



Politecnico di Torino

Corso di Laurea Magistrale in Ingegneria Gestionale

A.A. 2023/2024

Sessione di Laurea: luglio 2024

Effetti dei functional attributes sulle performance degli Airbnb: il caso studio su Berlino.

Relatori:

Raguseo Elisabetta

Milone Francesco Luigi

Candidata:

Stura Valentina

“Tutto posso in Colui che mi dà la forza.”

Filippesi 4:13

A Michele Stura

Sommario

1. Abstract.....	1
2. Introduzione.....	2
3. Analisi della letteratura.....	3
3.1. Effetti del Covid-19 su Airbnb	3
3.2. Resilienza di Airbnb rispetto agli hotel	5
3.3. Comportamento degli host	6
3.4. Comportamento dei guest.....	9
3.5. Focus sulle variabili spaziali e di performance	11
4. Obiettivo e contributo alla letteratura.....	15
5. Studio di ricerca.....	16
5.1. Dati in input.....	16
5.2. Dati aggiuntivi	17
5.2.1. Calcolo coordinate dei punti di riferimento.....	18
5.2.2. Calcolo distanze delle proprietà dai punti di riferimento	18
5.3. Strumenti software.....	20
6. Domanda di ricerca.....	21
6.1. Punto di Partenza	21
6.2. Research gap.....	22
6.3. Domanda di ricerca.....	22
7. Research Framework e ipotesi.....	23
8. Analisi statistiche descrittive	24
8.1. Caratteristiche della proprietà.....	24
8.1.1. Property ID	24
8.1.2. Listing Type	25
8.1.3. Neighborhood	29
8.1.4. Bedrooms.....	30
8.1.5. Bathrooms.....	31
8.2. Scelte strategiche dell'host	32
8.2.1. Max Guest	32
8.2.2. Minimum Stay	32
8.2.3. Number of Photos.....	33
8.3. Contributo dei guest	34
8.3.1. Number of Reviews	34
8.3.2. Overall Rating	35
8.4. Caratteristiche dell'host.....	37
8.4.1. Host type.....	37

8.4.2.	Superhost	39
8.5.	Variabili spaziali	40
8.5.1.	Metro station.....	40
8.5.2.	Metro Station Distance	42
8.5.3.	Monument.....	43
8.5.4.	Monument Distance.....	44
8.5.5.	Park.....	45
8.5.6.	Park Distance	46
8.5.7.	Parking.....	47
8.5.8.	Parking Distance.....	48
8.6.	Variabili di performance	49
8.6.1	Revenue	49
8.6.2.	Reservation Days	51
8.6.3.	RevPAN	52
8.6.4.	OCC	53
8.6.5.	ADR	54
8.7.	Variabili di performance per classi di distanza.....	55
8.7.1.	Revenue – Distance Class	55
8.7.2.	Reservation Days – Distance class	56
8.7.3.	RevPAN – Distance class	57
8.7.4.	OCC – Distance Class	58
8.7.5.	ADR - Distance Class.....	59
9.	Analisi di correlazione.....	60
10.	Analisi di regressione	65
10.1.	Regressioni Univariate	65
10.2.	Regressioni multivariate	68
10.2.1.	Regressioni multivariate con Superhost	75
10.2.2.	Regressioni multivariate con Long Stay.....	82
11.	Conclusioni.....	89
12.	Sviluppi futuri.....	90
13.	Ringraziamenti	91
14.	Bibliografia.....	92

1. Abstract

L'obiettivo della seguente tesi di laurea magistrale è analizzare e quantificare l'impatto dei functional attributes di location sulle performance degli Airbnb, con focus sulla città di Berlino negli anni tra il 2019 ed il 2023. I functional attributes oggetto di indagine sono le distanze da quattro tipologie di punti di interesse: la metropolitana, i monumenti, i parchi ed i parcheggi. Con Python sono state calcolate le coordinate geografiche di questi punti e le loro distanze dagli Airbnb. È stato costruito un nuovo dataset ottenuto dal merge dei dati sulle prenotazioni provenienti da AirDNA e di quelli calcolati con Python. Dalle analisi di correlazione e regressione realizzate su Stata tra le distanze spaziali ed il RevPAN, si dimostra che la vicinanza degli Airbnb dai punti di interesse noti fa aumentare in modo statisticamente significativo i ricavi per notte disponibile, e quindi le performance, e che per i superhost questo incremento è significativamente maggiore.

2. Introduzione

L'obiettivo principale della tesi è indagare l'esistenza di un effetto causale negativo della distanza degli Airbnb da punti di interesse noti, quali la stazione della metropolitana più vicina, il monumento, il parco ed il parcheggio più vicini, sulle performance.

L'obiettivo secondario è verificare se ci sono differenze significative per la categoria dei superhost e per i soggiorni di lungo periodo.

Il lavoro inizia con un'analisi ed una sintesi della letteratura ad oggi presente su Airbnb. Gli argomenti sono stati raggruppati in cinque principali cluster, ovvero l'impatto che ha avuto la pandemia sulle performance di Airbnb, la sua resilienza rispetto agli hotel, il comportamento degli host, quello dei guest ed infine un focus sul ruolo delle variabili spaziali.

La tesi mira a colmare una lacuna presente nella letteratura indagando l'esistenza effettiva di una correlazione tra le distanze da punti di interesse ed i ricavi per notte disponibile, studiando la città di Berlino.

Il calcolo delle distanze è stato eseguito con il supporto del linguaggio di programmazione Python. Il dataset utilizzato per le analisi statistiche è un foglio di calcolo Excel. Il software statistico usato per le regressioni è STATA.

3. Analisi della letteratura

3.1. Effetti del Covid-19 su Airbnb

Nel 2008 tre ragazzi decidono di affittare dei materassi gonfiabili nella loro casa a San Francisco per ospitare visitatori in città per una conferenza. Da una semplice idea di Brian Chesky, Joe Gebbia e Nathan Blecharczyk nasce un nuovo modello di business che si inserisce nel settore del turismo, nel segmento dell'ospitalità e che ottiene un forte ed inaspettato successo.

Quest'idea di alloggio alternativo e a basso costo poteva essere sfruttata in moltissime occasioni, soprattutto nei periodi di più forte flusso turistico, come alternativa al classico soggiorno in hotel.

I fondatori hanno allora creato una piattaforma digitale per poter mettere in affitto per brevi periodi una vasta gamma di alloggi e abitazioni o parte di essi, da parte di soggetti privati. La piattaforma peer to peer, che prese il nome di Airbnb, ha avuto negli anni una crescita ed un successo esplosivi, avendo portato una vera e propria rivoluzione nell'economia del turismo. Airbnb può quindi essere considerato un disruptor nel settore turistico.

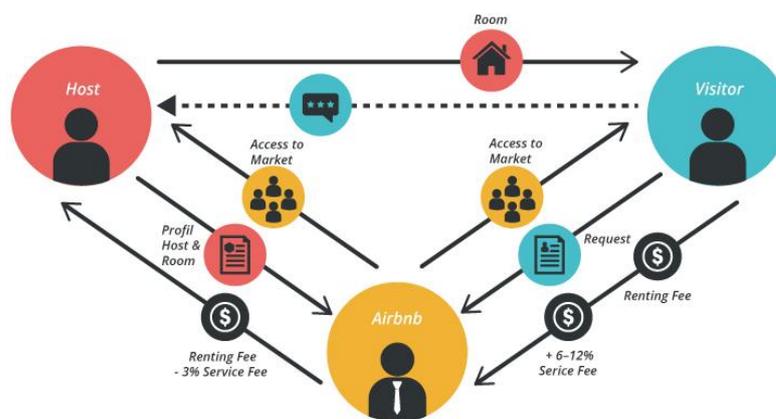


Figura 1: Business Model di Airbnb

L'esperienza offerta da Airbnb comprende la possibilità di interagire con persone del luogo, condividere con loro esperienze, cibo, abitudini, ricevere consigli autentici su cosa visitare, in quali ristoranti provare le pietanze del posto, il tutto con un approccio più familiare ed improntato ad un'immersione nella cultura e nelle tradizioni locali.

La piattaforma si inserisce quindi nel contesto della sharing economy, ovvero dell'economia della condivisione, in cui lo scopo principale è la massimizzazione dello sfruttamento delle risorse disponibili, nel caso analizzato si tratta della capacità abitativa in eccesso. Gli ambiti in cui questa modus operandi si può applicare sono molteplici: alloggio, trasporto, cibo, intrattenimento, finanza. Airbnb è la piattaforma online che permette di far incontrare la domanda di affitti di breve termine con l'offerta di alloggi, case o stanze private condivise con i turisti (Kevin Kam Fung So, Giampaolo Viglia, Stephanie Liu, Dan Wang, 2023).

Gli stakeholders coinvolti sono: gli host che mettono a disposizione l'unità abitativa, i guest che prenotano il soggiorno, i competitor, ovvero i proprietari di hotel, pensioni, bed&breakfast, gli abitanti del luogo che subiscono l'impatto dei cambiamenti ed i governi che, se non adottano regolamenti ad hoc, risultano responsabili delle conseguenze che si verificano. La piattaforma si basa essenzialmente su un rapporto di fiducia tra tutti questi stakeholders, che interagiscono come illustrato in figura 1.

Airbnb ha cambiato gran parte delle dinamiche di domanda e offerta, di prezzi e qualità del soggiorno, impattando fortemente anche il mercato immobiliare. Il fenomeno consiste in un'esternalità positiva poichè la piattaforma ha portato ad un aumento del turismo e quindi di ricavi. D'altro canto, ha anche portato a delle esternalità negative, come l'aumento della rumorosità e atti vandalici nei quartieri con conseguente necessità di aumentare i controlli e il livello di sicurezza per i residenti.

L'avvento della pandemia ha portato ad un cambiamento sostanziale nel modello di viaggio, aprendo la strada ad un modello di business che era nato pochi anni prima e che faceva leva su qualità, comfort, accessibilità e sicurezza.

La piattaforma digitale di Airbnb è un two-sided market, ovvero un modello di business che favorisce le transazioni tra due gruppi di utenti, quali host e guest, che hanno interessi complementari, ottenendo delle vere e proprie sinergie, che portano benefici ad entrambi i lati della piattaforma. L'offerta è rappresentata dagli host che forniscono, inserendo annunci sulla piattaforma, affitti per brevi periodi. La domanda è rappresentata dai guest che desiderano soggiornare in maniera innovativa, più agevole, economica e smart nel luogo desiderato.

Lo scopo principale della piattaforma è quindi favorire una riallocazione della capacità abitativa in eccesso, permettendo di affittare per brevi periodi alloggi, case, stanze o posti letto non sfruttati dai proprietari. Questo ha portato ad un aumento di valore degli immobili con conseguente aumento dei prezzi delle case e dei canoni di affitto. L'impatto di Airbnb quindi

non si limita solo al settore turistico, ma muta anche il settore immobiliare che osserva un aumento dei canoni di affitto e dei prezzi delle case in vendita. L'effetto risulta quindi economicamente positivo.

In generale si è registrata una tendenza ad adibire le abitazioni da affitti di lungo ad affitti di breve periodo. Questo ha portato ad un aumento dei ricavi per gli host, del valore degli immobili, dei prezzi e dei ricavi per il settore immobiliare in generale. (Kyle Barron, Edward Kung, Davide Proserpio, 2021). Tutto ciò ha anche portato ad un aumento della domanda di affitti di breve periodo e ad una riduzione del tasso di case sfitte.

La pandemia da Covid-19 è stata uno shock esogeno senza precedenti che il settore del turismo ha subito principalmente negli anni 2020 e 2021, portando con sé importanti cambiamenti nella gestione del business, soprattutto per le piattaforme digitali P2P.

Anche la pandemia è stato un vero e proprio disruptor che ha cambiato moltissimi aspetti della vita quotidiana delle persone: dal modo di lavorare, alle interazioni sociali, agli hobby, sport ed infine al modo di viaggiare.

Come per tutte le disruption, possono rimanere sul mercato solo coloro che rapidamente riescono ad adeguarsi ai cambiamenti e a rimanere competitivi.

Una possibile minaccia emersa in questo periodo è la piattaforma Zoom per videocall e meeting online, che ha diminuito la necessità di spostamenti fisici per affari, non essendo più necessaria la presenza in persona per partecipare a riunioni e conferenze (Daniel A. Guttentag, Stephen W. Litvin, Wayne W. Smith, 2023).

3.2. Resilienza di Airbnb rispetto agli hotel

In letteratura sono state comparate le performance di Airbnb con quelle degli hotel tradizionali ed è emerso che i due modelli di business differenti hanno avuto una reazione diversa allo shock pandemico, risultati economici differenti e anche una diversa ripresa nel periodo post-pandemico.

La domanda di mercato per gli affitti di breve periodo è drasticamente diminuita in tutto il mondo, ma i paesi maggiormente colpiti dal calo di flusso turistico e domanda sono stati quelli in cui i governi hanno adottato restrizioni più severe. Il maggior tasso di calo della domanda è stato registrato dagli hotel.

Anche all'interno dello stesso segmento del settore turistico, l'impatto economico è stato molto eterogeneo tra le varie tipologie di alloggi, di host e di guest.

È stato dimostrato da studi scientifici che Airbnb è stata più resiliente durante la pandemia rispetto agli hotel, registrando un minor calo della domanda a parità di altre condizioni. Terminata la pandemia, la piattaforma ha anche registrato un aumento delle performance molto più rapido rispetto agli hotel tradizionali (Tarik Dogru, Lydia Hanks, Courtney Suess, Nathan Line, Makarand Mody, 2023).

Per quanto riguarda il mercato azionario si è registrato un aumento del valore di mercato di Airbnb nei periodi in cui il trend dei contagi era in rialzo. Questo conferma che la tendenza generale è stata il passaggio dalle prenotazioni presso gli hotel a quelle su Airbnb.

Airbnb, più degli hotel, ha mostrato una forte capacità adattativa alle condizioni mutevoli del panorama turistico mondiale e questa è stata la chiave del successo che ha permesso alla piattaforma, non solo di sopravvivere, ma anche di rafforzarsi e di puntare su una value proposition unica e non imitabile dai competitor per diventare leader nel settore (Juan Luis Nicolau, Abhinav Sharma, Hakseung Shin, Juhyun Kang, 2023).

3.3. Comportamento degli host

Gli host di Airbnb possono mettere un annuncio relativo ad una casa intera, ad una stanza privata o ad una stanza condivisa con altri ospiti o con il proprietario stesso. Gli host che offrono un alloggio intero hanno registrato un impatto negativo causato dal Covid-19 inferiore rispetto a chi offriva una stanza condivisa. Gli host professionali sono stati in grado di reagire prima e meglio con strategie di differenziazione di prezzo che hanno permesso loro di contenere le perdite. Coloro che ospitano viaggiatori per mestiere hanno registrato un calo di prenotazioni inferiori rispetto a chi accoglie i viaggiatori per interesse personale (Francesco Luigi Milone, Ulrich Günter, Bozana Zekan, 2023).

La qualità degli annunci di host professionali è migliore ed essi sono percepiti dai guest come annunci che danno maggior sicurezza e migliori garanzie, per cui hanno registrato una diminuzione delle prenotazioni inferiore rispetto agli host non professionali.

Ecco che gli host che riescono maggiormente ad adattarsi alle nuove esigenze del consumatore, a segmentare i clienti e ad applicare meccanismi dinamici di aggiustamento di prezzi, hanno il

più alto tasso di sopravvivenza e riescono ad ottenere un margine di profitto ed un successo maggiori (David Boto-García, 2022).

Sono considerati single-host coloro che gestiscono una sola proprietà al mese, multi-host coloro che ne gestiscono da 2 a 4 e business-host coloro che ne gestiscono oltre 4. Considerando la professionalità dell'host, si evidenzia come ci sia una relazione tra management e tipologia di host. Ad esempio, ad Amsterdam e Berlino dove il mercato è più regolamentato sono presenti un maggior numero di annunci di single-host (Kristóf Gyódi e Łukasz Nawaro, 2021). Questa tipologia di host tende a specializzarsi nella gestione di una singola proprietà anziché gestirne diverse. D'altro canto, all'aumentare del numero di proprietà gestite da un utente è inevitabile che la cura con cui si seguirà il business sarà decrescente (Huihui Zhang, 2022).

Agli host di Airbnb è possibile attribuire il titolo di superhost qualora rispettino precisi requisiti richiesti dalla piattaforma stessa, come avere almeno 10 soggiorni completati negli ultimi 12 mesi, avere un tasso di risposta del 90% o superiore alle richieste di prenotazione e ai messaggi degli ospiti, avere un tasso di accettazione delle prenotazioni almeno del 90%, non aver cancellato prenotazioni (tranne in circostanze eccezionali), avere recensioni generalmente positive, che includono un punteggio di almeno 4,8 su 5 stelle. Il titolo di superhost quindi è attribuito all'host direttamente da Airbnb nel caso in cui questo abbia raccolto un elevato livello di soddisfazione da parte degli ospiti. Ottenere questo titolo è segno di qualità e reputazione (Kristóf Gyódi e Łukasz Nawaro, 2021). Il superhost fornisce solitamente informazioni aggiuntive, trova il tempo di interagire durante il soggiorno con gli ospiti, garantendo così un valore aggiunto. Molti studi scientifici dimostrano come l'assenza di interazione personale e comunicazione riducano per l'ospite la qualità del servizio. Si tratta infatti di aspetti fondamentali nell'ospitalità peer-to-peer (Madalyn A. et al., 2020).

Anche la piattaforma stessa influenza la comunicazione e l'interazione tra host e guest attraverso la richiesta di recensioni, fondamentali per mantenere il titolo e guadagni maggiori (Madalyn A. et al., 2020).

Gli host hanno reagito all'impatto dello shock pandemico adottando nuove strategie di marketing per far fronte alle nuove sfide e per non dover uscire dal mercato.

In base al trinomio consapevolezza-motivazione-capacità, gli host sono stati segmentati in speculatori, diplomatici ed imprenditori.

Gli speculatori sono coloro che utilizzano Airbnb solo per scopo di trarre profitto da variazioni di prezzo o fluttuazioni della domanda sul mercato. Questi hanno avuto una visione molto pessimistica dopo l'avvento del Covid e, a causa dell'incapacità di cambiare il modello di business, hanno deciso gradualmente di abbandonare l'attività.

I diplomatici sono coloro che gestiscono le proprietà con focus sull'ospitalità e sul rapporto con gli ospiti. Sono particolarmente attenti alle loro esigenze, offrono servizi aggiuntivi, consigli locali e un'esperienza autentica del luogo. Possono anche essere impegnati nella gestione delle relazioni con gli ospiti, garantendo un livello elevato di comunicazione e soddisfazione. I diplomatici cercano di creare un'esperienza gradita al fine di ottenere recensioni positive e costruire una reputazione solida sulla piattaforma. Dato che portano avanti il business con focus sulla motivazione con cui Airbnb è nata, hanno mantenuto il business invariato e si sono ripresi rapidamente dopo la pandemia.

Gli imprenditori sono invece host che utilizzano la piattaforma per gestire e promuovere attivamente la propria attività di ospitalità. Possiedono una o più proprietà destinate all'affitto di breve periodo e adottano strategie di marketing mirate ad attirare gli ospiti e a massimizzare l'occupazione delle loro unità abitative. Possono anche cercare opportunità di espansione, investendo in nuove proprietà o sviluppando partnership con altri host o servizi complementari. La gestione efficiente delle proprietà e la creazione di un'esperienza di soggiorno memorabile sono al centro dell'approccio degli imprenditori. Poiché possiedono capacità manageriali di gestione del business, sono destinati ad innovarlo e a farlo crescere nel tempo (Mo Zhang, Ruoqi Geng, Yuan Huang, Shengce Ren, 2021).

In sostanza, si può affermare che coloro che hanno da sempre sfruttato la piattaforma con lo spirito originario con cui era stata concepita dai fondatori e che pongono al centro del business l'accoglienza del viaggiatore con spirito di condivisione e familiarità sono anche coloro che meglio sono sopravvissuti allo shock e che hanno subito il minor numero di danni economici.

Le nuove strategie di marketing principalmente adottate dagli host sono state due. La prima consiste nell'aggiustamento dei prezzi in modo da seguire il più possibile la curva di domanda, abbassando i prezzi nei periodi di più forti restrizioni dei governi e alzandoli invece nei periodi con indice di stringenza più basso e con maggior flusso turistico. In questo modo, sono riusciti a compensare le perdite nei mesi di totale chiusura e a conquistare persino nuove quote di mercato. La segmentazione dei clienti sulla base dell'elasticità della domanda al prezzo è stata

fondamentale per creare un meccanismo di autoselezione dei clienti nel corretto segmento di mercato ed una massimizzazione dei ricavi per gli host. La seconda strategia di marketing adottata è stata l'inserimento di politiche di cancellazione flessibili che garantiscono la possibilità di cancellare la prenotazione senza costo aggiuntivo fino a 24 ore prima della data prenotata. Questo ha portato sicurezza nei clienti ed è stato un forte incentivo a continuare a prenotare tramite la piattaforma.

Le aree urbane sono per gli host ad alto potenziale per quanto riguarda i ricavi, ma anche ad alto rischio perché i clienti erano restii a prenotare in aree densamente popolate. Allo stesso modo le aree rurali sono a basso potenziale e a basso rischio di contagio.

Sia host che guest erano così costretti ad applicare dei trade-off di scelta sulla base delle nuove variabili che entrano in gioco.

3.4. Comportamento dei guest

La paura dei viaggiatori di contrarre il virus e quindi la loro avversità al rischio li ha portati a spostare il focus dell'attenzione ad un aspetto che prima non era mai stato preso in considerazione, ovvero il distanziamento sociale, e contemporaneamente è aumentata l'attenzione sul livello di igiene e pulizia offerto dalla struttura ricettiva. Il comportamento dei consumatori, quindi, comincia a mutare e, di conseguenza, mutano anche i driver di scelta.

A questo proposito è stata dimostrata la teoria della segnalazione, secondo cui un prezzo maggiore è indicatore di un più alto livello di servizio e quindi fa aumentare la willingness-to-pay degli ospiti. Anche la teoria della distanza psicologica è stata confermata, secondo cui la distanza fisica percepita dalle persone influisce sul comportamento e sulle scelte dell'individuo (Raffaele Filieri, Francesco Luigi Milone, Emilio Paolucci, Elisabetta Raguseo, 2023).

In generale, le aree geograficamente più colpite sono state quelle con elevata densità di popolazione, ovvero le grandi città. Sono state colpite meno gravemente le aree rurali o montane. La fascia di clienti lusso è quella che ha subito il maggior calo di prenotazioni, i clienti di fascia media sono coloro che le hanno ridotte meno drasticamente, mentre i clienti di tipo business hanno avuto una riduzione di media entità. Il settore lusso in effetti è quello che è tornato più rapidamente alle condizioni pre-Covid, mentre il settore budget ha registrato molta più inerzia.

È necessario distinguere i viaggiatori in due segmenti: business, ovvero coloro che viaggiano per affari, e leisure, coloro che viaggiano per interesse e svago personale. La tendenza comune durante la pandemia è stata passare dal soggiorno in un classico hotel, in cui la probabilità di contatto fisico con albergatore e altri clienti era molto alto, ad un soggiorno prenotato su Airbnb, con preferenza per intere case o appartamenti piuttosto che stanze condivise. Il livello di pulizia ed igiene offerti dalla struttura, insieme alla qualità deducibile da foto e recensioni dei precedenti visitatori sono stati i driver di scelta maggiormente utilizzati. Cambia inoltre la disponibilità a pagare dei clienti che associano ad un prezzo maggiore un livello di igiene e pulizia degli ambienti migliore e quindi una maggior sicurezza (Jongho Im, Jewoo Kim, Joon Yeon Choeh, 2021).

Tra i due diversi segmenti di clienti, il settore business è stato meno intaccato dallo shock a causa della maggior rigidità della domanda, rispetto al settore leisure, rallentato dalle restrizioni politiche sui viaggi per abbassare la curva dei contagi (Karima Kourtit, Peter Nijkamp, John Östh, Umut Turk, 2022).

Il principale driver di scelta per tutte le tipologie di ospiti è senza dubbio la sicurezza. L'ospite sceglie in base alla tipologia del soggiorno, fra aree urbane e rurali o montane, in base alla tipologia di viaggio che vuole effettuare, per affari o per piacere, in base all'ambiente proposto e alla percezione del rischio di contatto fisico, alimentare e di igiene delle stanze (Seongsoo Jang, Jungkeun Kim, Jinwon Kim, Seongseop Kim, 2021).

Una nuova opportunità di business nata proprio nello stesso periodo sul mercato lavorativo è quella legata ad una nuova tipologia di lavoratori chiamati nomadi digitali, ovvero coloro che possono lavorare con un computer portatile ed una connessione internet da qualsiasi posto nel mondo. Costoro hanno allora l'opportunità di lavorare viaggiando e soggiornare a basso costo presso famiglie sparse per il mondo che mettono a disposizione parte della loro casa (Tarik Dogru, Lydia Hanks, Courtney Suess, Nathan Line, Makarand Mody, 2023).

Dai dati sulle prenotazioni analizzate è emerso un forte spillover spaziale, per cui la probabilità di occupazione di appartamenti vicini ad altri considerati di successo e con ottime recensioni è molto più alta di altri, a parità di condizioni. Questo significa che le performance di attori vicini possono avere una forte influenza sugli altri.

Le preferenze dei viaggiatori cambiano con l'avvento del Covid: si registra un aumento delle prenotazioni in aree rurali, montane o sub-urbane ed un calo nelle aree urbane. La maggior parte di essi prenota in aree a bassa densità di popolazione. I turisti non optano più per appartamenti

vicini a mezzi pubblici di trasporto, in quanto si preferisce spostarsi all'aperto a piedi o in bicicletta. I clienti sono molto più attenti alle caratteristiche che indicano sicurezza e basso rischio di contagio (Raffaele Filieri, Francesco Luigi Milone, Emilio Paolucci, Elisabetta Raguseo, 2023).

Nel complesso, la scelta del consumatore è frutto di un'interazione di molti fattori. I segnali che portano il consumatore a scegliere un appartamento sono il numero di stanze, posizione degli annunci, informazioni sugli host (se sono superhost, il loro profilo e stato di verifica), il prezzo, i costi aggiuntivi (ad esempio di pulizia), la competitività e popolarità regionale, le regole della casa. Il servizio internet e cucina, avere letti veri e non divani, avere una villetta e non un appartamento rendono l'annuncio fortemente più desiderabile per il cliente.

Ad una riduzione dell'asimmetria informativa tra host e guest, aumenta la popolarità dell'annuncio la probabilità di essere occupato. La probabilità di prenotazione dipende anche dalle caratteristiche personali del viaggiatore come sesso, età, motivo del viaggio, gusti personali. Il tutto dev'essere anche valutato in relazione alle caratteristiche ambientali, sociali, governative del momento della prenotazione e del soggiorno.

In sostanza, l'avvento del Covid ha stravolto sia il business degli host che le abitudini di viaggio, le preferenze ed il comportamento dei guest.

3.5. Focus sulle variabili spaziali e di performance

Con l'avvento del Covid, alcune variabili che avevano basso impatto sulle scelte del consumatore hanno iniziato ad acquisire importanza. Si tratta delle performance degli host vicini e le caratteristiche dell'host stesso: professionale o privato, superhost o meno. Alcune variabili hanno mantenuto la stessa importanza prima, durante e dopo il Covid, sono le caratteristiche intrinseche delle proprietà: tipologia, numero di camere, numero di bagni, numero massimo di ospiti. Altre variabili hanno invece perso di importanza e sono quelle principalmente legate alla distribuzione spaziale, comprendono la vicinanza ai mezzi di trasporto e ai principali esercizi commerciali (Ruggero Sainaghi, Jorge Chica-Olmo, 2022).

In particolare, gli aspetti che compaiono negli annunci con maggiore frequenza riguardano: le attrazioni e i mezzi di trasporto vicini alla proprietà, l'ubicazione delle camere e dei servizi e la tipologia di visitatori tipici dell'appartamento (Yejin Chung e Surendra Sarnikar, 2021).

