



**Politecnico  
di Torino**

**Corso di Laurea Magistrale in Ingegneria Gestionale**

**A.A. 2024/2025**

**Sessione di Laurea: Marzo/Aprile**

***Il gender gap su Airbnb: un'analisi empirica sulla disparità di  
genere nelle performance degli host***

Relatori

*Elisabetta Raguseo*

*Francesco Luigi Milone*

Candidato

*Francesca Lanzara*

*Per aspera ad astra*

## **Abstract**

La discriminazione contro le donne rappresenta un problema economico e sociale. Sebbene questo tema sia stato ampiamente studiato in diversi ambiti socioeconomici e in diversi settori di attività, la letteratura non dà indicazioni chiare sull'interazione tra piattaforme digitali, un fenomeno in grande espansione negli ultimi decenni, e discriminazione femminile. Per offrire una prospettiva su questo tema relativamente poco studiato, il lavoro si concentra sugli annunci disponibili sulla piattaforma di Airbnb nella città di Torino. Viene utilizzato un dataset microeconomico dettagliato reso disponibile da AirDNA. L'analisi empirica si basa su modello econometrico per dati panel, che coniugano la variabilità cross-sezionale con quella temporale. Il lavoro mira ad identificare la presenza di un divario di genere tra gli host della piattaforma Airbnb e i risultati evidenziano che, sebbene le host donne ricevano valutazioni più alte nelle recensioni, il gender gap persiste, con gli host uomini che registrano valori superiori in termini di RevPAR, numero di prenotazioni e tasso di occupazione.

La tesi si articola di quattro capitoli principali.

Il primo capitolo introduce l'argomento dello studio, fornendo il contesto e gli obiettivi della ricerca. Il secondo mostra una panoramica del fenomeno di gender gap, esaminando le sue manifestazioni nei contesti tradizionali e digitali, con un focus specifico sulla sharing economy. Il terzo capitolo è dedicato all'analisi approfondita della piattaforma Airbnb evidenziando le caratteristiche specifiche di questo contesto economico. Infine, il quarto capitolo, presenta l'analisi empirica in cui vengono descritti i dati raccolti, spiegata la metodologia econometrica adottata per la loro analisi, mostrati i risultati ottenuti e le loro implicazioni, concludendo con riflessioni e suggerimenti per future ricerche.

# Indice

Capitolo 1 .....	4
1.1 Introduzione.....	4
1.2 Contributo alla letteratura.....	5
Capitolo 2 .....	7
2.1 La Discriminazione nella Letteratura Empirica.....	7
2.2 Discriminazione femminile .....	10
2.3 Global Gender Gap.....	11
2.3.1 Global Gender Gap: la situazione in Italia .....	14
2.4 Gender Pay Gap.....	16
2.5 Le Piattaforme Digitali .....	18
2.6 La sharing economy.....	19
2.7 Il Gender Gap nelle piattaforme digitali.....	19
Capitolo 3 .....	21
3.1 Introduzione ad Airbnb.....	21
3.1.1 Storia di Airbnb .....	21
3.2 Funzionamento di Airbnb .....	22
3.3 Il gender gap in Airbnb.....	22
4 Capitolo 4 .....	25
4.1 Analisi del database AirDNA .....	25
4.2 Misure di performance.....	27
4.3 Estrapolazione dei nomi degli Host e analisi del genere.....	28
4.4 Strumenti per l'analisi .....	32
4.4.1 Excel.....	32
4.4.2 Stata .....	32
4.5 Analisi descrittive preliminari .....	32

4.5.1	Distribuzione degli host per genere .....	33
4.5.2	Distribuzione degli annunci per genere dell'host .....	33
4.5.3	Distribuzione nel tempo delle prenotazioni in base al genere dell'host .....	34
4.5.4	Distribuzione delle entrate in base al genere dell'host .....	35
4.5.5	Punteggio delle recensioni in base al genere dell'host .....	36
4.5.6	Analisi dei Superhost in base al genere .....	37
4.5.7	Servizi offerti dagli host in base al genere .....	38
4.5.8	Distribuzione per fascia di prezzo in base al genere .....	39
4.5.9	ADR.....	40
4.5.10	Listing Type .....	41
4.5.11	Distribuzione del genere per tipologia di annuncio.....	42
4.5.12	Distribuzione giorni prenotati e disponibili per genere dell'host .....	43
4.5.13	Tasso di occupazione medio per genere dell'host .....	44
4.5.14	Distribuzione camere per annuncio .....	45
4.5.15	Distribuzione del RevPAN .....	46
4.6	Analisi Statistiche inferenziali.....	48
4.6.1	Analisi di regressione del Panel.....	48
4.6.2	Regressioni Multivariate .....	50
4.6.3	Output delle regressioni.....	51
4.6.3.1	Regressione multivariata lineare .....	55
4.6.3.2	Regressione multivariata logaritmica .....	62
4.6.3.3	Effetto di moderazione .....	64
4.6.3.4	Regressioni multivariata con interazioni .....	65
4.7	Interaction plot.....	81
	Conclusioni.....	85
	Implicazioni future .....	86

Bibliografia..... 87

# Capitolo 1

## 1.1 Introduzione

L'Italia si caratterizza per evidenti differenze di genere nel mercato del lavoro (Quintano et al., 2010; Pena-Boquete et al., 2010; Cutillo & Centra, 2017). La letteratura recente mette in evidenza il problema della discriminazione femminile come radice culturale delle disparità di genere (Carter et al., 2019; Agovino et al., 2023). In Italia, infatti, le posizioni socialmente conservatrici e gli stereotipi tradizionali di genere sono rafforzati dall'influenza storica della Chiesa Cattolica (Worthen et al., 2017). Le politiche pubbliche implementate nella direzione della parità di genere d'altro canto hanno avuto storicamente un carattere molto cauto, accumulando un ritardo significativo rispetto agli altri paesi europei (Severini et al., 2017; Furlotti et al., 2019).

Nonostante la persistenza di un lungo periodo del divario di genere, l'economia italiana degli ultimi 30 anni ha subito importanti trasformazioni. Un elemento importante riguarda la rapida diffusione delle piattaforme digitali, che hanno drasticamente ridefinito i modelli di business tradizionali (Hein et al., 2020). Parte del successo di queste imprese, che si sono da subito caratterizzate per un forte focus digitale, è legato alla disponibilità da parte degli utenti (consumatori e venditori) a partecipare a una community internazionale, in cui la condivisione di beni e servizi è stata sdoganata (Möhlmann, 2015). I modelli di sharing economy pongono tipicamente il consumatore in una posizione nuova: l'utente non possiede un bene a tempo indeterminato, ma ne gode i benefici in una finestra limitata, per poi passarlo al prossimo utente (Ritter & Schanz, 2019; Zhu & Liu, 2021).

La diffusione in Italia delle piattaforme digitali può rappresentare un momento di cambiamento e di rottura dei tradizionali schemi di perpetuazione delle differenze di genere, in un mondo internazionale ed interconnesso che procede ad una velocità superiore. Di conseguenza, questo studio si propone di indagare la presenza di discriminazione di genere nel mercato del lavoro delle piattaforme digitali, concentrandosi sulla piattaforma di Airbnb.

Come caso di studio è stata identificata la città di Torino, per la quale sono disponibili dati dettagliati sulle dinamiche del mercato degli affitti immobiliari a breve termine. In

particolare, il lavoro punta ad identificare l'eventuale presenza di discriminazione femminile, catturata dal gender gap, una metrica che verrà discussa nel Capitolo 2.

Dal punto di vista metodologico, grazie all'ampia disponibilità di dati di natura microeconomica, si propone un'analisi econometrica di tipo panel. I risultati rilevano la presenza di discriminazione in alcune dimensioni, ma mostrano al contempo che questa discriminazione viene mitigata dalle caratteristiche del bene considerato.

## **1.2 Contributo alla letteratura**

Il presente studio si pone l'obiettivo di esaminare il divario di genere su Airbnb, una delle piattaforme di sharing economy più riconosciute nel settore dell'ospitalità. (Adamiak, 2019; Airbnb, 2019; Li & Srinivasan, 2019; Zervas et al., 2017).

L'obiettivo di questo lavoro è quello di colmare la carenza di ricerche sull'argomento, con un particolare focus sulle differenze di performance nel genere sulla piattaforma di Airbnb.

La già menzionata piattaforma è stata scelta per le sue particolari caratteristiche che potrebbero condurre ad una riduzione del gender gap in quanto permette agli host di avere orari flessibili e autonomia, permettendo di stabilire autonomamente i prezzi degli annunci. (Davidson, A., & Gleim, M. - 2022).

È importante sottolineare che la piattaforma Airbnb si basa sull'ospitalità e sulla fiducia percepita dagli ospiti ed alcune ricerche (Ert e Fleischer, 2020) suggeriscono che le donne potrebbero avere un vantaggio in questi aspetti rispetto agli uomini. Inoltre, uno studio di "Ert e al. (2024)" ha dimostrato che spesso gli utenti preferiscono soggiornare in appartamenti gestiti dalle donne piuttosto che dagli uomini, un aspetto che potrebbe tradursi in una riduzione del divario di guadagno tra i generi. Dunque, in questo contesto le host donne potrebbero avere l'opportunità di una retribuzione più equa, riducendo le potenziali differenze di genere o addirittura favorendo queste ultime. (Fleischer, A., & Kopolovich, D. – 2024).

In contrapposizione ai già menzionati studi, emerge però la presenza di ricerche che dimostrano che il gender gap continua a persistere nelle piattaforme digitali e in particolare, in Airbnb, con host uomini che riescono a ottenere guadagni annuali più elevati grazie alla loro capacità di ottenere margini più consistenti, generare più soggiorni

e ospitare un numero maggiore di ospiti per soggiorno. (Davidson, A., & Gleim, M. R. (2022)).

I risultati contrastanti, dunque, hanno spinto ad approfondire la predetta dinamica attraverso un'analisi econometrica in un caso studio rilevante.

La città di Torino è stata presa in esame per il seguente studio in quanto è la quarta città in Italia per abitanti (ISTAT 2024) ed è una delle città italiane maggiormente interessate dal turismo, grazie ad un ricco patrimonio culturale, una forte tradizione industriale e una crescente vocazione all'innovazione.

Ex capitale d'Italia, rappresenta una città storica con un'attrattività consolidata dal punto di vista artistico-culturale. Intatti ospita alcune delle più importanti manifestazioni culturali e sportive come il Salone internazionale del Libro, gli ATP di tennis e partite di calcio di livello nazionale e internazionale

Alla luce di queste considerazioni, in aggiunta alla carenza di studi sulla disparità di genere nella città di Torino, ha portato a scegliere quest'ultima come caso studio rilevante offrendo nuovi spunti di riflessione.

Sulla base delle suddette considerazioni, si formula la seguente domanda di ricerca:

*Esiste una differenza nelle prestazioni tra host donne e host uomini su Airbnb per annunci comparabili?*

# Capitolo 2

## 2.1 La Discriminazione nella Letteratura Empirica

Il problema della discriminazione è stato affrontato da diverse prospettive nella letteratura scientifica (Becker, 1971; Ángeles Cea D'Ancona, 2017; Carter et al., 2019; Boring & Philippe, 2021). Se questo fenomeno riscuote interesse da decenni, il motivo è che rappresenta da un lato una minaccia alla coesione sociale e al benessere delle categorie discriminate e dall'altro lato un problema economico, che danneggia il livello di produzione (Sedgley and Elmslie, 2018; Heath and Di Stasio, 2019).

La teoria economica distingue la discriminazione basata sulle preferenze da quella di natura statistica (Lahey, 2008). Nel primo caso, i datori di lavoro, i colleghi, oppure i clienti subiscono una disutilità nell'interfacciarsi con membri delle minoranze, a causa di un costo psicologico che nasce dal pregiudizio. Questo è il caso del razzismo e della xenofobia, che distorcono il comportamento degli agenti economici, i quali finiscono per disprezzare certe categorie di persone indipendentemente dalle informazioni disponibili sulle loro caratteristiche diverse dal background etnico (Zschirnt & Ruedin, 2016).

Nel caso di discriminazione basata sulle preferenze, gli agenti economici sono disposti a rinunciare a una fetta di benessere o profitto pur di evitare le interazioni con i membri delle minoranze. Questo tipo di comportamento è assolutamente inefficiente ed in contrasto con il principio di razionalità delle scelte economiche.

La discriminazione statistica al contrario è basata su un utilizzo efficiente di informazioni incomplete. Poiché i datori di lavoro non conoscono tutte le caratteristiche dei potenziali lavoratori, ma osservano solo un sottoinsieme di caratteristiche sociodemografiche, sono costretti a fare un uso il più ampio possibile delle informazioni a disposizione. Così, se un individuo appartiene a gruppo caratterizzato da risultati socioeconomici meno favorevoli rispetto alla media, in assenza di informazioni a livello individuale, verrà trattato come la media del suo gruppo, e quindi discriminato (Phelps, 1972; Midtbøen, 2014). In questo, le caratteristiche osservabili assumono un peso preponderante rispetto a quelle inosservabili. Sebbene la discriminazione statistica danneggi i membri più virtuosi delle categorie meno virtuose, dal punto di vista sociale rappresenta un

meccanismo che migliora il benessere in presenza di un fallimento di mercato legato all'asimmetria informativa (Arrow, 1973).

Diversi tentativi sono stati fatti per misurare il livello di discriminazione che colpisce svariate categorie. Alcuni studi utilizzano misure dirette del fenomeno, basate sulla valutazione della percezione da parte della popolazione generale (Riach et al., 1991; Agovino et al., 2021). Altri studi invece si concentrano sulla percezione del fenomeno da parte delle categorie discriminate (Brondolo et al., 2005; Brenick et al., 2012). Il problema di questi lavori è che rappresentano principalmente casi studio di ampiezza e validità esterna limitati.

Altri lavori hanno tentato di dedurre il livello di discriminazione sulla base di variabili, soprattutto di natura economica, considerando ad esempio il tasso di occupazione, il reddito e i consumi delle categorie di interesse (Cain, 1986; Chambers, 1992; Wachter and Megbolugbe, 1992; Goldberg, 1996; Kiel and Zabel, 1996; Nordin & Rooth, 2009; Midtbøen, 2015; Bunel et al., 2016). Gli approcci pionieristici di Oaxaca (1973) consistevano nello scomporre la differenza tra il salario medio del gruppo non discriminato e il salario medio del gruppo discriminato in due componenti, di cui la prima è legata alla differenza nelle caratteristiche (osservabili) tra i due gruppi, mentre la seconda è data dalla pura discriminazione. Un esempio di questa strategia empirica è dato dal lavoro di Fitzenberger et al. (2011), in cui la differenza di salario tra i lavoratori sindacalizzati e quelli non sindacalizzati in Germania viene scomposta in (1) una parte legata alle caratteristiche dei lavoratori, tra cui genere, età, esperienza, grado di istruzione, stato civile, cittadinanza, tipo di contratto, dimensione dell'impresa, settore, orientamento politico e (2) una parte legata allo status di membro del sindacato. La seconda componente rappresenta l'effetto della discriminazione a parità di tutte le altre caratteristiche. La letteratura più recente può essere divisa in tre diversi tipi di approcci:

- Audit Studies: individui con caratteristiche del tutto simili, ma che presentano una sola differenza (ad esempio un gruppo di bianchi e un gruppo di persone di colore con i medesimi livelli di istruzione, esperienza lavorativa, età, ecc.) vengono inviati sul campo e si confrontano i risultati da loro ottenuti, al fine di identificare l'effetto della loro unica differenza. Questa letteratura mostra che per una serie di outcome socioeconomici emerge un problema sostanziale di trattamento diverso per i membri delle minoranze, sia che si tratti dell'ottenimento di sconti (Ayres &

Siegelman, 1995), di posizioni lavorative (Neumark et al., 1996) o delle conseguenze di una fedina penale sporca (Pager, 2003).

