441)	0.126* (0.0615)	0.0526 (0.0441)
superhost	4.073 (4.797)	10.91* (4.230)	4.791 (4.887)	11.67** (4.291)	4.547 (4.842)	11.59** (4.256)	4.921 (4.814)	11.62** (4.265)	-1.442 (5.150)	6.284 (5.430)	4.750 (4.832)	11.45** (4.305)
minimum_stay	8.474* (3.761)	5.154* (2.060)	8.281* (3.740)	5.043* (2.076)	8.298* (3.728)	5.076* (2.070)	8.405* (3.717)	5.051* (2.073)	8.571* (3.751)	5.351* (2.094)	8.266* (3.774)	5.101* (2.070)
num_reviews	-0.0331 (0.0668)	-0.0153 (0.0451)	-0.0401 (0.0626)	-0.0183 (0.0441)	-0.0418 (0.0616)	-0.0185 (0.0454)	-0.0507 (0.0617)	-0.0202 (0.0439)	-0.0275 (0.0601)	0.00000553 (0.0452)	-0.0431 (0.0608)	-0.0204 (0.0419)
number_photos	0.0100 (0.159)	0.130 (0.0954)	-0.0106 (0.164)	0.126 (0.0966)	-0.0122 (0.162)	0.118 (0.0994)	0.0120 (0.162)	0.121 (0.0968)	0.0302 (0.165)	0.156 (0.0993)	-0.0131 (0.162)	0.0941 (0.0998)
instantbook	20.02** (7.035)	6.246 (3.177)	21.18** (7.075)	6.262 (3.238)	21.33** (6.861)	6.350* (3.220)	20.37** (6.689)	6.390* (3.217)	21.01** (6.851)	5.939 (3.214)	21.28** (6.956)	6.657* (3.114)
_cons	-26.15 (15.37)	0.561 (5.831)	-29.20 (15.44)	-1.080 (6.033)	-29.05 (15.66)	-0.565 (6.083)	-28.25 (15.44)	-0.620 (6.068)	-31.52* (15.93)	-1.396 (6.120)	-28.99 (15.69)	-1.718 (6.061)
N	295	229	295	229	295	229	295	229	295	229	295	229
R2	0.188	0.171	0.178	0.162	0.178	0.162	0.183	0.161	0.187	0.177	0.178	0.176

I modelli appena riportati rappresentano quelli relativi alle parole più frequenti nei dataset di Erice. Tutte le parole, tranne una, risultano non significative in entrambi gli anni: evidentemente non si sono distinti andamenti ben delineati, tali da risultare dalle analisi econometriche.

Tuttavia, in questo primo gruppo di modelli risalta un modello di grande interesse, relativo alla parola “residence”: si nota che la variabile ad essa connessa risulta significativa con impatto positivo in entrambe le annualità, ma si evidenzia un calo di questa influenza. Considerato che tale parola indica un luogo di aggregazione, caratterizzato da numerosi alloggi e persone, anche questo ritrovamento risulta perfettamente in linea con l’andamento, precedentemente sottolineato, dettato dalla pandemia.

Nuovamente si ottiene una conferma che le persone si siano dimostrate disinteressate ai luoghi centrali e affollati. Anche questa evidenza si aggiunge all’ormai ricco quadro di ritrovamenti ottenuto nelle analisi relative ai dataset delle città di Torino e Gallipoli. Lo scenario delineato diviene ancora più chiaro: si può asserire con sempre più convinzione che gli host che non hanno reagito in modo adeguato alle necessità

dettate dal forte shock esterno dovuto al covid sono andati incontro a un peggioramento delle prestazioni economiche dei propri alloggi.

Per la restante parte delle parole, non sono presenti ulteriori casi di particolare interesse. Per questi dataset relativi ad Erice, relativamente al parametro di rappresentatività si notano valori molto stabili e simili sia nel 2019 che nel 2020, senza che ci sia un anno che mostri nettamente dei valori superiori come per i precedenti dataset relativi alle altre città. Ad ogni modo, gli ordini di grandezza sono in linea con i precedenti e i valori che assume l'R-squared risultano, in generale, poco elevati. Il coefficiente di intercetta risulta non significativamente diverso da zero come nel caso di Torino.

Per quanto riguarda i regressori di controllo, questa volta si notano delle differenze più marcate dai casi precedenti: soltanto due variabili risultano sempre significative: quelle relative al massimo numero di ospiti e al minimo numero di giorni di soggiorno. Il risultato relativo a quest'ultima variabile risulta questa volta contro intuitivo, riportando segno positivo: si potrebbe tradurre tale risultato in una preferenza degli ospiti per soggiorni più lunghi, cosicché rendere questo parametro più stringente rappresenta da parte dell'host un'azione che non impatta negativamente le performance, ma coadiuva il risultato economico. Probabilmente la presenza di prenotazioni più brevi, sparse, potrebbe impedire soggiorni più lunghi ai quali la maggior parte dei turisti sono interessati ad Erice.