Il prezzo finale degli Airbnb è frutto di numerose variabili. Tra queste troviamo le caratteristiche della proprietà come la tipologia di appartamento (casa indipendente, alloggio, stanza singola, stanza condivisa). Un'altra componente del prezzo è data dalla tipologia di host, gli host professionali sono in grado di differenziare maggiormente il prezzo segmentando i clienti, facendo leva sulle proprie competenze manageriali e fissando il prezzo in base ai costi e ricavi marginali mentre gli host privati invece sono più propensi ad imitare il prezzo degli Airbnb vicini e meno a seguire vere e proprie strategie di prezzo. Per gli host professionali si crea quindi un equilibrio monopolistico, mentre per gli host privati si crea la cosiddetta "concorrenza a colpo d'occhio". (David Boto-García, Matías Mayor, Pablo De la Vega, 2022).

Una differenza di prezzo significativa è dovuta alla presenza o meno del titolo di superhost. Coloro che lo possiedono, danno al cliente garanzia di un servizio di qualità apprezzato dai precedenti ospiti, per questo induce ad un incremento del prezzo che incentiva una maggior disponibilità a pagare del cliente che desidera un servizio di qualità.

Un'altra variabile determinante sulla creazione del prezzo è la tipologia di ospite e quindi gli host che riescono maggiormente a segmentare i guest riusciranno ad ottenere un prezzo più vicino all'ottimo teorico e quindi ad incrementare il margine netto di guadagno.

Sono presenti, inoltre, diversi fattori che creano spillover spaziali, come l'accessibilità a determinati servizi, il rumore del traffico, la percorribilità a piedi e l'etnia dei residenti nel quartiere in cui si trova l'annuncio (Jorge Chica-Olmo, Juan Gabriel González-Morales, José Luis Zafra-Gómez, 2020). Sono molti i fattori che creano effetti di spillover spaziale ed hanno un evidente impatto sul prezzo finale della proprietà in affitto.

È stato dimostrato che gli effetti legati alla posizione spaziale di alcuni punti di interesse, come l'accessibilità al centro città, ai luoghi di interesse storico-culturale, alle spiagge per i luoghi turistici balneari, la percorribilità a piedi e la vicinanza e punti di snodo di mezzi di trasporto, quali metropolitane e treni, hanno un impatto significativo sui prezzi delle proprietà.

In particolare, attraverso l'uso di modelli di regressione spaziale, è stata analizzata la correlazione tra variabili di distanza ed i prezzi degli Airbnb di 10 città europee: Amsterdam, Atene, Barcellona, Berlino, Budapest, Lisbona, Londra, Parigi, Roma, e Vienna (Krist of Gyodi, Lukasz Nawaro, 2021). Gli attributi relativi a dimensioni, qualità e posizione sono tutti fattori significativi delle tariffe applicate dagli host. La conclusione a cui si è arrivati è che queste dipendono fortemente dalla posizione geografica. Le variabili spaziali che giocano

un ruolo fondamentale sono la distanza dal centro città, la distanza dai monumenti, piazze, edifici di interesse storico-culturale come palazzi, ville, musei, parchi, punti di accesso ai mezzi pubblici di trasporto o ai parcheggi. La tipologia di città a cui si fa riferimento porta con sé differenze nell'impatto delle diverse variabili sulle performance, ad esempio città balneari come Lisbona e Barcellona hanno come punti di maggior interesse le spiagge, mentre città storiche e culturali come Roma e Parigi hanno come punti di maggior interesse i monumenti e gli edifici storici, città invece legate al business come Amsterdam e Berlino hanno come punti di riferimento gli snodi di mezzo di trasporto come la metropolitana che garantisce un facile e rapido raggiungimento di tutti i punti della città.

La teoria della segnalazione sostiene che ci sono alcuni attributi nell'annuncio che danno un chiaro segnale all'utente su una o più caratteristiche dell'appartamento in modo tale da aumentare notevolmente la probabilità di essere prenotato. È stata analizzata una grande quantità di attributi delle inserzioni sulla piattaforma ed è emerso che quelle relative ad intere case con letti veri, con più camere, bagni e servizi come internet e cucina hanno maggiori probabilità di essere prenotate (Bin Yao, Richard T.R. Qiu, Daisy X.F. Fan, Anyu Liu, Dimitrios Buhalis, 2019).

Gli attributi di una proprietà possono essere classificati in tre categorie: funzionali, di segnale e duali (Bin Yao, Richard T.R. Qiu, Daisy X.F. Fan, Anyu Liu, Dimitrios Buhalis, 2019). Gli attributi funzionali riguardano caratteristiche intrinseche della proprietà, principalmente sulla facilità di prenotazione (instant booking e verifica dell'host), di localizzazione (vicinanza al centro città o alla metropolitana), alle funzionalità dell'alloggio (numero di stanze, numero di bagni). Gli attributi di segnale riguardano la piattaforma (verifica dell'ID), l'host (profilo, descrizione delle proprietà), il mercato (proprietà nelle vicinanze) e gli ospiti (recensioni, punteggio totale). Gli attributi duali, ovvero quelli che hanno entrambe le funzionalità, sono quelli di prezzo (prezzo per notte, prezzo per pulizia aggiuntiva, sconto settimanale, sconto mensile).

Sono stati analizzati anche fattori geografici come la concorrenza regionale e la popolarità che influenzano le performance. È stato però dimostrato che l'importanza degli attributi di segnale diminuisce con l'aumentare delle recensioni accumulate. Questo significa che, se un appartamento è molto ben recensito sulla piattaforma, le informazioni pubblicate dall'host passano in secondo piano. Da questa osservazione emerge una grande fiducia nelle esperienze degli host precedenti e grande flessibilità nelle scelte del soggiorno.

In letteratura sono stati confermati alcuni effetti spaziali come la vicinanza ai mezzi di trasporto, la vicinanza agli esercizi commerciali e gli effetti di spillover spaziale (Ruggero Sainaghi, Jorge Chica-Olmo, 2022). Questi hanno un effetto positivo misurato sulle performance degli Airbnb tramite l'analisi del RevPAR (Revenue Per Available Room), come illustrato in figura 2.

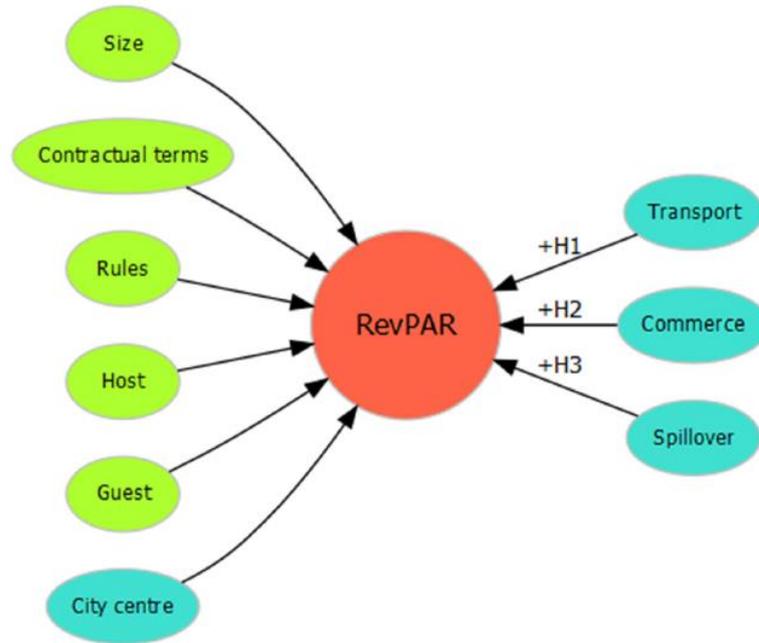


Figura 2: Ipotesi e nessi causali, Sainaghi

4. Obiettivo e contributo alla letteratura

Il paper intitolato “Determinants of Airbnb prices in European cities: a spatial econometrics approach” di Gyodi e Nawaro, pubblicato sul *Tourism Management* nel 2021, effettua un’analisi di correlazione tra le variabili spaziali ed il prezzo delle proprietà in 10 città europee. Poiché l’influenza delle variabili spaziali sulle performance degli Airbnb è stata trattata in pochissimi articoli di ricerca, si è deciso di indagare se esiste un effetto causale tra le distanze da punti di interesse noti e le variabili di performance e di quantificarlo. In un secondo tempo, si vuole indagare l’esistenza di possibili differenze tra le tipologie di host e tra le tipologie di soggiorno, ovvero se per i superhost ed i soggiorni di lungo periodo l’entità della correlazione è differente.

La città presa in esame è Berlino, una delle dieci città introdotte dall’articolo di Nawaro. La capitale della Germania ha una storia ricca e complessa, che include periodi come la Prima e la Seconda Guerra Mondiale, la divisione della città durante la guerra fredda e la riunificazione tedesca, è stata un importante centro culturale per molti decenni poiché ha ospitato artisti, scrittori, musicisti e intellettuali di fama mondiale, è una città con una varietà di stili architettonici e una storia urbanistica unica, è stata al centro di importanti eventi politici ed è una delle principali città economiche d'Europa e ha un ruolo significativo nell'economia globale.

Avendo a disposizione un dataset contenente gli annunci relativi alle 54048 proprietà attive a Berlino tra il 2019 ed il 2023, si vuole capire se gli attributi funzionali di location, ovvero le distanze spaziali da punti di interesse noti hanno effettivamente un’influenza sulle performance, in particolare sui ricavi per notte disponibile.

5. Studio di ricerca

5.1. Dati in input

Il dataset fornito in input è un file Excel csv scaricato dal sito web AirDNA, che fornisce i dati registrati da Airbnb sulle prenotazioni. Il dataset è di tipo panel e comprende gli annunci a Berlino tra il gennaio 2019 ed il dicembre 2023. In totale sono presenti 647811 osservazioni, ognuna delle quali è relativa ad un annuncio, di una singola proprietà, in un determinato mese. Questo permette di fare analisi comparate delle performance registrate dagli host nei periodi pre-Covid (2019), Covid (2020-2021) e post-Covid (2022-2023).

Le variabili a disposizione sono le seguenti.

Caratteristiche dell'annuncio:

- Property ID: codice alfanumerico identificativo della proprietà
- Reporting Month: mese di osservazione
- Year: anno di osservazione
- Airbnb Host ID: codice numerico identificativo dell'host

Variabile di performance:

- Revenue (USD): ricavi mensili
- Number of reservations: numero di prenotazioni in un mese
- Reservation Days: numero di giorni prenotati al mese
- Available Days: numero di giorni disponibili al mese

Caratteristiche della proprietà:

- Listing Type: tipologia
- Bedrooms: numero di letti disponibili
- Bathrooms: numero di bagni disponibili
- Max Guests: numero massimo di ospiti ammessi
- Latitude: latitudine
- Longitude: longitudine
- Neighborhood: quartiere di appartenenza

Scelte strategiche dell'host:

- Minimum Stay: numero minimo di giorni prenotabili

- Number of Photos: numero di foto presenti sull'annuncio

Valutazione degli host:

- Number of Reviews: numero di recensioni presenti sull'annuncio
- Overall Rating: punteggio complessivo da 1 a 100
- Airbnb Superhost: attributo che indica un'ospitalità eccezionale dell'host che soddisfa determinati criteri stabiliti da Airbnb

Le variabili di performance sono state calcolate a partire dal dataset e sono:

OCC (Occupation Rate): tasso di occupazione

[%]

$$OCC = \frac{Reservation\ Days}{Reservation\ Days + Available\ Days}$$

ADR (Average Daily Rate): ricavo medio per giorno prenotato

[\$/giorno]

$$ADR = \frac{Revenue}{Reservation\ Days}$$

RevPAN (Revenue Per Available Night): ricavo medio per notte disponibile

[\$/giorno]

$$RevPAN = OCC * ADR$$

5.2. Dati aggiuntivi

Sul sito del comune di Berlino sono stati scaricati quattro ulteriori dataset contenenti le seguenti informazioni:

- 173 fermate della metropolitana

- 44 monumenti storici
- 11 parchi
- 17 parcheggi

Le informazioni sono state raccolte in un file Excel in forma tabellare.

5.2.1. Calcolo coordinate dei punti di riferimento

Sfruttando il linguaggio di programmazione Python sono state calcolate tutte le coordinate geografiche in formato (latitudine, longitudine) dei 245 punti di riferimento sopra elencati. Sono stati utilizzati la libreria Geopy e l'oggetto Geocator. L'output ottenuto è un file Excel con l'elenco dei punti di interesse e le loro coordinate geografiche).

5.2.2. Calcolo distanze delle proprietà dai punti di riferimento

È stato utilizzato Python anche per il calcolo della distanza minima tra gli Airbnb e le quattro tipologie di punti di interesse. L'obiettivo era calcolare per ogni proprietà la distanza minima da ciascun punto. Per ogni coppia di proprietà e punto di interesse è stata quindi calcolata la distanza e ad ogni calcolo è stata tenuta in memoria la distanza minima.

Nella pratica si è inizializzata la distanza minima tra proprietà e punto di interesse a 0 e poi, per ognuna delle 54048 Property ID del database, è stato creato un ciclo che confrontasse la distanza da ogni punto di interesse alla proprietà e che per ogni categoria di punto confrontasse la nuova distanza con il valore minimo. Se il nuovo valore fosse stato inferiore sarebbe stato memorizzato come distanza minima, se superiore sarebbe stato scartato.

In questo modo, l'output ottenuto riporta le seguenti informazioni:

- Property ID
- Stazione della metropolitana più vicina
- Distanza dalla stazione più vicina
- Monumento più vicino
- Distanza dal monumento più vicino
- Parco più vicino
- Distanza dal parco più vicino
- Parcheggio più vicino

- Distanza dal parcheggio più vicino

Per il calcolo della distanza è stata usata la formula di Haversine, che assume che la Terra sia una sfera perfetta, non tiene conto delle sue irregolarità, ma fornisce comunque una buona approssimazione per distanze relativamente brevi.

La formula è la seguente:

$$d = 2 * r * \arcsin \left(\sqrt{\sin^2 \left(\frac{\varphi_2 - \varphi_1}{2} \right) + \cos \varphi_1 \cos \varphi_2 \sin^2 \left(\frac{\lambda_2 - \lambda_1}{2} \right)} \right)$$

d è la distanza tra i due punti, r è il raggio della sfera (per la Terra, il suo valore medio è di circa 6371 km), φ_1 e φ_2 sono le latitudini delle proprietà (in radianti), λ_1 e λ_2 sono le longitudini delle stazioni metro (in radianti).

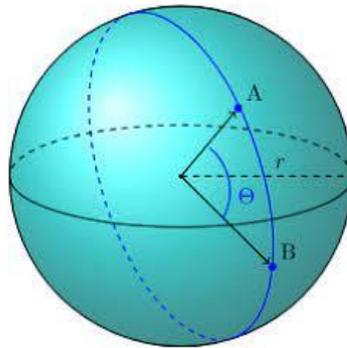


Figura 3: Latitudine e longitudine terrestri

Latitudini e longitudini in input, essendo in gradi, sono state convertite precedentemente in radianti, moltiplicando il valore in gradi per $\frac{\pi}{180}$.

Di seguito è riportato un esempio del codice Python utilizzato:

```
def haversine(lat1, lon1, lat2, lon2):
    # Raggio della Terra in metri
    R = 6371000.0
    # Conversione delle coordinate da gradi a radianti
    lati_rad = math.radians(lat1)
    lon1_rad = math.radians(lon1)
    lat2_rad = math.radians(lat2)
```

```

lon2_rad = math.radians(lon2)
# Differenze nelle coordinate
dlat = lat2_rad - lati_rad
dlon = lon2_rad - lon1_rad
# Formula di Haversine
a = math.sin(dlat / 2)*2 + math.cos(lati_rad) * math.cos(lat2_rad) * math.sin(dlon / 2)*2
c = 2 * math.atan2(math.sqrt(a), math.sqrt(1 - a))
# Distanza in metri
distance_meters = R * c
return distance_meters

```

Successivamente è stato fatto un merge del dataset iniziale con le nuove colonne di dati ottenute con Python.

5.3. Strumenti software

Il software utilizzato per le successive analisi statistiche ed econometriche dei dati è Stata, che ha permesso di effettuare in primis delle analisi descrittive delle variabili di interesse appartenenti al dataset e poi di correlazione e di regressione, univariate e multivariate, per poter stabilire la presenza di un nesso causale tra una variabile indipendente (le distanze) e una o più variabili dipendenti (le performance), con l'inserimento di variabili di controllo e moderanti.

Con Stata ed Excel sono stati anche realizzati diagrammi a torta, diagrammi a barre, istogrammi e scatterplot, che consentono di ottenere una visione immediata delle informazioni di percentuali, distribuzioni di densità e trend di variabili.

6. Domanda di ricerca

6.1. Punto di Partenza

Lo spunto per arrivare alla domanda di ricerca è arrivato inizialmente con la lettura di due articoli. Il primo è intitolato “Determinants of Airbnb prices in European cities: a spatial econometrics approach” di Gyodi e Nawaro e che analizza la correlazione tra variabili spaziali e prezzi degli Airbnb in 10 città europee. Il secondo è “Standing out from the crowd – an exploration of signal attributes of Airbnb listings” di Buhalis, in cui sono presentati gli attributi delle proprietà e degli annunci che influiscono sulla decisione di prenotazione e sui prezzi per notte disponibile.

Questi attributi sono suddivisi in tre categorie.

La prima è quella dei functional attributes, ovvero delle caratteristiche della proprietà che influenzano il soggiorno del guest (ad esempio la facilità di prenotazione, la localizzazione geografica e le funzioni della proprietà).

La seconda è quella dei signal attributes, ovvero caratteristiche che indicano qualità o valore per gli ospiti (ad esempio gli attributi della piattaforma, degli host, del mercato o dei guest).

La terza è quella dei dual attributes, ovvero le caratteristiche della proprietà o dell’annuncio che influenzano la modalità di consumo del servizio e allo stesso tempo indicano alta qualità (ad esempio il prezzo).

Al fondo dell’articolo è presentata la Signal Theory secondo cui gli host usano i signal attributes per comunicare in modo efficace la qualità della loro proprietà agli ospiti potenziali (ad esempio con lo status di super host, le recensioni degli ospiti precedenti, il prezzo o le spese aggiuntive).

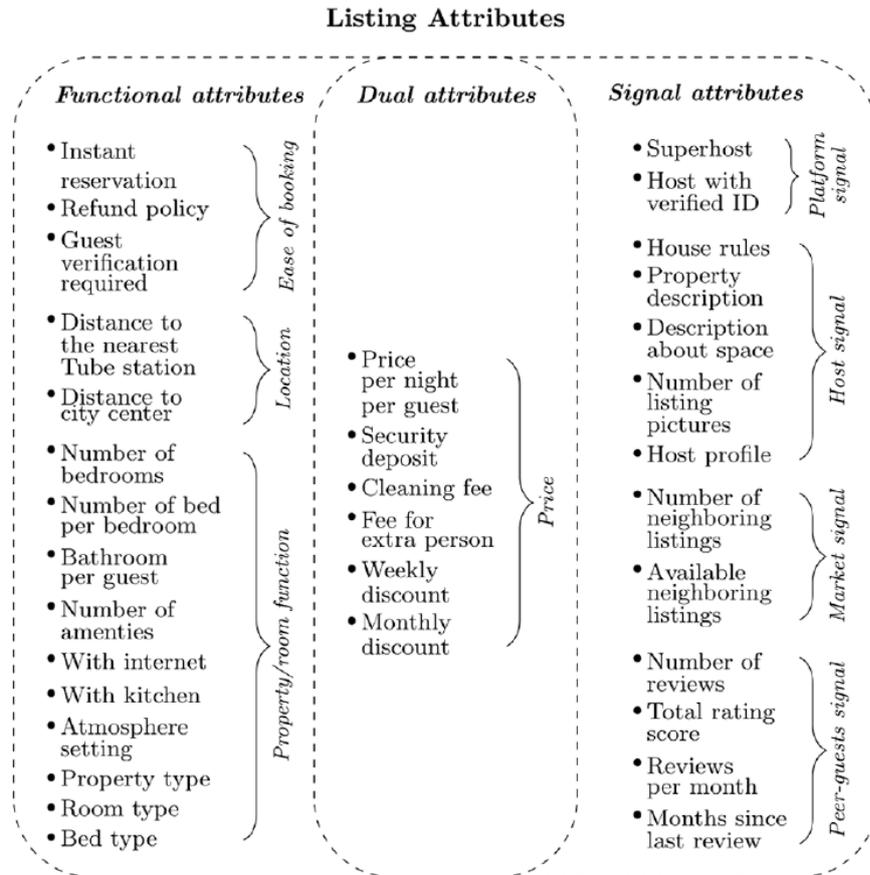


Figura 4: Listing Attributes, Buhalis 2019

6.2. Research gap

Il gap di presente in letteratura e che ci si propone di colmare è allora l'influenza dei functional attributes, in particolare quelli di location sulle performance degli Airbnb, con focus sulla città di Berlino. I location attributes considerati sono le distanze spaziali da metropolitana, monumenti, parchi e parcheggi perché considerati di interesse turistico.

6.3. Domanda di ricerca

Dall'approfondita analisi di questi articoli l'idea scaturita è stata quella di analizzare l'impatto delle distanze dalle quattro categorie di punti noti, sui ricavi per notte disponibile, ovvero il RevPAN. Si vuole anche indagare l'eventuale presenza di differenze per i superhost e per le prenotazioni di lungo periodo, ovvero con soggiorno minimo di tre mesi.

7. Research Framework e ipotesi

Il quadro di ricerca è costituito da quattro variabili indipendenti, che sono le distanze dalla stazione metro più vicina, dal monumento più vicino, dal parco più vicino e dal parcheggio più vicino, misurate in metri. La variabile dipendente è il RevPAN, misurato in dollari per notte. Le variabili di controllo inserite nelle analisi di regressione sono Max Guests, Overall Rating, Listing Type, Host Type, Neighborhood, Year e Month. Le variabili moderanti prese in considerazione sono le variabili dummy Superhost e Long Stay. I valori di queste possono essere 0 se l'attributo è falso oppure 1 se è vero.

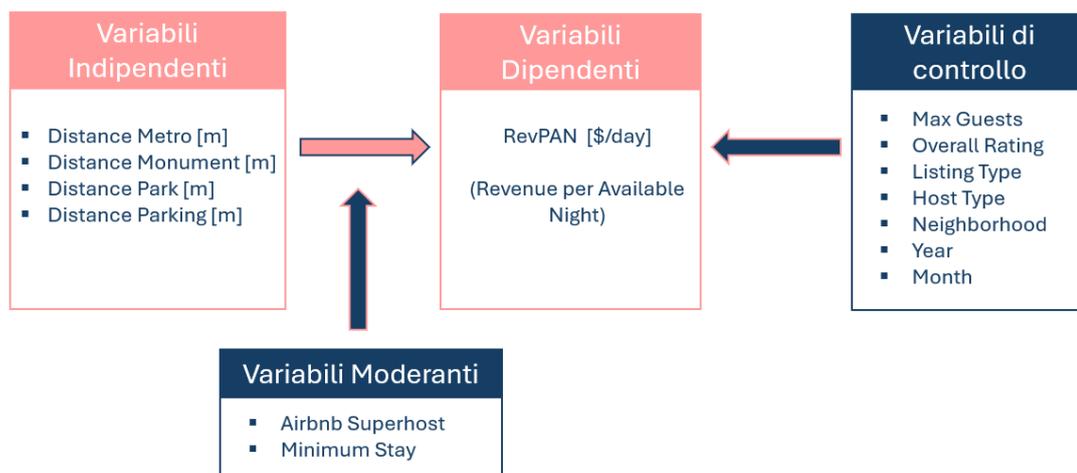


Figura 5: Research Framework

Da questo schema sono state realizzate tutte le regressioni eseguite su Stata per poter confermare o confutare le ipotesi seguenti:

Hp 1) La distanza dalla metropolitana più vicina ha un impatto negativo sul RevPAN.

Hp 2) La distanza dal monumento più vicino ha un impatto negativo sul RevPAN.

Hp 3) La distanza dal parco più vicino ha un impatto sul RevPAN.

Hp 4) La distanza dal parcheggio più vicino ha un impatto negativo sul RevPAN.

Hp 5-6) Per i superhost e per i soggiorni di lungo periodo la magnitudine dell'impatto negativo è maggiore.

8. Analisi statistiche descrittive

La seconda parte dello studio si occupa di fare un'analisi preventiva sulle variabili presenti nel dataset di input, che possono essere classificate in: caratteristiche della proprietà, caratteristiche dell'host, variabili strategiche e di performance. È approfondito l'andamento generale, tramite l'osservazione di operatori statistici come la media e la mediana, che sono misure di centratura, deviazione standard e percentili, come misure di dispersione, valori minimo, massimo e la distribuzione dei dati. Essendo il dataset di tipo panel è stato possibile comparare le performance in anni successivi trovando le differenze tra i periodi pre-Covid, Covid e post-Covid.

Con l'aiuto di grafici a torta, diagrammi a barre e istogrammi si possono rendere queste informazioni, raccolte ed analizzate, facilmente visibili ed interpretabili.

Sono di seguito riportate le analisi eseguite.

8.1. Caratteristiche della proprietà

8.1.1. Property ID

Nel grafico a barre è riportato il numero di proprietà attive sulla piattaforma per ogni anno. Con l'avvento del Covid il numero di proprietà è diminuito del 57% nel 2021. Dopo la pandemia il numero ha continuato a decrescere, registrando nel 2023 un calo del 70,6% rispetto al 2019.

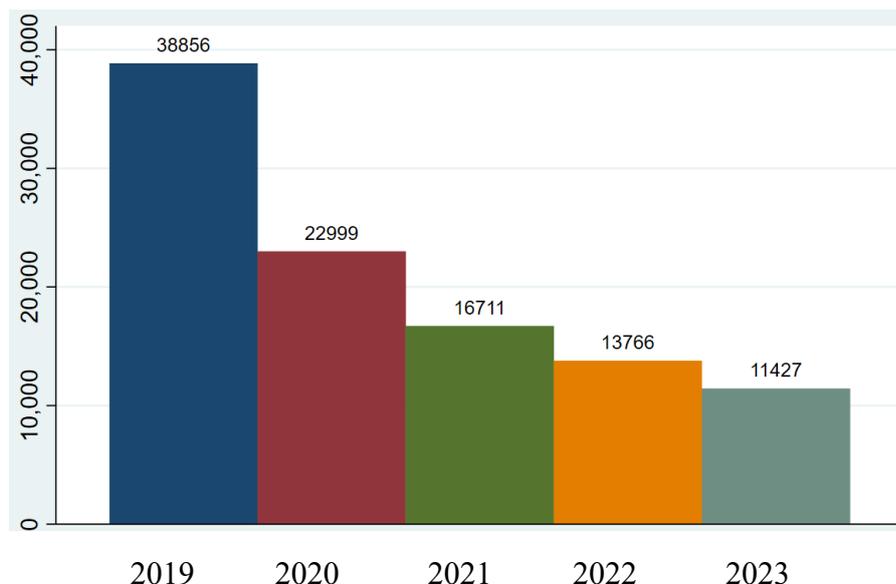


Figura 6: Numero di Property ID attive suddivise per anno

8.1.2. Listing Type

Le tipologie di proprietà sono quattro: appartamento/casa intero/a, stanza privata, stanza condivisa, stanza d'hotel. In generale, è evidente che la casa/ appartamento intero/a siano la scelta preferita (il 58,47% del totale), seguita dalla stanza privata (37,94%), mentre solo una percentuale molto bassa di utenti prenota in una stanza privata (1,82%) o stanza d'hotel (1,77%).

Listing Type	Freq. Ass.	Freq. Rel.	Freq. Cum.
Entire home/apt	378,806	58,47%	58,47%
Hotel Room	11,446	1,77%	60,24%
Private Room	245,779	37,94%	98,18%
Shared Room	11,78	1,82%	100,00%
Tot.	647,811	100,00%	100,00%

Tabella 1: Frequenza assoluta, relativa e cumulata del numero di proprietà, suddivise per Listing Type

Dall'analisi dei dati si evince che, durante e dopo la pandemia, parte dei clienti che optava per una stanza privata è passata a prenotare una intera casa o appartamento. Nel 2019 il 52,34% di ospiti optava per una casa intera, nel 2020-21 il 58,64% ed infine il 65,99% nel 2022-23.