- Correspondence Studies: in questo caso piuttosto che persone reali, si ricorre a dei gruppi fittizi molto più numerosi, ad esempio inviando curricula alle imprese, oppure richieste di affitto di appartamenti ai proprietari. Questo tipo di approccio garantisce numeri molto più grandi rispetto agli audit studies e non richiede di utilizzare persone reali, le cui caratteristiche inosservabili potrebbero rappresentare un ostacolo all'identificazione dell'effetto della discriminazione. Alcuni esempi legati direttamente al problema del razzismo sono rappresentati da studi in cui le persone fittizie si caratterizzano per un nome straniero oppure tipicamente legato a una minoranza etnico-linguistica (Blommaert et al., 2014; Carlsson & Eriksson, 2014; Galarza & Yamada, 2014; Flage, 2018). Bertrand & Mullainathan, (2004) mostrano come la probabilità di essere invitati per un colloquio di lavoro diminuisca sensibilmente per candidati che portano un nome di battesimo tipicamente adottato dalla minoranza afroamericana degli Stati Uniti. Un esperimento simile condotto in Italia mostra risultati analoghi, sottolineando il problema della natura della discriminazione (Busetta et al., 2018). In particolare, differenziando tra immigrati di prima e di seconda generazione, gli autori trovano un minore probabilità di ottenere un colloquio per gli immigrati di seconda generazione (nati, istruiti e cresciuti in Italia) rispetto ai candidati italiani.
- Association Studies: utilizzando i test IAT, mirati ad identificare il pregiudizio inconscio, alcuni studi hanno proposto questa misura diretta della discriminazione a livello individuale per valutarne l'impatto economico e sociale (Carlsson & Agerström, 2016; Bertrand & Duflo, 2017). Questo tipo di approccio si basa sul ruolo della discriminazione implicita nella determinazione delle scelte al livello individuale (Neumark, 2018) Ad esempio, Alesina et al. (2018) utilizzano questo test per valutare il livello di discriminazione degli insegnanti nelle scuole pubbliche della provincia di Milano e trovano un effetto persistente che penalizza i voti degli studenti immigrati di seconda generazione. La buona notizia è che una volta rivelato questo pregiudizio inconscio agli insegnanti, gli stessi riescono nel tempo a prendere delle contromisure, eliminando la distorsione che caratterizzava le loro valutazioni.

Questo filone di letteratura corrobora l'idea secondo cui la discriminazione delle minoranze etniche in Italia non è un problema di natura statistica, ma piuttosto emerge da una questione di preferenze da parte dei datori di lavoro. Sebbene questo approccio presenti dei vantaggi in termini di ampiezza e validità esterna, consente di ottenere non già una misura della discriminazione, ma piuttosto una misura dei suoi effetti. Alcune categorie potrebbero d'altronde esibire una maggiore 'resilienza' alla discriminazione e manifestarne in maniera più leggera gli effetti.

## **2.2 Discriminazione femminile**

Una categoria sociale che è stata oggetto di diverse forme di discriminazione nel corso della storia è costituita dalle donne, che rappresentano circa la metà della popolazione mondiale. Utilizzando le metriche disponibili, la letteratura economica riporta un significativo calo della discriminazione femminile nell'ultimo secolo, soprattutto nei paesi industrializzati, dove la crescita economica è stata accompagnata da un graduale empowerment femminile (Kleven et al., 2019). Nonostante questa tendenza al ribasso, la discriminazione femminile rimane una delle grandi sfide globali di natura sociale, la cui portata presenta un'ampia eterogeneità tra i paesi. Questo è particolarmente vero per l'Italia, un paese in cui posizioni socialmente conservatrici e gli stereotipi di genere tradizionali sono rafforzati dall'influenza storica della Chiesa cattolica (Worthen et al., 2017). In questo senso, la cultura dominante viene identificata come una delle principali cause della discriminazione femminile (Carter et al., 2019).

Ferrant e Kolev (2016) stimano che la perdita di reddito globale associata alla discriminazione femminile sia pari a circa il 16%. Allo stesso modo, secondo l'Organizzazione Internazionale del Lavoro (ILO), la discriminazione di genere costa fino al 10% del PIL totale in alcuni paesi (ILO, 2018). Hunt et al. (2016) affermano che colmare i divari di genere nella partecipazione alla forza lavoro, nell'occupazione e nei salari potrebbe aggiungere 12 trilioni di dollari al PIL globale entro il 2025, una stima sostanzialmente confermata da Woetzel et al. (2018). Sulla stessa frequenza d'onda, Neyer e Stempel (2021) mostrano che, a livello macroeconomico, la discriminazione femminile aggrava le recessioni in presenza di shock della domanda, riducendo la produzione e aumentando l'effetto inflazionistico delle politiche monetarie.

Complessivamente, dalla letteratura economica emerge un chiaro consenso sul fatto che ridurre la discriminazione femminile possa generare significativi benefici economici, tra cui maggiore produttività, migliori performance aziendali e una crescita economica più elevata (Shoreibah et al., 2019).

## **2.3 Global Gender Gap**

La gender equality è uno dei grandi obiettivi che gli Stati intendono raggiungere, consapevoli del fatto che una società fondata su tale principio è destinata a crescere in maniera responsabile e sostenibile sotto tutti i punti di vista.

Tale strategicità, da ultimo, è stata riconosciuta anche dalle Nazioni Unite, che hanno inserito proprio la gender equality tra i Sustainable Development Goals dell'Agenda 2030.

Iniziativa si registrano, in ogni caso, già dagli inizi degli anni 2000. Con l'obiettivo di incrementare la sensibilità sul tema, il "World Economic Forum", a partire dal 2006, pubblica ogni anno il Global Gender Gap Report, documento che fornisce una classifica mondiale di 146 paesi, secondo quattro parametri principali:

- Economia: in termini di partecipazione alla forza lavoro, opportunità in ambito lavorativo, parità salariale per mansioni simili;
- Educazione: analizzando tasso di alfabetizzazione, obbligo di frequenza scolastica e accessibilità ai percorsi di formazione;
- Politica: determinato dal suffragio universale e presenza femminile nei ruoli istituzionali;
- Salute: analizzando aspettativa di vita dei due generi, possibilità di accedere a cure di base e/o specialistiche.

La valorizzazione di questi KPIs consente di classificare, all'interno del Global Gender Gap Index e lungo una scala percentuale da 0% a 100% (Gender Parity Score), gli Stati partecipanti all'indagine.

La Figura 1 analizza i punteggi del "Global Gender Gap Index" e dei suoi sottocomponenti principali, misurati su una scala da 0 a 100, dove 100% rappresenta la piena parità di genere.

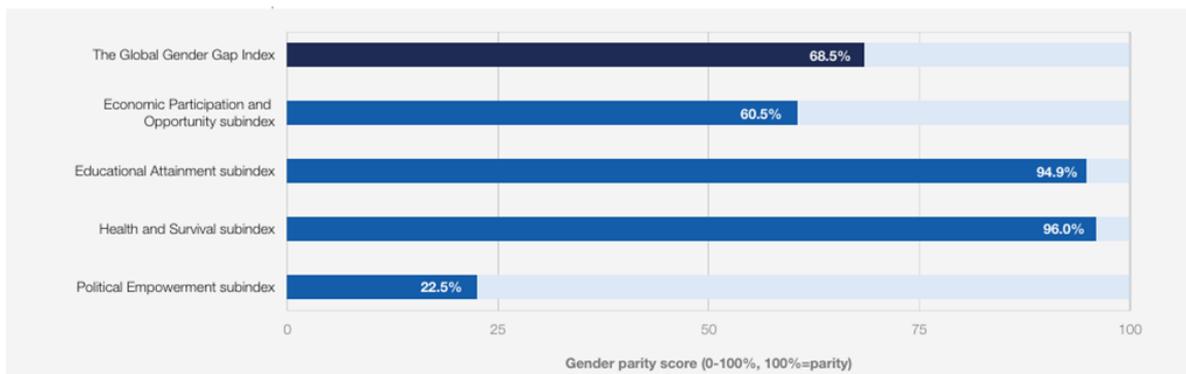


Figura 1: World Economic Forum, Global Gender Gap 2024, Gender Equality Index by subindex

Rispetto all'anno precedente, il 2024 ha registrato un complessivo miglioramento a livello globale dello 0.1%, attestandosi al 68.5% (Figura 2).

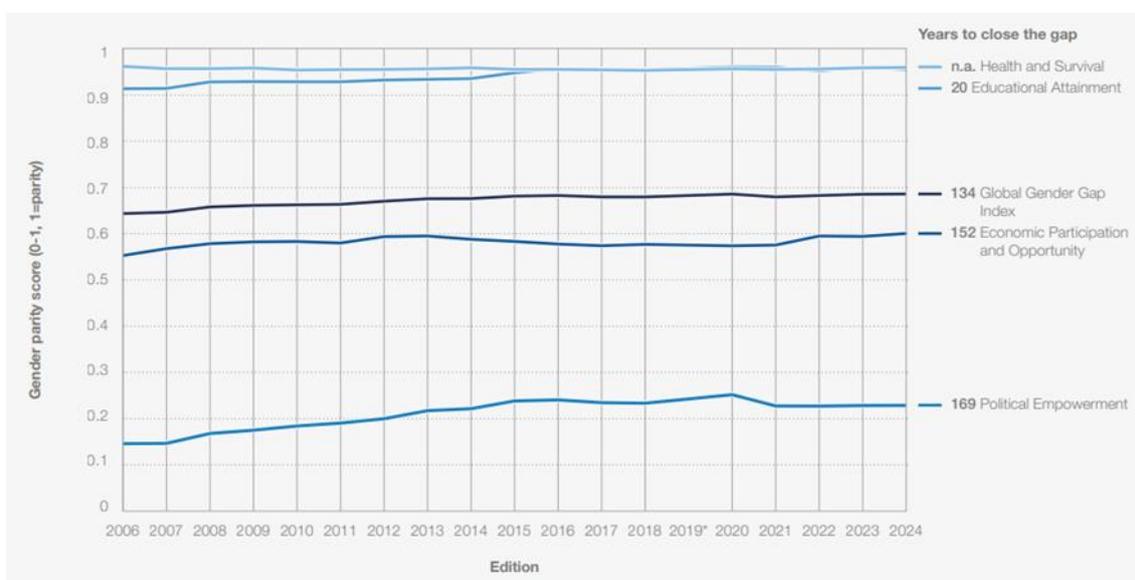


Figura 2: World Economic Forum, Global Gender Gap 2024, Evolution of the Global Gender Gap Index and subindexes over time

Esaminando i singoli KPIs, si scopre che l'aspetto economico è quello maggiormente interessato dal gender gap, attestandosi al 60.5% (nonostante un increase vs 2023 del 6%), secondo solo all'esposizione politica femminile, ferma al 22.5% (in aumento del 0.1% rispetto all'anno precedente). Bene, invece, la dimensione educativa e la dimensione sanitaria, rispettivamente al 94.9% (-0.5% vs 2023) e al 96% (+0.1% vs 2023), in cui, di fatto, si è quasi raggiunta la tanto agognata gender equality.

La Figura 3, invece, consente di approfondire la dimensione regionale.

	Subindexes				
	Global Gender Gap Index	Economic Participation and Opportunity	Educational Attainment	Health and Survival	Political Empowerment
Central Asia	69.1%	66.8%	98.7%	96.4%	14.6%
Eastern Asia and the Pacific	69.2%	71.8%	95.3%	95.0%	14.6%
Europe	75.0%	68.0%	99.5%	97.0%	35.7%
Latin America and the Caribbean	74.2%	65.7%	99.5%	97.6%	34.0%
Middle East and Northern Africa	61.7%	42.7%	96.8%	96.4%	10.8%
Northern America	74.8%	76.3%	100.0%	96.9%	26.0%
Southern Asia	63.7%	38.8%	94.5%	95.4%	26.0%
Sub-Saharan Africa	68.4%	67.4%	86.7%	97.2%	22.4%
<b>Global average</b>	<b>68.5%</b>	<b>60.5%</b>	<b>94.9%</b>	<b>96.0%</b>	<b>22.5%</b>

Parity 0% 50% 100%

Figura 3: World Economic Forum, Global Gender Gap 2024, Regional result for subindexes

Come anticipato, l'Europa guida la classifica, con un punteggio del 75% (+6.2% rispetto al 2006) e con ben 5 Stati (Islanda, Finlandia, Norvegia, Svezia e Germania) che si attestano nella top 10 mondiale. Sebbene 21 su 40 Stati hanno superato la soglia del 75% di gender gap, resta importante la distanza tra i best in class e i peggiori, con un divario di 29 punti percentuali tra la prima (Islanda, appunto) e l'ultimo (Turchia), a testimonianza di come la questione sia trattata diversamente a seconda del contesto preso in considerazione.

Dopo l'Europa, si posiziona il Nord America a 74.8%, con un incremento di 4.3% rispetto al 2006. In particolare, il NA risulta la regione più virtuosa per il KPI economico, lo stesso ha registrato un lieve declino a 76.3%, soprattutto a causa delle poche donne in ruoli di leadership. Tale mancanza si registra anche in ambito politico, dove la regione si ferma al 26%, quasi 9 punti in meno rispetto all'Europa.

Chiude il podio delle regioni più virtuose America Latina e Caraibi, con un gender parity score del 74.2%, guadagnando ben 8.5 punti percentuali rispetto al 2006 e migliorando del 0.5% l'average del 2023. Tale miglioramento è frutto di imponenti campagne per la gender equity, soprattutto in contesto lavorativo ed istituzionale. Non sorprende, quindi, che tale regione si attesti al secondo posto per il Political Empowerment subindex, con un 34% secondo solo al 35.7% europeo, ma con un significativo +22.4% rispetto al 2006.

In ritardo di più di 14 punti percentuali, chiudono la classifica il Medio Oriente ed il Nord Africa, con un index del 61.7%, nonostante un trend positivo dal 2006 (+3.9). Considerando l'Economic Participation and Opportunity subindex, i paesi MENA si fermano ad una media del 43.1%, con una partecipazione ancora bassa, seppur in aumento. Bene anche l'aspetto educativo (97.2%) e sanitario (96.4%), che confermano i dati del 2023. Ampi margini di miglioramento si registra, invece, in ambito politico, fermo all'11.7% (dato peggiore cross countries), nonostante i quasi 9 punti percentuali recuperati dal 2006.

### 2.3.1 Global Gender Gap: la situazione in Italia

L'Italia, dal canto suo, si colloca all'87esimo posto, con un punteggio medio di 70.3% (-0.2% vs 2023) e ampi margini di miglioramento.

La Figura 4 restituisce una panoramica approfondita del nostro Paese, individuando i punteggi ottenuti per ogni subindex e la classificazione, per ognuno di essi, su scala mondiale.