Anche la possibilità di prenotazione immediata risulta essere importante, seppur con una sola eccezione, mentre il tasso di risposta dell'host e la qualifica di superhost risultano positivamente impattanti in modo saltuario. Poco si può asserire invece relativamente al numero di recensioni e di foto, sempre non significative.

(se)	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020
vista	3.836 (12.59)	-0.300 (5.582)										
B&B			-5.285 (4.517)	-9.935* (4.139)								
house					-5.541 (4.406)	-4.044 (4.202)						
antico							-7.905 (5.268)	-3.088 (4.055)				
garden									-5.628 (4.867)	-13.89* (6.951)		
home											-11.87 (7.531)	6.980 (17.18)
max_guests	4.952* (2.073)	2.075* (1.011)	4.904* (2.083)	1.999 (1.018)	4.925* (2.070)	2.071* (1.013)	4.932* (2.075)	2.088* (1.014)	4.963* (2.083)	2.333* (1.064)	4.982* (2.082)	1.987* (0.943)
response_rate	0.128* (0.0623)	0.0432 (0.0437)	0.128* (0.0613)	0.0507 (0.0444)	0.129* (0.0622)	0.0438 (0.0440)	0.128* (0.0616)	0.0433 (0.0439)	0.128* (0.0624)	0.0492 (0.0439)	0.111 (0.0621)	0.0472 (0.0416)
superhost	4.881 (4.928)	11.61** (4.267)	4.470 (4.852)	10.94* (4.331)	4.436 (4.892)	11.71** (4.265)	4.740 (4.844)	11.86** (4.330)	4.620 (4.877)	11.13* (4.289)	4.438 (4.888)	11.79** (4.300)
minimum_stay	8.309* (3.719)	5.049* (2.070)	8.199* (3.773)	4.797* (2.121)	8.243* (3.729)	5.009* (2.080)	8.293* (3.729)	5.004* (2.082)	8.274* (3.725)	4.846* (2.069)	8.392* (3.740)	4.912* (1.968)
num_reviews	-0.0412 (0.0611)	-0.0205 (0.0444)	-0.0467 (0.0617)	-0.0287 (0.0441)	-0.0403 (0.0611)	-0.0190 (0.0445)	-0.0377 (0.0615)	-0.0182 (0.0445)	-0.0430 (0.0614)	-0.0264 (0.0436)	-0.0301 (0.0632)	-0.0252 (0.0464)
number_photos	-0.0140 (0.162)	0.122 (0.0975)	-0.0100 (0.163)	0.130 (0.0954)	-0.0155 (0.162)	0.122 (0.0970)	-0.0146 (0.162)	0.121 (0.0970)	-0.0147 (0.163)	0.127 (0.0959)	-0.0131 (0.161)	0.122 (0.0963)
instantbook	21.11** (7.062)	6.404* (3.248)	21.33** (6.868)	6.521* (3.198)	21.50** (6.915)	6.138 (3.305)	21.18** (6.870)	6.270 (3.253)	21.27** (6.852)	6.546* (3.179)	20.99** (6.804)	6.426* (3.246)
_cons	-29.29 (15.65)	-0.575 (6.038)	-28.56 (15.86)	0.0560 (6.144)	-28.88 (15.68)	-0.262 (6.123)	-28.88 (15.69)	-0.412 (6.095)	-28.99 (15.69)	-1.343 (6.095)	-27.71 (15.40)	-0.453 (5.953)
N	295	229	295	229	295	229	295	229	295	229	295	229
R2	0.178	0.161	0.178	0.168	0.178	0.162	0.178	0.162	0.178	0.170	0.180	0.163

Nel secondo gruppo di modelli i risultati sono esigui, però è presente una conferma di quanto visto in precedenza: nuovamente, come per Torino e Gallipoli, solamente nel 2020, la parola “B&B” compare negativamente correlata alle prestazioni economiche. Sulle restanti parole non è possibile sbilanciarsi, considerata la non significatività, mentre restano invariate le considerazioni relative alle variabili di controllo, la rappresentatività e il coefficiente di intercetta.