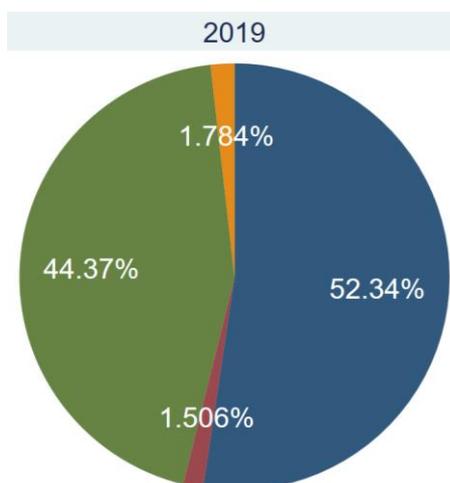


Figura 7: Percentuali di Property ID, suddivise per listing type nel 2019

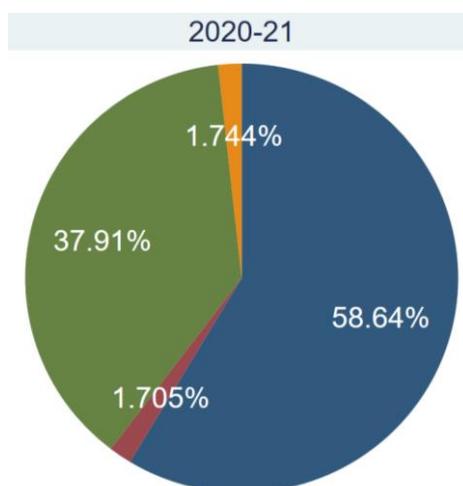


Figura 8: Percentuale di Property ID, suddivise per listing type nel 2020-21

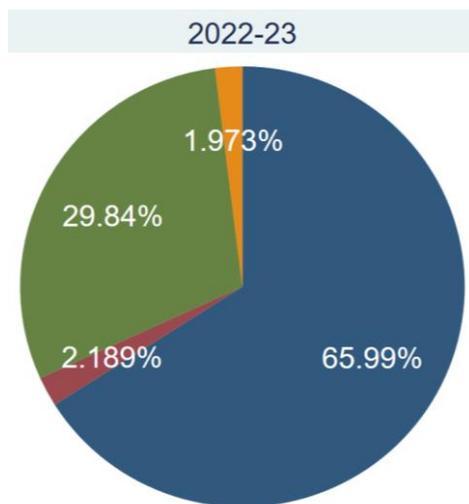
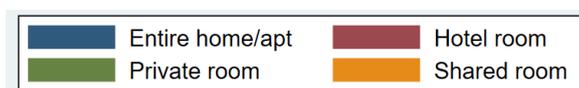


Figura 9: Percentuali di Property ID, suddivise per listing type nel 2022-23



Nei grafici seguenti sono riportati il numero medio di prenotazioni, di giorni prenotati e di giorni disponibili mensili, suddivisi per tipologia di proprietà e raggruppati per periodo. Da questi si evince che prima del Covid il numero medio di giorni prenotati era molto alto per casa intera, stanza d'hotel e stanza privata, con preferenza massima per la stanza d'hotel. Nel periodo pandemico, tutti i valori si abbassano e diventa la casa intera la scelta con massimo numero di giorni prenotati. Questa rimane anche nel post-Covid. Il numero di giorni disponibili ha andamento opposto: la stanza condivisa durante la pandemia ha maggior disponibilità di giorni e numero di prenotazioni più basso. Nel pre-Covid la scelta preferita è la stanza d'hotel, nel periodo Covid e nel post-Covid la casa intera.

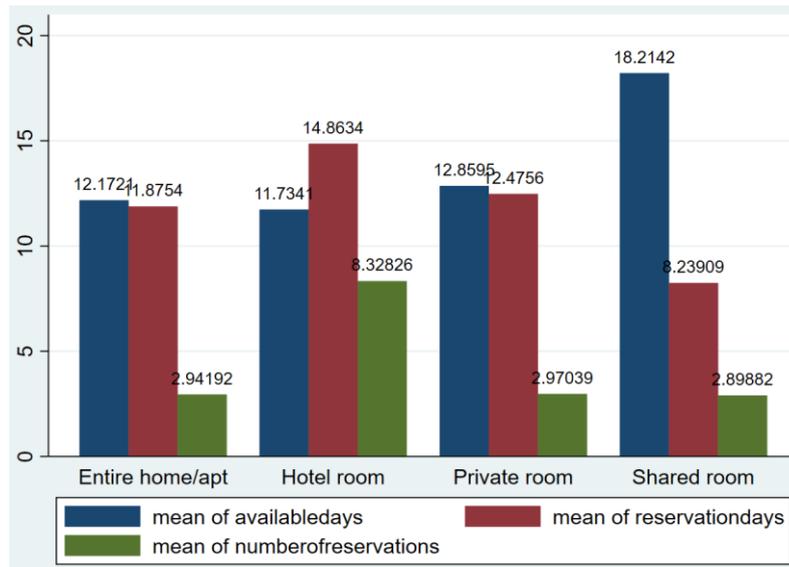


Figura 10: Numero medio di Available Days, Reservation Days, Number of Reservation Days 2019

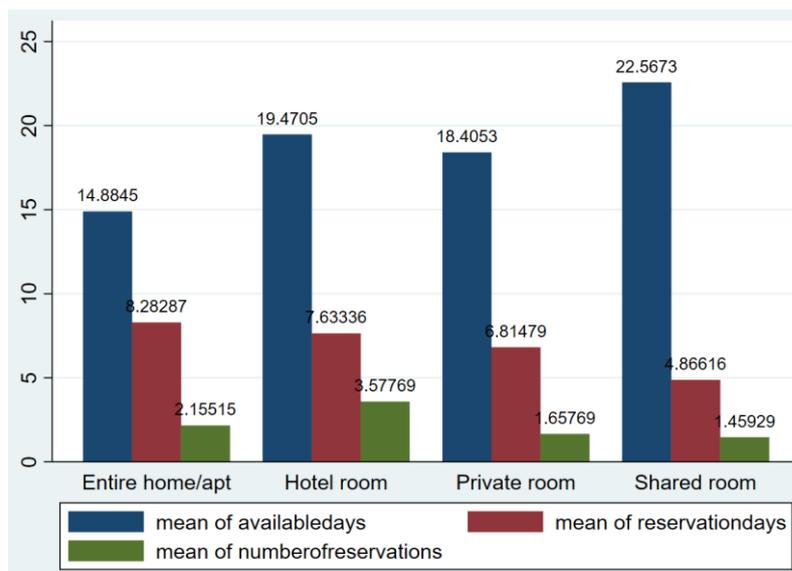


Figura 11: Numero medio di Available Days, Reservation Days, Number of Reservation Days nel 2020-21

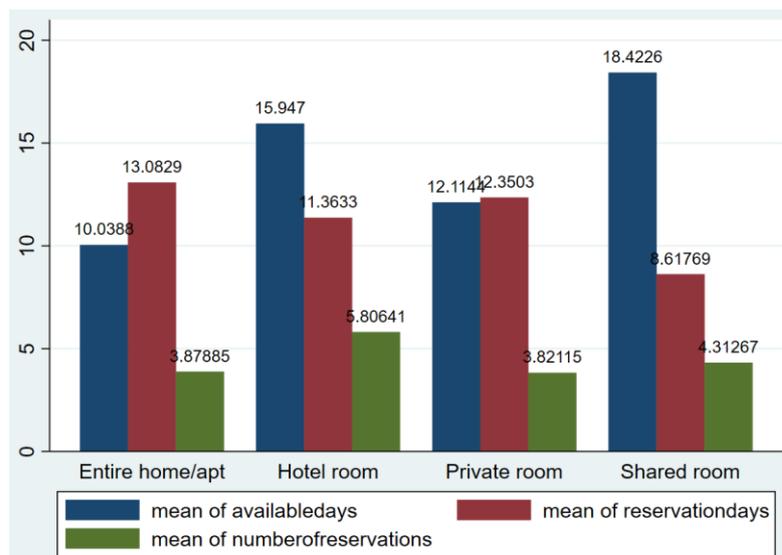


Figura 12: Numero medio di Available Days, Reservation Days, Number of Reservation Days nel 2022-23

8.1.3. Neighborhood

Di seguito è possibile vedere una distribuzione delle proprietà nei quartieri di Berlino. Il 22,27% si trova nel quartiere Friedrichshain-Kreuzberg, il 21,35% in Pankow il 13,20% in Neukolln ed il 15,19% in Mitte, che sono quartieri centrali. Nei restanti quartieri la percentuale di Airbnb presenti risulta essere piuttosto bassa.

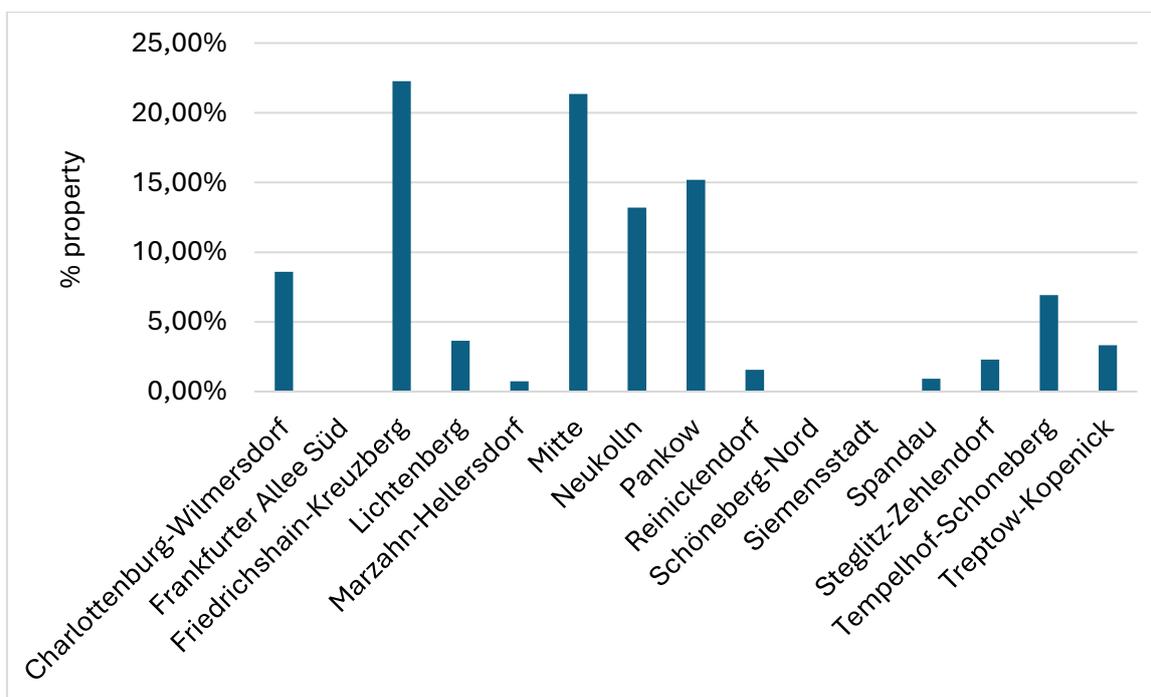


Figura 13: Percentuale di Property ID, suddivise per Neighborhood

La mappa rappresenta la diffusione degli Airbnb a Berlino, aggiornata al 2023 con i dati di AirDNA. Il numero di proprietà è stato suddiviso per quartieri e con tonalità più scure di blu si indicano i quartieri con densità maggiore di Airbnb, mentre con tonalità più chiare i quartieri con densità minore.

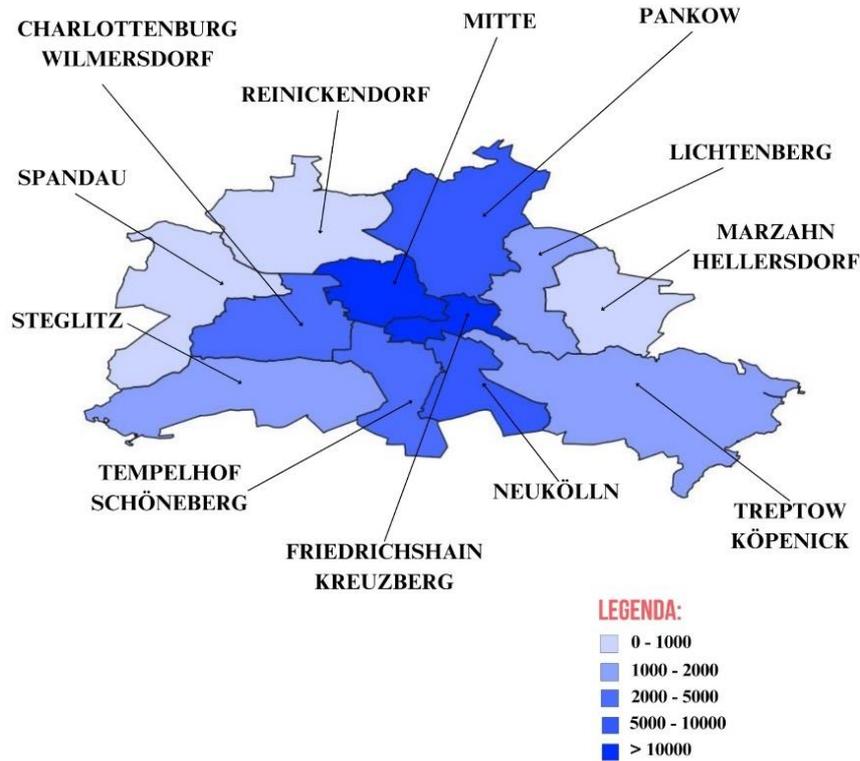


Figura 14: Diffusione Airbnb a Berlino nel 2023

8.1.4. Bedrooms

Il numero di camere da letto per il 77,07% degli annunci è una, per il 16,41% è due, solo per il 4,88% è tre e l'1,18% è quattro. Percentuali residuali di annunci hanno più di quattro camere da letto.

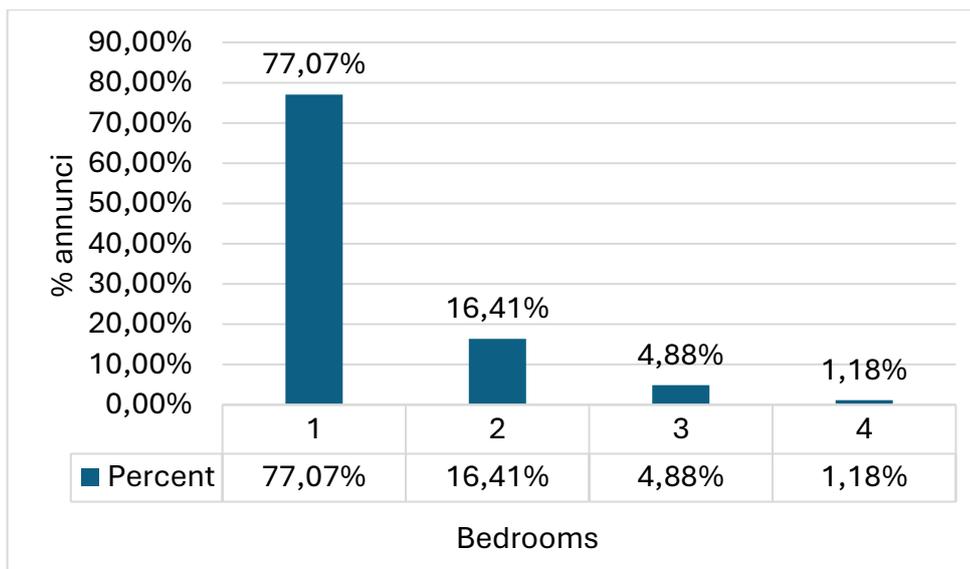


Figura 15: Percentuali di annunci, suddivise per Bedrooms

8.1.5. Bathrooms

Nell'84,02% di proprietà c'è a disposizione un bagno, nel 14,23% due bagni, nell'1,09% tre bagni e in percentuali residuali (non rappresentate nel grafico a barre) più di tre bagni.

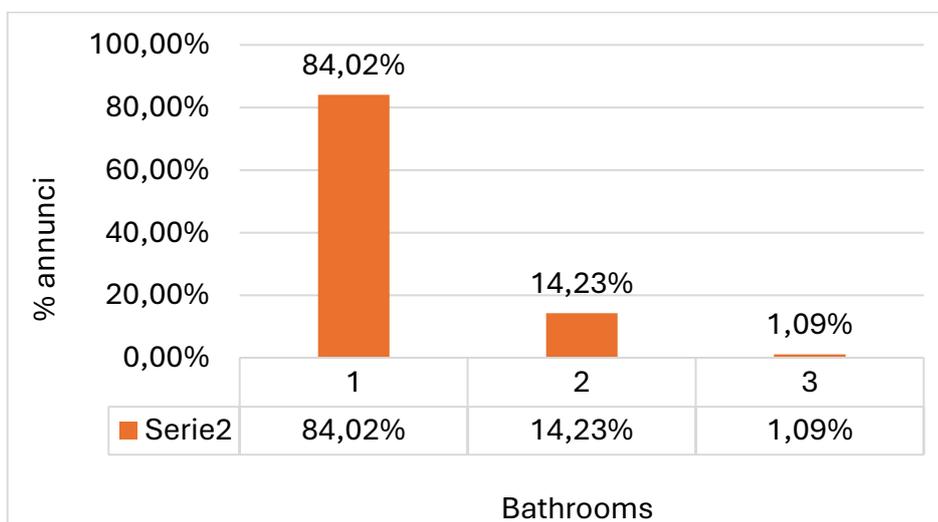


Figura 16: Percentuali di annunci, suddivisi per numero di Bathrooms

8.2. Scelte strategiche dell'host

8.2.1. Max Guest

Dal sottostante grafico a barre, è evidente che la tendenza del valore medio di Reservation Days è crescente con Max Guest.

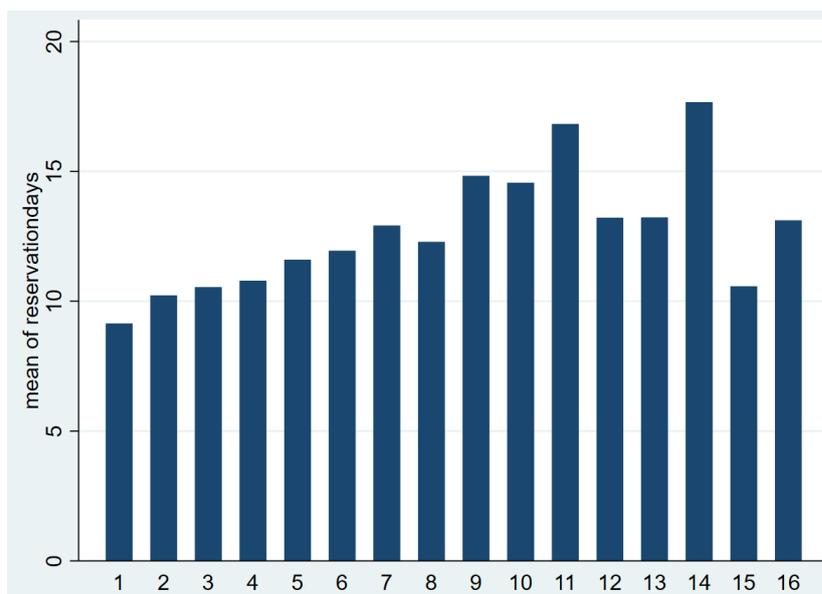


Figura 17: Valore medio di Reservation Days rispetto a Max Guests

Questo trend è rimasto sostanzialmente costante sia nel periodo pre, durante e post pandemia.

8.2.2. Minimum Stay

La variabile Minimum Stay indica il numero minimo di giorni di soggiorno richiesti dall'host. Dall'analisi descrittiva si evince che solo nel 2020 c'è stata una decrescita del 15,74%. Nel post-Covid si assiste ad un notevole incremento del valore con un picco dell'86,38% nel 2022.

Minimum Stay	Pre-Covid	Covid		Post-Covid	
	2019	2020	2021	2022	2023
Media	9,2974	7,8413	10,8361	17,3281	16,9025
Dev. Std.	87,541	27,7428	34,77	39,9232	35,04
Min	1	1	1	1	1
Max	9999	1000	1000	1000	365
25°percentile	2	1	1	1	1
50°percentile (mediana)	2	2	3	3	3
75°percentile	4	4	5	6	5

Figura 18: Statistiche descrittive di Minimum Stay

8.2.3. Number of Photos

Il numero di foto pubblicate dall'host sull'annuncio è in media superiore del 26,24% per gli host professionali e del 29,57% superiore per i superhost.

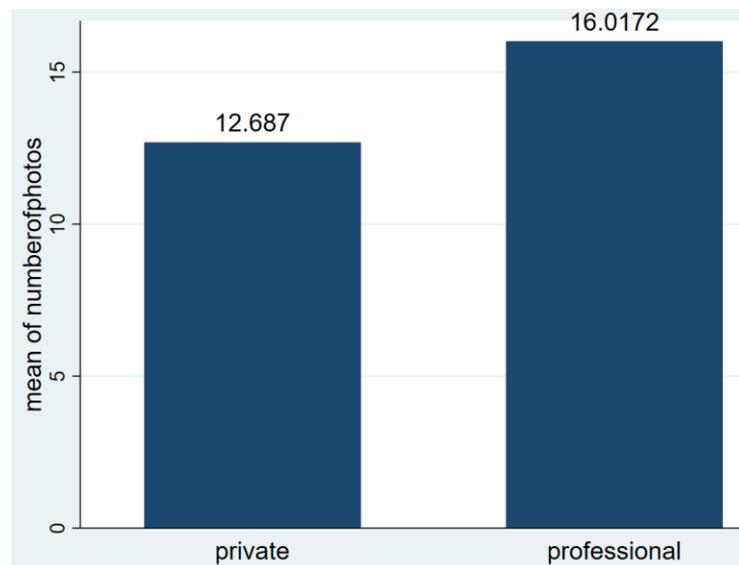


Figura 19: Valore medio di Number of Photos, suddiviso per Host Type

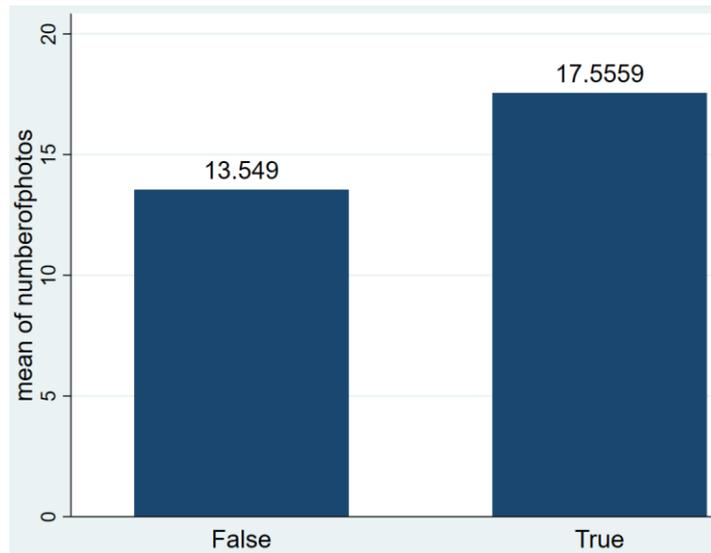


Figura 20: Valore medio di Number of Photos, suddiviso per superhost e non superhost

8.3. Contributo dei guest

8.3.1. Number of Reviews

Il numero di recensioni per annuncio ha un minimo di 0 ed un massimo di 2420. Il valore medio è circa 45 con una deviazione standard di 76. I professional host hanno in media il 32,58% di recensioni in più dei private host. In media, rispetto alle stanze d'hotel, le case intere hanno il 37,14% di recensioni in più, le stanze private il 6,67% e le stanze condivise il 22,85%.

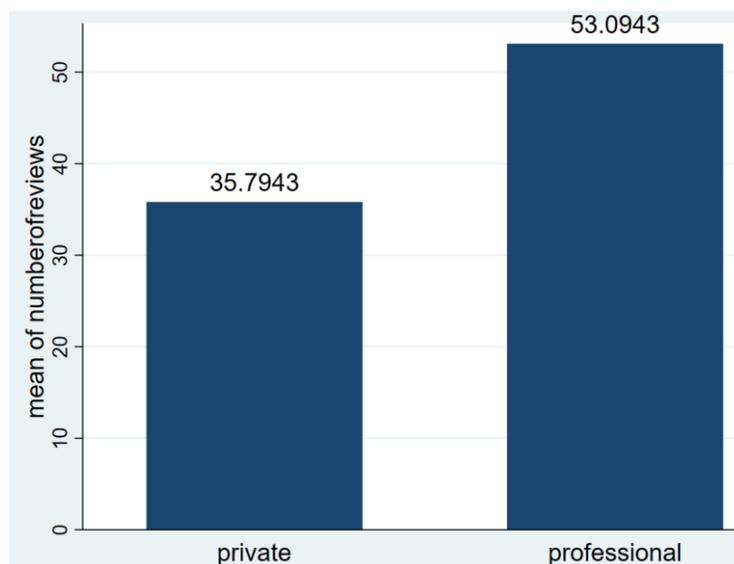


Figura 21: Valore medio di Number of Reviews, suddiviso per Host Type

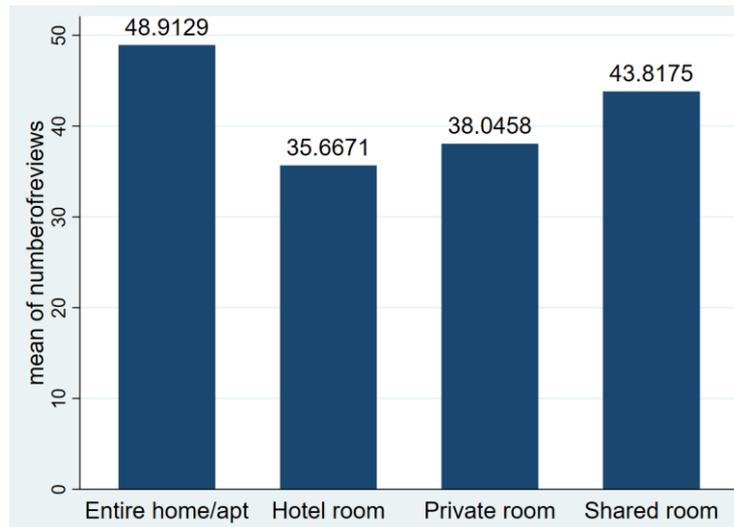


Figura 22: Valore medio di Number of Reviews, suddiviso per Listing Type

8.3.2. Overall Rating

L'Overall Rating è il punteggio assegnato dagli utenti ai soggiorni e rispecchia il loro livello di soddisfazione. Il rating va da 1 a 100. Il valore medio è 29,40 con una deviazione standard di 40,07.

L'Overall Rating è in media superiore del 29,6% per gli host professionali, rispetto ai privati. Il valore più basso è assegnato alle stanze private (24,29), il più alto alle stanze d'hotel (36,79), con uno scarto di 12,14 punti.