Figura 4: World Economic Forum, Global Gender Gap 2024, Italy

È bene sottolineare, sin da subito, che l'Italia rientra nel primo 50% degli Stati presi in considerazione solo per 2 KPIs (Educazione e Partecipazione Politica) su 4, evidenziando una generale stagnazione sul tema.

Per il subindex Educational Attainment, l'Italia si colloca al 56° posto, con un incremento di 0.1% rispetto al 99.5% del 2023. E' bene evidenziare come, nel nostro sistema scolastico e universitario, sono sempre di più le donne che completano un percorso di studio specialistico, molto spesso arricchito da esperienze all'estero.

E' di 0.2 punti percentuali, invece, l'increase per la partecipazione politica, per il quale l'Italia si colloca al 67esimo posto.

Passando alle note dolenti, l'Italia si colloca al 111° posto (60.8% vs 61.8% del 2023) per la partecipazione economica, con gap del 17.4% tra lavoratrici e lavoratori (40.7% vs 58.1%, 96esimi a livello globale) e poco più del 40% di donne in posizioni apicali (102esimi worldwide).

Per quanto riguarda la dimensione medico – sanitaria, l'Italia si colloca al 94° posto, confermando la valorizzazione del 2023.

Global Gender Gap Index Indicators						2024	
Indicator	Rank	Score*	Compare with Global average	Difference F-M	◆ Female vs ◆ Male		Min Max
<b>Economic Participation and Opportunity</b>	111th	0.608		-	Min	Max	-
Labour-force participation rate %	96th	0.701		-17.39	40.69	58.08	0-100
Wage equality for similar work 1-7 (best)	95th	0.601		-	-	-	-
Estimated earned income int'l \$ 1,000	108th	0.539		-26.73	31.26	57.99	0-150
Legislators, senior officials and managers %	102nd	0.387		-44.24	27.88	72.12	0-100
Professional and technical workers %	88th	0.866		-7.20	46.40	53.60	0-100
<b>Educational Attainment</b>	56th	0.996		-	-	-	-
Literacy rate %	64th	0.997		-	-	-	-
Enrolment in primary education %	82nd	0.998		-0.17	98.32	98.49	0-100
Enrolment in secondary education %	98th	0.987		-1.34	101.64	102.98	0-200
Enrolment in tertiary education %	1st	1.000		23.54	60.07	83.61	0-200
<b>Health and Survival</b>	94th	0.967		-	-	-	-
Sex ratio at birth** %	1st	0.944		-	-	-	-
Healthy life expectancy** years	105th	1.020		-	-	-	-
<b>Political Empowerment</b>	67th	0.243		-	-	-	-
Women in parliament %	51st	0.477		-35.40	32.30	67.70	0-100
Women in ministerial positions %	65th	0.333		-50.00	25.00	75.00	0-100
Years with female/male head of state (last 50)	62nd	0.028		-47.28	1.36	48.64	0-50

Figura 5: Global Gender Gap Index Indicators, 2024, Italia

## 2.4 Gender Pay Gap

Come anticipato, il maggior divario di genere si registra in ambito professionale (Özgür Önday, 2020) dove ancora permane un significativo divario in termini di accesso al mondo del lavoro, opportunità di carriera e riconoscimenti tra lavoratori e lavoratrici. (Olsen et al., 2018)

Evidente è la connessione tra il gender gap e tale divario, in quanto il primo crea i presupposti sociali, culturali e strutturali per cui il secondo persiste e, se non arginato, si alimenta.

Il divario professionale, ovviamente, si sostanzia in un differenza salariale tra uomini e donne, che prende il nome di gender pay gap.

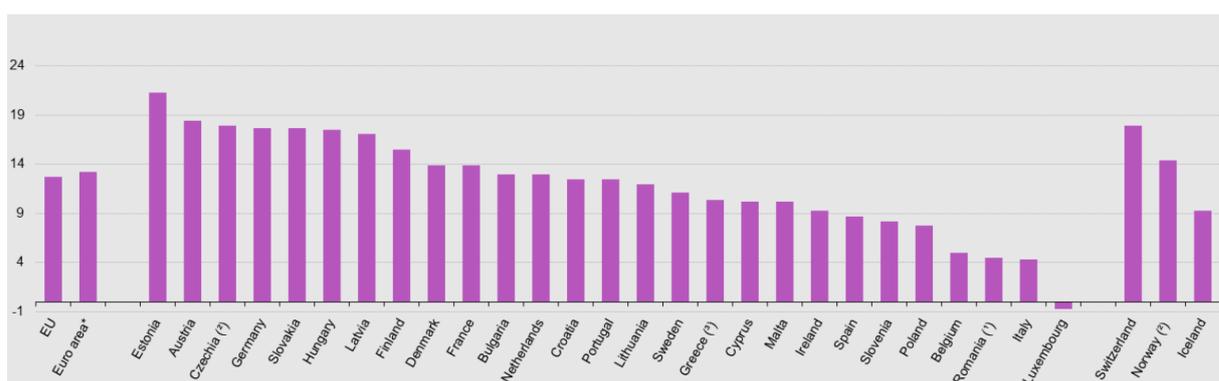
Ulteriormente, il gender pay gap può distinguersi in : “unadjusted gender pay gap”, nei casi in cui il calcolo si basa esclusivamente sul salario medio, senza prendere in esame le differenze dei lavoratori, e “adjusted gender pay gap”, qualora si tengano in considerazione, nel calcolo del divario, le differenze nell’istruzione, nella tipologia di impiego e nelle esperienze già maturate. Tale KPI consente di confrontare lavoratori e lavoratrici, ponderando percorsi di istruzione simili e con analoghe mansioni. (Eurostat)

In generale, un gender pay gap pari a 0% indica equità salariale. In caso di positività, vorrebbe significare una retribuzione maggiore per le donne; viceversa, un valore negativo riflette un divario salariale a favore degli uomini.

$$\text{Gender Pay Gap}(\%) = \frac{(\text{Salario medio uomini} - \text{Salario medio donne})}{(\text{Salario medio uomini})} \times 100$$

*Equazione 1: Gender Pay Gap Unadjusted*

Nel 2022, Eurostat riportava una retribuzione oraria lorda femminile inferiore, in media, del 12,7% rispetto a quella percepita dagli uomini all’interno dell’Unione Europea e del 13,2% nell’area Euro (Figura 6). Tra gli Stati membri dell’UE, il gender pay gap varia significativamente, passando da un valore negativo del -0,7% in Lussemburgo, dove le donne guadagnano mediamente più degli uomini, al 21,3% in Estonia, il Paese con la disparità salariale più elevata.



*Figura 6: Eurostat, Gender Pay gap based on hourly gross salary*

L'Italia rivela un duplice volto: da una parte, risulta tra i paesi più virtuosi in termini di gender gap pay con 4.1%; dall'altra, però, in termini di RAL media, il gap sale al 43%, contro il 36.5% medio europeo. (Eurostat)

Da questo punto di vista, i dati dell'Osservatorio sui lavoratori dipendenti del settore privato dell'INPS nel 2022 registravano una retribuzione media annua maschile pari a +€ 7.9k rispetto a quella femminile. Si ricorda che la Strategia nazionale per la parità di genere intende ridurre il salary gap dal 17% al 10% entro il 2026.

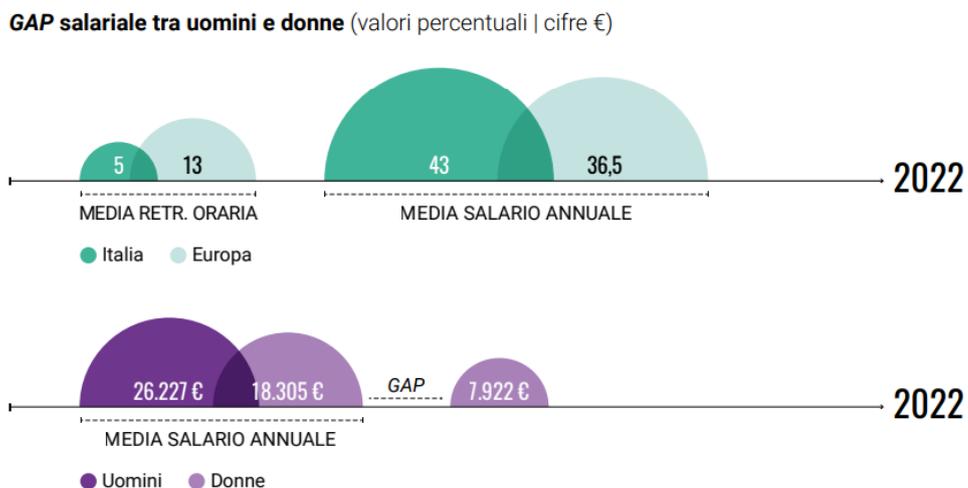


Figura 7: Eurostat, Salary gap man vs woman

## 2.5 Le Piattaforme Digitali

Negli ultimi decenni, l'economia globale ha assistito a una trasformazione radicale guidata dall'ascesa delle piattaforme digitali. Grandi aziende multinazionali come Apple, Google, Meta e Netflix hanno ridefinito non solo i modelli di business tradizionali, ma anche il modo in cui si misurano e si interpretano le dinamiche economiche e finanziarie (Hein et al, 2020). Le piattaforme digitali rappresentano ecosistemi complessi che integrano consumatori, produttori, reti e tecnologie, generando un impatto economico significativo su scala globale (Gawer, 2022). Il progressivo affermarsi delle cosiddette piattaforme digitali pone una serie di questioni significative dal punto di vista economico e sociale. L'economia delle piattaforme digitali rappresenta un vasto campo di ricerca, che affronta molteplici questioni sotto profili differenti (si veda ad esempio Belleflame &

Peitz, 2021). L'operare delle piattaforme digitale coinvolge, tra gli altri, aspetti che riguardano effetti di network, economie di scala, e scelta dei prodotti.

Il commercio online – di cui le piattaforme digitali rappresentano una particolare espressione – ha determinato una maggiore capacità da parte dei produttori di beni e di servizi di acquisire informazioni sul lato della domanda, dal momento che i dati sui consumatori possono essere raccolti, conservati, ed elaborati con un grado di dettaglio prima non immaginabile. Questo significa che con il commercio online l'offerta è sempre più in grado di distinguere all'interno del lato della domanda. In altre parole, il commercio online mette a disposizione delle piattaforme e, indirettamente, delle imprese che impiegano i servizi delle piattaforme, un'informazione più dettagliata riguardante i consumatori. Per le piattaforme che fanno un lavoro di matching tra domanda e offerta, come Amazon, e-Bay, ma anche Booking e Airbnb, la disponibilità di informazioni dettagliate riguarda entrambi i lati del mercato, cioè sia i consumatori che i “venditori”, cioè nel caso del turismo, gli host.

## **2.6 La sharing economy**

Strettamente connessa alla nascita e allo sviluppo di tali piattaforme, la *sharing economy* sfrutta il concetto stesso di condivisione per incrementare l'accessibilità e la flessibilità, la collaborazione e la sostenibilità, la fiducia e le relazioni umane, sfidando e mutando i canoni tradizionali di consumo e proprietà individuale.

Ed ecco che categorie tradizionalmente più emarginate, come le donne e i giovani, hanno individuato proprio nella *sharing economy* un bacino alternativo al mercato tradizionale del lavoro per potersi inserire.

## **2.7 Il Gender Gap nelle piattaforme digitali**

I mercati digitali e le piattaforme peer-to-peer offrono un percorso promettente per colmare il divario di genere nei guadagni. Queste piattaforme attenuano molte sfide tradizionali sul posto di lavoro, come pregiudizi e diseguaglianze di opportunità.

Molte piattaforme presentano infatti, un'opportunità per ridurre le differenze di guadagno tra i generi poiché, garantiscono maggiore flessibilità e consentono ai lavoratori di

stabilire i propri prezzi, introducendo così un contesto unico con cui esaminare i fenomeni (Lin et al., 2019; Ozbal et al., 2020).

In questo contesto le donne hanno la possibilità di guadagnare una retribuzione equa, consentendo loro di compensare potenziali differenze di genere nei guadagni.

Pertanto, il reddito generato dall'economia della condivisione può consentire alle donne di integrare i salari part-time, investire nelle loro famiglie e sostenere le spese domestiche. Le piattaforme digitali hanno però reso possibile una conoscenza dettagliata sui due lati del mercato che prima non esisteva.

Nel caso specifico di Airbnb, il servizio offre ai potenziali consumatori una serie di informazioni sull'host, cioè sul proprietario dell'immobile. Tra le diverse caratteristiche rese note, si sottolinea il genere dell'host. A parità di caratteristiche dell'immobile, sarebbe lecito aspettarsi che i consumatori non guardino al genere dell'host al momento di effettuare una scelta di acquisto.

Alcuni recenti studi empirici, tuttavia, trovano risultati diversi. Marchenko (2019), considerando un campione di 40 città in giro per il globo, trova che a parità di caratteristiche le donne guadagnano meno degli uomini e le donne di colore ancora meno. Questo risultato sembra fornire evidenza in favore della teoria della discriminazione basata sulle preferenze lanciata originariamente da Becker (1961).

Ert et al. (2024) guardano alla città di Tel Aviv, con in mente un filone di studi secondo cui le donne sono tipicamente percepite come più accoglienti ed affidabili (ad esempio Ert & Fleischer, 2020), e trovano che le donne guadagnano effettivamente più degli uomini. Anche in questo caso, il framework teorico di riferimento è quello della discriminazione basata sulle preferenze, dove, al contrario, le scelte vengono declinate come favorevoli alle donne.

Se a livello internazionale esistono quindi alcuni studi che cercano di rilevare un eventuale fenomeno discriminatorio, misurato con il gender gap e con altre metriche, per l'Italia non sono disponibili lavori recenti basati su dati microeconomici dettagliati. Sulla base di queste considerazioni, si prosegue nell'analisi della domanda di ricerca che vede il genere come predittore significativo degli incassi degli host.

Per rispondere alla domanda di ricerca, si propone una regressione panel, la cui descrizione dettagliata verrà offerta nel Capitolo 4.

# Capitolo 3

## 3.1 Introduzione ad Airbnb

Airbnb, unicorno della sharing economy, ha senza dubbio rivoluzionato il segmento dell'hospitality e in particolare, l'offerta degli affitti a breve termine.

Mediante un'offerta ampiamente personalizzabile, per rispondere alle esigenze di host e ospiti, Airbnb rappresenta una "all in one box", in cui domanda e offerta si incontrano, in una soluzione scalabile ed altamente remunerativa.

Grazie a questa innovazione, i proprietari di spazi inutilizzati come stanze o appartamenti, possono interagire direttamente con viaggiatori in cerca di alloggi temporanei. (Hall et al., 2022).

Airbnb non solo ha reso possibile un approccio più flessibile e scalabile alla gestione degli alloggi, ma ha anche ridefinito l'esperienza del soggiorno, offrendo soluzioni personalizzate e autentiche che rispondono alle esigenze sia degli host che degli ospiti.

Il modello peer-to-peer di Airbnb offre un sistema di recensioni per aggiungere valore ai potenziali hosts e, come già analizzato in precedenza, nei mercati online i partecipanti si fidano delle recensioni altrui e proprio quest'ultime divengono essenziali per il buon funzionamento del servizio offerto.

### 3.1.1 Storia di Airbnb

Airbnb nasce nel 2007 a San Francisco, dall'intuizione di Brian Chesky e Joe Gebbia, due giovani designer con difficoltà economiche.