(se)	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020
suite	8.137 (11.48)	-7.646 (7.728)										
spiaggia			1.400 (10.73)	7.235 (5.691)								
relax					7.928 (11.44)	-13.26*** (2.431)						
giardino							-27.65** (10.53)	1.703 (15.03)				
vetta									-13.08 (12.28)	-8.665 (8.425)		
beachside											0 (.)	-10.09* (4.501)
max_guests	4.967* (2.098)	2.016* (1.016)	4.949* (2.075)	2.102* (1.012)	4.963* (2.076)	2.065* (1.008)	5.094* (2.104)	2.063* (1.028)	4.991* (2.091)	2.095* (1.014)	4.947* (2.072)	2.040* (1.013)
response_rate	0.125* (0.0614)	0.0451 (0.0440)	0.126* (0.0616)	0.0415 (0.0440)	0.124* (0.0617)	0.0391 (0.0439)	0.122* (0.0615)	0.0441 (0.0439)	0.126* (0.0614)	0.0456 (0.0437)	0.126* (0.0614)	0.0444 (0.0440)
superhost	4.789 (4.833)	11.43** (4.276)	4.773 (4.860)	11.77** (4.271)	4.791 (4.839)	11.43** (4.271)	6.131 (4.494)	11.54** (4.323)	4.412 (4.920)	11.40** (4.278)	4.752 (4.835)	11.49** (4.275)
minimum_stay	8.305* (3.748)	5.022* (2.074)	8.290* (3.733)	5.035* (2.072)	8.270* (3.733)	5.062* (2.070)	8.531* (3.758)	5.036* (2.107)	8.357* (3.751)	5.076* (2.075)	8.286* (3.724)	5.015* (2.075)
num_reviews	-0.0418 (0.0609)	-0.0242 (0.0440)	-0.0424 (0.0613)	-0.0185 (0.0443)	-0.0427 (0.0612)	-0.0230 (0.0440)	-0.0399 (0.0598)	-0.0201 (0.0441)	-0.0391 (0.0611)	-0.0186 (0.0445)	-0.0427 (0.0611)	-0.0236 (0.0439)
number_photos	-0.0176 (0.167)	0.138 (0.101)	-0.0131 (0.162)	0.120 (0.0969)	-0.00971 (0.163)	0.125 (0.0964)	0.000324 (0.161)	0.120 (0.0962)	-0.0169 (0.163)	0.116 (0.0977)	-0.0129 (0.162)	0.138 (0.101)
instantbook	21.37** (6.933)	6.211 (3.238)	21.30** (6.881)	6.375* (3.210)	21.46** (6.880)	6.403* (3.205)	20.98** (6.801)	6.336 (3.239)	21.52** (6.937)	6.599* (3.228)	21.31** (6.859)	6.140 (3.257)
_cons	-29.07 (15.79)	-0.436 (6.068)	-29.03 (15.71)	-0.655 (6.063)	-29.13 (15.69)	0.000392 (6.021)	-29.73 (15.78)	-0.543 (6.198)	-29.11 (15.72)	-0.790 (6.080)	-28.98 (15.67)	-0.481 (6.067)
N	295	229	295	229	295	229	295	229	295	229	295	229
R2	0.178	0.163	0.177	0.163	0.178	0.166	0.184	0.161	0.179	0.163	0.180	0.163

Nel terzo gruppo di modelli, purtroppo, si evince ancora meno: la maggioranza delle parole non mostrano significatività e risultano di scarso interesse per le analisi. Le poche significative riportano casi di difficile lettura e interpretazione per via del segno, mentre non si verificano casi da attenzionare anche per le variabili di controllo.

(se)	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020
dimora	-18.93*	3.174										
	(9.439)	(7.379)										
terrazza			-14.04	0.821								
			(7.697)	(8.640)								
privata					59.17	-9.260						
					(42.19)	(5.120)						
cuore							9.531*	18.01				
							(4.079)	(24.92)				
super									-3.769	35.83***		
									(8.547)	(5.020)		
terrace											-2.393	3.692
											(12.46)	(3.661)
max_guests	4.996*	2.065*	4.970*	2.077*	5.132*	2.009*	4.939*	2.030*	4.942*	2.118*	4.944*	2.087*
	(2.087)	(1.017)	(2.083)	(1.011)	(2.080)	(1.019)	(2.075)	(1.004)	(2.074)	(1.014)	(2.076)	(1.014)
response_rate	0.130*	0.0431	0.127*	0.0434	0.140*	0.0441	0.125*	0.0472	0.125*	0.0521	0.126*	0.0433
	(0.0622)	(0.0438)	(0.0618)	(0.0439)	(0.0551)	(0.0438)	(0.0616)	(0.0444)	(0.0616)	(0.0430)	(0.0615)	(0.0439)
superhost	4.275	11.69**	4.999	11.61**	5.673	11.32**	4.776	11.73**	4.732	11.78**	4.716	11.65**
	(4.919)	(4.278)	(4.792)	(4.277)	(4.632)	(4.286)	(4.843)	(4.266)	(4.849)	(4.265)	(4.851)	(4.276)
minimum_stay	8.245*	5.055*	8.277*	5.053*	8.334*	5.032*	8.311*	5.125*	8.291*	4.945*	8.283*	5.047*
	(3.717)	(2.070)	(3.726)	(2.070)	(3.719)	(2.069)	(3.734)	(2.062)	(3.733)	(2.078)	(3.731)	(2.070)
num_reviews	-0.0476	-0.0196	-0.0443	-0.0202	-0.0580	-0.0216	-0.0423	-0.0191	-0.0429	-0.0186	-0.0421	-0.0213
	(0.0620)	(0.0443)	(0.0617)	(0.0442)	(0.0565)	(0.0442)	(0.0613)	(0.0441)	(0.0612)	(0.0442)	(0.0613)	(0.0447)
number_photos	-0.00687	0.120	-0.0114	0.122	-0.0581	0.138	-0.0121	0.122	-0.0137	0.132	-0.0131	0.122
	(0.162)	(0.0969)	(0.162)	(0.0970)	(0.133)	(0.101)	(0.162)	(0.0969)	(0.162)	(0.0967)	(0.162)	(0.0969)
instantbook	21.80**	6.291	21.50**	6.366	21.48**	6.659*	21.42**	6.382*	21.32**	6.388*	21.32**	6.429*
	(6.986)	(3.245)	(6.933)	(3.231)	(6.866)	(3.245)	(6.879)	(3.207)	(6.874)	(3.178)	(6.870)	(3.245)
_cons	-29.46	-0.500	-29.20	-0.609	-30.57	-0.662	-29.07	-1.099	-28.88	-1.898	-28.96	-0.664
	(15.79)	(6.121)	(15.76)	(6.064)	(15.73)	(6.045)	(15.70)	(6.084)	(15.66)	(6.004)	(15.70)	(6.072)
N	295	229	295	229	295	229	295	229	295	229	295	229
R2	0.180	0.161	0.178	0.161	0.190	0.163	0.178	0.166	0.178	0.180	0.178	0.161