Figura 23: Valore medio di Overall Rating, suddiviso per Host Type

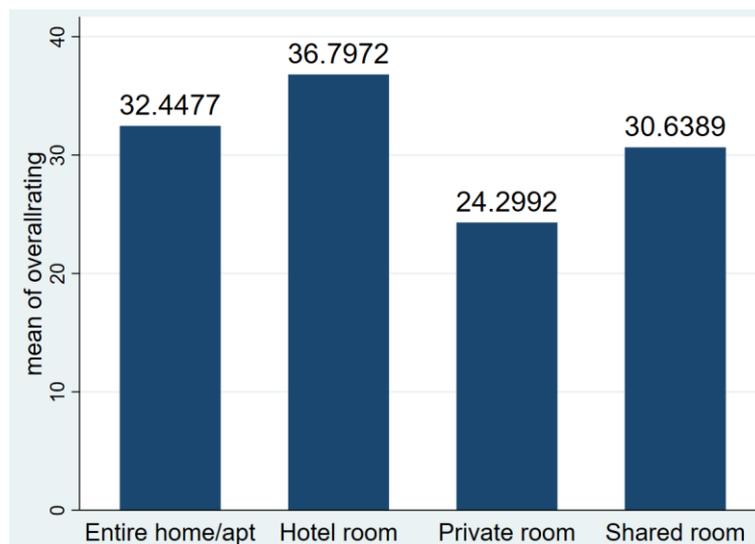


Figura 24: Valore medio di Overall Rating, suddiviso per Listing Type

La distribuzione di densità della variabile è molto dispersa, con molti valori inferiori al 10 e oltre 80.

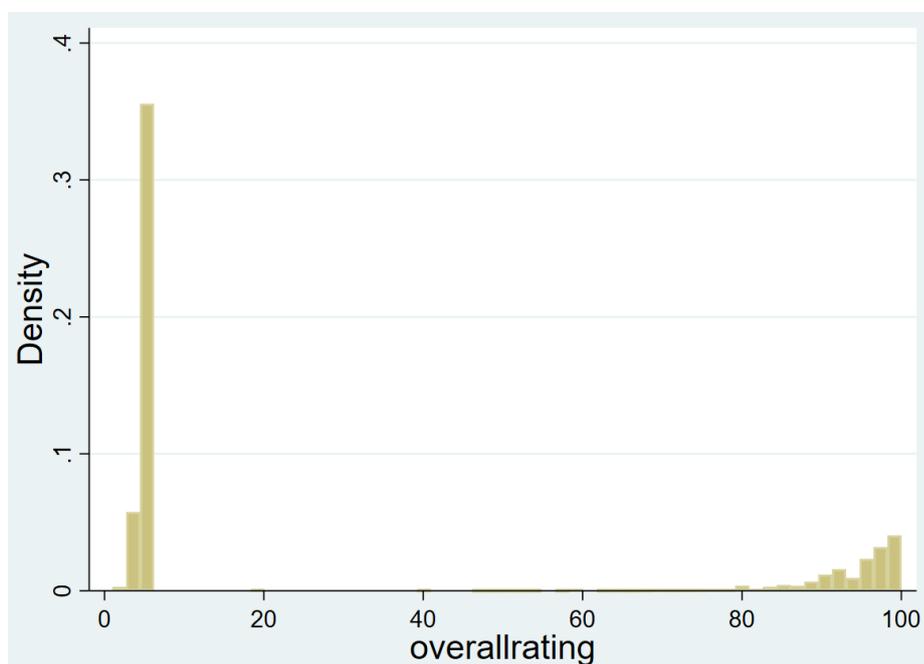


Figura 25: Distribuzione di densità discreta della variabile Overall Rating

8.4. Caratteristiche dell'host

8.4.1. Host type

Nel dataset sono presenti annunci pubblicati da 38680 host identificati dalla variabile Airbnb Host ID.

Gli host possono essere suddivisi in due cluster: “professional”, ovvero coloro che gestiscono più di una proprietà al mese, e “private”, che gestiscono una sola proprietà al mese.

Nel 2019 la maggior parte degli host è private, il 58,06% del totale. Durante la pandemia la percentuale diminuisce: nel 2020 è appena il 50,65% e nel 2021 si inverte la tendenza e la maggioranza diventa di tipo professional, il 53,13% del totale. La tendenza nel periodo post pandemico rimane la medesima, ovvero la percentuale di host professionali aumenta a discapito di quella dei privati. I professionali sono il 57,11% nel 2022 ed il 59,59% nel 2023.

Host Type		pre-Covid	Covid		post-Covid	
		2019	2020	2021	2022	2023
Frequenza Assoluta	Private	126,76	72,368	53,489	37,627	43,253
	Professional	91,555	70,52	60,626	50,109	50,508
Frequenza Percentuale	Private	58,06%	50,65%	46,87%	42,89%	40,41%
	Professional	41,94%	49,35%	53,13%	57,11%	59,59%

Tabella 2: Frequenze assoluta e relativa di Host Type, suddivisi per periodo

I ricavi medi di host privati e professionali sono molto diversi: i professionali guadagnano mediamente l'87% in più.

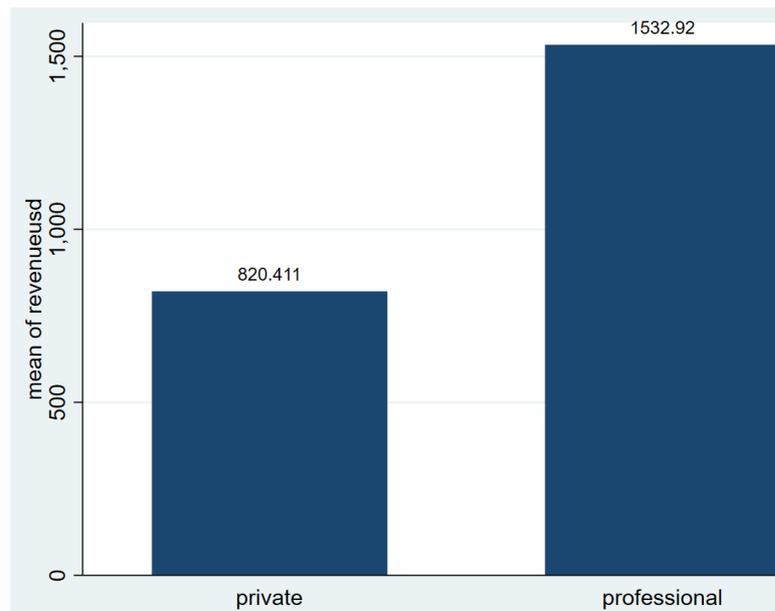


Figura 26: Valore medio di Revenue, suddiviso per Host Type

Nel grafico sottostante sono riportati i valori medi di ADR e RevPAN, rispettivamente per host privati e professionali. I professionali hanno un ADR in media maggiore del 30,86% ed un RevPAN maggiore del 26,22%, rispetto ai privati.

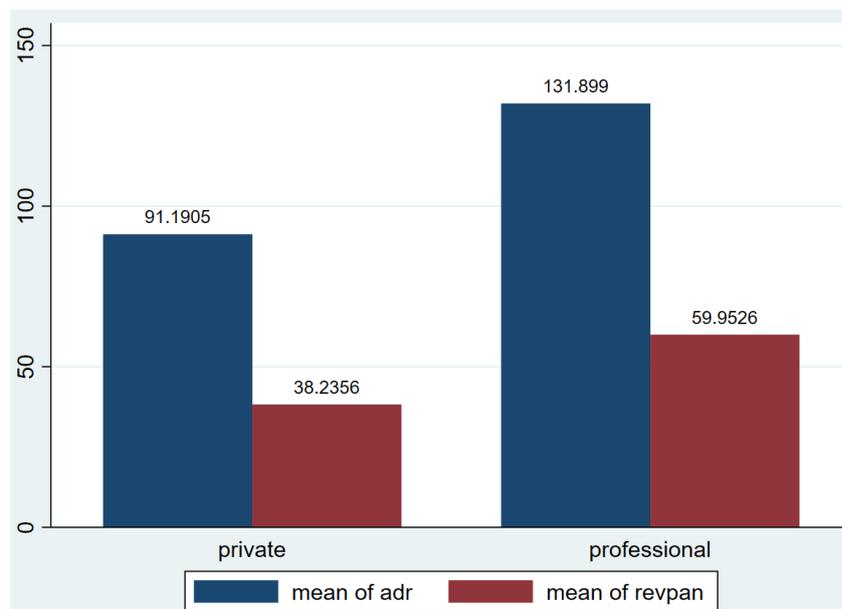


Figura 27: Valori medi di ADR e RevPAN, suddivisi per Host Type

Nel seguente grafico si riporta invece, con la stessa suddivisione, il valore medio del numero di prenotazioni mensili, del numero di giorni prenotati e disponibili. I professional host registrano un numero medio di giorni prenotati maggiore del 21,71% rispetto ai private ed un numero medio di prenotazioni maggiore del 38,52%.

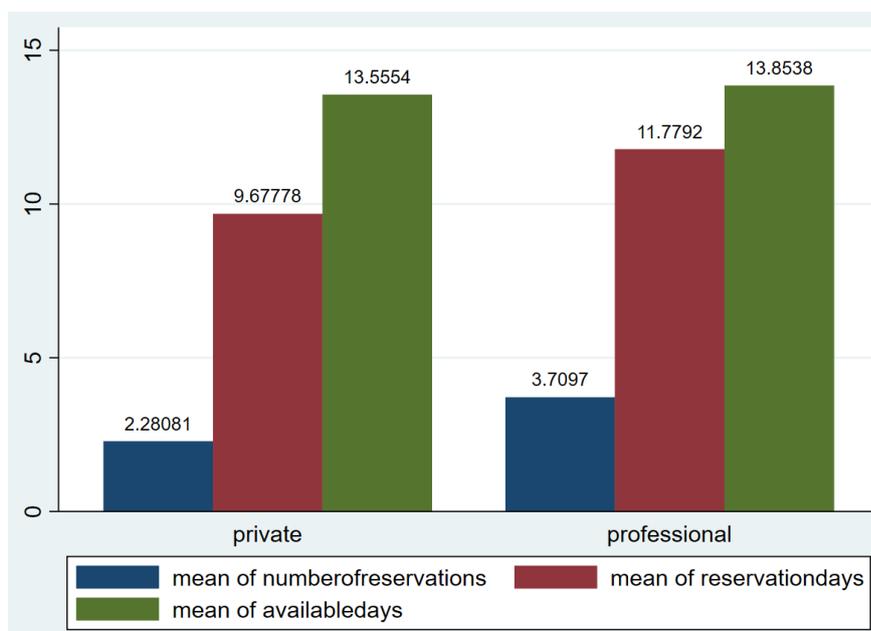


Figura 28: Valore medio di Number of Reservations, Reservation Days e Available Days, suddivisi per Host Type

8.4.2. Superhost

Gli host possono essere anche classificati in base al possesso o meno del titolo di superhost. La percentuale di superhost è il 20,69% del totale nel 2019, cresce al 25,12% nel 2020. Si arresta intorno al 13% circa nel 2021 e 2022, per poi tornare a crescere ad un livello maggiore nel 2023, ovvero al 34,53%. In tutti e cinque gli anni si può vedere che la maggior parte degli host non possiede il badge di superhost. Solo nel 2023 la differenza tra le due percentuali diminuisce: i non superhost scendono al 65,47%.

Nell'istogramma seguente si riportano le percentuali di host con e senza il titolo di superhost, suddivisi per periodo.

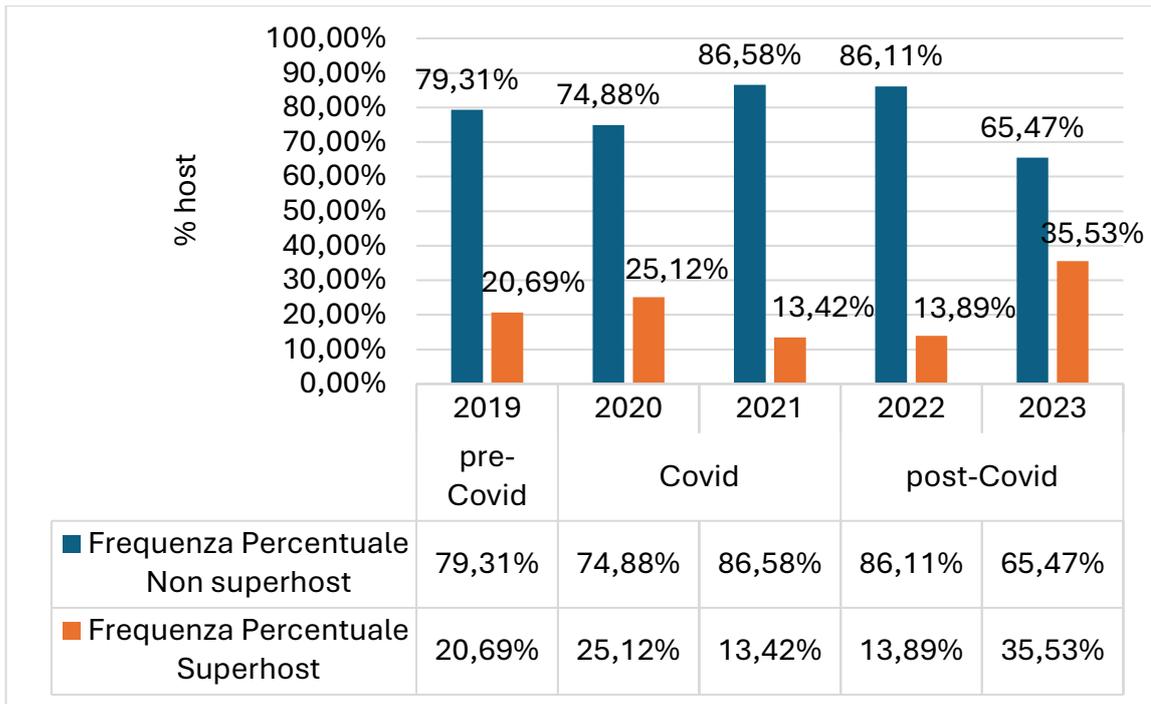


Figura 29: Frequenze percentuali di host, suddivise tra superhost e non superhost e per periodo

8.5. Variabili spaziali

8.5.1. Metro station

La rete della metropolitana di Berlino è composta da 173 fermate raggruppate in 10 linee. La variabile Metro Station indica per ogni proprietà la fermata della metropolitana più vicina.



Figura 30: Rete della metropolitana di Berlino

Nel grafico sono rappresentate le stazioni della metropolitana con una percentuale di proprietà vicine non trascurabile.

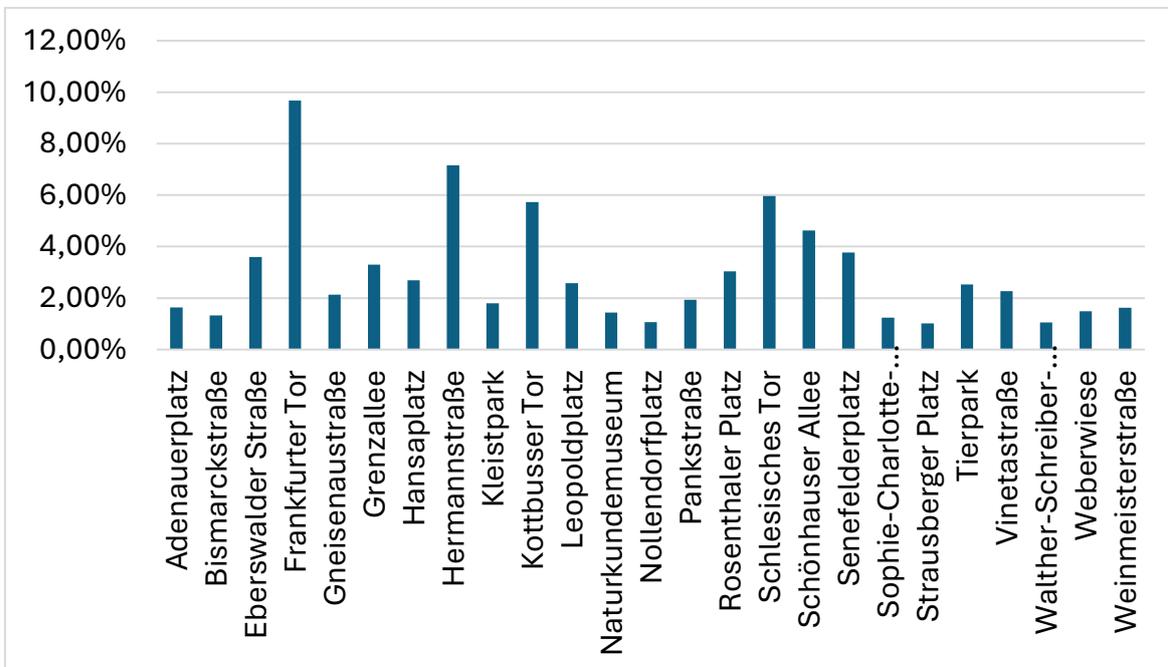


Figura 31: Percentuali di Property ID suddivise per stazione metro più vicina

8.5.2. Metro Station Distance

La distribuzione della distanza in metri delle proprietà dalla stazione della metropolitana più vicina è rappresentata nel seguente istogramma, da cui si vede un'alta concentrazione tra i valori 0 e 500 m e bassa altrove.

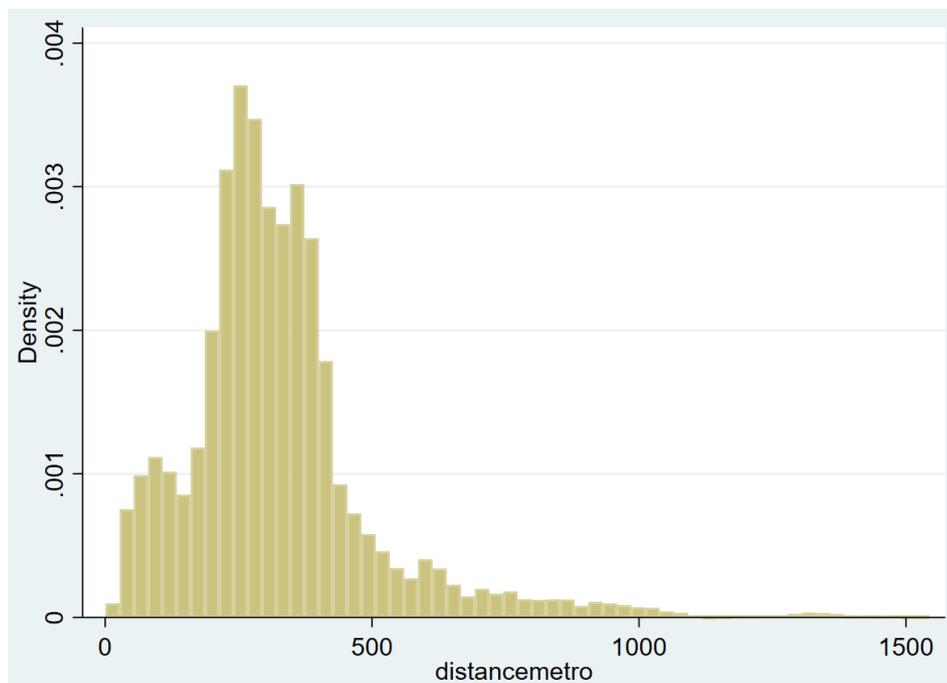


Figura 32: Distribuzione di densità discreta della variabile Distance Metro

Nel grafico a barre è invece riportato il numero di proprietà per fasce di distanza dalla metropolitana. La maggior parte di esse si trova nelle prime due fasce (A e B), ovvero a meno di 500 metri da una fermata. Il 31,95% delle proprietà dista meno di 250 metri dalla metropolitana, il 58,54% tra i 250 ed i 500 m e le proprietà rimanenti distano oltre 500 metri. Soltanto una percentuale trascurabile di proprietà si trova oltre i 2 km di distanza dalla metro.

Questo indica che la città di Berlino è ben fornita rete dalla metropolitana e gli Airbnb sono distribuiti in modo da essere sufficientemente vicini.

Nel seguente grafico a barre è riportato il numero di proprietà a Berlino suddivise per fasce di distanza.

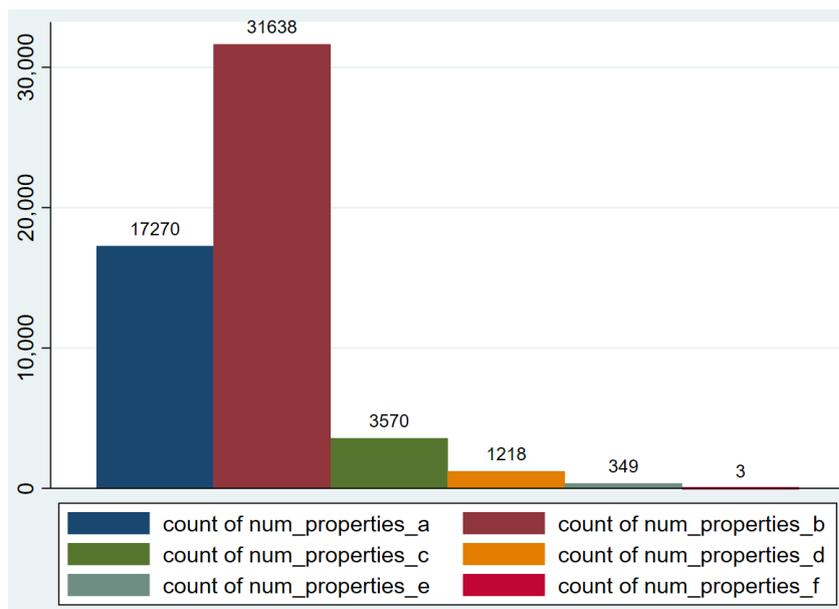


Figura 33: Numero di Property ID suddivise per fasce di distanza dalla metropolitana

8.5.3. Monument

La variabile Monument assegna ad ogni proprietà il monumento, piazza, edificio, museo o punto di interesse storico più vicino ad essa. Nel seguente diagramma a barre è mostrata la percentuale di proprietà che hanno distanza minima dai monumenti principali di Berlino. La concentrazione è massima nelle vicinanze dell'Alexanderplatz con il 18,67% e dell'Esta Side Gallery con il 37,7%.

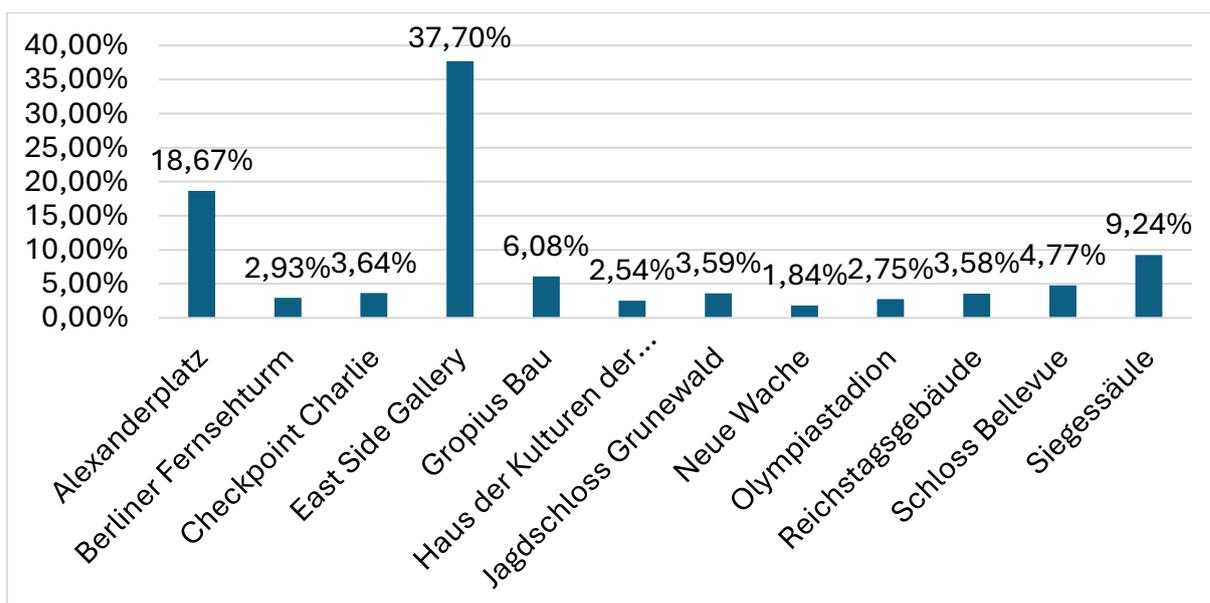


Figura 34: Percentuali di Property ID nelle vicinanze di monumenti

8.5.4. Monument Distance

L'istogramma seguente mostra che la maggior parte degli Airbnb si trova a meno di 500 metri da un monumento.

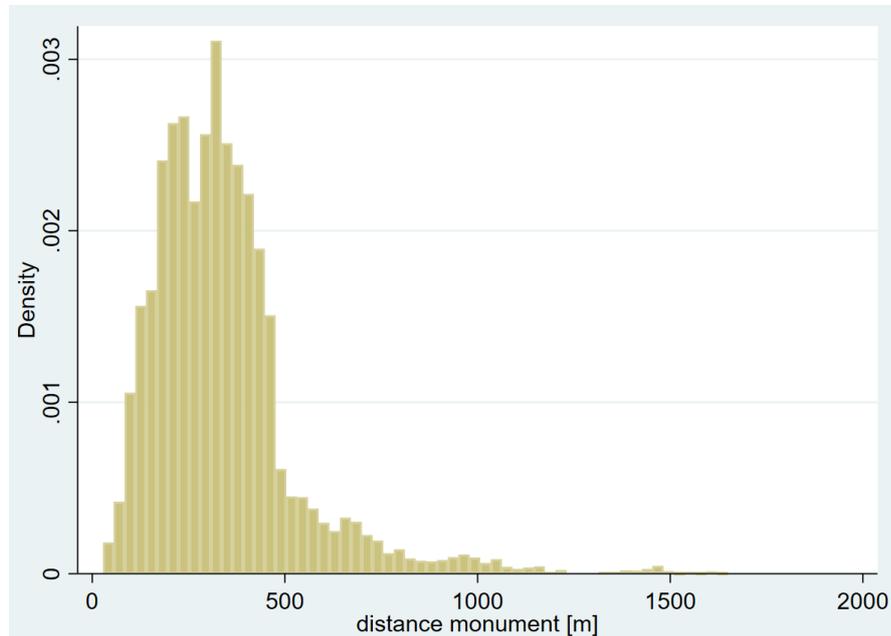


Figura 35: Distribuzione di densità discreta della variabile Monument Distance

Nel grafico a barre è invece riportato il numero di proprietà per fasce di distanza dai principali monumenti. Il 32,89% delle proprietà si trova a meno di 250 m da un monumento ed il 56,41% si trova ad una distanza tra i 250 ed i 500 m.

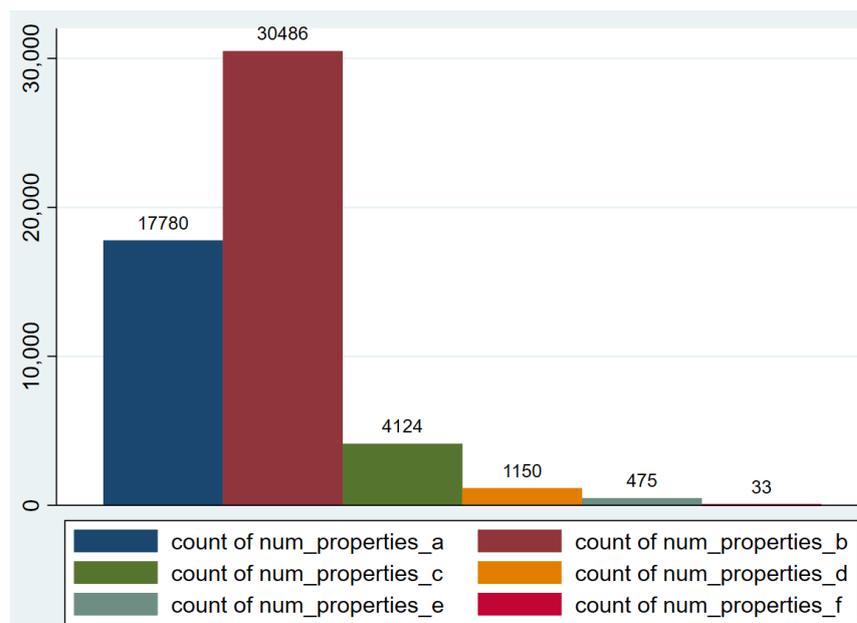


Figura 36: Numero di Property ID per fasce di distanza dai monumenti

8.5.5. Park

La variabile Park indica per ogni proprietà il parco più vicino. Il 50,26% degli Airbnb si trova nei pressi di Lustgarten, il 24,58% di Viktoriapark ed il 10,93% di Tiergarten.