In occasione della riunione annuale della Industrial Design Society of America, Brian e Joe, sfruttando il tutto esaurito nelle strutture alberghiere della città, misero a disposizione posti letto su materassi gonfiabili all'interno del proprio appartamento. (Airbnb)

In questo modo non solo riuscirono ad affrontare serenamente le spese dell'affitto, ma l'esperimento fu un tale successo da dare vita a quella che oggi è la piattaforma di home sharing più conosciuta ed utilizzata al mondo. (Zervas et al., 2017)

Non ancora ventennale, Airbnb opera, offrendo varie soluzioni d'alloggio, in oltre 220 paesi, con oltre 8 milioni di annunci attivi e più di 8 miliardi di dollari di fatturato.

### **3.2 Funzionamento di Airbnb**

Il funzionamento della piattaforma, in linea con un approccio user friendly, è semplice ed intuitivo. Dopo essersi registrato, l'host pubblica immagini e descrizioni dello spazio che intende mettere a disposizione, completando con tariffe e indicazione dei periodi in cui è possibile fruire dello spazio.

I guest ricercano autonomamente l'alloggio di proprio interesse e successivamente procedono alla prenotazione. Airbnb media anche il pagamento, trattenendo quanto dovuto fino a 24 ore dopo il check-in, momento in cui trasferisce la somma, al netto delle commissioni, all'host.

Al termine del soggiorno, la piattaforma dà la possibilità agli ospiti di recensire l'alloggio e l'host, al fine di alimentare un rapporto di trasparenza e fiducia all'interno dell'ecosistema.

### **3.3 Il gender gap in Airbnb**

Come anticipato nel capitolo precedente, la sharing economy e le relative piattaforme possono rappresentare una soluzione importante per ridurre il gender gap, creando un ambiente non solo inclusivo, ma anche democratico, in cui chiunque può agire come un imprenditore. (Airbnb, 2019; Farmaki, 2019).

Questo studio esplora il divario di genere nei guadagni su Airbnb, scelto per le sue caratteristiche uniche che lo distinguono dalle altre piattaforme.

In primo luogo, Airbnb consente ai consumatori di identificare facilmente il genere di un host. In secondo luogo, opera in un sistema di prezzi guidato dal mercato (a differenza dei prezzi guidati da algoritmi o offerte). In terzo luogo, l'unico indicatore del capitale umano è l'esperienza dell'host, al contrario di ampi background professionali. Inoltre, è importante sottolineare che Airbnb si basa sull'ospitalità e sulla fiducia percepita e ricerche precedenti (Ert e Fleischer, 2020) suggeriscono che le donne potrebbero avere un vantaggio in questi aspetti.

Infine, l'accesso alla piattaforma non ha particolari sbarramenti. Chiunque abbia a disposizione uno spazio nella propria disponibilità può diventare host; di fatto, un imprenditore. In questo modo, si evita qualsiasi tipo di discriminazione sistemica e soggettiva.

Essere lavoratore autonomo significa anche poter bilanciare, in maniera del tutto personale, vita professionale e lavorativa; l'host, infatti, ha piena libertà nel decidere se e quando rendere disponibile agli altri i propri spazi. Inoltre, l'assenza di un datore di lavoro permette una riduzione di possibili discriminazioni, un problema che spesso persiste nei mercati di lavoro tradizionali.

La piattaforma, inoltre, è strutturata per focalizzare l'attenzione sulla user experience, piuttosto che sull'host, e su un sistema di pubblicità meritocratico, grazie alle recensioni rese dagli ospiti.

Tale ecosistema ha consentito negli anni di rendere Airbnb uno spazio sicuro per le donne che intendano avviare la propria carriera in ambito hospitality. Come confermato direttamente dalla piattaforma, infatti, al 2023 il 60% degli host è di genere femminile, con una media di revenue nettamente superiore a quelli registrati nei comparti professionali "tradizionali". Per incrementare la propria immagine di women support, la piattaforma ha avviato programmi di mentoring e avviamento all'imprenditorialità dedicata alle donne, cercando di limare ulteriormente la differenza di genere.

Nonostante questo, permangono ancora dei potenziali limiti alla libera espressione femminile su Airbnb.

Le indagini sociologiche evidenziano come, nella scelta di un'experience, l'utente medio consideri non solo l'offerta (descrizione, immagini, tariffe, recensioni) (Cheng e Jin, 2019; Liang et al., 2019), ma anche l'offerente, soffermandosi, in questo caso, non solo sulle recensioni, ma anche su dati anagrafici (banalmente, nome e cognome) e foto profilo (Linag et al. 2017;2020). Purtroppo, il mezzo digitale, se, da un lato, rappresenta un contesto informale, che alimenta una dimensione di familiarità, dall'altro, potrebbe rappresentare terreno fertile per il proliferare di un approccio basato su processi euristici, pregiudizi e bias, che influenzerebbero negativamente gli utenti in fase di prenotazione. (Edelman et al., 2017; Ert e Fleischer, 2019).

Di conseguenza, la struttura della piattaforma, pur favorendo trasparenza e fiducia, può involontariamente contribuire a perpetuare disuguaglianze di genere.

# Capitolo 4

## 4.1 Analisi del database AirDNA

Lo studio si basa su un dataset fornito da AirDNA ed è relativo alla città di Torino, nel periodo compreso tra il 2017 ed il 2023.

Il dataset è di tipo Panel in quanto i dati sono organizzati in serie temporali e include un'ampia gamma di variabili che permettono di analizzare performance economiche, caratteristiche degli alloggi e comportamento degli host.

Le variabili a disposizione sono le seguenti:

### 1. *Caratteristiche dell'annuncio:*

- Property ID: stringa identificativa che caratterizza in modo univoco una singola proprietà. Ogni proprietà ha un proprio codice alfanumerico che serve a distinguerla dalle altre nel database.
- Year: indica l'anno di riferimento, con valori compresi tra il 2017 e il 2023.

### 2. *Variabili di performance:*

- Number of Reservations: rappresenta il numero totale di prenotazioni ricevute per un annuncio in un determinato arco temporale.
- Revenue (USD): ricavi totali (in dollari USD) generati da un determinato annuncio nell'anno di riferimento.
- Reservation Days: numero di giorni in cui l'appartamento è stato effettivamente prenotato rispetto alla sua disponibilità nell'anno esaminato.
- Available Days: numero di giorni in cui l'appartamento non è stato prenotato rispetto alla sua disponibilità nell'anno di riferimento.

### 3. *Caratteristiche della proprietà:*

- Listing Type: tipologia di sistemazione, che può includere intero appartamento, hotel, stanza condivisa, stanza privata e stanza singola.

- Overall Rating: rappresenta la valutazione media dell'alloggio espressa dagli ospiti su una scala da 0 a 100. Questo punteggio è una media delle valutazioni che gli ospiti danno a diversi aspetti dell'alloggio, come la pulizia, la comunicazione con l'host, la posizione, il valore per il prezzo, e l'esperienza complessiva.
- Latitudine: coordinata geografica che indica la latitudine dell'alloggio.
- Longitudine: coordinata geografica che indica la longitudine dell'alloggio.
- Bedrooms: numero di camere da letto presenti all'interno dell'alloggio.
- Max Guests: numero massimo di ospiti che l'alloggio può ospitare.
- Comune: comune in cui si trova l'alloggio.
- Property Manager: indica se l'annuncio è gestito da un'agenzia.
- Price Tier: fascia di prezzo a cui appartiene la proprietà, suddivisa in budget, economy, midscale, upscale e luxury.
- Has Pool: indica se l'alloggio dispone di una piscina ("1 = vero" o "0 = falso").
- Has Hot Tub: indica se l'alloggio dispone di vasca idromassaggio ("1 = vero" o "0 = falso").
- Has Air Condition: indica se l'alloggio dispone di aria condizionata ("1 = vero" o "0 = falso").
- Has Gym: indica se l'alloggio dispone di una palestra ("1 = vero" o "0 = falso").
- Has Kitchen: indica se l'alloggio dispone di una cucina ("1 = vero" o "0 = falso").
- Has Parking: indica se l'alloggio dispone di un parcheggio ("1 = vero" o "0 = falso").

#### 4. *Caratteristiche dell'host:*

- Host ID: codice univoco identificativo associato a ciascun host.
- Airbnb Superhost: indica se l'host è un Superhost su Airbnb ("1 = vero" o "0 = falso").
- Annunci host: numero di annunci pubblicati dall'host.
- Host type: classificazione host in base al numero di alloggi gestiti (small, mid, large).

## 5. Scelte strategiche dell'host:

- Instantbook: indica se l'alloggio può essere prenotato in maniera istantanea senza alcuna comunicazione tra ospite e host.
- Minimum Stay: pernottamento minimo richiesto dall'host.

## 4.2 Misure di performance

Per verificare l'eventuale esistenza di un divario di genere, si procede all'analisi delle performance ottenute cioè l'analisi dei ricavi degli host sulla piattaforma, dunque si individuano i KPIs corrispondenti:

1. **ADR:** metrica tipica nell'industria dell'ospitalità, misura il prezzo medio per notte prenotata. L'ADR si calcola con la seguente equazione:  
[\$/giorno]

$$ADR = Revenue/ReservationDays$$

*Equazione 2: Tariffa media Giornaliera*

Dove:

- *Revenue: ricavo totale dell'host.*
  - *Reservation Days: numero totale di notti occupate*
2. **OCC:** percentuale di notti in cui l'alloggio è prenotato rispetto al totale delle notti in cui è stato disponibile. Più è elevata la percentuale, maggiore è la capacità dell'host di massimizzare i propri utili.

L'OCC si calcola con la seguente formula:

[%]

$$OCC = Reservation\ Days / (ReservationDays + AvailableDays)$$

*Equazione 3: Tasso di Occupazione*

Dove:

- ReservationDays: numero di notti effettivamente prenotate.
- AvailableDays: notti disponibili per l'affitto durante il timeframe considerato

**3. RevPAN:** Rapporto tra i guadagni e il totale delle notti in cui l'alloggio è disponibile (che sia prenotato o meno). Tale metrica consente di valutare l'efficacia nell'utilizzo delle risorse disponibili; in altre parole, permette di verificare il potenziale di guadagno dello spazio.

Una misura del parametro elevato, mostra che l'host sta riuscendo a generare più ricavi per ogni notte disponibile. Viceversa, un valore basso potrebbe indicare un utilizzo inefficiente della piattaforma e delle sue relative prenotazioni.

RevPAN si calcola con la seguente formula:

[\$/giorno]

$$RevPAN = ADR * OCC = Revenue / (ReservationDays + AvailableDays)$$

*Equazione 4: RevPAN*

Dove:

- ADR: prezzo medio per notte
- OCC: rapporto tra notti prenotate e notti disponibili.

### **4.3 Estrapolazione dei nomi degli Host e analisi del genere**

Per effettuare un'analisi del divario di genere tra gli host di Airbnb nella città di Torino, è stato necessario ottenere i nominativi a partire dagli identificativi unici degli host disponibili nel database. A tale scopo, è stato implementato uno script di Python. Questo utilizzando richieste HTTP e la libreria BeautifulSoup accede alle pagine profilo degli host su Airbnb per estrarre il nome visualizzato. Lo script "extraazione\_nomi.py" legge un file CSV contenente gli ID degli host, effettua richieste automatiche al sito e raccoglie i

nomi in un nuovo file CSV. Vengono implementati anche dei meccanismi di restrizione di gestione degli errori e accesso per garantire la robustezza del processo, rispettare le linee guida del server e introdurre ritardi tra una richiesta e un'altra.

Successivamente sono stati analizzati i nomi estratti e il genere è stato determinato usando la libreria "Gender\_Guasser". Questa associa un genere probabilistico basato su database preesistenti. Lo script "gender\_from\_name.py" legge il file CSV in cui sono presenti i nomi degli host, utilizza un algoritmo di riconoscimento di genere e genera un nuovo file che include una colonna aggiuntiva "Genere" con il risultato della classificazione ("Maschio", "Femmina", "Incerto" o "Sconosciuto").

Durante questa analisi, tuttavia, è emerso che per molti host non era possibile identificare il genere a causa della natura privata dei loro profili che, non fornivano informazioni sufficienti.

Inoltre, sono stati esclusi gli alloggi che consistevano in:

- un host con un nome ambiguo in termini di genere;
- un host con un nome commerciale;
- un host che possiede più di un nome (es. Anna e Alberto).

L'esclusione delle suddette ambiguità ha raffinato il database considerando un numero totale di 3.897 host.

Una volta effettuato questo processo, è stato possibile ottenere la distribuzione iniziale di genere degli host, rendendo possibile la successiva analisi del divario di genere.

```

import pandas as pd
import requests
from bs4 import BeautifulSoup
import time

# Funzione per estrarre il nome dall'HTML della pagina
def get_host_name(host_id):
    url = f"https://www.airbnb.com/users/show/{host_id}"
    headers = {
        "User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/114.0.0.0 Safari/537.36"
    }

    try:
        response = requests.get(url, headers=headers, timeout=10)
        if response.status_code == 200:
            soup = BeautifulSoup(response.text, 'html.parser')
            # Selettore aggiornato per cercare il nome nella pagina
            name_tag = soup.find("span", class_="t1gpcl1t atm_w4_16rzvi6 atm_9s_1o8liyq atm_gi_idpfg4 dir dir-ltr")
            if name_tag:
                return name_tag.text.strip()
            else:
                return "Name not found"
        elif response.status_code == 429: # Too many requests
            print("Raggiunto il limite di richieste, attendendo...")
            time.sleep(60) # Aspetta un minuto e riprova
            return get_host_name(host_id)
        else:
            return f"Error {response.status_code}"
    except requests.exceptions.RequestException as e:
        return f"Error: {e}"

```

Figura 8: Codice Python Estrazione Nomi parte 1

```

# File CSV di input
input_file = "ID1.CSV"
output_file = "host_names.csv"
# Legge il file CSV con il separatore corretto
try:
    data = pd.read_csv(input_file, sep=";", on_bad_lines='skip') # Specifica il separatore ";"
    print("Colonne trovate nel file CSV:", data.columns.tolist()) # Stampa i nomi delle colonne
    ids = data['ID'].tolist() # Usa il nome corretto della colonna
except KeyError:
    print("Errore: La colonna 'Airbnb HOST ID' non è stata trovata nel file.")
    exit()
except FileNotFoundError:
    print(f"File {input_file} non trovato!")
    exit()
except Exception as e:
    print(f"Errore durante la lettura del file: {e}")
    exit()
# Estrazione dei nomi
results = []
for host_id in ids:
    name = get_host_name(host_id)
    results.append({"id": host_id, "name": name})
    print(f"ID: {host_id} ok")
    time.sleep(3) # Rispetta il server con un ritardo tra le richieste

# Salva i risultati in un file CSV
output_df = pd.DataFrame(results)
output_df.to_csv(output_file, index=False)

print(f"Processo completato. Risultati salvati nel file {output_file}")

```

Figura 9: Codice Python Estrazione Nomi parte 2

```

import csv # Importa la libreria csv per leggere e scrivere file CSV.
import gender_guesser.detector as gender # Importa il modulo gender_guesser per determinare il genere basato sul nome.
import sys # Importa il modulo sys per configurare l'output del terminale.

# Configura stdout per supportare caratteri UTF-8
# Questo garantisce che i caratteri speciali (es. accenti) vengano gestiti correttamente in stampa.
sys.stdout.reconfigure(encoding='utf-8')

# Inizializza il rilevatore di genere.
# Il modulo `gender.Detector` utilizza un database per fare supposizioni sul genere in base al nome.
detector = gender.Detector()

# Funzione per determinare il genere basato sul nome
def determine_gender(name):
    # Usa il detector per identificare il genere basato sul nome fornito.
    gender = detector.get_gender(name)
    # Confronta il risultato con possibili valori restituiti dalla libreria.
    if gender in ["male"]: # Se il genere è "male", restituisci "Maschio".
        return "Maschio"
    elif gender in ["female"]: # Se il genere è "female", restituisci "Femmina".
        return "Femmina"
    elif gender in ["mostly_male", "mostly_female"]: # Per generi "incerti", restituisci "Incerto".
        return "Incerto"
    else:
        # Per altri casi o nomi non riconosciuti, restituisci "Sconosciuto".
        return "Sconosciuto"

# File di input e output
input_file = "host_name.csv" # Nome del file CSV di input che contiene i nomi.
output_file = "host_name_con_gender.csv" # Nome del file CSV di output che conterrà i nomi con i generi.