Il quarto e ultimo gruppo di modelli, invece, mostra uno dei risultati di maggiore importanza delle ricerche su Erice: la parola “cuore” sembrerebbe essere andata incontro a una marcata riduzione di attrattività tra i due anni in esame: la stessa risulta positivamente impattante nel 2019 e risulta non significativa nell’anno successivo.

Tale evidenza è l’ennesima conferma di quanto delineato nei modelli precedenti di Torino e Gallipoli, relativamente alle parole legate alla centralità e al disinteressamento delle persone per i luoghi centrali e affollati in conseguenza agli effetti del covid. Gli host che non hanno reagito in maniera adeguata allo shock esterno dovuto alla pandemia sono andati incontro a un peggioramento delle prestazioni economiche dei propri alloggi anche a Erice.

Per la restante parte dei dati non si evidenziano casi di particolare interesse, mentre si ritrovano gli stessi andamenti per tutti gli altri fattori non semantici della regressione.

Una volta definite le variabili indipendenti relative alla parte semantica e agli attributi di controllo è possibile effettuare i confronti tra le differenti città, che sono già in parte stati messi in luce durante le analisi per le singole città.

Il confronto più importante e ricco di evidenze riguarda le parole legate alla centralità dell'alloggio in ottica di periodo pandemico: in tutti e tre i dataset delle città si nota che gli annunci nei quali compaiono determinate parole, aventi la valenza appena sottolineata, risultano effettivamente essere andati incontro a una riduzione statisticamente provata di attrattività tra il 2019 al 2020.

Tale fenomeno avviene per le parole “central” e “cuore” a Torino, per “center” a Gallipoli e per “residence” e “cuore” a Erice: il confronto tra le tre città diverse per svariati fattori aiuta a generalizzare gli effetti che il covid ha avuto sul settore. Non si parla di un caso isolato, bensì si ritrovano delle tracce significative in ogni città analizzata.

Relativamente alle altre parole risultate di interesse si evidenzia che la parola suite, a conferma di quanto presentato, ha mostrato gli andamenti sottolineati in precedenza sia nella città di Torino che in quella di Gallipoli. La parola “terrace” invece non ha riscontro nei dataset esterni a quelli di Torino ma risulta comunque di interesse. La parola “B&B”, invece, conferma i propri andamenti, seppur non in tutti gli anni, per tutte le città analizzate: la lettura di questo risultato è analoga a quella della parola suite ma con segno opposto.

Passando ai confronti delle variabili di controllo solamente un regressore mostra sempre lo stesso andamento, con altissimi livelli di significatività e per tutte le città: sembrerebbe che il massimo numero di ospiti è un aspetto fondamentale in termini di risultati economici. Effettivamente i prezzi salgono nettamente in funzione del numero di persone che soggiornano ed è un risultato ragionevole, parlando di ordini di grandezza comuni (da un solo ospite a salire) e non addentrandosi nelle casistiche puntuali, pensare che un numero di ospiti maggiore possa tradursi in maggiori

prestazioni economiche. Altri fattori che dal confronto escono comunque forti sono la qualifica di superhost, il numero di foto pubblicate nell'annuncio e la possibilità di prenotazione istantanea. Tali aspetti hanno generalmente un apporto, che assume segno positivo e pertanto da tenere in considerazione nei panni di host. Meno allineato il confronto per quanto riguarda la variabile del minimo numero di giorni di soggiorno: sembrerebbe che in base alle località ci possa essere un interesse diverso per la flessibilità dei soggiorni e la lunga durata degli stessi.