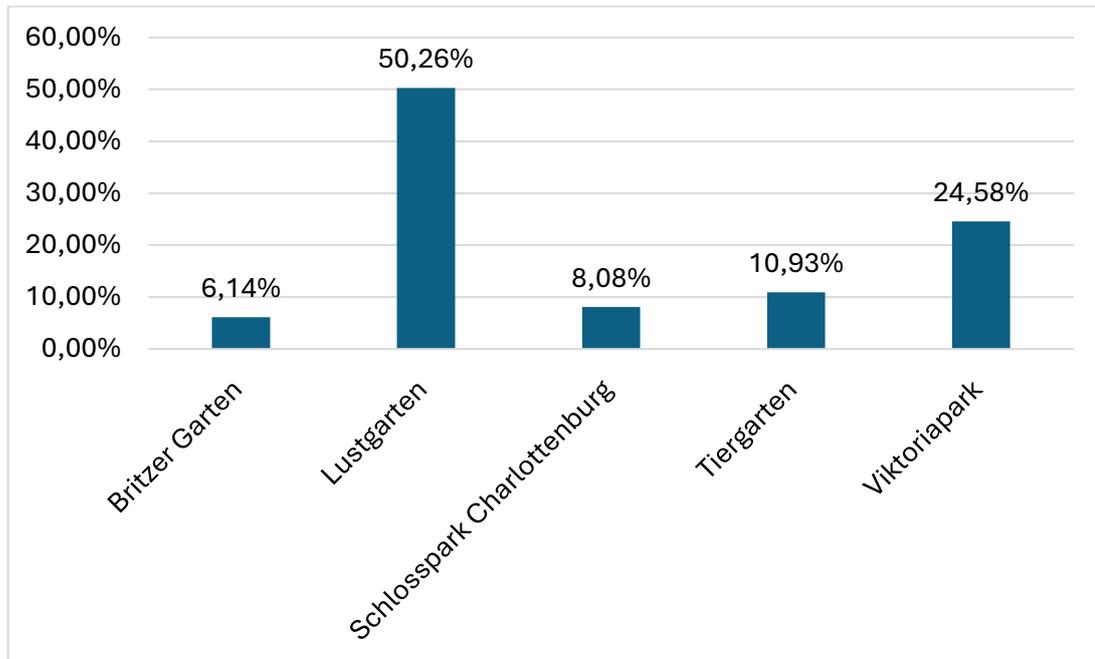


Figura 37: Percentuale di Property ID nelle vicinanze dei parchi

8.5.6. Park Distance

La distribuzione delle distanze è molto concentrata tra gli 0 ed i 500 metri dai parchi.

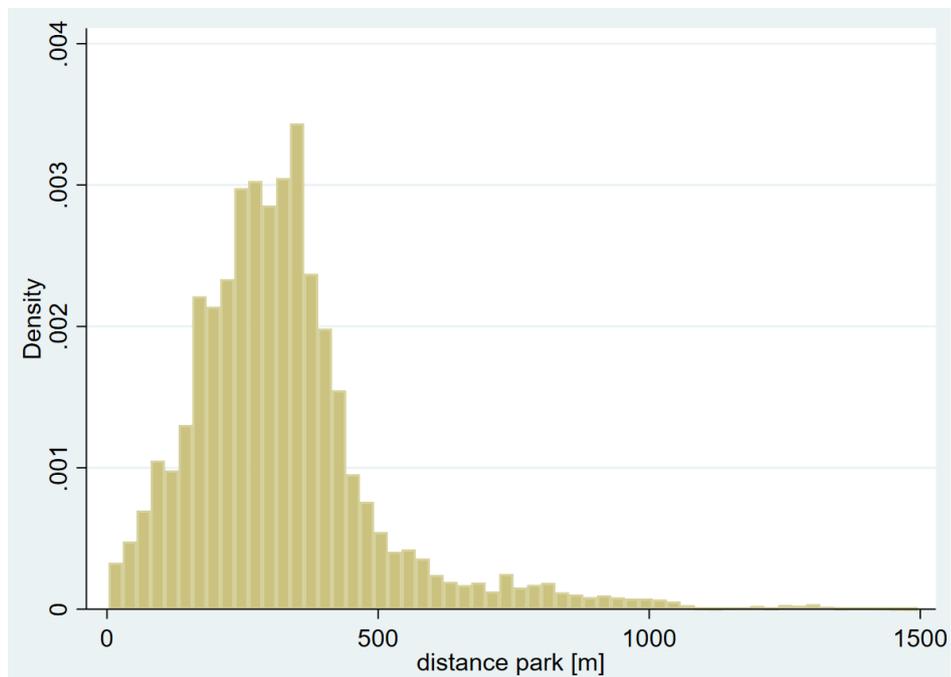


Figura 38: Distribuzione di densità discreta della variabile Park Distance

Nel grafico a barre seguente è riportato il numero di proprietà per fasce di distanza dai parchi. Il 56,19% si trova tra i 250 ed i 500 metri, mentre il 34,48% si trova a meno di 250 metri da un parco della città.

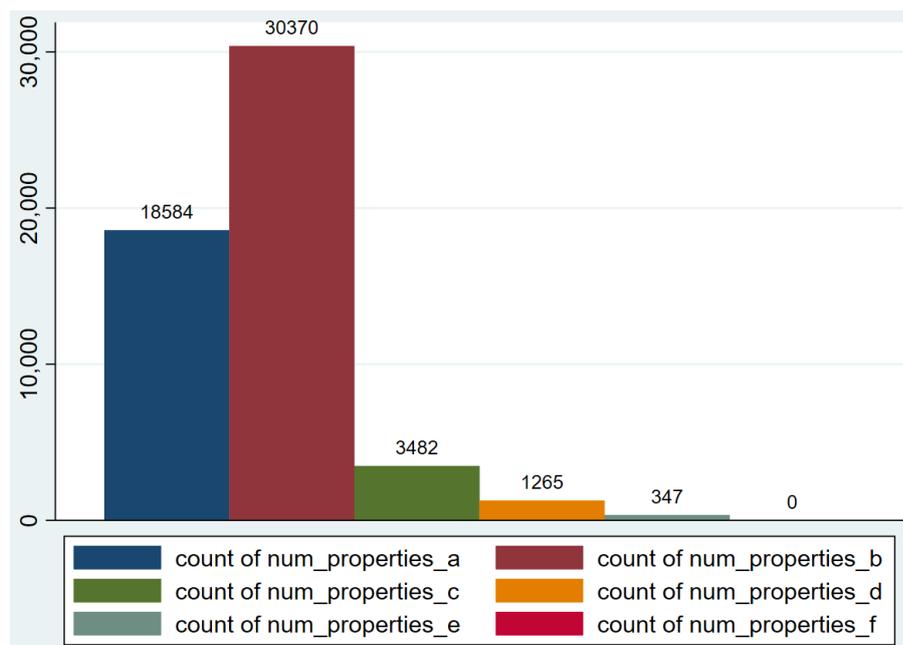


Figura 39: Numero di proprietà per fasce di distanza dai parchi

8.5.7. Parking

La variabile Parking indica, per ogni proprietà, il parcheggio più vicino, la maggior parte dei quali sono grandi parcheggi multipiano e automatici. Nel diagramma a barre si vede che il 41,56% delle proprietà si trova presso il Parkhaus Alexa, il 20,87% presso il Parkhaus Am Bahnhof Friedrichstraße e l'11,23% presso il Parkhaus Gendarmenmarkt e percentuali inferiori al 10% si trovano nelle vicinanze degli altri parcheggi.

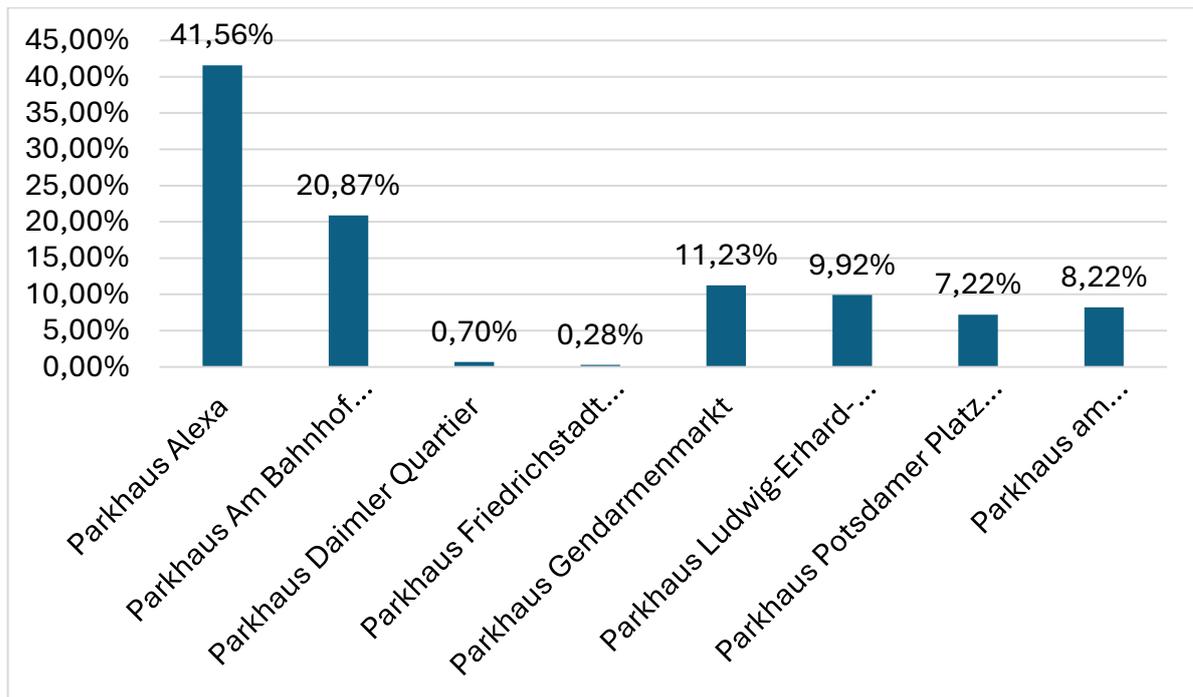


Figura 40: Percentuale di proprietà nelle vicinanze dei parcheggi

8.5.8. Parking Distance

La distribuzione delle distanze è molto concentrata tra gli 0 ed i 500 metri dai parcheggi.

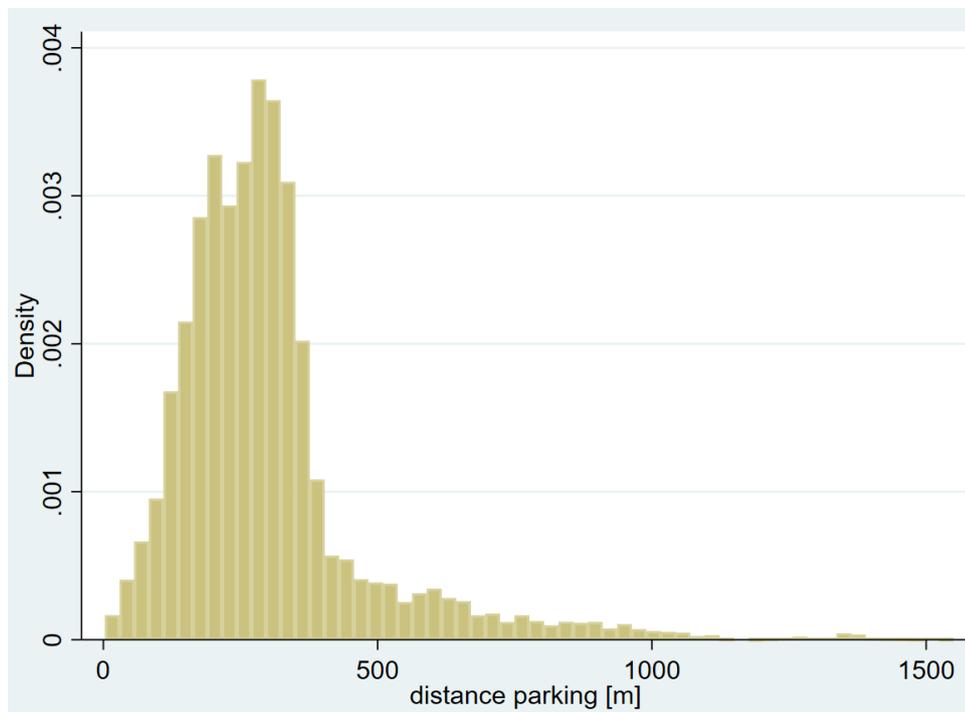


Figura 41: Distribuzione di densità della variabile Parking Distance

Nel grafico a barre è riportato il numero di proprietà per fasce di distanza dai parcheggi. Il 51,98% si trova tra i 250 ed i 500 metri, mentre il 39,39% si trova a meno di 250 metri da un parcheggio della città.

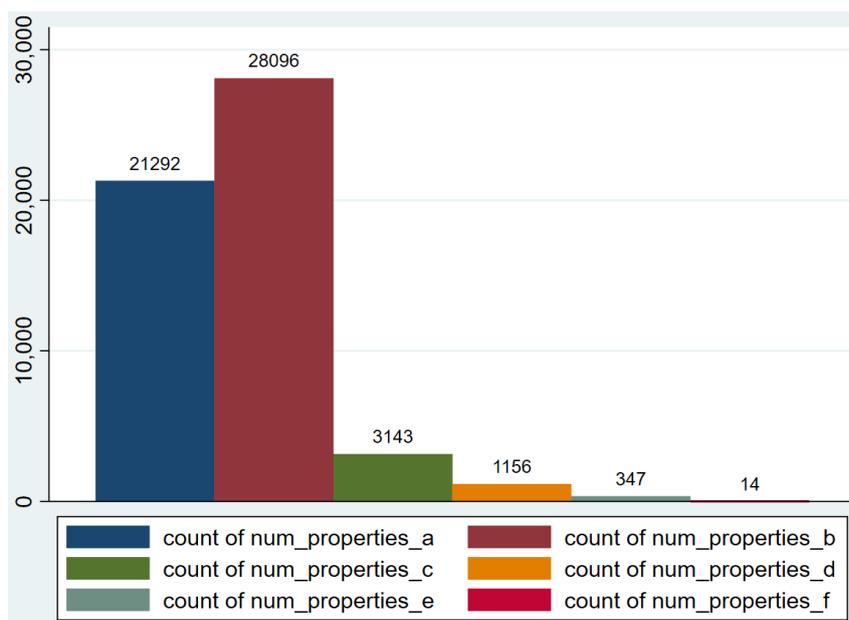


Figura 42: Numero di proprietà per fasce di distanza dai parcheggi

8.6. Variabili di performance

8.6.1 Revenue

I ricavi medi, espressi in \$ americani, mostrano un andamento nettamente decrescente negli anni della pandemia. Nel 2020 si registra una decrescita del 34%. Nel periodo post pandemico i valori sono in forte crescita: nel 2022 si registra un aumento del 67% e nel 2023 dell'82%.

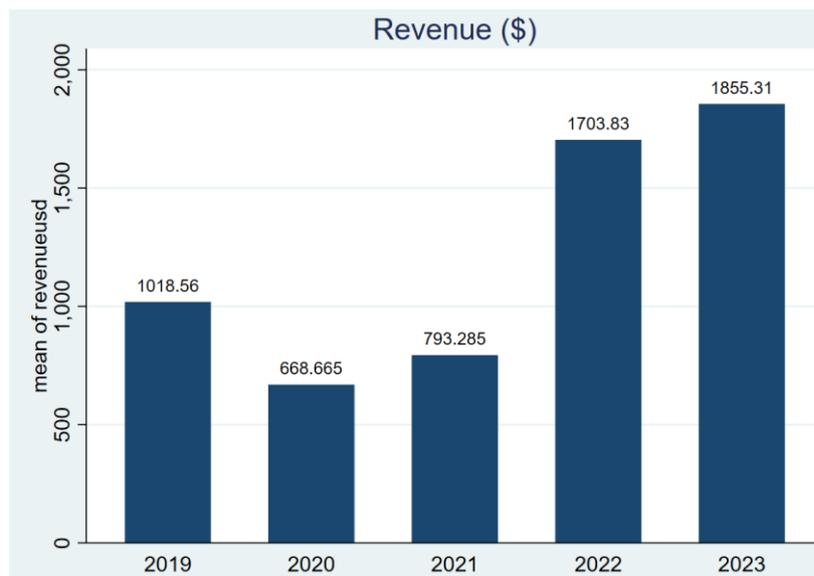


Figura 43: Valore medio di Revenue, suddiviso per anni

Revenue [\$]	Pre-Covid	Covid		Post-Covid	
	2019	2020	2021	2022	2023
Media	1018,559	668,66	793,28	1703,83	1855,31
Dev. Std.	1470,289	1175,229	1410,18	2200,96	2515,2
Tot [M\$]	222,362	95,554	90,526	149,487	157,258
Min	0	0	0	0	0
Max	89860	64975	49279	86815	94700
25°percentile	0	0	0	145	0
50°percentile (mediana)	608	235	178	1091	1083
75°percentile	1463	931	1087	2410	2748

Tabella 3: Statistiche descrittive di Revenue

Il totale dei ricavi, diminuito fortemente nel 2020 e 2021, è tornato a crescere nel 2022 e nel 2023, ma senza arrivare al valore registrato nel 2019. La somma dei ricavi è analizzata nel seguente grafico a barre.

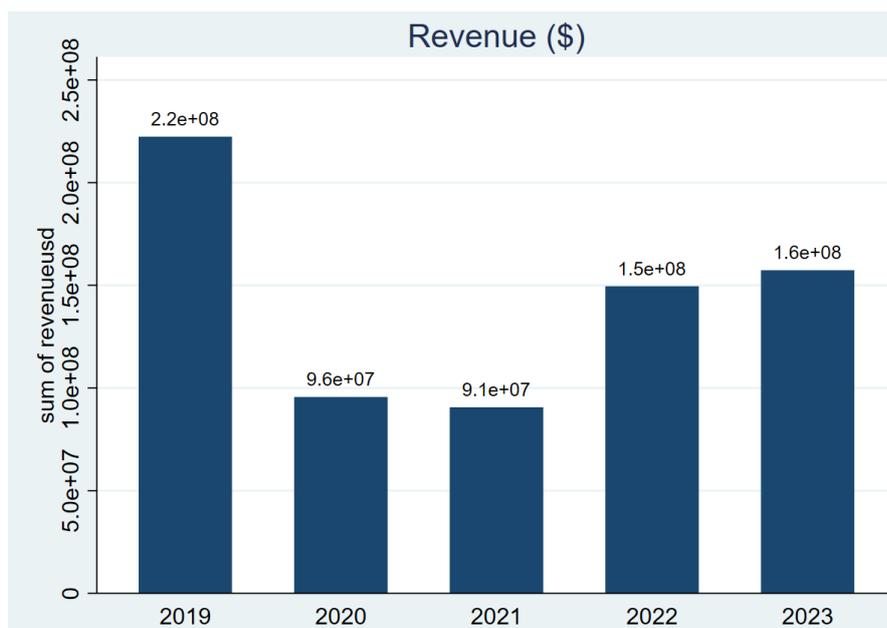


Figura 44: Valore Totale di Revenue, suddiviso per anno

8.6.2. Reservation Days

Il numero medio di giorni prenotati al mese è un buon indicatore dell'andamento del business. Dalla statistica descrittiva è evidente che durante la pandemia il numero medio delle prenotazioni ha subito un forte calo, registrando nel 2021 -38,4%. Nel periodo immediatamente successivo alla fine della pandemia, il valore medio di Reservation Days ha ripreso a salire, superando quello del 2019 del 9,44% nel 2022 e del 0,58% nel 2023.

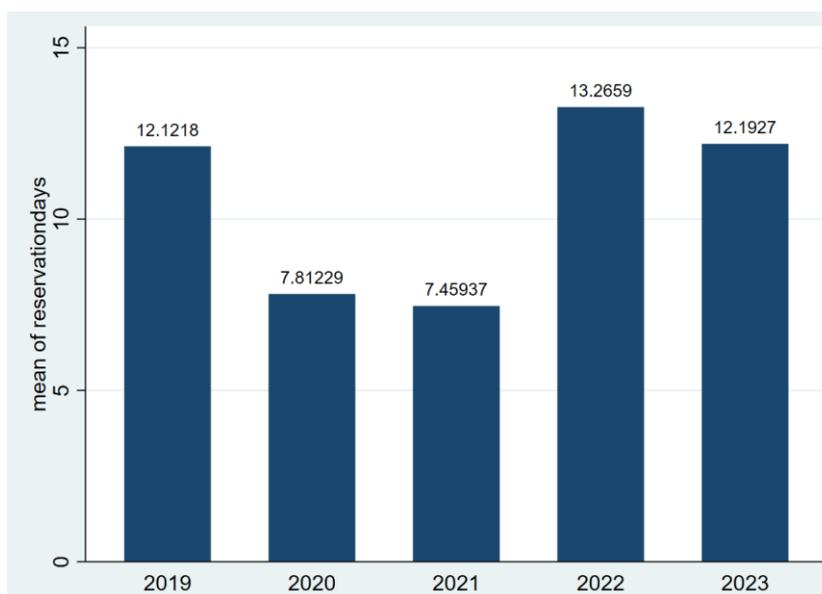


Figura 45: Valore medio di Reservation Days, suddiviso per anno

Reservation Days	Pre-Covid	Covid		Post-Covid	
	2019	2020	2021	2022	2023
Media	12,1218	7,8123	7,4594	13,2659	12,1927
Dev. Std.	11,3965	9,4988	9,5773	11,1065	11,4682
Min	0	0	0	0	0
Max	31	31	31	31	31
25°percentile	0	0	0	2	0
50°percentile (mediana)	50	3	2	12	10
75°percentile	23	14	14	24	24

Tabella 4: Statistiche descrittive di Reservation Days

8.6.3. RevPAN

I ricavi medi per notte disponibile subiscono una forte decrescita in piena pandemia, ma dal 2022 ritornano rapidamente a salire registrando un +80% nel 2023.

RevPAN [\$/day]	Pre-Covid	Covid		Post-Covid	
	2019	2020	2021	2022	2023
Media	42,069	29,217	35,649	73,469	76,663
Dev. Std.	55,125	47,114	58,944	82,079	93,261
Min	0	0	0	0	0
Max	3435	2095,968	2989,5	3333,039	3103
25°percentile	0	0	0	11,4	0
50°percentile (mediana)	30,61	12,42	10,84	56,79	55,97
75°percentile	61	43	54	103,72	114,31

Tabella 5: Statistiche descrittive di RevPAN

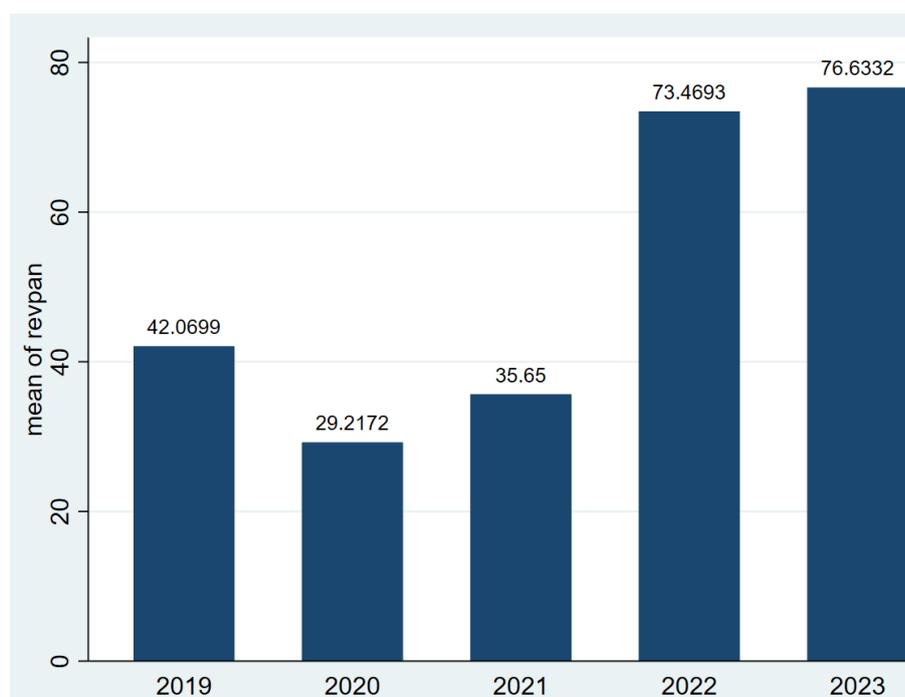


Figura 46: Valore medio di RevPAN, suddiviso per anno

8.6.4. OCC

Il valore medio dell'Occupation Rate passa dal 50,12% nel 2019, al 33% durante il Covid, per arrivare al 58,2% nel 2022 e decrescere al 51,2% nel 2023. Dopo la pandemia in media si registra circa lo stesso tasso di occupazione del periodo pre-pandemico.

OCC	pre-Covid	Covid		post-Covid	
	2019	2020	2021	2022	2023
Media	0,5012	0,3389	0,335	0,5819	0,5115
Dev. Std.	0,4111	0,3722	0,3821	0,3394	0,4213
Min	0	0	0	0	0
Max	1	1	1	1	1
25°percentile	0	0	0	0,111	0
50°percentile (mediana)	0,59	0,167	0,129	0,7241	0,591
75°percentile	0,91	0,677	0,7096	0,9677	0,9655

Tabella 6: Statistiche descrittive di OCC

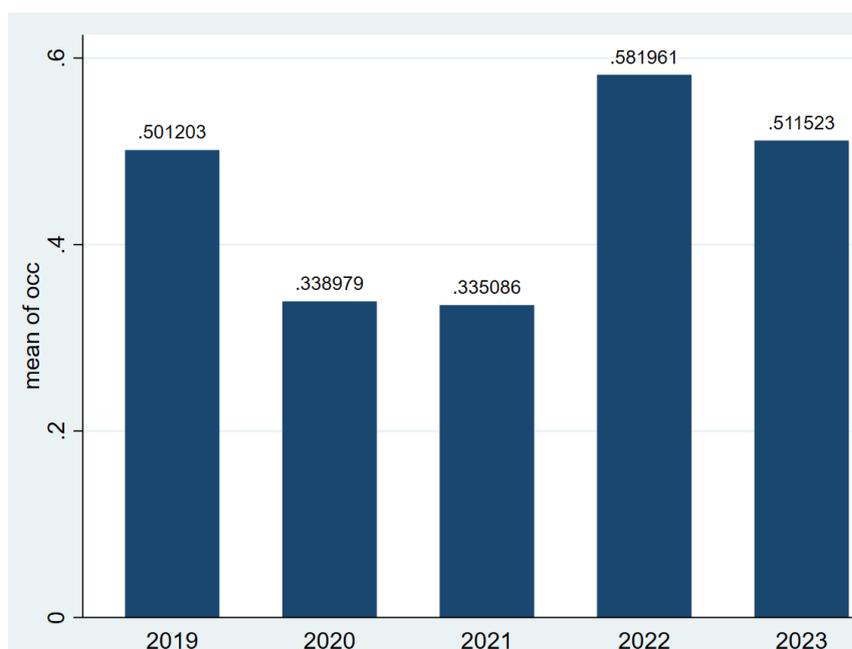


Figura 47: Valore medio di OCC, suddiviso per anno

8.6.5. ADR

Il ricavo medio per notte prenotata è cresciuto ogni anno ed in proporzione sempre maggiore. Si passa da un aumento del 5,78% nel 2020, al 25,05% nel 2021, al 49,49% nel 2022 ed al 77,97% nel 2023 rispetto al 2019.