```

Figura 10: Codice Python per ricavare il genere dal nome parte 1

```

# Apertura dei file CSV di input e output
# Usa la modalità "r" per leggere il file di input e "w" per scrivere il file di output.
with open(input_file, mode="r", encoding="utf-8") as infile, open(output_file, mode="w", encoding="utf-8", newline="") as outfile:
    # Crea un oggetto DictReader per leggere il file di input come una lista di dizionari.
    reader = csv.DictReader(infile)
    # Aggiungi una nuova colonna "gender" ai fieldnames (nomi delle colonne) esistenti.
    fieldnames = reader.fieldnames + ["gender"]
    # Crea un oggetto DictWriter per scrivere nel file di output, includendo la nuova colonna "gender".
    writer = csv.DictWriter(outfile, fieldnames=fieldnames)

    # Scrive l'intestazione (header) nel file di output.
    writer.writeheader()

    # Itera su ogni riga del file di input.
    for row in reader:
        # Recupera il valore della colonna "name" e rimuove eventuali spazi extra.
        name = row.get("name", "").strip()
        # Determina il genere utilizzando la funzione `determine_gender`.
        row["gender"] = determine_gender(name)
        # Scrive la riga aggiornata (includendo il genere) nel file di output.
        writer.writerow(row)
        # Stampa un messaggio di elaborazione per tenere traccia del progresso.
        print(f"Elaborato: {name} -> {row['gender']}")

```

Figura 11: Codice Python per ricavare il genere dal nome parte 2

## **4.4 Strumenti per l'analisi**

### **4.4.1 Excel**

Per l'organizzazione, l'elaborazione e la gestione del dataset, la scelta è ricaduta sul software Excel. Nello specifico, il software è stato utilizzato per pulire ed ottimizzare i dati, mediante l'impiego di filtri.

Si è, quindi, provveduto alla creazione di tabelle pivot, utili a svolgere le analisi descrittive e di sintesi (somme, media e conteggi) del database.

Istogrammi, grafici a barre e torta hanno permesso di visualizzare gli output delle diverse fasi in maniera chiara.

### **4.4.2 Stata**

Il software Stata è un programma avanzato per l'analisi statistica e la gestione dei dati e trova applicazione per lo svolgimento delle analisi econometriche, finalizzate a verificare l'impatto delle diverse variabili sulle performances del gender.

Nel contesto di questa tesi, Stata è stato utilizzato in primo luogo per la pulizia e manipolazione dei dati, consentendo di gestire dataset complessi e prepararli per l'analisi.

Tra le analisi più rilevanti, sono stati sviluppati modelli di regressione lineare multipla, per esplorare la relazione tra la variabile indipendente, rappresentata dal genere dell'host, e le variabili dipendenti legate alle performance economiche.

## **4.5 Analisi descrittive preliminari**

L'analisi esplorativa ha permesso di verificare la distribuzione dei dati rispetto alle variabili chiave, come la distribuzione degli host, le caratteristiche degli alloggi e le variabili di performance. Per facilitare la comprensione dei risultati, le informazioni sono state rappresentate mediante visualizzazioni grafiche, quali istogrammi, grafici a barre e diagrammi a torta.

#### 4.5.1 Distribuzione degli host per genere

Per approfondire la distribuzione degli host attivi sulla piattaforma di Airbnb a Torino si esamina il database e si osserva la seguente ripartizione:

- le host donne ammontano a 2077;
- gli host uomini sono 1820.

Questi risultati suggeriscono la tendenza nell'industria dell'ospitalità di una sovra rappresentazione del genere femminile. Inoltre, la composizione demografica degli host potrebbe riflettere tendenze socioculturali, come l'incremento della partecipazione femminile nel settore dell'economia digitale.

Di seguito si riporta la distribuzione del genere degli host in un grafico a torta:

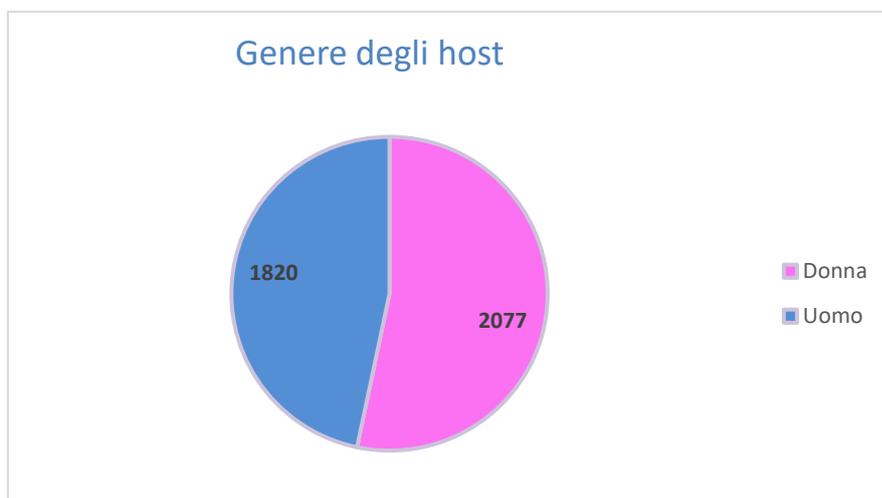
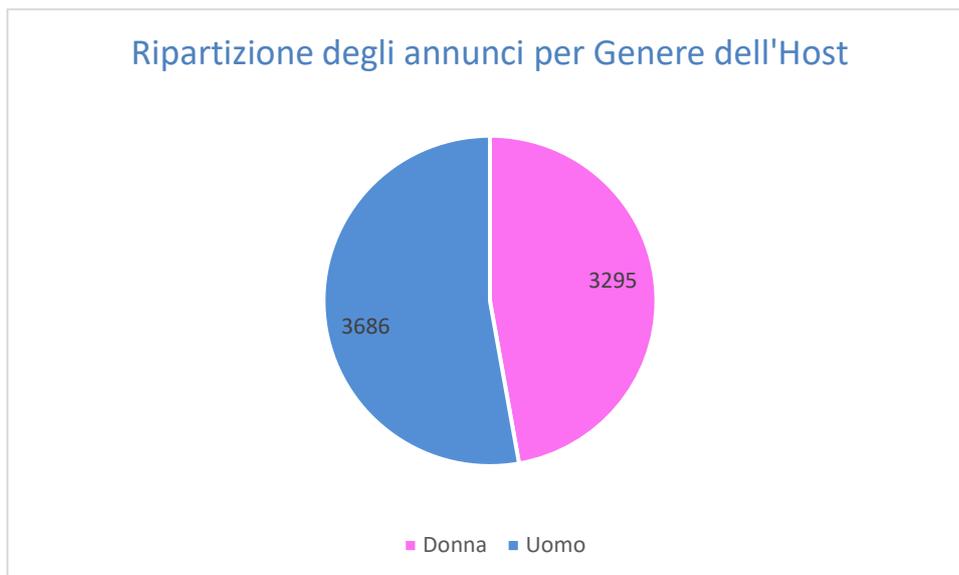


Grafico 1: Distribuzione degli host per genere

#### 4.5.2 Distribuzione degli annunci per genere dell'host

Valutando il numero di annunci pubblicati per ogni host, si rileva una sensibile differenza: per ogni host di genere maschile sono presenti in media 2,2 annunci mentre per quelli femminili 1,6. Si conta dunque un totale annunci pari a 3.686 nel primo caso e 3.295 nel secondo come rappresentato nel grafico seguente.



*Grafico 2: Distribuzione degli annunci per genere*

### 4.5.3 Distribuzione nel tempo delle prenotazioni in base al genere dell'host

Il numero di prenotazioni che un host raggiunge annualmente affittando alloggi, permette di analizzare la sua capacità di attrarre ospiti. L'analisi si incentra sull'osservare se esiste una differenza di genere nel numero di prenotazioni ottenute nell'arco di tempo analizzato.

A primo impatto si evidenzia un vantaggio del genere maschile nel numero di prenotazioni, infatti, si raggiunge un picco nel 2023 con una media di 56 prenotazioni per gli host uomini e 48 per quelli delle donne.

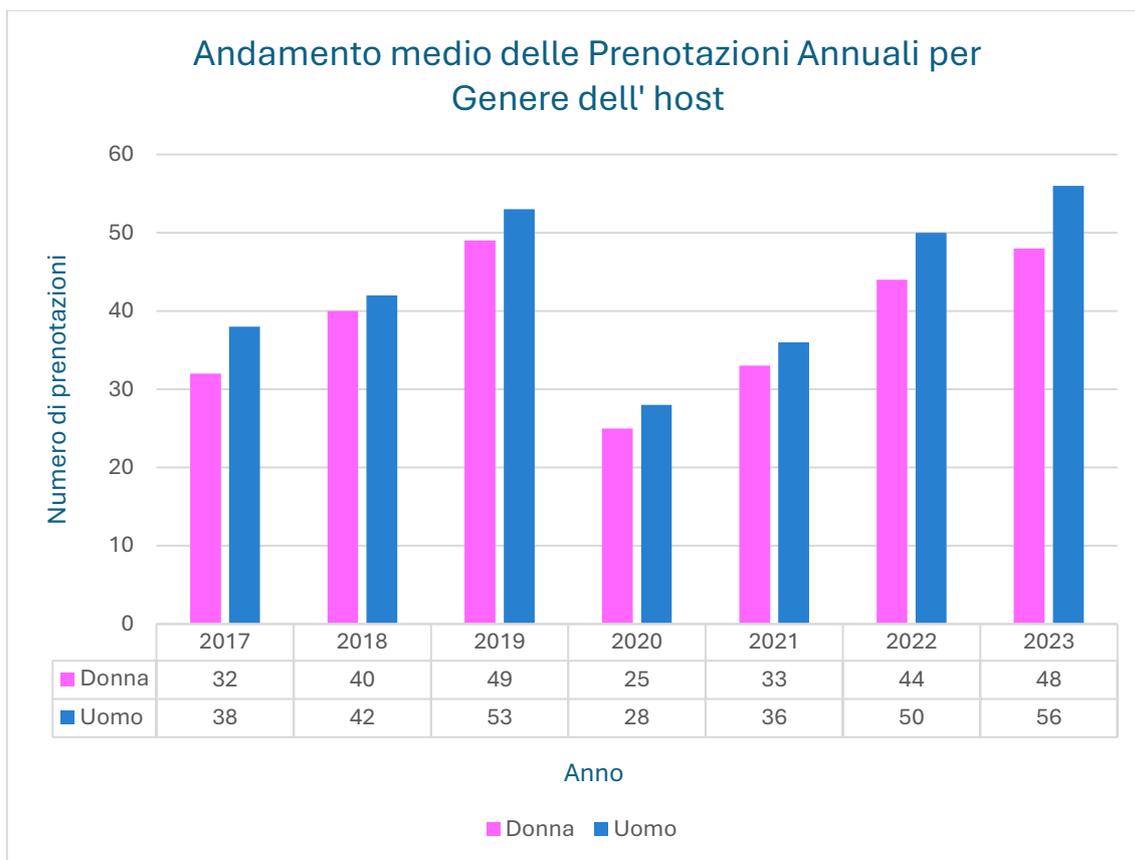
Questo andamento, pertanto, potrebbe rappresentare un possibile fenomeno di disparità di genere che va approfondito.

Per analizzare se effettivamente la differenza della media delle prenotazioni sia statisticamente significativa deve essere effettuato il test  $t$  per la differenza delle medie.

Effettuando il test si ottiene che:

- $t = 6.62$
- $q = 3.70$  quantile con livello di significatività dell'1%

Il valore  $t = 6.62$  è superiore rispetto al quantile, si rifiuta l'ipotesi nulla confermando una significatività nella differenza di genere nel numero di prenotazioni non dovuta al caso.



*Grafico 3: Prenotazione medie annuali per genere*

#### 4.5.4 Distribuzione delle entrate in base al genere dell'host

Fin da subito è evidente dal grafico sottostante che non vi è un netto vantaggio per nessuno dei due gruppi: nel lasso temporale tra il 2017 e il 2019 si osserva un sostanziale aumento delle entrate probabilmente dovuto a una crescita della domanda del mercato, registrando nel 2019 un reddito pari a 10.399 USD per le donne e 9609 USD per gli uomini.

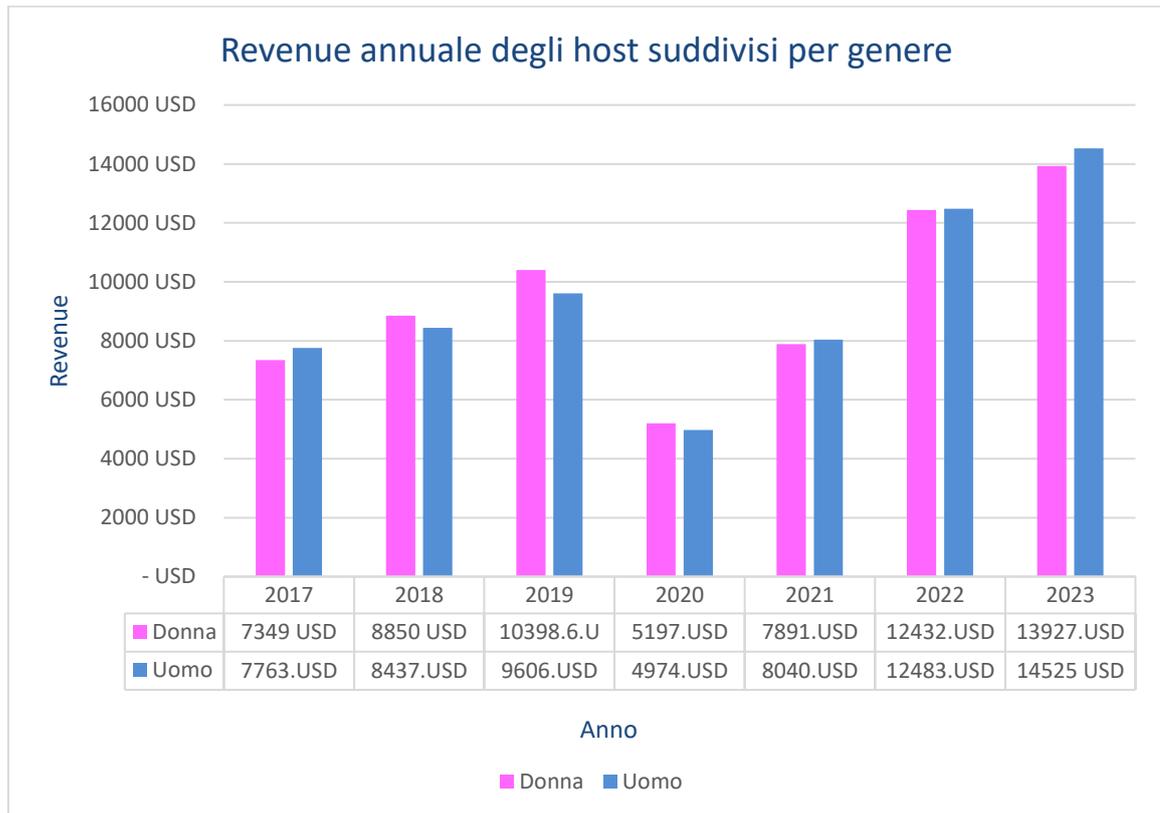
Dal 2020 in poi, le differenze dei ricavi ottenuti tra i gruppi sono sostanzialmente nulle fino al 2023, anno in cui si registra un lieve vantaggio per gli host maschi con entrate pari a 14525 USD rispetto ai 13927 USD delle host donne.