Ulteriori analisi, nelle quale si è direttamente effettuato un confronto tra le città, hanno riguardato i modelli di regressione relativi al cambio del titolo: come anticipato, nella prospettiva semantica è stato analizzato pure se in generale i cambiamenti dei titoli abbiano portato a un effetto preciso sulle performance degli annunci. Di seguito si riportano tali modelli a confronto.

	TORINO	GALLIPOLI	ERICE
cambio_titolo	0.712	7.544*	-4.596
	(0.996)	(3.710)	(3.141)
max_guests	4.700***	7.534***	2.065*
	(0.519)	(0.768)	(1.012)
response_rate	-0.00194	0.178***	0.0457
	(0.0158)	(0.0295)	(0.0440)
superhost	4.210***	17.20***	11.68**
	(0.675)	(4.436)	(4.269)
minimum_stay	-0.0496*	-0.449	4.998*
	(0.0226)	(0.427)	(2.072)
num_reviews	0.00484	-0.0153	-0.0205
	(0.00454)	(0.0779)	(0.0442)
number_photos	0.167***	0.488***	0.125
	(0.0307)	(0.106)	(0.0973)
instantbook	5.522***	9.001***	6.604*
	(0.584)	(2.370)	(3.225)
_cons	1.982	-8.275	-0.640
	(1.325)	(4.605)	(6.055)
N	4829	1567	229
R2	0.193	0.1924	0.163

Dai risultati, quindi, non è possibile affermare che ci sia una tendenza definita ma sembrerebbe che gli host di Gallipoli che hanno cambiato titolo si siano mossi nella

giusta direzione: effettivamente osservando i singoli casi si nota l'aggiunta della parola suite nei nuovi titoli oltre che sparuti casi di un'attenzione a livello semantico al covid (poche parole indicanti centralità e rare eccezioni di parole covid-related). Tale risultato è in linea con i ritrovamenti delle analisi, secondo cui attuare azioni del genere comporti un miglioramento delle prestazioni economiche. Inoltre, gli andamenti delle variabili di controllo vengono confermati anche in questi modelli.

Anche per le analisi econometriche, nel periodo durante il quale si sono svolte le ricerche, sono state testate più metodologie, portando a determinati procedimenti che ritenuti validi in partenza si sono dimostrati inadeguati per specifici fattori emersi durante il proseguo delle analisi, ma di grande valore per mettere in risalto le criticità che le metodologie finora presentate riescono a superare con successo e altre evidenze relative ai dataset considerati. In questa sezione si riportano i principali riscontri delle ulteriori analisi che si sono svolte a livello econometrico.

Il primo approccio è di tipo top down e prevede la creazione di un modello che metta in relazione sia le parole che le caratteristiche delle abitazioni e degli host con le prestazioni economiche. Tale modello ha una formulazione del tipo seguente:

$$ann\_rev = c + \sum(b_i \times word_i) + \sum(a_i \times X_i) + e$$

Il metodo consiste nell'inserire nel modello tutte le variabili definite, eliminando, passo dopo passo, le variabili aventi p-value più elevato e di conseguenza non significative, osservando attentamente il variare dell'R-squared e dei parametri di regressione. Ad ogni iterazione si giunge a un modello migliore, fino al risultato che si ritiene adeguato.

Difatti, questa metodologia non è adeguata e mette in luce delle criticità superate attraverso il metodo utilizzato nelle analisi precedenti. La principale problematica riguarda la multicollinearità delle variabili semantiche, che possono effettivamente presentare delle correlazioni significative tra di loro e generare delle distorsioni nei dati. Inoltre, attraverso un procedimento del genere si va incontro a una continua perdita di informazione per ogni passaggio in cui viene eliminato un regressore: non è scontato, come dimostrato durante le analisi, che una variabile non significativa sia un risultato di scarso interesse.

Si sottolinea, quindi, la bontà della metodologia alla base delle analisi effettuate in riferimento alla solidità nei confronti di tali problematiche.

## **PARTE 3 - FASE FINALE**

## Conclusioni

### Ritrovamenti

I ritrovamenti più significativi delle analisi svolte si riportano in questa sezione, dopo aver nuovamente chiarito delle premesse fondamentali: ci si focalizza sui risultati relativi al livello semantico delle analisi svolte. In tal senso, è necessario sottolineare che le analisi osservano degli aspetti di dimensione parziale rispetto all'intera cornice del fenomeno della sharing economy e degli scambi su piattaforma digitale di AirBnb definita nel corso della tesi.