ADR [\$/day]	Pre-Covid	Covid		Post-Covid	
	2019	2020	2021	2022	2023
Media	88,03	93,12	110,07	131,59	156,677
Dev. Std.	83,91	83,13	92,25	108,14	119,37
Min	1	1,86	0	0	9,79
Max	8987	3124,67	2989,5	4823,06	3936
25°percentile	45	47,79	57,5	68,5	83,59
50°percentile (mediana)	67	72,2	88	104	126,5
75°percentile	104	399	132,74	161,29	195,18

Tabella 7: Statistiche descrittive di ADR

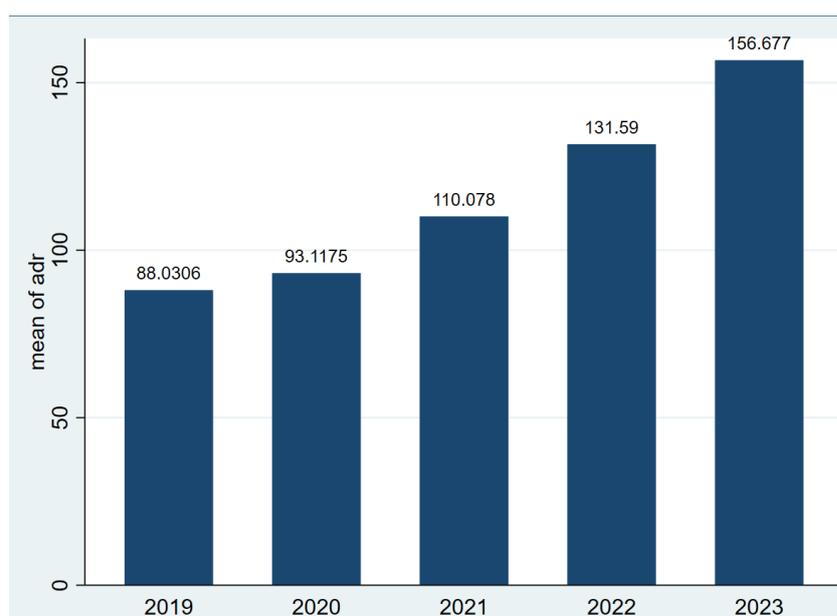


Figura 48: Valore medio di ADR, suddiviso per anno

8.7. Variabili di performance per classi di distanza

L'analisi della performance ha come obiettivo l'individuazione delle differenze di performance in base alle fasce di distanza dai punti di interesse.

8.7.1. Revenue – Distance Class

Analizzando i ricavi medi mensili suddivisi per fasce di distanza si evince che le fasce A, B, E sono le più redditizie. Gli Airbnb in fasce C, D, E di distanza dai punti di interesse registrano ricavi inferiori.

Nel seguente grafico si riporta l'andamento dei ricavi medi mensili suddivisi per fascia di distanza e per tipologia di punti di interesse. L'andamento registrato è molto simile anche tra tipologie diverse di punti attrattivi.

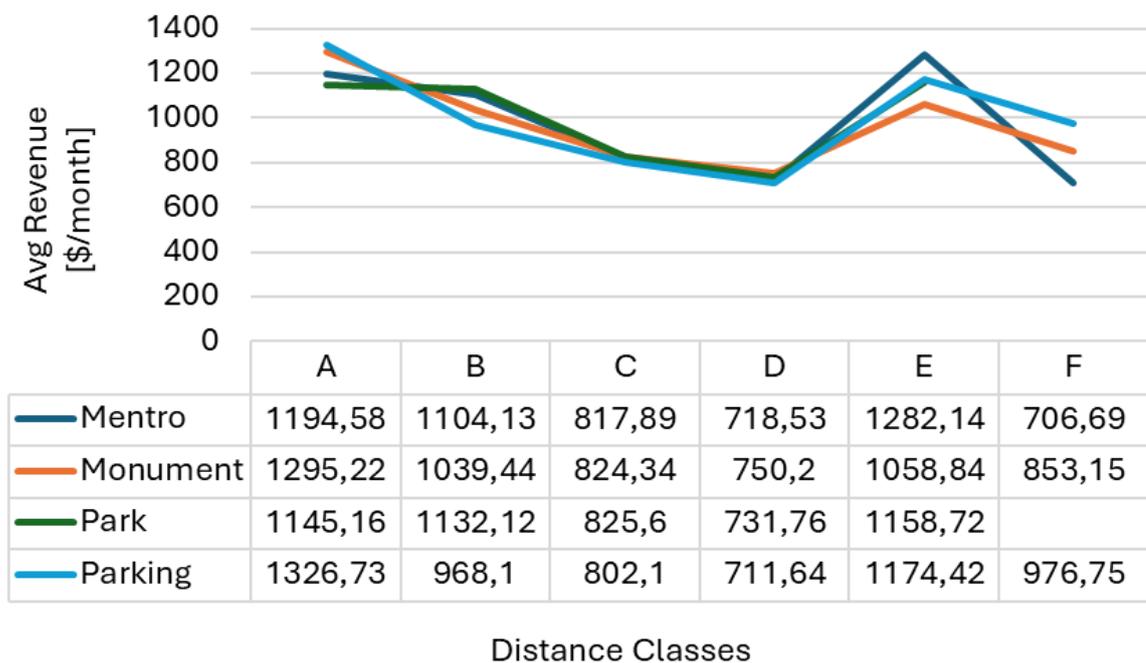


Figura 49: Valore medio di Revenue, suddiviso per classi di distanza

8.7.2. Reservation Days – Distance class

Il numero medio di giorni prenotati al mese ha poca variazione tra le fasce di distanza. Le A, B, E hanno una media di oltre 10 giorni al mese, le B e C oltre 9, mentre la fascia più lontana dalle attrazioni, la F ha una media di oltre 7 giorni al mese.

Nel grafico sottostante si vede bene che il numero medio di giorni prenotati ha poco scostamento tra le diverse tipologie di attrazione.

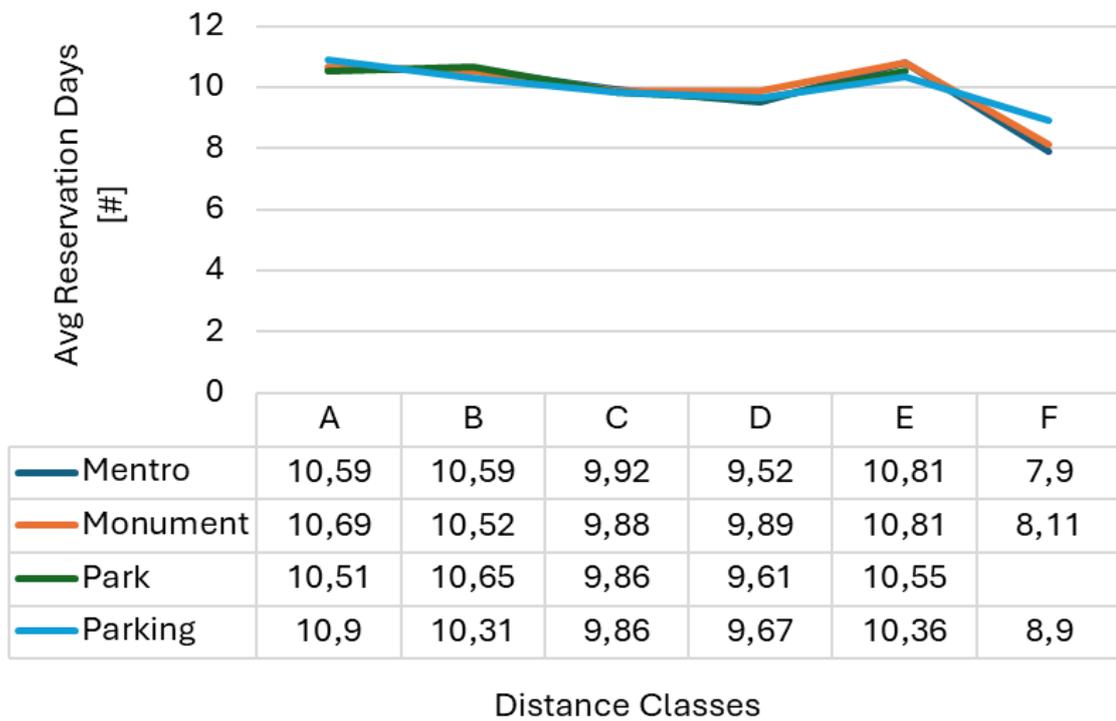


Figura 50: Valore medio di Reservation Days, suddiviso per classi di distanza

8.7.3. RevPAN – Distance class

Il valore medio dei ricavi per notte disponibile è massimo per la fascia E, seguito subito dopo dalla A e dalla B. Le fasce C, D, F hanno valori in media inferiori quasi del 20%. Il valore minimo si trova in corrispondenza della fascia D.

Per tutte e quattro le tipologie di punti di interesse l'andamento rispecchia quello medio del grafico a barre. Dal seguente grafico si vede che le performance in base alla distanza da metropolitana e parcheggi ha una varianza maggiore tra le diverse fasce.

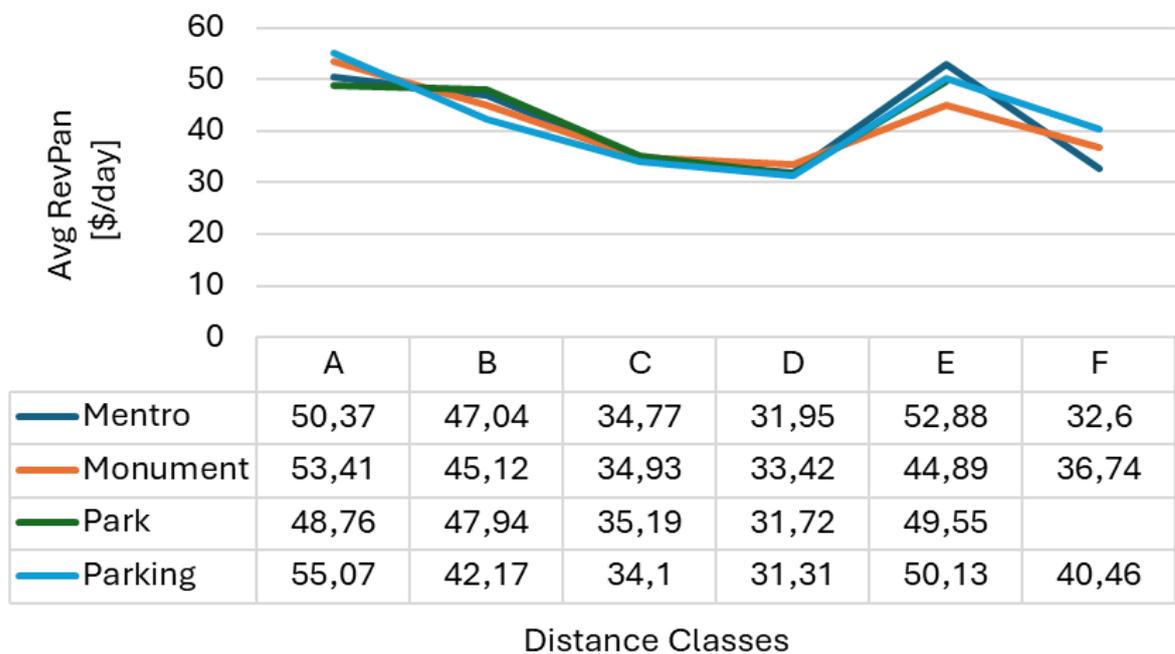


Figura 51: Valore medio di RevPAN, suddiviso per classi di distanza

8.7.4. OCC – Distance Class

Il tasso medio di occupazione mensile è massimo nelle fasce A, B ed E con il 45%, poi ci sono la C e la D con il 41%. La fascia F registra un valore del 36%.

Andando a trovare i valori di OCC per tipologia di punto di interesse si vede che la differenza tra le tipologie di attrazioni è trascurabile e tutte registrano valori medi molto simili fra loro.

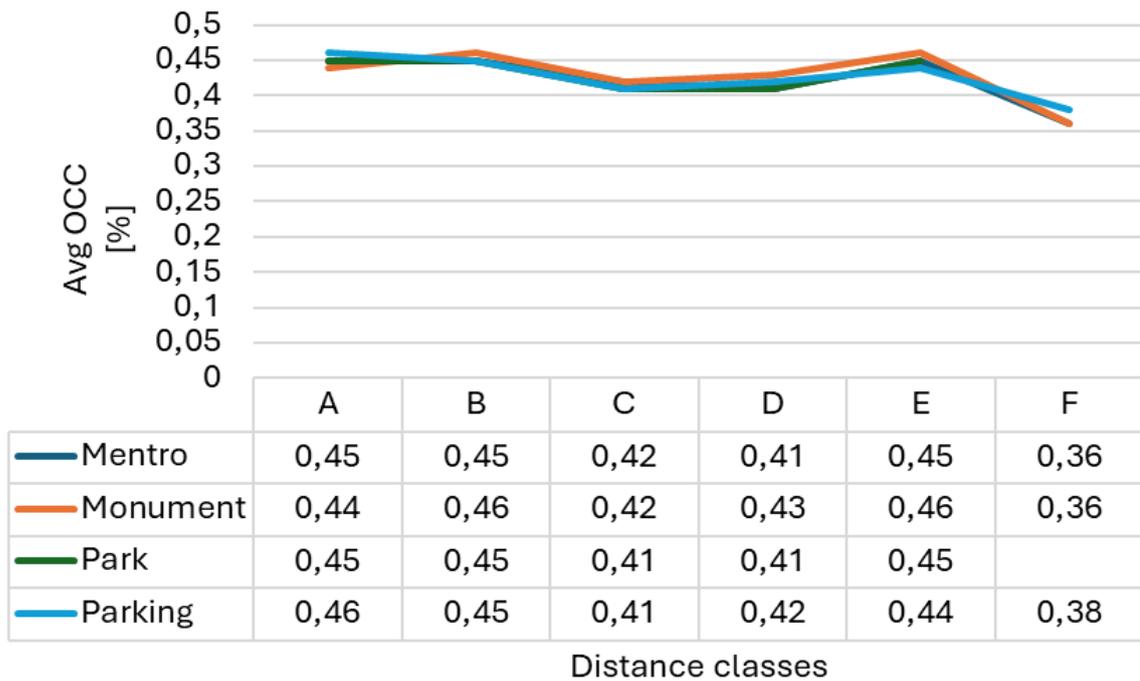


Figura 52: Valore medio di OCC, suddiviso per classi di distanza

8.7.5. ADR - Distance Class

Il valore medio dell'ADR è massimo per la fascia E, seguito dalla A e poi dalla B. La C e la F registrano un valore inferiore. La D ha il valore minimo.

Anche l'ADR ha andamento simile pur differenziando le fasce tra tipologie di attrazione. Le distanze da metropolitana e parcheggi sono quelle che registrano una maggiore varianza.

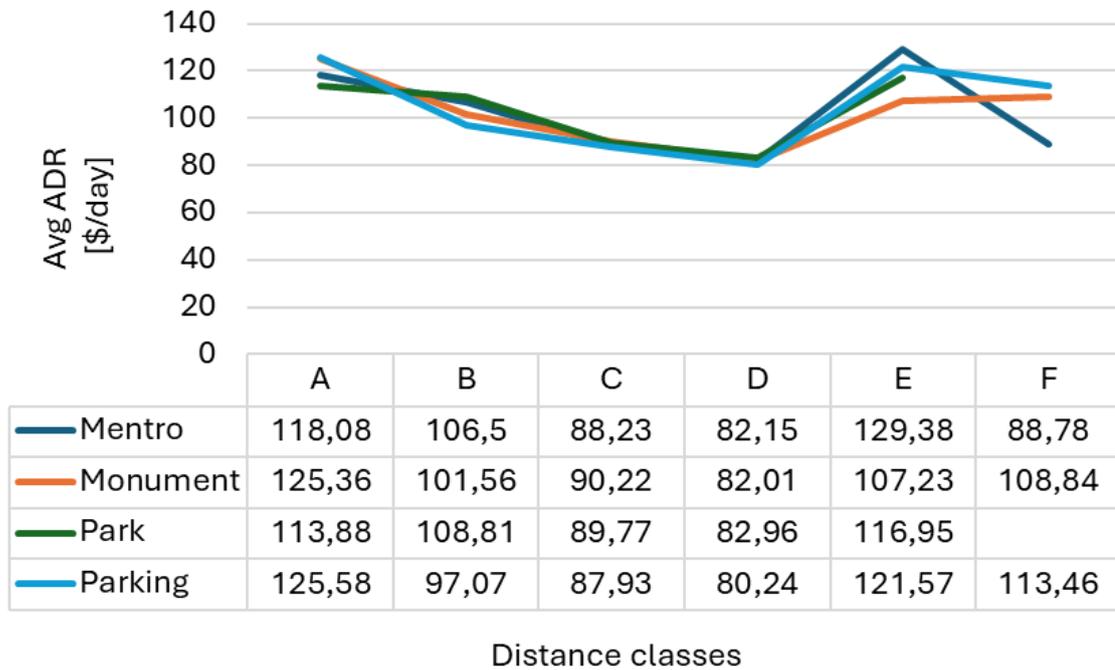


Figura 53: Valore medio di ADR, suddiviso per classi di distanza

Dalle analisi svolte per fasce di distanza, in conclusione, si evince che la performance più alte si possono attribuire agli Airbnb che si trovano in una fascia di distanza dai punti di interesse A, B ed E, ovvero a meno di 500 m oppure tra 1 km e 1,5 km. In queste fasce i ricavi ed il tasso di occupazione sono in media maggiori.

9. Analisi di correlazione

L'obiettivo di ricerca è indagare se esiste una correlazione tra le performance degli Airbnb di Berlino e la distanza dai punti di interesse. Le variabili considerate nell'analisi sono le distanze da metropolitana, monumenti, parchi e parcheggi in metri.

L'obiettivo è capire se esiste una correlazione tra le variabili per poi applicare correttamente le regressioni. Nel caso si trovi una correlazione, si vuole anche determinare il segno e l'intensità.

Lo strumento utilizzato a tal fine è l'Indice di Correlazione di Pearson che ha un valore compreso tra -1 e +1 per indicare intensità e segno della correlazione. Se l'indice vale 0 significa che le due variabili non sono correlate, se è un valore tra -1 e 0 che sono correlate negativamente, tra 0 e +1 che sono correlate positivamente.

L'indice è solitamente indicato con una r e l'espressione per il calcolo è la seguente:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

n è il numero di coppie di dati

x e y sono i valori delle variabili

$\sum x$ e $\sum y$ sono le somme di tutti i valori di x e y

$\sum xy$ è la sommatoria di tutti i prodotti tra x e y

$\sum x^2$ e $\sum y^2$ sono le somme dei quadrati dei valori di x e y

$r=0$ indica assenza di correlazione, $r=-1$ indica perfetta correlazione negativa, $r=1$ indica perfetta correlazione positiva. Valori compresi tra -1 e 0 o tra 0 e +1 indicano una correlazione imperfetta.

Su Stata è calcolato l'indice di correlazione di Pearson tra le quattro variabili di distanza e si è visto che sono tutte altamente correlate fra loro. Questo significa che nei modelli di regressione andrà inserita una sola variabile di distanza alla volta per evitare il problema della multicollinearità.

L'output ottenuto è il seguente:

Pearson Correlation Index	Distance Metro	Distance Monument	Distance Park	Distance Parking
Distance Metro	1	0,88	0,98	0,95
Distance Monument	0,88	1	0,79	0,94
Distance Park	0,98	0,79	1	0,92
Distance Parking	0,95	0,94	0,92	1

Figura 54: Indici di Correlazione di Pearson tra le variabili di distanza

Successivamente si calcolerà lo stesso indice di correlazione tra le distanze e le misure di performance, utilizzando un modello di tipo lineare del tipo:

$$\text{performance} = \alpha + \beta \text{ distance}$$

Nella seguente tabella sono sintetizzati i valori ottenuti:

Pearson Correlation Index				
LIN model	Distance Metro	Distance Monuments	Distance Parks	Distance Parkings
Revenue	-0,058	-0,078	-0,054	-0,082
Reservation Days	-0,017	-0,02	-0,018	-0,026
OCC	-0,016	-0,014	-0,02	-0,022
RevPAN	-0,06	-0,075	-0,057	-0,083
ADR	-0,075	-0,104	-0,068	-0,106

Tabella 8: Indici di correlazione di Pearson tra distanze e performance con modello lineare

Dai risultati si evince che la distanza dai punti di interesse è correlata negativamente alle variabili di performance degli Airbnb e l'intensità della correlazione risulta piuttosto bassa. Prendendo ad esempio l'indice di Pearson tra Distance Metro e Revenue, si può affermare che all'aumentare della distanza di 1 metro, i ricavi diminuiscono in media di 00,583\$/mese, i Reservation Days di 0,0173 giorni/mese, l'OCC dell'1,62%, il RevPAN di 0,06 \$/giorno e l'ADR di 0,0754 \$/giorno.

Di seguito è mostrato lo scatterplot ottenuto per visualizzare la correlazione tra la Distance Metro e i Revenue:

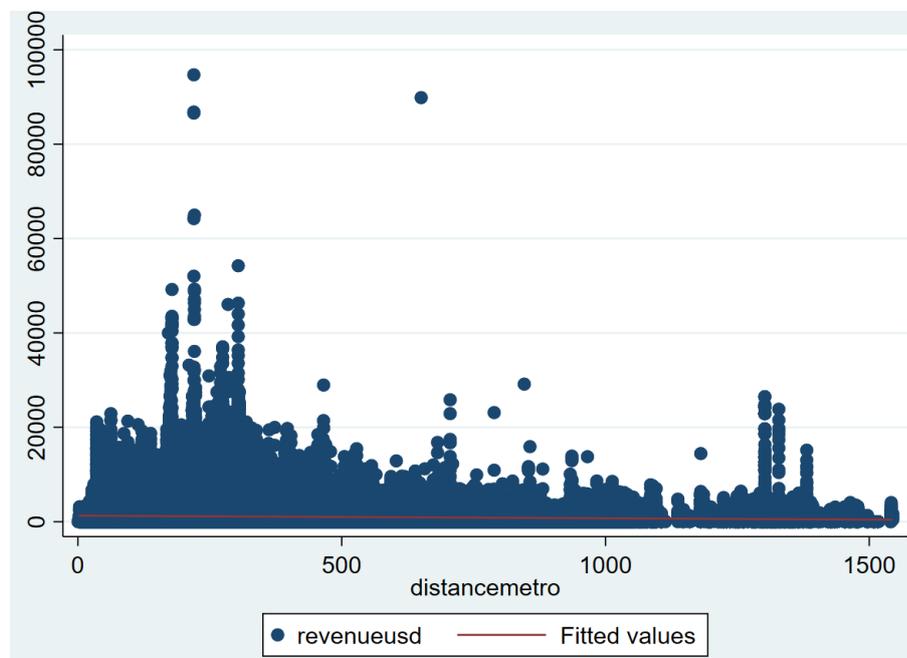


Figura 55: Scatterplot Revenue - Distance Metro

Successivamente si passa ad un modello di tipo logaritmico. L'equazione corrispondente è la seguente:

$$\ln(\text{performance}) = \alpha + \beta \text{ distance}$$

I valori dell'indice ottenuti su stata sono i seguenti:

Pearson Correlation Index				
LOG Model	Distance Metro	Distance Monuments	Distance Parks	Distance Parkings
Ln (Revenue)	-0,0733	-0,0928	-0,0655	-0,0981
Ln (Reservation Days)	-0,0168	-0,0201	-0,0164	-0,023
Ln (OCC)	-0,0182	-0,009	-0,0239	-0,0197
Ln (RevPAN)	-0,0816	-0,0927	-0,0799	-0,1048
Ln (ADR)	-0,1052	-0,1338	-0,0925	-0,1394

Tabella 9: Indici di correlazione di Pearson tra distanze e performance con modello logaritmico

In questo caso i coefficienti trovati indicano una variazione delle performance in base alle distanze. Prendendo sempre la coppia Distance Metro e Ln(Revenue) si può affermare che all'aumentare della distanza di 1 metro, in media i ricavi diminuiscono del 7,33%, i Reservation Days dell'1,68%, l'OCC dell'1,82%, il RevPAN dell'8,16% e l'ADR del 10,52%.

Lo scatterplot ottenuto con modello lineare è il seguente:

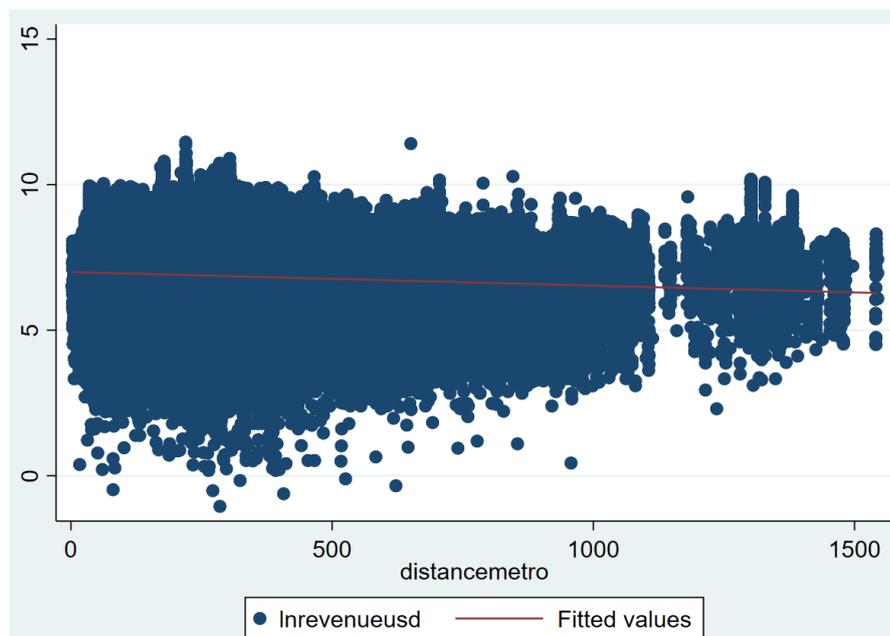


Figura 56: Scatterplot Ln(Revenue) - Distance Metro

Con il modello logaritmico è graficamente più evidente la correlazione negativa interpolata dalla retta di correlazione in rosso.

In generale, si può affermare da queste prime analisi che è presente una correlazione negativa tra le distanze spaziali e le performance.

Per poter indagare e confermare o rifiutare le ipotesi fatte è necessario procedere con analisi di regressione.

10. Analisi di regressione

L'obiettivo dell'analisi di regressione è capire se esiste effettivamente un rapporto di causa-effetto tra la vicinanza ai punti di interesse e le performance degli Airbnb. Se esiste, si vuole comprendere e quantificare questa relazione per poter fare delle previsioni o estrapolazioni di informazioni.

Dall'analisi statistica si può comprendere il complesso legame esistente tra le variabili del dataset costruito e intuire che le distanze dai punti di interesse hanno un'influenza sulle performance.

Per effettuare le analisi di regressione si utilizza un modello lineare, che presuppone quindi una relazione di tipo lineare tra le variabili indipendenti e le variabili dipendenti.

Il modello in generale può essere sintetizzato con la seguente equazione:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon$$

In cui:

y è la variabile dipendente, x_1, x_2, \dots, x_n sono le variabili indipendenti, $\beta_0, \beta_1, \dots, \beta_n$ sono i coefficienti di correlazione lineare ed ε è il residuo o errore.

Dapprima sono state eseguite delle analisi di regressione univariate e successivamente multivariate. In ultimo sono state inserite le variabili moderanti.

10.1. Regressioni Univariate

L'analisi di regressione univariata coinvolge una sola variabile indipendente (detta predittiva) e una variabile dipendente. L'obiettivo è modellare la relazione tra la variabile indipendente e la variabile dipendente, comprendendo come la variabile indipendente influenzi la variabile dipendente.

Nella pratica quindi si analizza una coppia variabili indipendente-dipendente, prendendo come variabili indipendenti la distanza dalla metropolitana, dai monumenti, dai parchi e dai parcheggi e come variabile dipendente il RevPAN.

I modelli di regressione utilizzati sono quattro e ciascuno analizza l'impatto della distanza in metri dal punto di interesse più vicino.