Si esamina tramite il test  $t$  la differenza del reddito tra i due gruppi per verificare se quest'ultima sia statisticamente significativa. L'ipotesi nulla ( $H_0$ ) assume una differenza del ricavo tra host uomini e donne pari a zero.

La statistica  $t$  calcolata per il confronto delle due medie è pari a:

- $t = -0.1697$
- il quantile critico al livello di significatività dell'1% è pari a  $q = -3.7074$ .

I risultati confermano l'assenza di una differenza significativa tra i due gruppi di host.



*Grafico 4: Andamento della Revenue annuale in base al genere dell' host*

#### 4.5.5 Punteggio delle recensioni in base al genere dell'host

Anche in termini di overall rating non si rilevano particolari differenze, pur segnalando il vantaggio delle host donne (95.4) rispetto agli uomini (94.8).

Tale “vantaggio” viene confermato dal test  $t$ , per cui  $t = -94.87$ , nettamente inferiore al valore soglia  $-3.7072$ .

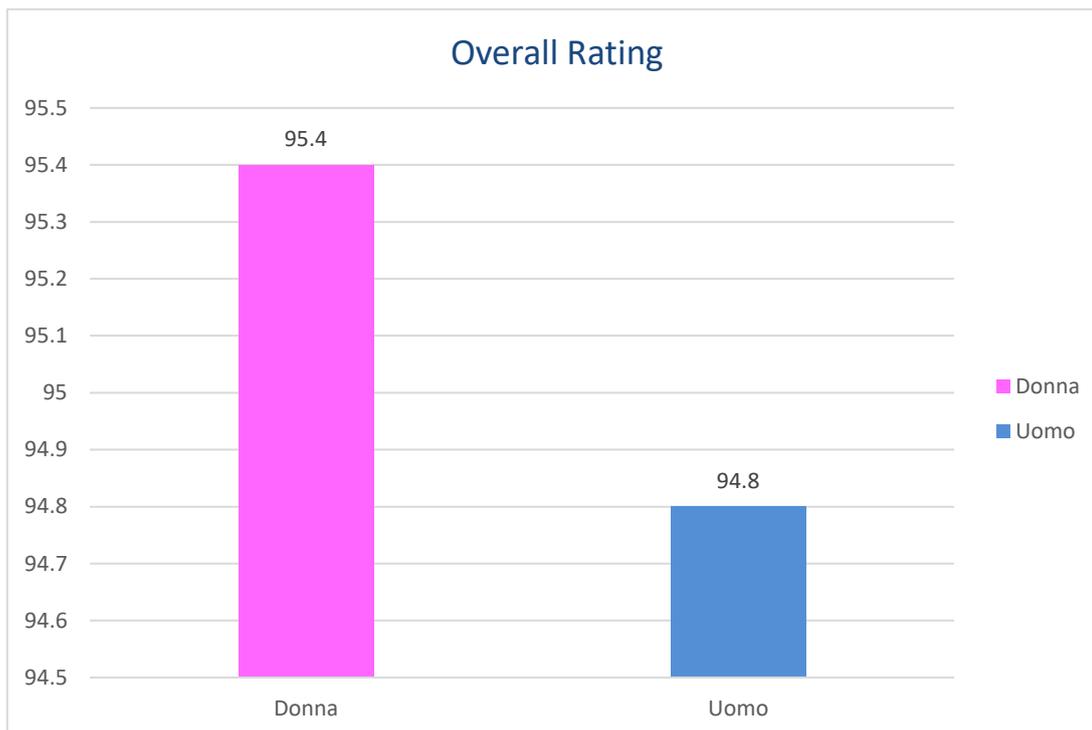


Grafico 5: Overall Rating per genere

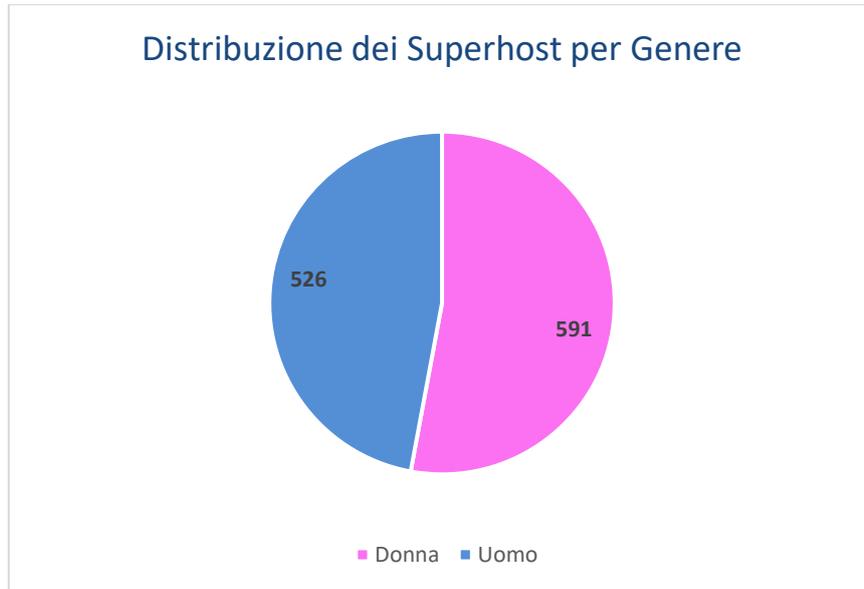
#### 4.5.6 Analisi dei Superhost in base al genere

Il badge di Superhost indica un host particolarmente orientato ad offrire un servizio di qualità. I requisiti per l'ottenimento del badge sono i seguenti:

- Aver mantenuto una valutazione complessiva di almeno 4.8 stelle su 5;
- Aver offerto almeno 10 soggiorni in un anno o in alternativa almeno 3 soggiorni di lunga durata per un totale di almeno 100 pernottamenti;
- Aver mantenuto un tasso di risposta minimo del 90% entro le 24h;
- Aver mantenuto un tasso di cancellazione inferiore all'1%, fatta eccezione per cancellazioni dovute a circostanze eccezionali o altri motivi validi.

In termini di distribuzione, l'analisi rivela che solo 1.117 host su 3.897 (28.7%) possiedono questo badge, con una leggera predominanza delle donne (591 - 53%) rispetto agli uomini (526 - 47%).

La differenza seppur minima, potrebbe riflettere una propensione per il genere femminile ad offrire un servizio qualitativamente migliore dedicando ad esempio, una maggior attenzione agli ospiti.



*Grafico 6: Distribuzione dei Superhost per genere*

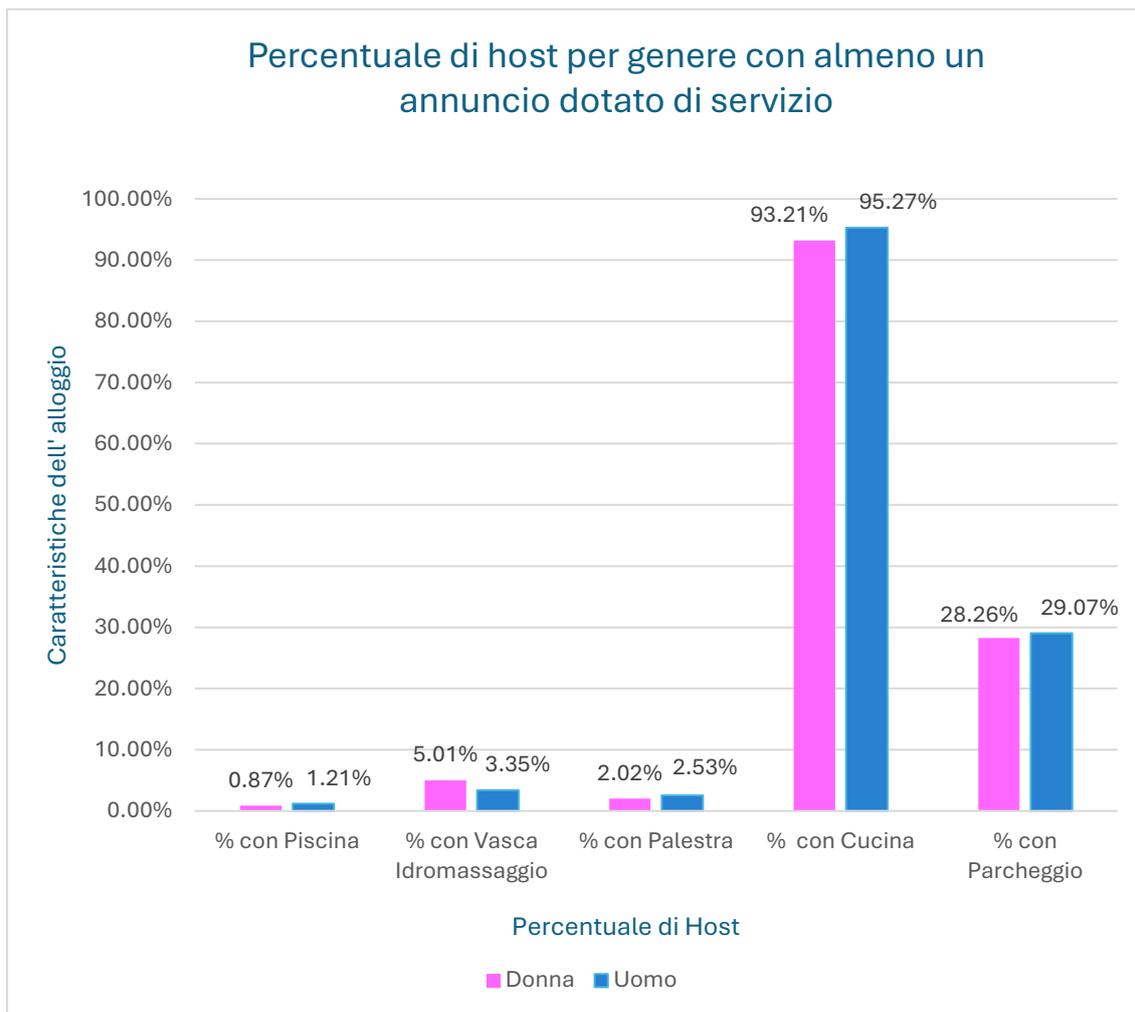
#### **4.5.7 Servizi offerti dagli host in base al genere**

L'obiettivo di quest'analisi è comprendere se esistono differenze significative tra uomini e donne nella tipologia di alloggi offerti.

Il grafico permette di visualizzare la percentuale di host su Airbnb che possiedono almeno un alloggio con specifici servizi tra cui: piscina, vasca idromassaggio, palestra, cucina e parcheggio.

Non si evincono particolari differenze tra i due generi. In entrambi i casi, infatti, il servizio maggiormente offerto è la cucina, seguito da parcheggio; di contro, è difficile trovare nell'offerta piscina, vasca idromassaggio e palestra.

In conclusione, l'offerta di servizi negli annunci è abbastanza uniforme tra i due generi, con lievi differenze che non risultano essere significative.



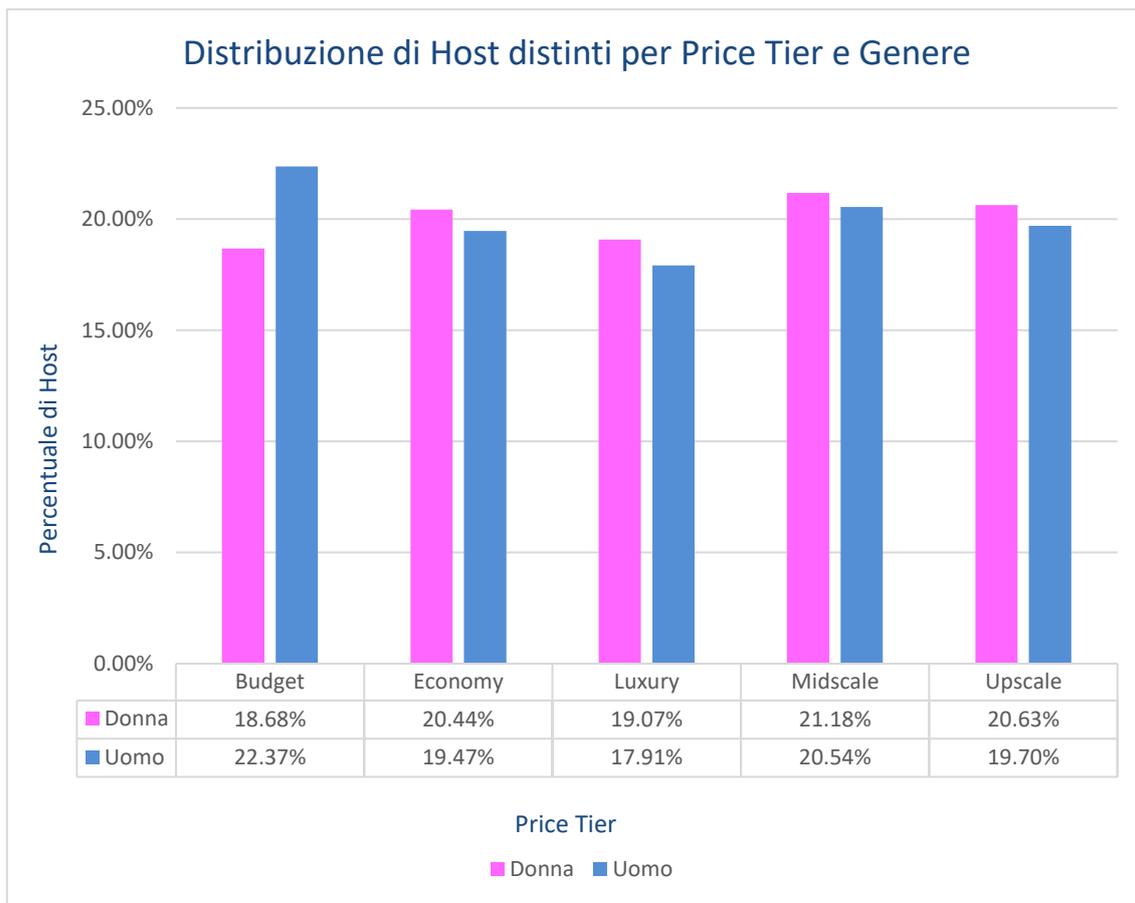
*Grafico 7: Distribuzione dei servizi offerti per genere dell'host*

#### **4.5.8 Distribuzione per fascia di prezzo in base al genere**

Quest'analisi ha l'obiettivo di esaminare in quale fascia di prezzo si collocano gli alloggi offerti dagli host e verificare se esistono differenze significative nelle strategie di prezzo adottate dai due gruppi.