Guardando l'insieme con uno sguardo ampio, potrebbero apparire trascurabili, soprattutto se confrontati con tutti gli attributi modellabili, o meno, che caratterizzano il settore. La dimostrazione di quanto detto è banalmente l'ordine di grandezza degli R-squared ritrovati, non solo durante le presenti analisi, ma in quelle dell'intero settore: si tratta di fenomeni sociali complessi, difficili da modellare e in certi aspetti non modellabili. Nonostante ciò, è possibile modellarne delle parti ed ottenere dei risultati di dettaglio significativi. L'ottica è proprio questa: i ritrovamenti che si riportano a seguire possono essere trascurabili se si guarda agli aspetti che maggiormente influenzano le prestazioni economiche degli annunci, ma rappresentano delle evidenze rilevanti se calate nell'ambito semantico, guardando ad aspetti secondari, ma importanti, degli scambi economici su piattaforma. Tali risultati rappresentano una risposta valida alle domande di ricerca poste all'inizio del lavoro di tesi.

I primi ritrovamenti di forte interesse si sono riscontrati già durante la fase di analisi semantica. In prima istanza, bisogna sottolineare che le parole utilizzate di anno in anno si mantengono, andando incontro a poche variazioni. In aggiunta, tale fenomeno non si presenta soltanto a livello di singola città, ma le parole più frequenti ricorrono in tutti i dataset presi in esame: le stesse parole compaiono con simile frequenza percentuale nelle diverse città, evidenziando un lessico e un vocabolario comune che caratterizza la pubblicazione degli annunci su piattaforma digitale.

In seconda istanza, in relazione agli esigui cambiamenti semantici appena sottolineati, si riscontra una inerzia degli host diffusa, per lo meno a livello semantico, nella risposta

al cambiamento delle condizioni esterne in seguito a un forte shock come il diffondersi del virus e delle misure di contenimento correlate.

Inoltre, grazie agli studi semantici, è stato possibile effettivamente mettere in luce il forte bias di positività che caratterizza il settore e che si era già nettamente evidenziato durante lo studio della letteratura relativo ad altre località fuori dal territorio italiano.

Passando ai riscontri econometrici, il più importante riguarda l'impatto della pandemia sul settore. Infatti, risulta confermato che i turisti prediligono luoghi non centrali e meno affollati a causa del covid. Difatti, varie parole legate alla centralità dell'alloggio, in tutti e tre i dataset delle città, mostrano tra il 2019 e il 2020 una perdita di attrattività statisticamente verificata. Di conseguenza gli host che non hanno risposto in modo adeguato al forte shock esterno, dettato dal covid, hanno avuto un peggioramento delle prestazioni economiche dei propri annunci.

Il fatto che tale fenomeno si presenta in tutte le città analizzate, diverse per svariati fattori, rappresenta una forte conferma e permette di affermare quanto detto con sicurezza. Gli host, in seguito al delinearsi della pandemia, avrebbero dovuto limitare i vocaboli legati alla centralità dell'abitazione o alla vicinanza a punti di interesse di grande aggregazione sociale.

In secondo luogo, sono stati ritrovati interessanti anche casi legati a singole parole aventi correlazioni significative e definite, in più città, con le prestazioni economiche. Nel dettaglio le parole in questione sono "suite" e "B&B", alle quali si può dare una lettura analoga ma di segno opposto.

La parola "suite" sembra avere un'accezione positiva, probabilmente connessa generalmente al campo del lusso e alla cura degli ambienti, che comporta una certa attrattività della stessa nel titolo. Gli host dovrebbero ispirarsi a tale tipologia di impostazione dell'alloggio mettendo in risalto gli aspetti che la caratterizzano (cura e bellezza degli ambienti). Similmente, "B&B" potrebbe esser letto con un'accezione negativa se comparato all'immagine lussuosa appena presentata per "suite": nonostante si stia parlando proprio di Bed&Breakfast, sembrerebbe meglio non

sottolineare tale aspetto: essi vengono generalmente scelti per la loro praticità ed economicità ma in ottica di pubblicizzazione, più che sottolineare queste caratteristiche, può esser meglio far riferimento ad alloggi di maggior levatura.

Altra parola interessante è “terrace”, in modo simile alla parola “suite”: pubblicizzare nell’annuncio la presenza di una terrazza risulta positivo. Negli alloggi che possiedono questo tipo di caratteristica gli host dovrebbero assolutamente mettere in risalto tale aspetto, che risulta positivamente correlato con le prestazioni economiche dell’alloggio.

Tra le variabili di controllo risalta il caso del massimo numero di ospiti, sempre correlato positivamente ai risultati economici. Risulta ragionevole pensare che un numero di ospiti maggiore possa tradursi in maggiori prestazioni economiche. Altri fattori da attenzionare da parte degli host sono la qualifica di superhost, il numero di foto pubblicate nell’annuncio e la possibilità di prenotazione istantanea. Tali variabili mostrano generalmente un’influenza, che assume segno positivo e pertanto è opportuno tenerle in considerazione nella posizione di host.