Le rette di regressione sono le seguenti:

$$\text{RevPAN} = \beta_0 + \beta_1 \text{ distancemetro} + \varepsilon$$

$$\text{RevPAN} = \beta_0 + \beta_1 \text{ distancemonument} + \varepsilon$$

$$\text{RevPAN} = \beta_0 + \beta_1 \text{ distancepark} + \varepsilon$$

$$\text{RevPAN} = \beta_0 + \beta_1 \text{ distanceparking} + \varepsilon$$

Dall'output di regressione, il coefficiente β_1 dà una stima della variazione del RevPAN al variare di 1 metro della distanza. La costante β_0 è l'intercetta della retta di regressione e rappresenta il valore previsto della y quando la x è 0.

Sotto i coefficienti di regressione β , il primo valore tra parentesi è lo Standard Error (SE) che è una misura della variabilità dei coefficienti stimati del modello di regressione rispetto ai veri valori dei parametri della popolazione. In altre parole, rappresenta la precisione con cui il coefficiente stimato approssima il vero coefficiente della popolazione.

Il secondo valore tra parentesi è il p-value, che indica se il coefficiente è statisticamente significativo. Nell'output di Stata il livello di significatività, ovvero la probabilità di rifiutare erroneamente l'ipotesi nulla è indicato da uno, due o tre asterischi nel modo seguente:

(*) livello di significatività: 90%, p-value: 10%

(**) livello di significatività: 95%, p-value: 5%

(***) livello di significatività: 99%, p-value: 1%

N indica il numero di osservazioni utilizzate per effettuare la regressione.

R^2 è una misura di bontà dell'adattamento, ovvero indica quanto bene il modello di regressione lineare aderisce ai dati osservati. In sostanza, fornisce una misura della proporzione di variazione nella variabile dipendente che è spiegata dalle variabili indipendenti nel modello.

F si riferisce alla statistica o test F (Fisher) ed è utilizzato per valutare la significatività complessiva di un modello di regressione, ovvero per determinare se almeno una delle variabili indipendenti ha un effetto significativo sulla variabile dipendente. In pratica, si

confronta il modello di regressione completo, che include quindi tutte le variabili indipendenti, con un modello nullo, che include solo l'intercetta. Il test F valuta se il modello completo spiega la variazione nei dati in modo significativamente migliore rispetto al modello nullo. Il valore della statistica F viene confrontato con una distribuzione F sotto l'ipotesi nulla che tutte le variabili indipendenti non abbiano alcun effetto significativo. Se il valore della statistica F è sufficientemente grande da superare una soglia critica, si può concludere che il modello di regressione complessivo è significativo.

L'output di regressione con modello lineare è il seguente:

RevPAN	M1	M2	M3	M4
	b/se/p	b/se/p	b/se/p	b/se/p
distance metro	-0.022***			
	(0.000)			
	(0.000)			
distance monument		-0.026***		
		(0.000)		
		(0.000)		
distance park			-0.021***	
			(0.000)	
			(0.000)	
distance parking				-0.030***
				(0.000)
				(0.000)
_cons	54.055***	55.717***	53.727***	56.085***
	(0.177)	(0.178)	(0.171)	(0.172)
	(0.000)	(0.000)	(0.000)	(0.000)
N	6.48e+05	6.48e+05	6.48e+05	6.48e+05
R2	0.004	0.006	0.003	0.007
F	2.520.210	4.001.066	2.515.497	4.799.968

Tabella 10: Output di Regressione Univariata

Le rette di regressione corrispondenti sono:

$$\text{RevPAN} = 54,055 - 0,022 \text{ Distance Metro}$$

$$\text{RevPAN} = 55,717 - 0,026 \text{ Distance Monument}$$

$$\text{RevPAN} = 53,727 - 0,021 \text{ Distance Park}$$

$$\text{RevPAN} = 56,085 - 0,030 \text{ Distance Parking}$$

Si può quindi affermare che, all'aumentare di 1 m della distanza tra una proprietà e la metropolitana più vicina, il RevPAN in media diminuisce di 0,022 \$/giorno, dai monumenti di 0,026 \$/giorno, dai parchi di 0,021 \$/giorno e dai parcheggi di 0,030 \$/giorno.

Il valore dell' R^2 è piuttosto basso e va dallo 0,4% allo 0,7%.

10.2. Regressioni multivariate

La regressione multivariata coinvolge due o più variabili indipendenti e una variabile dipendente. L'obiettivo è modellare la relazione tra le variabili indipendenti e la variabile dipendente, comprendendo come le variabili indipendenti influenzino la variabile dipendente. La regressione multivariata è più complessa rispetto a quella univariata perché considera degli effetti congiunti delle variabili.

Con queste regressioni si vuole indagare in modo approfondito e completo la relazione tra le distanze ed il RevPAN. Nei modelli di regressione si inseriscono le variabili di controllo Max Guest, Overall Rating, Listing Type, Host Type, Neighborhood, Year e Month, che influenzano la variabile dipendente. Successivamente in inseriscono le variabili moderanti Airbnb Superhost e Minimum Stay che hanno influenza sulla relazione tra le variabili indipendenti e la variabile dipendente.

I modelli di regressione multivariata utilizzati sono di tipo lineare e sono in totale cinque.

Il modello M1 considera tutte le variabili di controllo.

I modelli M2, M3, M4, M5 considerano le variabili di controllo e rispettivamente Distance Metro, Distance Monument, Distance Park e Distance Parking.

La retta di regressione corrispondente al modello M1 è:

$$\text{RevPAN} = \beta_0 + \beta_1 \text{maxguest} + \beta_2 \text{overallrating} + \beta_3 \text{professionalhost} + \beta_4 \text{hotelroom} + \beta_5 \text{privateroom} + \beta_6 \text{sharedroom} + \beta_7 \text{Charlottenburg-Wilmersdorf} + \beta_8 \text{Frankfurter Allee Süd} + \beta_9 \text{Friedrichshain-Kreuzberg} + \beta_{10} \text{Lichtenberg} + \beta_{11} \text{Marzahn-Hellersdorf} + \beta_{12} \text{Mitte} + \beta_{13} \text{Neukolln} + \beta_{14} \text{Pankow} + \beta_{15} \text{Reinickendorf} + \beta_{16} \text{Reuterstraße} + \beta_{17} \text{Schöneberg-Nord} + \beta_{18} \text{Siemensstadt} + \beta_{19} \text{Spandau} + \beta_{20} \text{Steglitz-Zehlendorf} + \beta_{21} \text{Tempelhof-Schöneberg} + \beta_{22} \text{Tiergarten Süd} + \beta_{23} \text{Treptow-Kopenick} + \beta_{24} \text{January} + \beta_{25} \text{February} + \beta_{26} \text{March} + \beta_{27} \text{April} + \beta_{28} \text{May} + \beta_{29} \text{June} + \beta_{30} \text{July} + \beta_{31} \text{August} + \beta_{32} \text{September} + \beta_{33} \text{October} + \beta_{34} \text{November} + \beta_{35} \text{December} + \beta_{36} \text{2020} + \beta_{37} \text{2021} + \beta_{38} \text{2022} + \beta_{39} \text{2023} + \varepsilon$$

Nei modelli M2, M3, M4, M5 è:

$$\text{RevPAN} = \beta_0 + \beta_1 \text{maxguest} + \beta_2 \text{overallrating} + \beta_3 \text{professionalhost} + \beta_4 \text{hotelroom} + \beta_5 \text{privateroom} + \beta_6 \text{sharedroom} + \beta_7 \text{Charlottenburg-Wilmersdorf} + \beta_8 \text{Frankfurter Allee Süd} + \beta_9 \text{Friedrichshain-Kreuzberg} + \beta_{10} \text{Lichtenberg} + \beta_{11} \text{Marzahn-Hellersdorf} + \beta_{12} \text{Mitte} + \beta_{13} \text{Neukolln} + \beta_{14} \text{Pankow} + \beta_{15} \text{Reinickendorf} + \beta_{16} \text{Reuterstraße} + \beta_{17} \text{Schöneberg-Nord} + \beta_{18} \text{Siemensstadt} + \beta_{19} \text{Spandau} + \beta_{20} \text{Steglitz-Zehlendorf} + \beta_{21} \text{Tempelhof-Schöneberg} + \beta_{22} \text{Tiergarten Süd} + \beta_{23} \text{Treptow-Kopenick} + \beta_{24} \text{January} + \beta_{25} \text{February} + \beta_{26} \text{March} + \beta_{27} \text{April} + \beta_{28} \text{May} + \beta_{29} \text{June} + \beta_{30} \text{July} + \beta_{31} \text{August} + \beta_{32} \text{September} + \beta_{33} \text{October} + \beta_{34} \text{November} + \beta_{35} \text{December} + \beta_{36} \text{2020} + \beta_{37} \text{2021} + \beta_{38} \text{2022} + \beta_{39} \text{2023} + \beta_{40} \text{Distance} + \varepsilon$$

Stata considera la variabile numero 1 come modello base ed indica le variazioni rispetto a questo, considerando tutti i possibili valori che la variabile può assumere oltre ai valori di base considerato secondo lo schema seguente:

Host Type	Listing Type	Year	Month	Neighborhood
1. Private host	1. Entire home/apt	1. 2019	1. January	1. Charlottenburg-Wilmersdorf
2. Professional host	2. Hotel room	2. 2020	2. February	2. Frankfurter Allee Süd
	3. Private room	3. 2021	3. March	3. Friedrichshain-Kreuzberg
	4. Shared room	42.022	4. April	4. Lichtenberg
		52.023	5. May	5. Marzahn-Hellersdorf
			6. June	6. Mitte
			7. July	7. Neukolln
			8. August	8. Pankow
			9. September	9. Reinickendorf
			10. October	10. Reuterstraße
			11. November	11. Schöneberg-Nord
			12. December	12. Siemensstadt
				13. Spandau
				14. Steglitz-Zehlendorf
				15. Tempelhof-Schöneberg
				16. Tiergarten Süd
				17. Treptow-Kopenick

Tabella 11: Schema dei Modelli di Regressione

Il modello base ha come valori di riferimento i seguenti:

Host Type: Private host

Listing Type: Entire home/apt.

Year: 2019

Month: January

Neighborhood: Charlottenburg-Wilmersdorf

Per evitare il problema della multicollinearità perfetta, il software per ogni variabile categorica che può assumere x valori diversi creare $x-1$ variabili dummy che rappresentino le categorie. Se la variabile dummy ha valore 1 significa che quell'attributo è vero, se ha valore 0 significa che è falso.

Nell'output ottenuto è riportato in ordine il coefficiente di regressione, lo standard error ed il p-value.

L'output ottenuto dalla regressione lineare è il seguente:

RevPAN	M1	M2	M3	M4	M5
	b/se/p	b/se/p	b/se/p	b/se/p	b/se/p
max guests	9.752***	9.838***	9.843***	9.830***	9.857***
	(0.057)	(0.057)	(0.057)	(0.057)	(0.057)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
overall rating	0.980***	0.973***	0.974***	0.973***	0.972***
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
professional host	2.431***	2.146***	2.223***	2.143***	2.089***
	(0.128)	(0.128)	(0.128)	(0.128)	(0.128)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
hotel room	4.317***	3.271***	3.636***	3.362***	3.222***
	(0.705)	(0.703)	(0.701)	(0.703)	(0.702)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
private room	-27.922***	-27.808***	-27.854***	-27.791***	-27.770***
	(0.130)	(0.129)	(0.129)	(0.129)	(0.129)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
shared room	-70.052***	-69.744***	-69.334***	-69.899***	-69.639***
	(0.490)	(0.490)	(0.488)	(0.491)	(0.490)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Friedrichshain- Kreuzberg	7.091***	6.835***	10.500***	4.851***	6.005***
	(0.247)	(0.248)	(0.257)	(0.251)	(0.248)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Lichtenberg	-6.622***	0.323	3.316***	-1.694***	0.148
	(0.348)	(0.372)	(0.402)	(0.360)	(0.367)
	(0.000)	(0.385)	(0.000)	(0.000)	(0.686)
Marzahn- Hellensdorf	-12.992***	1.710**	4.415***	-0.247	2.608***
	(0.629)	(0.694)	(0.728)	(0.678)	(0.690)
	(0.000)	(0.014)	(0.000)	(0.715)	(0.000)

Mitte	9.871***	9.580***	9.543***	8.931***	7.011***
	(0.259)	(0.259)	(0.259)	(0.258)	(0.261)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Neukolln	-2.148***	-1.695***	3.376***	-4.271***	-1.003***
	(0.265)	(0.265)	(0.291)	(0.266)	(0.266)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Pankow	6.222***	9.760***	10.014***	8.763***	7.487***
	(0.263)	(0.274)	(0.277)	(0.269)	(0.265)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Reinickendorf	-13.572***	-3.652***	-6.250***	-3.697***	-5.046***
	(0.425)	(0.473)	(0.458)	(0.472)	(0.460)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Schöneberg-Nord	4.220	-0.568	1.004	-0.431	-1.439
	-4.320	-4.351	-4.322	-4.352	-4.353
	(0.329)	(0.896)	(0.816)	(0.921)	(0.741)
Siemensstadt	-4.711	0.383	-1.747	0.586	-0.324
	-6.230	-6.237	-6.236	-6.236	-6.238
	(0.450)	(0.951)	(0.779)	(0.925)	(0.959)
Spandau	-6.291***	7.339***	5.767***	7.512***	8.661***
	(0.794)	(0.833)	(0.831)	(0.834)	(0.834)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Steglitz-Zehlendorf	-5.338***	0.017	3.415***	-1.030**	2.831***
	(0.422)	(0.437)	(0.459)	(0.433)	(0.448)
	(0.000)	(0.968)	(0.000)	(0.018)	(0.000)
Tempelhof-Schöneberg	0.602**	-3.038***	1.163***	-3.915***	-1.840***
	(0.301)	(0.306)	(0.300)	(0.311)	(0.302)
	(0.045)	(0.000)	(0.000)	(0.000)	(0.000)
Tiergarten Süd	14.989	8.920	12.674	8.880	8.161
	-12.949	-12.942	-12.947	-12.948	-12.948
	(0.247)	(0.491)	(0.328)	(0.493)	(0.528)
Treptow-Kopenick	-10.009***	2.963***	7.327***	0.509	5.046***
	(0.363)	(0.422)	(0.486)	(0.400)	(0.430)

	(0.000)	(0.000)	(0.000)	(0.204)	(0.000)
febbraio	4.675***	4.640***	4.647***	4.642***	4.636***
	(0.286)	(0.285)	(0.285)	(0.285)	(0.284)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
marzo	1.958***	1.957***	1.953***	1.961***	1.954***
	(0.278)	(0.277)	(0.277)	(0.277)	(0.277)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
aprile	11.436***	11.489***	11.485***	11.492***	11.496***
	(0.298)	(0.297)	(0.297)	(0.297)	(0.297)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
maggio	13.525***	13.606***	13.608***	13.610***	13.624***
	(0.298)	(0.297)	(0.297)	(0.297)	(0.297)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
giugno	19.207***	19.310***	19.312***	19.314***	19.331***
	(0.302)	(0.301)	(0.301)	(0.301)	(0.301)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
luglio	21.840***	21.923***	21.923***	21.929***	21.943***
	(0.292)	(0.291)	(0.291)	(0.291)	(0.291)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
agosto	23.849***	23.940***	23.939***	23.945***	23.959***
	(0.287)	(0.286)	(0.286)	(0.286)	(0.286)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
settembre	25.597***	25.680***	25.679***	25.686***	25.698***
	(0.291)	(0.290)	(0.290)	(0.290)	(0.290)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ottobre	21.703***	21.774***	21.777***	21.774***	21.790***
	(0.298)	(0.297)	(0.297)	(0.297)	(0.296)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
novembre	14.557***	14.618***	14.625***	14.614***	14.630***
	(0.319)	(0.317)	(0.318)	(0.317)	(0.317)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
dicembre	9.257***	9.340***	9.346***	9.337***	9.357***
	(0.302)	(0.301)	(0.301)	(0.301)	(0.301)

	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
2020	-16.752***	-16.637***	-16.625***	-16.633***	-16.608***
	(0.160)	(0.160)	(0.160)	(0.160)	(0.159)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
2021	-6.288***	-6.143***	-6.127***	-6.151***	-6.104***
	(0.189)	(0.189)	(0.189)	(0.189)	(0.189)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
2022	-65.598***	-64.822***	-64.842***	-64.779***	-64.647***
	-1.957	-1.957	-1.954	-1.957	-1.956
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
2023	-55.294***	-54.474***	-54.475***	-54.453***	-54.292***
	-1.966	-1.965	-1.963	-1.966	-1.964
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
distance metro		-0.027***			
		(0.001)			
		(0.000)			
distance monument			-0.026***		
			(0.001)		
			(0.000)		
distance park				-0.028***	
				(0.001)	
				(0.000)	
distance parking					-0.031***
					(0.001)
					(0.000)
_cons	20.744***	27.893***	25.913***	29.267***	29.097***
	(0.361)	(0.380)	(0.371)	(0.389)	(0.382)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	4.00e+05	4.00e+05	4.00e+05	4.00e+05	4.00e+05
R2	0.433	0.437	0.437	0.437	0.438
F	6.964.980	6.838.096	6.824.734	6.841.787	6.848.660

Tabella 12: Output di Regressione Multivariata

Dall'output è evidente che tutti i coefficienti delle variabili inserite nel modello sono statisticamente significativi.

La variazione di 1 m della distanza dalla metropolitana più vicina fa diminuire il RevPAN in media di 0,027 \$/giorno, dal monumento più vicino fa diminuire il RevPAN in media di 0,026 \$/giorno, dal parco più vicino fa diminuire il RevPAN in media di 0,028 \$/giorno, dal parcheggio più vicino fa diminuire il RevPAN in media di 0,031 \$/giorno.

Il valore dell'R2 aumenta progressivamente da 0,433 nel modello M1 a 0,438 in M5.

Nel grafico è riassunto il risultato:

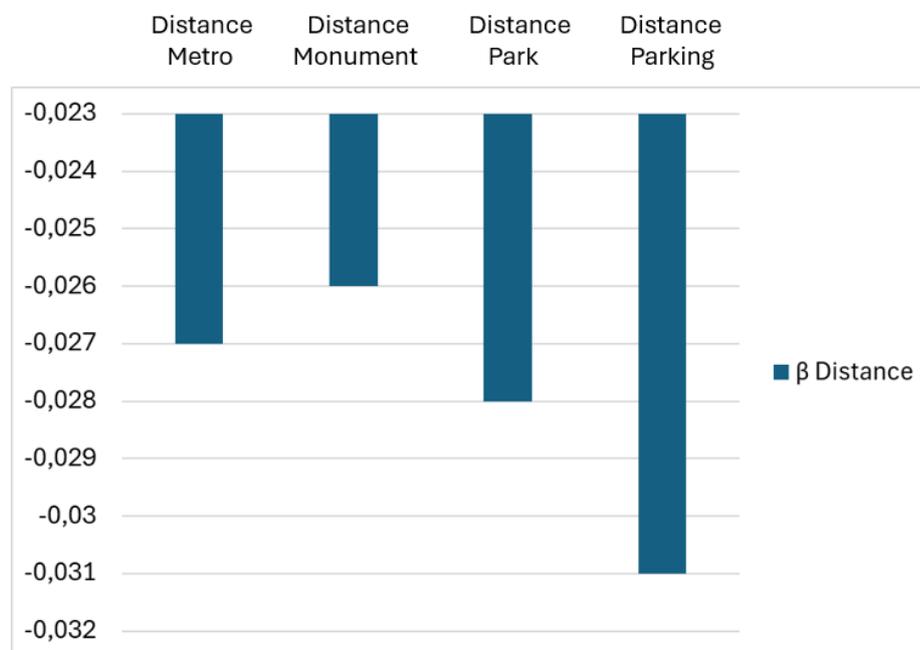


Figura 57: Grafico di Regressione Multivariata

10.2.1. Regressioni multivariate con Superhost

Sono state successivamente effettuate le stesse analisi di regressione multivariata introducendo la variabile dummy Airbnb Superhost come moderante.

Le equazioni di regressione risultano allora modificate nel seguente modo:

Modello M1:

$$\text{RevPAN} = \beta_0 + \beta_1 \text{maxguest} + \beta_2 \text{overallrating} + \beta_3 \text{professionalhost} + \beta_4 \text{hotelroom} + \beta_5 \text{privateroom} + \beta_6 \text{sharedroom} + \beta_7 \text{Charlottenburg-Wilmersdorf} + \beta_8 \text{Frankfurter Allee Süd} + \beta_9 \text{Friedrichshain-Kreuzberg} + \beta_{10} \text{Lichtenberg} + \beta_{11} \text{Marzahn-Hellersdorf} + \beta_{12} \text{Mitte} + \beta_{13} \text{Neukolln} + \beta_{14} \text{Pankow} + \beta_{15} \text{Reinickendorf} + \beta_{16} \text{Reuterstraße} + \beta_{17} \text{Schöneberg-Nord} + \beta_{18} \text{Siemensstadt} + \beta_{19} \text{Spandau} + \beta_{20} \text{Steglitz-Zehlendorf} + \beta_{21} \text{Tempelhof-Schöneberg} + \beta_{22} \text{Tiergarten Süd} + \beta_{23} \text{Treptow-Kopenick} + \beta_{24} \text{January} + \beta_{25} \text{February} + \beta_{26} \text{March} + \beta_{27} \text{April} + \beta_{28} \text{May} + \beta_{29} \text{June} + \beta_{30} \text{July} + \beta_{31} \text{August} + \beta_{32} \text{September} + \beta_{33} \text{October} + \beta_{34} \text{November} + \beta_{35} \text{December} + \beta_{36} \text{2020} + \beta_{37} \text{2021} + \beta_{38} \text{2022} + \beta_{39} \text{2023} + \beta_{40} \text{AirbnbSuperhost} + \varepsilon$$

Modelli M2, M3, M4, M5:

$$\text{RevPAN} = \beta_0 + \beta_1 \text{maxguest} + \beta_2 \text{overallrating} + \beta_3 \text{professionalhost} + \beta_4 \text{hotelroom} + \beta_5 \text{privateroom} + \beta_6 \text{sharedroom} + \beta_7 \text{Charlottenburg-Wilmersdorf} + \beta_8 \text{Frankfurter Allee Süd} + \beta_9 \text{Friedrichshain-Kreuzberg} + \beta_{10} \text{Lichtenberg} + \beta_{11} \text{Marzahn-Hellersdorf} + \beta_{12} \text{Mitte} + \beta_{13} \text{Neukolln} + \beta_{14} \text{Pankow} + \beta_{15} \text{Reinickendorf} + \beta_{16} \text{Reuterstraße} + \beta_{17} \text{Schöneberg-Nord} + \beta_{18} \text{Siemensstadt} + \beta_{19} \text{Spandau} + \beta_{20} \text{Steglitz-Zehlendorf} + \beta_{21} \text{Tempelhof-Schöneberg} + \beta_{22} \text{Tiergarten Süd} + \beta_{23} \text{Treptow-Kopenick} + \beta_{24} \text{January} + \beta_{25} \text{February} + \beta_{26} \text{March} + \beta_{27} \text{April} + \beta_{28} \text{May} + \beta_{29} \text{June} + \beta_{30} \text{July} + \beta_{31} \text{August} + \beta_{32} \text{September} + \beta_{33} \text{October} + \beta_{34} \text{November} + \beta_{35} \text{December} + \beta_{36} \text{2020} + \beta_{37} \text{2021} + \beta_{38} \text{2022} + \beta_{39} \text{2023} + \beta_{40} \text{AirbnbSuperhost} + \beta_{41} \text{Distance} + \beta_{42} \text{AirbnbSuperhost Distance} + \varepsilon$$

L'output di regressione ottenuto è il seguente:

RevPAN	M1	M2	M3	M4	M5
	b/se/p	b/se/p	b/se/p	b/se/p	b/se/p
max guests	10.183***	10.288***	10.290***	10.278***	10.308***
	(0.068)	(0.068)	(0.068)	(0.068)	(0.068)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
overall rating	0.933***	0.927***	0.927***	0.927***	0.926***
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
airbnb superhost	9.373***	12.270***	13.154***	11.910***	12.780***
	(0.156)	(0.305)	(0.315)	(0.302)	(0.301)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
professional host	1.649***	1.323***	1.411***	1.319***	1.256***
	(0.139)	(0.138)	(0.138)	(0.138)	(0.138)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
hotel room	5.614***	4.556***	4.917***	4.663***	4.523***
	(0.821)	(0.820)	(0.818)	(0.820)	(0.819)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
professional host	-28.169***	-28.006***	-28.037***	-27.996***	-27.946***
	(0.137)	(0.137)	(0.137)	(0.137)	(0.136)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
shared room	-69.216***	-68.811***	-68.358***	-68.977***	-68.615***
	(0.501)	(0.501)	(0.499)	(0.502)	(0.501)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Friedrichshain- Kreuzberg	6.713***	6.384***	10.128***	4.354***	5.526***
	(0.265)	(0.266)	(0.277)	(0.269)	(0.266)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Lichtenberg	-6.767***	0.302	3.311***	-1.759***	0.118
	(0.377)	(0.404)	(0.438)	(0.390)	(0.398)
	(0.000)	(0.454)	(0.000)	(0.000)	(0.767)
Marzahn-Hellersdorf	-13.625***	1.724**	4.525***	-0.332	2.743***

	(0.683)	(0.753)	(0.793)	(0.735)	(0.748)
	(0.000)	(0.022)	(0.000)	(0.651)	(0.000)
Mitte	9.666***	9.334***	9.302***	8.671***	6.696***
	(0.279)	(0.279)	(0.279)	(0.279)	(0.282)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Neukolln	-1.886***	-1.426***	3.676***	-4.045***	-0.746***
	(0.282)	(0.282)	(0.311)	(0.284)	(0.283)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.008)
Pankow	5.919***	9.526***	9.749***	8.510***	7.181***
	(0.284)	(0.296)	(0.299)	(0.290)	(0.285)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Reinickendorf	-14.769***	-4.382***	-7.174***	-4.419***	-5.794***
	(0.451)	(0.502)	(0.486)	(0.501)	(0.487)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Schöneberg-Nord	-1.469	-7.730*	-6.649*	-7.143*	-8.809**
	-4.113	-4.081	-4.025	-4.103	-4.073
	(0.721)	(0.058)	(0.099)	(0.082)	(0.031)
Siemensstadt	-4.899	-0.084	-1.935	0.068	-0.886
	-6.295	-6.302	-6.301	-6.302	-6.303
	(0.436)	(0.989)	(0.759)	(0.991)	(0.888)
Spandau	-7.637***	6.395***	4.734***	6.572***	7.784***
	(0.804)	(0.853)	(0.850)	(0.853)	(0.854)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Steglitz-Zehlendorf	-5.439***	0.020	3.507***	-1.044**	2.985***
	(0.492)	(0.504)	(0.525)	(0.501)	(0.513)
	(0.000)	(0.968)	(0.000)	(0.037)	(0.000)
Tempelhof-Schöneberg	0.311	-3.513***	0.801**	-4.396***	-2.274***
	(0.343)	(0.349)	(0.343)	(0.354)	(0.343)
	(0.364)	(0.000)	(0.019)	(0.000)	(0.000)
Tiergarten Süd	17.073	11.535	15.383	11.314	10.844
	-12.877	-12.871	-12.874	-12.877	-12.876
	(0.185)	(0.370)	(0.232)	(0.380)	(0.400)

Treptow-Kopenick	-10.767***	2.508***	6.881***	-0.003	4.645***
	(0.388)	(0.447)	(0.519)	(0.424)	(0.456)
	(0.000)	(0.000)	(0.000)	(0.994)	(0.000)
febbraio	4.092***	4.055***	4.063***	4.058***	4.049***
	(0.312)	(0.311)	(0.311)	(0.311)	(0.310)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
marzo	0.983***	0.983***	0.979***	0.988***	0.978***
	(0.300)	(0.299)	(0.299)	(0.299)	(0.299)
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
aprile	10.718***	10.778***	10.775***	10.781***	10.786***
	(0.334)	(0.333)	(0.334)	(0.333)	(0.333)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
maggio	12.724***	12.808***	12.810***	12.814***	12.827***
	(0.327)	(0.326)	(0.326)	(0.326)	(0.325)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
giugno	18.405***	18.511***	18.512***	18.517***	18.533***
	(0.325)	(0.324)	(0.324)	(0.324)	(0.324)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
luglio	20.816***	20.898***	20.898***	20.905***	20.917***
	(0.314)	(0.313)	(0.313)	(0.313)	(0.313)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
agosto	22.987***	23.079***	23.079***	23.085***	23.100***
	(0.309)	(0.309)	(0.309)	(0.309)	(0.308)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
settembre	25.469***	25.554***	25.557***	25.561***	25.575***
	(0.323)	(0.322)	(0.322)	(0.322)	(0.322)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ottobre	20.800***	20.872***	20.876***	20.872***	20.889***
	(0.322)	(0.321)	(0.321)	(0.321)	(0.321)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
novembre	13.867***	13.929***	13.935***	13.926***	13.943***
	(0.348)	(0.347)	(0.347)	(0.347)	(0.347)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

dicembre	8.697***	8.786***	8.790***	8.782***	8.802***
	(0.347)	(0.347)	(0.347)	(0.346)	(0.346)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
2020	-16.872***	-16.747***	-16.747***	-16.742***	-16.720***
	(0.170)	(0.170)	(0.170)	(0.170)	(0.170)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
2021	-4.817***	-4.682***	-4.662***	-4.693***	-4.646***
	(0.209)	(0.209)	(0.209)	(0.209)	(0.208)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
2022	-59.450***	-58.726***	-58.660***	-58.749***	-58.570***
	-2.134	-2.130	-2.127	-2.131	-2.128
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
2023	-49.573***	-48.727***	-48.654***	-48.776***	-48.559***
	-2.138	-2.134	-2.132	-2.135	-2.133
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
distance metro		-0.026***			
		(0.001)			
		(0.000)			
distance metro*superhost		-0.009***			
		(0.001)			
		(0.000)			
distance monument			-0.024***		
			(0.001)		
			(0.000)		
distance monument*superhost			-0.011***		
			(0.001)		
			(0.000)		
distance park				-0.026***	
				(0.001)	
				(0.000)	

distance park*superhost				-0.008***	
				(0.001)	
				(0.000)	
distance parking					-0.029***
					(0.001)
					(0.000)
distance parking*superhost					-0.011***
					(0.001)
					(0.000)
_cons	18.562***	25.105***	22.823***	26.614***	26.206***
	(0.399)	(0.428)	(0.418)	(0.438)	(0.430)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	4.00e+05	4.00e+05	4.00e+05	4.00e+05	4.00e+05
R2	0.408	0.411	0.411	0.411	0.412
F	5.916.541	5.663.949	5.653.941	5.665.359	5.671.255

Tabella 13: Output di Regressione Multivariata con Superhost

I coefficienti sono tutti statisticamente significativi.