A seconda della fascia analizzata, si evince una forbice più o meno ampia.

Nel segmento Luxury, si riscontra una leggera predominanza (1,2%) delle donne rispetto agli uomini. Di contro, nella fascia budget, si registra lo scenario contrapposto con una differenza del 3,7%. Nelle altre categorie, non si segnalano differenze sostanziali.



*Grafico 8: Divisione per fascia di prezzo in base al genere dell'host*

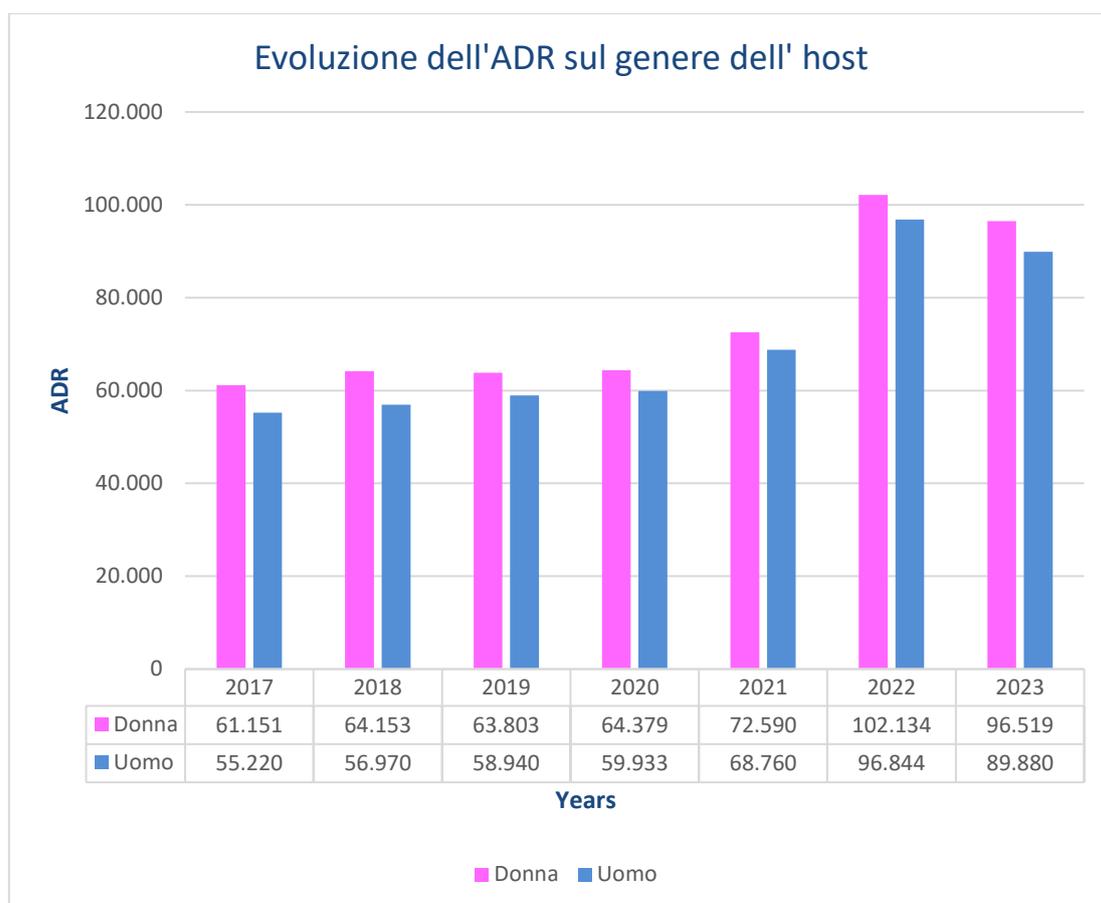
#### 4.5.9 ADR

Un altro fattore importante da tenere in considerazione nell'analisi delle performance degli host è l'Average Daily Rate, che rappresenta il prezzo medio per notte a cui è affittato l'alloggio.

Il tutto il periodo temporale considerato le donne si collocano in posizione di vantaggio rispetto agli uomini, con un picco nel biennio 2022-2023.

Per studiare se la differenza tra l'ADR medio degli host uomini e quello delle host donne sia statisticamente significativa è stato condotto un test  $t$  in cui si assume come ipotesi nulla ( $H_0$ ) che la differenza di prezzo per notte tra host uomini e host donne sia uguale a zero.

Considerando un valore  $t$  pari a  $-12.02$  e un quantile critico al livello di significatività dell'1% pari a  $-3.7074$ , si può affermare che tale differenza risulta, ai fini statistici, significativamente rilevante.



*Grafico 9: Andamento dell'ADR per genere*

#### 4.5.10 Listing Type

Le tipologie di proprietà sono quattro: appartamento/casa intera, stanza privata, stanza condivisa, stanza d’hotel. Si evidenzia che la casa/ appartamento intero siano la scelta preferita (il 79% del totale), seguita dalla stanza privata (18.87%), mentre solo una percentuale molto bassa di utenti prenota in una stanza condivisa (1.89%) o stanza d’hotel (0.24%).

Listing Type	Freq. Assoluta	Freq. Relativa	Freq. Cumulata
Entire home/apt	15,843	79%	79%
Hotel Room	49	0.24%	79.24%
Private Room	3,784	18.87%	98.11%
Shared Room	379	1.89%	100%

Tot.	20,055	100%	100%
------	--------	------	------

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

#### 4.5.11 Distribuzione del genere per tipologia di annuncio

Entrambi i gruppi tendono ad offrire interi appartamenti, mentre solo una bassa percentuale di host offre come soluzione di alloggio le camere condivise: 1.33% delle host donne contro un 2.50% degli host uomini. Ciò potrebbe indicare che gli uomini potrebbero essere leggermente più propensi a mettere a disposizione spazi condivisi.

Nel complesso, il grafico evidenzia una forte predilezione per le intere case o appartamenti tra tutti gli host, una distribuzione equilibrata per le stanze private, e differenze più marcate per le categorie meno comuni, come le stanze condivise e le camere d'hotel.

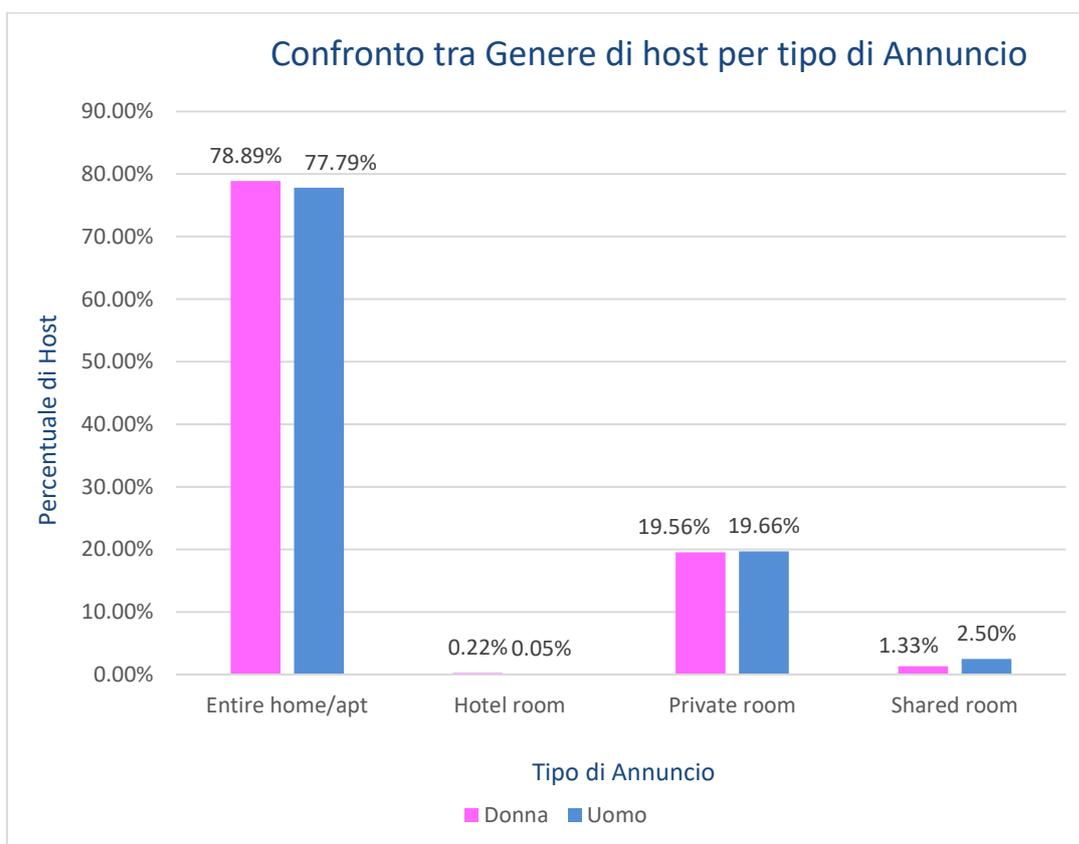


Grafico 10: Tipologia di annuncio per genere

#### 4.5.12 Distribuzione giorni prenotati e disponibili per genere dell'host

Gli host uomini tendono ad avere un numero leggermente superiore di giorni occupati. Questo potrebbe suggerire una gestione più efficiente delle disponibilità o una maggiore richiesta per gli annunci degli host uomini.

Il test  $t$  ( $t = 4.32$ ) conferma la rilevanza statistica di tale gap.

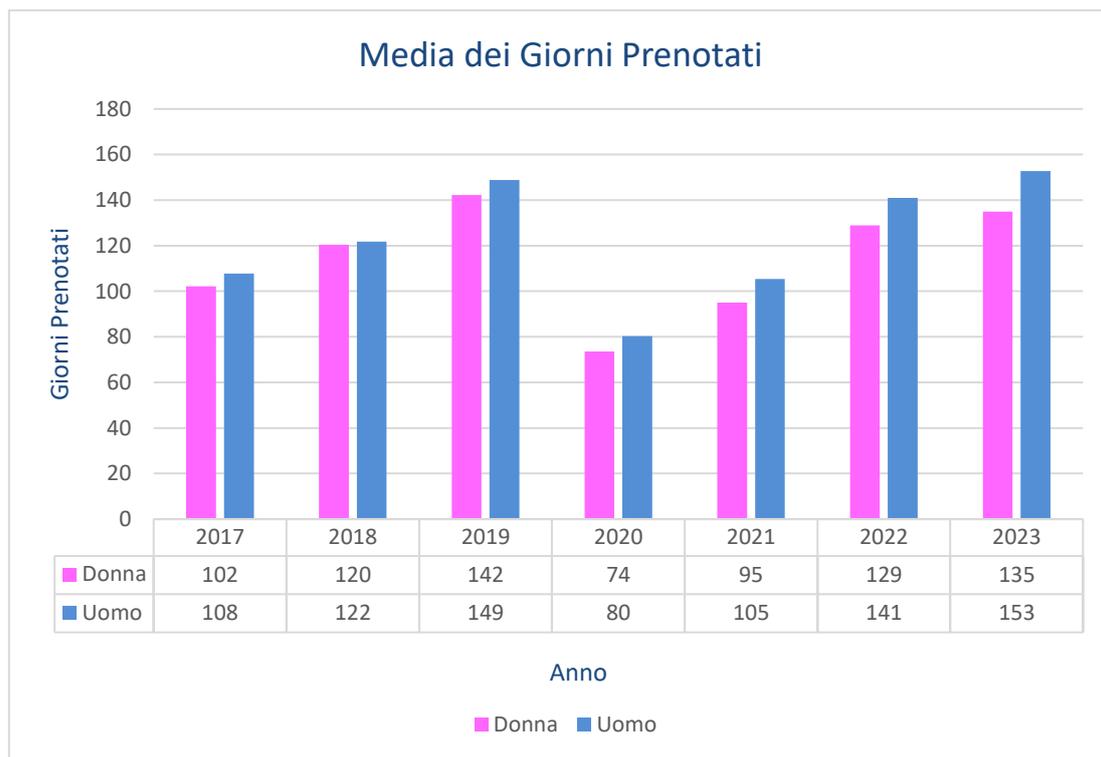
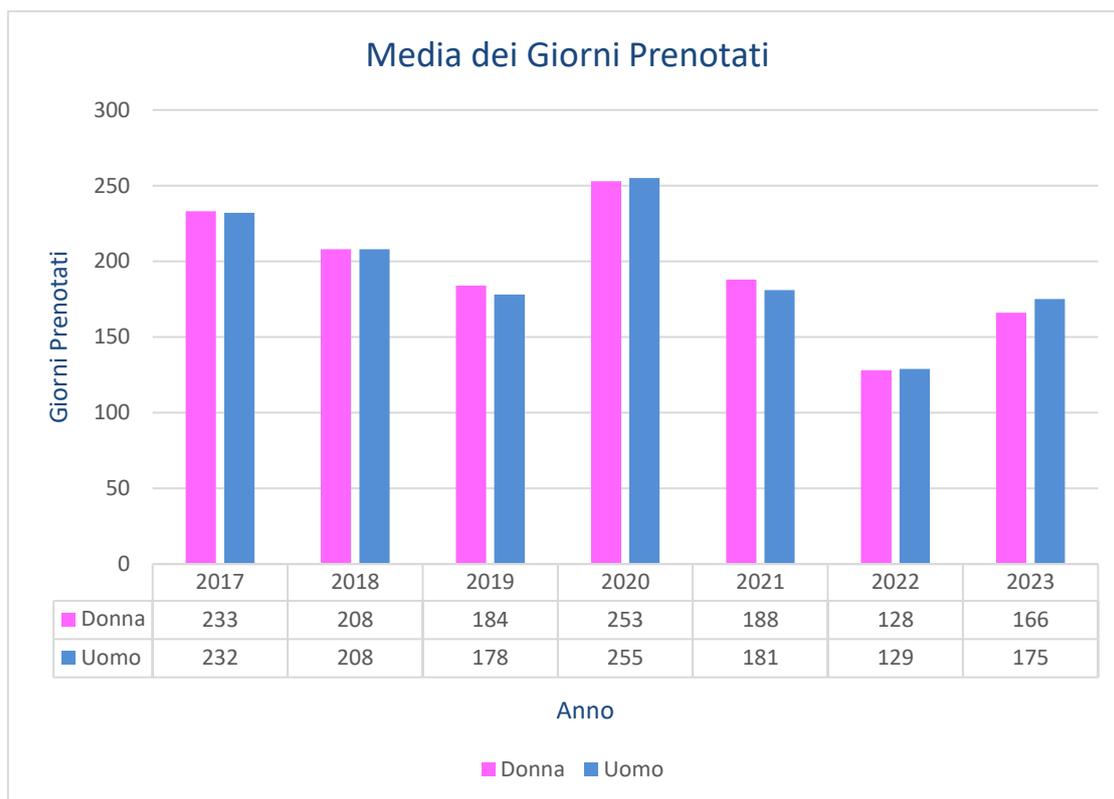


Grafico 11: Media dei giorni prenotati per genere

Di contro, nessuna tendenza significativa si ottiene, analizzando la media di giorni disponibili.