Importanti anche i ritrovamenti relativi al cambio del titolo dal 2019 al 2020 nel caso di Gallipoli: sembrerebbe che gli host che hanno cambiato titolo si siano mossi nella giusta direzione. Nei nuovi titoli si notano azioni manageriali dal punto di vista semantico in linea con i ritrovamenti appena esposti nei paragrafi precedenti, come l’aggiunta della parola suite e la bassa frequenza di parole indicanti la centralità dell’alloggio. Il risultato è una correlazione positiva tra il cambio del titolo e le prestazioni economiche.

Tirando le somme, si può affermare che le prestazioni economiche delle abitazioni posizionate sulla piattaforma di AirBnb, in ambito hospitality, sono maggiormente influenzate da attributi esterni al campo semantico. Tuttavia, non è possibile negare che l’utilizzo di una parola piuttosto che un’altra nell’annuncio dell’abitazione abbia un effetto sulle prestazioni. L’aspetto semantico non è il primo su cui focalizzarsi ma certamente da un contributo alla buona riuscita degli scambi su piattaforma.

## Opportunità di ricerca futura e punti aperti

Si evidenzia che le analisi svolte hanno riguardato un ristretto periodo temporale di due anni, e potrebbe essere di interesse, in lavori futuri, estendere l'orizzonte di ricerca. Inoltre, le analisi sono state svolte su singoli dataset relativi a città, geograficamente concentrate e ristrette, e di conseguenza potrebbe essere utile, in progetti di maggiore estensione temporale e non solo, effettuare delle ricerche a scala nazionale, con gli opportuni provvedimenti, così da estendere le numerosità campionarie e il campo di applicazione.

Si porta all'attenzione, in aggiunta, che nonostante siano stati trattati anche dati relativi al 2020, non siano comparse nelle analisi semantiche, parole fortemente connesse alla situazione pandemica, per esempio legate alla pulizia o alla sicurezza. Verosimilmente, i dati potrebbero essere relativi a un periodo troppo vicino all'inizio della pandemia, e di conseguenza sarebbe interessante riproporre le analisi più in là nel tempo, con dati successivi, relativi al 2021 e persino al 2022.

Un altro spunto per future ricerche riguarda la formulazione dei modelli di regressione: potrebbe essere interessante testare formulazioni non lineari, per comprendere se tale tipologia di funzioni più complesse possa mettere maggiormente in luce dei fattori significativi e rappresentare meglio la grandezza sotto osservazione.

Ancora, si è evidenziato dalle word clouds che spesso sono presenti i nomi dell'host nell'annuncio: potrebbe essere interessante creare una variabile semantica legata alla presenza di un nome nel dato annuncio, e validare in generale se la presenza del nome dell'host nell'annuncio possa impattare le prestazioni. Sarebbe chiaramente necessario mappare i nomi presenti nel dataset, così da rendere la variabile "nome", definita come prima, pari a uno se un nome è presente nell'annuncio e zero altrimenti.

## Bibliografia, sitografia e citazioni

1. Abbie-Gayle, J., & Barbara, N. (2017). Airbnb – An exploration of value co-creation experiences in Jamaica. *International Journal of Contemporary Hospitality Management*, 29(9), 2361–2376.
2. Airbnb. About us. Ottenuto da <https://www.airbnb.com.au/about/about-us>.
3. Airbnb. How it works. Ottenuto da <https://www.airbnb.co.nz/help/getting-started/how-it-works>.
4. Belarmino, A., Whalen, E., Koh, Y., Bowen, J.T., 2017. Comparing guests' key attributes of peer-to-peer accommodations and hotels: mixed-methods approach. *Curr. Issues Tour.* 1–7.
5. Blei D. (2012). Probabilistic topic models.
6. Bowlby, S., Gregory, S., & McKie, L. (1997). “Doing home”: Patriarchy, caring, and space. *Women's Studies International Forum*, 20(3), 343–350.
7. Bridges, J., Vásquez, C., 2016. If nearly all Airbnb reviews are positive, does that make them meaningless? *Curr. Issues Tour.* 1–19.
8. Brochado, A., Troilo, M., & Shah, A. (2017). Airbnb customer experience: Evidence of convergence across three countries. *Annals of Tourism Research*, 63, 210–212.
9. Cheng M., Edwards D., 2017. A comparative automated content analysis approach on the review of the sharing economy discourse in tourism and hospitality. *Curr. Issues Tour.* 1–15.
10. Cheng M., Jin X., 2019. What do Airbnb users care about? An analysis of online review comments - *International Journal of Hospitality Management* 76 (2019) 58-70.
11. Cohen, B., & Kietzmann, J. (2014). Ride on! Mobility business models for the sharing economy. *Organization & Environment*, 27(3), 279–296.
12. Dogru T., Majid K., Laroche M., Mody M., Courtney Suess. Communicating quality while evoking loss – How consumers assess extra charges in the Airbnb marketplace. *Tourism Management*.