La variazione di 1 m della distanza dalla metropolitana più vicina fa diminuire il RevPAN in media di 0,026 \$/giorno. Per i superhost il valore diminuisce ancora di 0,009 e quindi l'impatto totale negativo diventa pari a -0,035\$/giorno. La variazione di 1 m della distanza dal monumento più vicino fa diminuire il RevPAN in media di 0,024 \$/giorno e di 0,035\$/giorno ai superhost. La variazione di 1 m della distanza dal parco più vicino fa diminuire il RevPAN in media di 0,026 \$/giorno e di 0,034\$/giorno per i superhost. Infine, la variazione di 1 m della distanza dal parcheggio più vicino fa diminuire il RevPAN in media di 0,029 \$/giorno e di 0,040\$/giorno ai superhost.

Da questi risultati si può affermare che per i superhost l'impatto negativo delle distanze sul RevPAN è maggiore.

Il risultato ottenuto è schematizzato come segue:

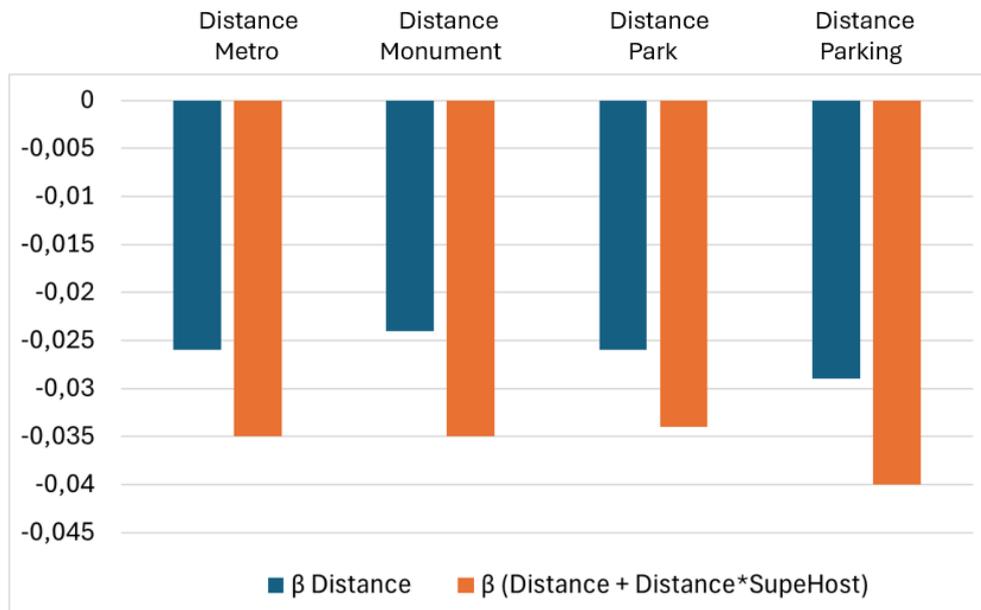


Figura 58: Grafico di Regressione Multivariata con Superhost

10.2.2. Regressioni multivariate con Long Stay

Sono state successivamente effettuate le stesse analisi di regressione ma con la variabile dummy Long Stay come moderante, che ha valore 1 qualora la durata minima del soggiorno richiesta dall'host ed espressa tramite la variabile Minimum Stay sia maggiore o uguale a 90 giorni e 0 altrimenti. Le equazioni di regressione risultano allora modificata nel seguente modo:

Modello M1:

$$\text{RevPAN} = \beta_0 + \beta_1 \text{maxguest} + \beta_2 \text{overallrating} + \beta_3 \text{professionalhost} + \beta_4 \text{hotelroom} + \beta_5 \text{privateroom} + \beta_6 \text{sharedroom} + \beta_7 \text{Charlottenburg-Wilmersdorf} + \beta_8 \text{Frankfurter Allee Süd} + \beta_9 \text{Friedrichshain-Kreuzberg} + \beta_{10} \text{Lichtenberg} + \beta_{11} \text{Marzahn-Hellersdorf} + \beta_{12} \text{Mitte} + \beta_{13} \text{Neukolln} + \beta_{14} \text{Pankow} + \beta_{15} \text{Reinickendorf} + \beta_{16} \text{Reuterstraße} + \beta_{17} \text{Schöneberg-Nord} + \beta_{18} \text{Siemensstadt} + \beta_{19} \text{Spandau} + \beta_{20} \text{Steglitz-Zehlendorf} + \beta_{21} \text{Tempelhof-Schöneberg} + \beta_{22} \text{Tiergarten Süd} + \beta_{23} \text{Treptow-Kopenick} + \beta_{24} \text{January} + \beta_{25} \text{February} + \beta_{26} \text{March} + \beta_{27} \text{April} + \beta_{28} \text{May} + \beta_{29} \text{June} + \beta_{30} \text{July} + \beta_{31} \text{August} + \beta_{32} \text{September} + \beta_{33} \text{October} + \beta_{34} \text{November} + \beta_{35} \text{December} + \beta_{36} \text{2020} + \beta_{37} \text{2021} + \beta_{38} \text{2022} + \beta_{39} \text{2023} + \beta_{40} \text{LongStay} + \varepsilon$$

Modelli M2, M3, M4, M5:

$RevPAN = \beta_0 + \beta_1 \text{ maxguest} + \beta_2 \text{ overallrating} + \beta_3 \text{ professionalhost} + \beta_4 \text{ hotelroom} + \beta_5 \text{ privateroom} + \beta_6 \text{ sharedroom} + \beta_7 \text{ Charlottenburg-Wilmersdorf} + \beta_8 \text{ Frankfurter Allee Süd} + \beta_9 \text{ Friedrichshain-Kreuzberg} + \beta_{10} \text{ Lichtenberg} + \beta_{11} \text{ Marzahn-Hellersdorf} + \beta_{12} \text{ Mitte} + \beta_{13} \text{ Neukolln} + \beta_{14} \text{ Pankow} + \beta_{15} \text{ Reinickendorf} + \beta_{16} \text{ Reuterstraße} + \beta_{17} \text{ Schöneberg-Nord} + \beta_{18} \text{ Siemensstadt} + \beta_{19} \text{ Spandau} + \beta_{20} \text{ Steglitz-Zehlendorf} + \beta_{21} \text{ Tempelhof-Schöneberg} + \beta_{22} \text{ Tiergarten Süd} + \beta_{23} \text{ Treptow-Kopenick} + \beta_{24} \text{ January} + \beta_{25} \text{ February} + \beta_{26} \text{ March} + \beta_{27} \text{ April} + \beta_{28} \text{ May} + \beta_{29} \text{ June} + \beta_{30} \text{ July} + \beta_{31} \text{ August} + \beta_{32} \text{ September} + \beta_{33} \text{ October} + \beta_{34} \text{ November} + \beta_{35} \text{ December} + \beta_{36} \text{ 2020} + \beta_{37} \text{ 2021} + \beta_{38} \text{ 2022} + \beta_{39} \text{ 2023} + \beta_{40} \text{ LongStay} + \beta_{41} \text{ Distance} + \beta_{42} \text{ LongStay Distance} + \varepsilon$

L'output di regressione ottenuto è il seguente:

RevPAN	M1	M2	M3	M4	M5
	b/se/p	b/se/p	b/se/p	b/se/p	b/se/p
max guests	10.098***	10.187***	10.192***	10.179***	10.206***
	(0.069)	(0.069)	(0.069)	(0.069)	(0.069)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
overall rating	1.076***	1.069***	1.070***	1.069***	1.068***
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
long stay	-11.063***	-12.273***	-11.511***	-12.441***	-11.351***
	(0.487)	(0.968)	(0.988)	(0.958)	(0.976)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
professional host	2.068***	1.771***	1.853***	1.768***	1.713***
	(0.139)	(0.139)	(0.139)	(0.139)	(0.139)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
hotel room	6.119***	5.025***	5.407***	5.121***	4.973***
	(0.823)	(0.822)	(0.820)	(0.822)	(0.820)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
professional host	-28.184***	-28.072***	-28.119***	-28.054***	-28.032***
	(0.138)	(0.137)	(0.137)	(0.137)	(0.137)

	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
shared room	-71.128***	-70.822***	-70.407***	-70.981***	-70.711***
	(0.501)	(0.502)	(0.499)	(0.503)	(0.501)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Friedrichshain- Kreuzberg	6.785***	6.516***	10.276***	4.468***	5.661***
	(0.266)	(0.266)	(0.278)	(0.270)	(0.267)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Lichtenberg	-7.003***	0.163	3.180***	-1.910***	-0.018
	(0.377)	(0.405)	(0.440)	(0.391)	(0.399)
	(0.000)	(0.688)	(0.000)	(0.000)	(0.963)
Marzahn- Hellensdorf	-13.099***	2.056***	4.738***	0.051	3.000***
	(0.695)	(0.769)	(0.811)	(0.751)	(0.765)
	(0.000)	(0.008)	(0.000)	(0.946)	(0.000)
Mitte	9.613***	9.316***	9.282***	8.645***	6.662***
	(0.280)	(0.280)	(0.280)	(0.280)	(0.283)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Neukölln	-2.503***	-2.038***	3.158***	-4.696***	-1.322***
	(0.283)	(0.283)	(0.312)	(0.285)	(0.284)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Pankow	6.030***	9.675***	9.912***	8.652***	7.332***
	(0.285)	(0.297)	(0.301)	(0.292)	(0.286)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Reinickendorf	-14.037***	-3.806***	-6.527***	-3.842***	-5.228***
	(0.447)	(0.500)	(0.483)	(0.499)	(0.485)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Schöneberg-Nord	3.922	-1.043	0.616	-0.904	-1.925
	-4.416	-4.450	-4.420	-4.451	-4.452
	(0.375)	(0.815)	(0.889)	(0.839)	(0.665)
Siemensstadt	-6.606	-1.363	-3.605	-1.133	-2.108
	-6.279	-6.286	-6.286	-6.286	-6.287
	(0.293)	(0.828)	(0.566)	(0.857)	(0.737)

Spandau	-7.488***	6.580***	4.883***	6.774***	7.962***
	(0.802)	(0.851)	(0.848)	(0.852)	(0.851)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Steglitz- Zehlendorf	-4.624***	0.909*	4.363***	-0.166	3.829***
	(0.500)	(0.508)	(0.527)	(0.506)	(0.517)
	(0.000)	(0.074)	(0.000)	(0.743)	(0.000)
Tempelhof- Schöneberg	0.480	-3.277***	1.052***	-4.186***	-2.043***
	(0.343)	(0.349)	(0.343)	(0.355)	(0.343)
	(0.162)	(0.000)	(0.002)	(0.000)	(0.000)
Tiergarten Süd	20.051	14.295	17.874	14.235	13.136
	-13.083	-13.098	-13.088	-13.105	-13.086
	(0.125)	(0.275)	(0.172)	(0.277)	(0.315)
Treptow- Kopenick	-10.820***	2.581***	6.954***	0.066	4.725***
	(0.387)	(0.451)	(0.524)	(0.427)	(0.460)
	(0.000)	(0.000)	(0.000)	(0.876)	(0.000)
febbraio	4.319***	4.283***	4.291***	4.286***	4.280***
	(0.311)	(0.310)	(0.310)	(0.310)	(0.309)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
marzo	1.311***	1.311***	1.308***	1.316***	1.309***
	(0.300)	(0.299)	(0.299)	(0.299)	(0.299)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
aprile	11.043***	11.101***	11.096***	11.105***	11.109***
	(0.335)	(0.334)	(0.334)	(0.334)	(0.334)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
maggio	13.115***	13.200***	13.200***	13.205***	13.218***
	(0.327)	(0.326)	(0.326)	(0.326)	(0.326)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
giugno	18.811***	18.919***	18.919***	18.924***	18.941***
	(0.325)	(0.324)	(0.324)	(0.324)	(0.324)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

luglio	21.139***	21.227***	21.226***	21.234***	21.248***
	(0.315)	(0.314)	(0.314)	(0.314)	(0.314)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
agosto	23.292***	23.387***	23.386***	23.392***	23.408***
	(0.310)	(0.309)	(0.309)	(0.309)	(0.309)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
settembre	25.770***	25.854***	25.853***	25.861***	25.872***
	(0.324)	(0.323)	(0.323)	(0.323)	(0.322)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ottobre	21.235***	21.309***	21.312***	21.309***	21.326***
	(0.322)	(0.321)	(0.321)	(0.321)	(0.321)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
novembre	14.438***	14.503***	14.508***	14.499***	14.515***
	(0.348)	(0.346)	(0.347)	(0.346)	(0.346)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
dicembre	9.013***	9.098***	9.104***	9.095***	9.116***
	(0.347)	(0.346)	(0.346)	(0.346)	(0.346)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
2020	-16.509***	-16.391***	-16.379***	-16.387***	-16.360***
	(0.172)	(0.171)	(0.171)	(0.171)	(0.171)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
2021	-5.607***	-5.450***	-5.434***	-5.459***	-5.410***
	(0.210)	(0.209)	(0.209)	(0.209)	(0.209)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
2022	-72.424***	-71.578***	-71.596***	-71.541***	-71.356***
	-2.151	-2.149	-2.147	-2.150	-2.148
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
2023	-62.536***	-61.654***	-61.654***	-61.639***	-61.424***
	-2.160	-2.158	-2.156	-2.158	-2.157
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
distance metro		-0.028***			
		(0.001)			
		(0.000)			

distance metro*longstay		0.003			
		(0.003)			
		(0.252)			
distance monument			-0.027***		
			(0.001)		
			(0.000)		
distance monument*long stay			0.000		
			(0.003)		
			(0.942)		
distance park				-0.029***	
				(0.001)	
				(0.000)	
distance park*long stay				0.004	
				(0.003)	
				(0.158)	
distance parking					-0.032***
					(0.001)
					(0.000)
distance parking*long stay					-0.000
					(0.003)
					(0.951)
_cons	20.372***	27.786***	25.679***	29.219***	29.007***
	(0.402)	(0.424)	(0.414)	(0.434)	(0.426)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	4.00e+05	4.00e+05	4.00e+05	4.00e+05	4.00e+05
R2	0.399	0.403	0.402	0.403	0.404
F	5.767.509	5.511.105	5.502.601	5.513.761	5.518.186

Tabella 14: Output di Rregressione Multivariata con Long Stay

I coefficienti sono tutti statisticamente significativi.

La variazione di 1 m della distanza dalla metropolitana più vicina fa diminuire il RevPAN in media di 0,028 \$/giorno. Se il soggiorno è maggiore di 90 giorni il valore aumenta di 0,03, ma in modo non statisticamente significativo e quindi l'impatto totale negativo diventa pari a -0,025\$/giorno. La variazione di 1 m della distanza dal monumento più vicino fa diminuire il RevPAN in media di 0,027 \$/giorno sia per soggiorni brevi che lunghi. La variazione di 1 m della distanza dal parco più vicino fa diminuire il RevPAN in media di 0,029 \$/giorno, 0,025\$/giorno per i soggiorni di lungo periodo. Infine, la variazione di 1 m della distanza dal parcheggio più vicino fa diminuire il RevPAN in media di 0,032 \$/giorno per qualsiasi durata del soggiorno.

Da queste analisi, non si può affermare che l'impatto negativo delle distanze sul RevPAN sia influenzato dalla durata del soggiorno.

I risultati delle regressioni sono schematizzati come segue:

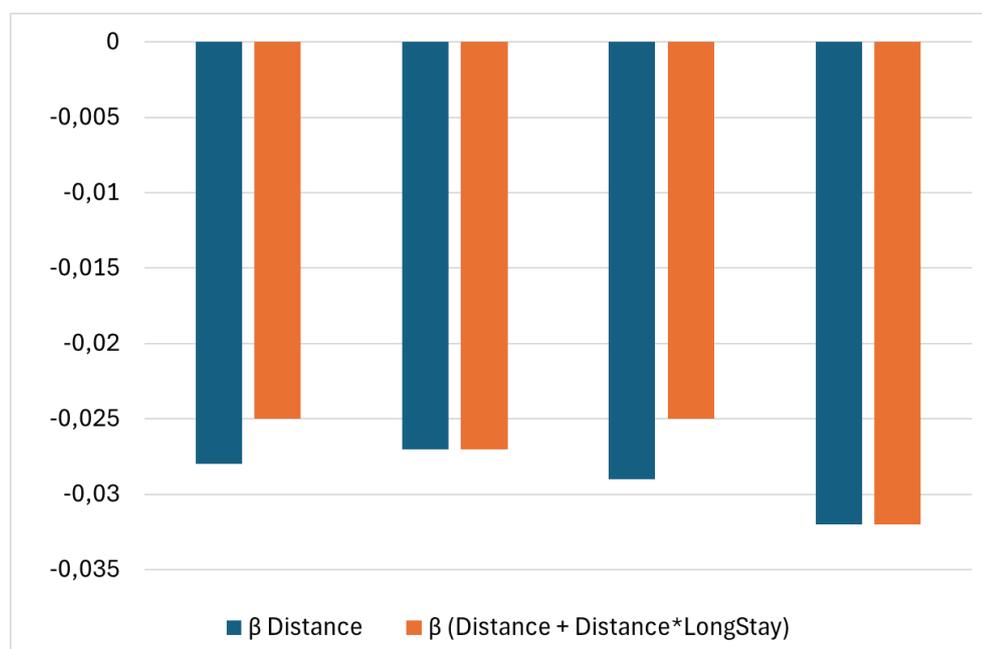


Figura 59: Grafico di Regressione Multivariata con Long Stay

11. Conclusioni

Le analisi di regressione multivariate sono lo strumento con cui è stato possibile confermare le ipotesi di ricerca formulate inizialmente.

Si è dimostrato che la vicinanza alle stazioni della metropolitana, ai monumenti, ai parchi e ai parcheggi causa un aumento dei ricavi per notte disponibile, e quindi in generale un miglioramento delle performance degli Airbnb.

Il titolo di superhost è effettivamente correlato ad un aumento delle dimensioni dell'impatto delle distanze sulle performance, cioè a parità di aumento di distanza da un appartamento, i superhost registrano una perdita di RevPAN maggiore e quindi, allo stesso tempo, a parità di riduzione della distanza registrano invece un incremento del RevPAN maggiore. Questo significa che i superhost sono stati più abili a cogliere l'importanza strategica della posizione degli appartamenti vicina a punti di interesse turistico.

La durata del soggiorno invece non ha un'influenza significativa sulla relazione tra le distanze e le performance. Non ci sono quindi differenze degne di nota nell'importanza della vicinanza a punti di interesse tra gli ospiti che soggiornano per brevi o lunghi periodi.

Il risultato finale della tesi è quindi in linea con i risultati dimostrati dallo studio di Buhalis, che aveva trovato un rapporto di causa effetto tra la vicinanza al centro città e alla metropolitana e la probabilità di prenotazione, in 10 città europee.

Questo studio sui dati di Berlino conferma quindi il risultato dello studio di Buhalis e lo amplia approfondendo l'influenza dei functional attributes di location sui ricavi per notte disponibile, analizzando una singola città.

12. Sviluppi futuri

Lo studio di ricerca condotto sui dati relativi agli Airbnb di Berlino tra il 2019 ed il 2023 può essere ampliato ed approfondito sotto molti aspetti.

Un possibile sviluppo futuro potrebbe essere l'analisi degli anni successivi al 2024.

Altrimenti si potrebbero costruire ulteriori analisi di regressione considerando aspetti del dataset non approfonditi in questo lavoro di tesi, ad esempio le differenze tra i quartieri della città.

Per quanto riguarda la pandemia, le due variabili su cui hanno puntato gli host per aumentare il numero di prenotazioni ed i ricavi sono la Cancellation Policy e l'abilitazione dell'Instantbook. Si potrebbe allora indagare come queste variabili abbiano un impatto sulle performance oppure considerare altre variabili strategiche.

Se consideriamo invece l'analisi effettuata sull'impatto delle distanze spaziali sulle performance, un possibile sviluppo futuro potrebbe essere l'analisi delle performance all'aumentare della distanza dal centro della città oppure da altri punti di interesse come bar, ristoranti, cinema, palestre, piscine, supermercati o negozi, sui cui si può ipotizzare un impatto negativo ed un risultato simile a quello ora trovato. Oppure si potrebbe ipotizzare un impatto positivo se si andasse ad analizzare la distanza da aree degradate, discariche, quartieri disagiati o con alto tasso di criminalità.

Lo stesso studio potrebbe essere sviluppato sulle altre 9 città europee dell'articolo di Nawaro, quindi su Amsterdam, Atene, Barcellona, Budapest, Lisbona, Londra, Parigi, Roma, Vienna oppure su altre grandi città nel mondo.

13. Ringraziamenti

Ringrazio dal profondo del cuore la mia famiglia che mi ha amata e sostenuta ogni giorno della mia vita, fino al raggiungimento di questo importante traguardo.

A loro ed al mio fratellone Michele dedico questa tesi di laurea magistrale che corona il mio percorso universitario.

Che Dio benedica tutti coloro che mi sono stati accanto.

14. Bibliografia

- Anna Farmaki, Cristina Miguel, Maria Hadjielia Drotarova, Ana Alekski, Anita Ceh Casni, Fani Efthymiadou, 2020, "Impacts of Covid-19 on peer-to-peer accommodation platforms: Host perceptions and responses", *International Journal of Hospitality Management*
- Daniel A. Guttentag, Stephen W. Litvin, Wayne W. Smith, 2023, "To Airbnb or not to Airbnb: Does Airbnb feel safer than hotels during a pandemic?", *International Journal of Hospitality Management*
- David Boto-García, 2022, "Heterogeneous price adjustments among Airbnb hosts amid Covid-19: Evidence from Barcelona", *International Journal of Hospitality Management*
- Francesco Luigi Milone, Ulrich Günter, Bozana Zekan, 2023, "The pricing of European airbnb listings during the pandemic: A difference-in-differences approach employing Covid-19 response strategies as a continuous treatment", *Tourism Management*
- Jongho Im, Jewoo Kim, Joon Yeon Choeh, 2021, "Covid-19, social distancing, and risk-averse actions of hospitality and tourism consumers: A case of South Korea." *Journal of Destination Marketing & Management*
- Juan Luis Nicolau, Abhinav Sharma, Hakseung Shin, Juhyun Kang, 2023, "Airbnb vs hotel? Customer selection behaviors in upward and downward Covid-19 trends", *International Journal of Contemporary Hospitality Management*
- Karima Kourtit, Peter Nijkamp, John Östh, Umut Turk, 2022, "Airbnb and Covid-19: space-time vulnerability effects in six world-cities", *Tourism Management* Kevin Kam
- Fung So, Giampaolo Viglia, Stephanie Liu, Dan Wang, 2023, "Guest editorial: The sharing economy in a post-pandemic world", *International Journal of Hospitality Management*
- Kyle Barron, Edward Kung, Davide Proserpio, 2021, "The Effect of Home-Sharing on House Prices and Rents: Evidence from Airbnb", *Marketing Science*
- Luigi Buzzacchi, Francesco Luigi Milone, Emilio Paolucci, Elisabetta Raguseo, 2023, "How to react to a shock? Effects of Airbnb hosts' choices and market segmentation at the time of Covid-19", *Information and management*
- Mo Zhang, Ruoqi Geng, Yuan Huang, Shengce Ren, 2021, "Terminator or accelerator? Lessons from the peer-to-peer hosts in China in responses to Covid-19 accommodation", *International Journal of Hospitality Management*
- Raffaele Filieri, Francesco Luigi Milone, Emilio Paolucci, Elisabetta Raguseo, 2023, "A big data analysis of Covid-19 impacts on Airbnbs' bookings behavior applying construal level and signaling theories", *International Journal of Hospitality Management*
- Richard T.R. Qiu, Daisy X.F. Fan, Anyu Liu, Dimitrios Buhalis, 2019, "Standing out from the crowd – an exploration of signal attributes of Airbnb listings", Bin Yao, *International Journal of Contemporary Hospitality Management*

Ruggero Sainaghi, Jorge Chica-Olmo, 2022, “The effects of location before and during Covid-19: Impacts on Revenue of Airbnb listings in Milan (Italy)”, *Annals of Tourism Research*

Sara Dolnicar, Samira Zare, 2020, “Covid19 and Airbnb – Disrupting the Disruptor”, *Annals of Tourism Research*

Seongsoo Jang, Jinwon Kim, 2021, “Remedying Airbnb COVID-19 disruption through tourism clusters and community resilience”, *Journal of Business Research*

Seongsoo Jang, Jungkeun Kim, Jinwon Kim, Seongseop Sam Kim, 2021, “Spatial and experimental analysis of peer-to-peer accommodation consumption during COVID-19”, *Journal of Destination Marketing & Management*

Tarik Dogru, Lydia Hanks, Courtney Suess, Nathan Line, Makarand Mody, 2023, “The resilience of the lodging industry during the pandemic: Hotels vs. Airbnb”, *International Journal of Hospitality Management*

Krist'of Gy'odi, Łukasz Nawaro, 2021, “Determinants of Airbnb prices in European cities: a spatial econometrics approach”, *Tourism Management*

David Boto-García, Matías Mayor, Pablo De la Vega, 2022, “Spatial price mimicking on Airbnb: Multi-host vs singlehost”, *Tourism Management*

Jorge Chica-Olmo, Juan Gabriel González-Morales, José Luis Zafra-Gómez, 2020, “Effects of location on Airbnb apartment pricing in Málaga”, *Tourism Management*