13. Edwards, D., Cheng, M., Wong, A., Zhang, J., Wu, Q., 2017. Ambassadors of knowledge sharing: co-produced travel information through tourist-local social media exchange. *International J. Contemp. Hosp. Manag.* 29 (2), 690–708.
14. Festila, M., Müller, S., 2017. The impact of technology-Mediated consumption on identity: the case of airbnb. 50th Hawaii International Conference on System Sciences.
15. Fuchs, M., Höpken, W., Lexhagen, M., 2014. Big data analytics for knowledge generation in tourism destinations—a case from Sweden. *J. Destin. Mark. Manag.* 3 (4), 198–209.
16. Guttentag, D., 2015. Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector. *Curr. Issues Tour.* 18 (12), 1192–1217.
17. Guttentag, D., 2016. Why Tourists Choose Airbnb: A Motivation-based Segmentation Study Underpinned by Innovation Concepts (PhD. University of Waterloo, Waterloo).
18. Guttentag, D., Smith, S., 2017. Assessing Airbnb as a disruptive innovation relative to hotels: substitution and comparative performance expectations. *Int. J. Hosp. Manag.* 64, 1–10.
19. Henten, A. H., & Windekilde, I. M. (2016). Transaction costs and the sharing economy. *Info*, 18(1), 1–15.
20. Kai Ding, Wei Chong Choo, Keng Yap Ng, Siew Imm Ng e Pu Song, 2021. Exploring sources of satisfaction and dissatisfaction in Airbnb accommodation using unsupervised and supervised topic modeling. *Front Psychol.*
21. Kozinets, R. V. (2010). *Netnography: Doing ethnographic research online*. London: Sage.
22. Kumar S., Zymbler M. (2019). A machine learning approach to analyze customer satisfaction from airline tweets. *J. Big Data*.
23. Lampinen, A., Cheshire, C., 2016. Hosting via Airbnb: motivations and financial assurances in monetized network hospitality. 2016 CHI Conference on Human Factors in Computing Systems.
24. Lee E. J. , Cho Y., 2021. Investigating Citizen Perceptions and Business Performance of Airbnb in Korea - *Journal of Asian Finance, Economics and Business* Vol 8 No 8 (2021) 0167–0180.

25. Lessig, L. (2008). *REMIK: Making art and commerce thrive in the hybrid economy*. New York, NY: The Penguin Press.
26. Litvin S., Goldsmith R., Pan B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tour. Manag.* 29, 458–468.
27. Mingming Cheng, Xin Jin, 2018. What do Airbnb users care about? An analysis of online review comments. *International Journal of Hospitality Management* 76 (2019) 58-70.
28. Natassha Selvaraj, 2020. A Beginner's Guide to Sentiment Analysis with Python - An end to end guide on building word clouds, beautiful visualizations, and machine learning models using text data" presente sul sito [www.towardsdatascience.com](http://www.towardsdatascience.com)
29. Saeed, J. I. (2015). *Semantics* (4th ed.). New York: Wiley-Blackwell.
30. Sigala, M. (2015). Collaborative commerce in tourism: Implications for research and industry. *Current Issues in Tourism*, 20(4), 346–355.
31. Sundararajan, A. (2014). Peer-to-peer businesses and the sharing (collaborative) economy: Overview, economic effects, and regulatory issues.
32. Tussyadiah, L., 2016. Factors of satisfaction and intention to use peer-to-peer accommodation. *Int. J. Hosp. Manag.* 55, 70–80.
33. Tussyadiah, L., Pesonen, J., 2016. Impacts of peer-to-Peer accommodation use on travel patterns. *J. Travel Res.* 55 (8), 1022–1040.
34. Tussyadiah, L., Zach, F., 2016. Identifying salient attributes of peer-to-peer accommodation experience. *J. Travel Tour. Mark.* 1–17.
35. [www.airbnb.com](http://www.airbnb.com)
36. [www.wikipedia.com](http://www.wikipedia.com)
37. Xiang, Z., Schwartz, Z., Gerdes, J.H., Uysal, M., 2015. What can big data and text analytics tell us about hotel guest experience and satisfaction? *Int. J. Hosp. Manag.* 44, 120–130.
38. Yannopoulou, N., 2013. User-generated brands and social media: couchsurfing and airbnb. *Contemp. Manag. Res.* 9 (1), 85–90.
39. Zervas, G., Proserpio, D., & Byers, J. (2015). A first look at online reputation on Airbnb, where every stay is above average. *Marketing Letters*, 1–16.

40. Zervas, G., Proserpio, D., & Byers, J. W. (2017). The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry. *Journal of Marketing Research*, 54(5), 687–705.
41. Zhu Y., Cheng M., Wang J., Ma L., Ruochen Jiang. The construction of home feeling by Airbnb guests in the sharing economy: A semantics perspective - *Annals of Tourism Research* 75 (2019) 308-321.