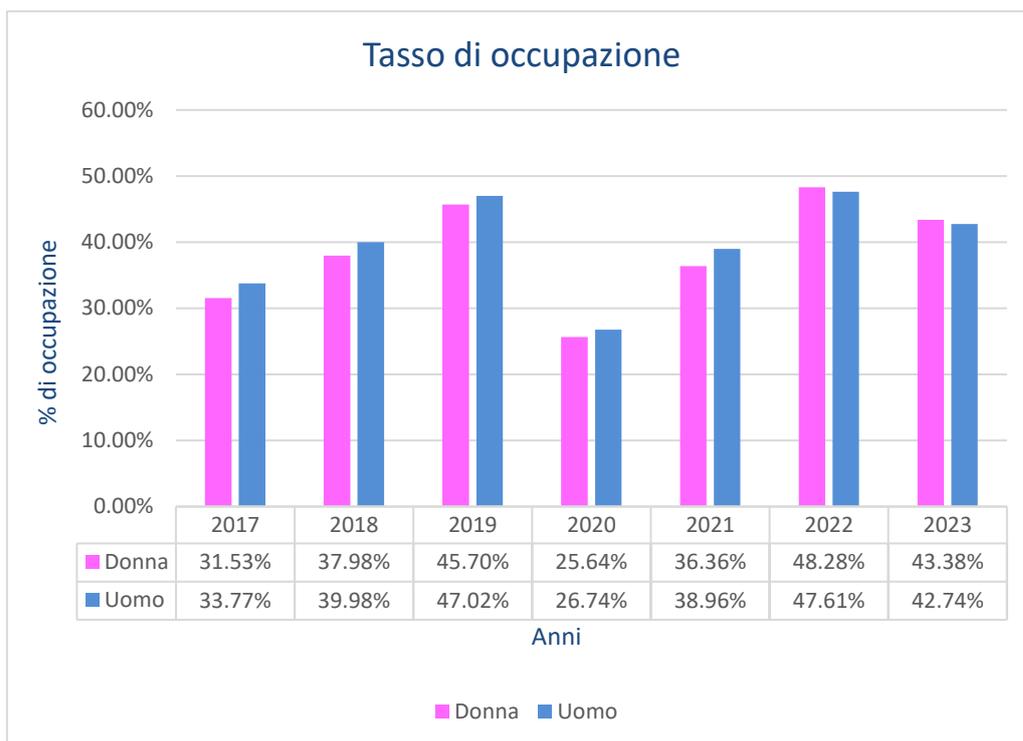


*Grafico 12: Media dei giorni disponibili per genere*

#### **4.5.13 Tasso di occupazione medio per genere dell'host**

Il tasso di occupazione è un'importante misura di performance, calcolato considerando sia i giorni in cui l'appartamento è stato prenotato sia quelli in cui è rimasto non occupato rispetto al totale della sua disponibilità.

Dall'analisi del grafico si osserva un tasso di occupazione inferiore al 50% che indica che c'è ancora una porzione significativa di giorni disponibili che non viene prenotata, suggerendo opportunità di ottimizzazione nella gestione delle tariffe.



*Grafico 13: Tasso di Occupazione per Genere dell' host*

#### **4.5.14 Distribuzione camere per annuncio**

Il grafico esamina se vi è un evidente differenza nel numero di camere da letto offerte negli annunci dagli host.

Si osserva che la maggior parte degli host indipendentemente dal genere, offre alloggi con una sola camera da letto in particolare, il 74.18% delle host donne e il 72.30% degli host uomini. Per quanto riguarda gli annunci con due camere da letto, si nota un'inversione della tendenza, con una quota leggermente superiore tra gli host uomini (23.78%) rispetto alle donne (21.73%).

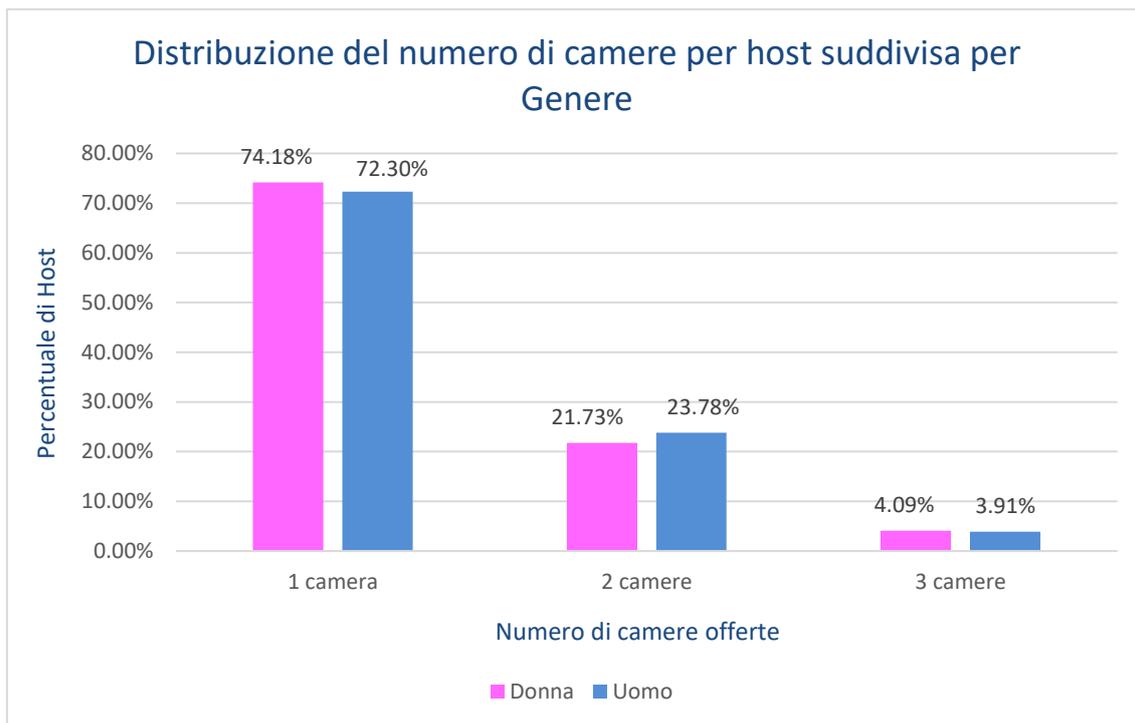


Grafico 14: Distribuzione del numero di camere per host suddivisa per genere

#### 4.5.15 Distribuzione del RevPAN

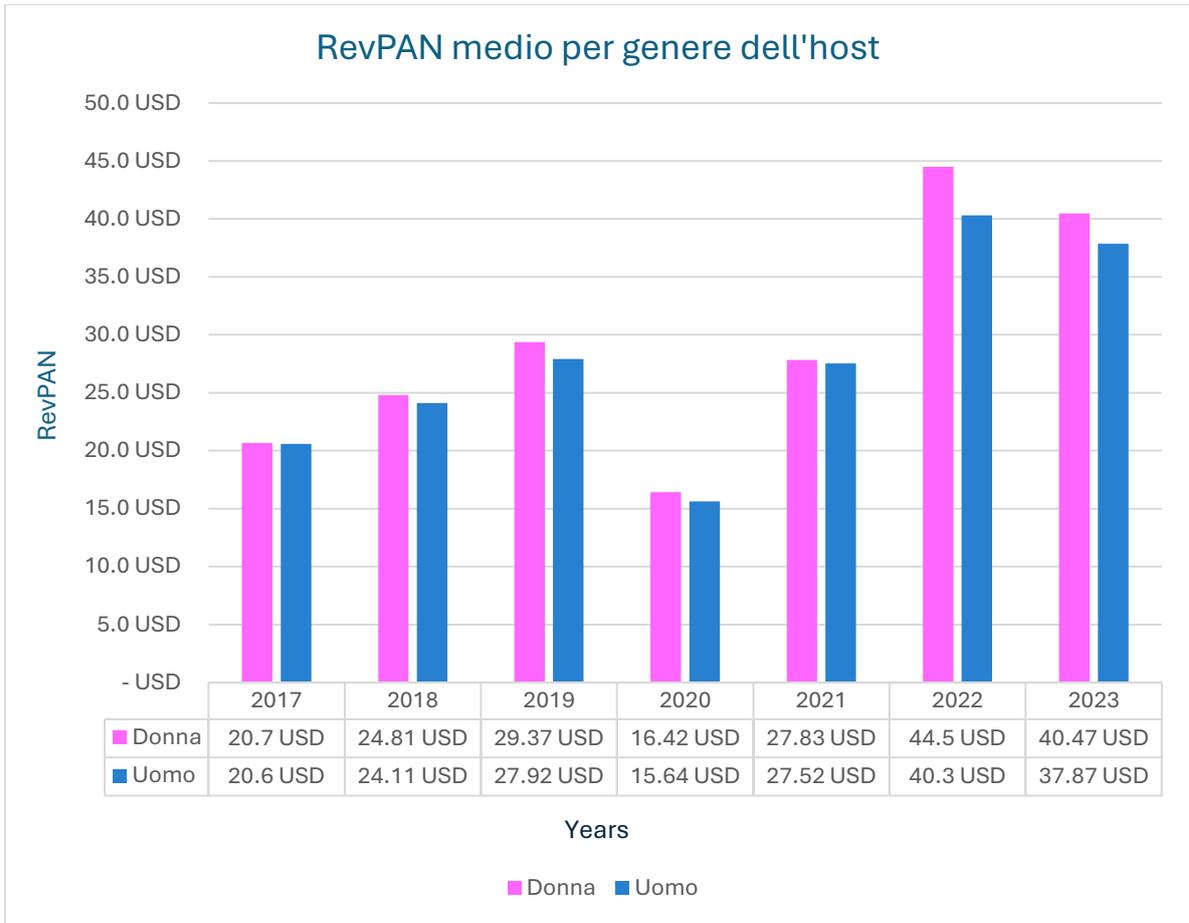
Il RevPAN è un indicatore chiave della performance economica, poiché misura il ricavo medio ottenuto per ogni notte in cui una proprietà è stata disponibile su Airbnb, includendo sia le notti effettivamente prenotate che quelle rimaste libere.

Per tutto il periodo preso in esame emerge una leggera differenza del fatturato a favore delle host donne.

Per valutare se questa minima differenza è significativa viene effettuato il test t che rileva i seguenti risultati:

- $t = -2.58$
- quantile con livello di significatività dell'1% è pari a  $-3.707$

Poiché  $t = -2.58$  non è inferiore a  $-3.70$  non possiamo rifiutare l'ipotesi nulla al livello di significatività dell'1% dunque, non ci sono prove sufficienti per affermare che il fatturato medio per notte delle host donne sia significativamente maggiore rispetto a quello degli host uomini.



*Grafico 15: Distribuzione del ricavo medio per notte suddiviso per genere*

## 4.6 Analisi Statistiche inferenziali

### 4.6.1 Analisi di regressione del Panel

Per analizzare la relazione tra il genere dell'host e le performance su Airbnb, è stata implementata una regressione su dati Panel a livello di annuncio.

Questo approccio consente di tenere conto della struttura longitudinale dei dati, monitorando l'evoluzione delle performance degli annunci nel tempo e controllando per eventuali eterogeneità non osservabili tra gli annunci stessi.

In particolare, possono essere considerati diversi tipi di modello che combinano in modi diversi informazioni sia tra le unità di individui, qui rappresentate dagli Host, sia nel tempo, in questo caso gli anni, tenendo conto della eterogeneità non osservabile, vale a dire di fattori specifici delle unità o del tempo che non sono inclusi esplicitamente come variabili nel modello.

Questi modelli sono denominati:

- *Modello a Effetti fissi (FE)*: tecnica di regressione utilizzata nell'analisi Panel che ipotizza che ci sia una componente non osservabile che è specifica per ogni unità di individuo, vale a dire l'Host, o per il tempo, cioè gli anni.

I modelli a effetti fissi si dividono in:

- *Modello a effetti fissi per unità* nel caso considerato, volendo analizzare il dataset di host Airbnb per diversi anni si suppone che i singoli host si differenzino uno dall'altro per specifiche caratteristiche quali, ad es. le competenze, la reputazione, lo spirito imprenditoriale etc. e che tali caratteristiche non cambiano nel tempo ma possono influenzare variabili come il numero di prenotazioni o i ricavi. Con il modello FE, queste caratteristiche non osservabili specifiche dell'host vengono rimosse dal modello. Un Esempio in formula per questo modello è il seguente:

$$Y_{it} = \alpha_i + X_{it} \beta + \epsilon_{it}$$

dove:

- $Y_{it}$ : variabile dipendente per l'unità  $i$  nel tempo  $t$ .

- $\alpha_i$ : effetto specifico per l'unità  $i$  (host).
- $X_{it}$ : matrice delle variabili indipendenti.
- $\epsilon_{it}$ : errore residuo.

Le caratteristiche principali dei modelli FE consistono nel rimuovere gli effetti non osservabili che si ipotizza siano costanti.

Infine, in questi modelli, le variabili che non cambiano nel tempo (es. genere, livello di istruzione) non possono essere incluse.

- *Effetti casuali (RE)*: questo modello assume che le caratteristiche non osservabili specifiche del Panel, per unità o tempo, siano casuali e siano non correlate con le variabili indipendenti. A esempio, nel caso considerato, si può analizzare come le variabili indipendenti (Genere, Years, ecc.) influenzano le variabili di performance controllando per le caratteristiche fisse. L'obiettivo del loro utilizzo è quello di includere gli effetti specifici suddetti come parte dell'errore e stimare la loro varianza. Un Esempio in formula per questo modello è il seguente:

$$Y_{it} = \gamma + X_{it} \beta + u_i + \epsilon_{it}$$

dove:

- $Y_{it}$ : variabile dipendente per l'unità  $i$  nel tempo  $t$ .
- $u_i$ : effetto specifico per l'unità  $i$ , trattato come casuale e indipendente da  $X_{it}$
- $X_{it}$ : matrice delle variabili indipendenti.
- $\epsilon_{it}$ : errore residuo.

Le caratteristiche principali dei modelli RE sono quelle di stimare gli effetti specifici del tempo o degli individui e di includere  $u_i$  come parte dell'errore composto. Questi modelli vanno utilizzati quando si ritiene che suddetti effetti specifici non siano correlati con le variabili esplicative. Gli stimatori RE sono più efficienti (varianza più bassa) rispetto a quelli FE, ma solo se l'ipotesi di assenza di correlazione è valida. Inoltre, tali modelli possono includere variabili invariante nel tempo come genere o istruzione.

Nei modelli empirici che si vogliono qui studiare le variabili legate al genere sono considerate oggetto di studio e entrano nel modello stesso come variabili indipendenti: questo preclude l'utilizzo dei modelli a effetti fissi e impone di far ricorso ai modelli ad effetti random. Inoltre, per i modelli testati in STATA, è stato scelto di utilizzare l'opzione `vce(robust)`, opzione che viene aggiunta ai modelli di regressione per calcolare errori standard robusti, che correggono per problemi di eteroschedasticità (varianza non costante degli errori correlata alle variabili esplicative) e correlazione tra errori.

#### 4.6.2 Regressioni Multivariate

La regressione multivariata è una tecnica statistica utilizzata per analizzare la relazione tra una variabile dipendente e più variabili indipendenti. A differenza della regressione semplice, che esamina l'effetto di una sola variabile esplicativa sulla variabile dipendente, la regressione multivariata permette di isolare e quantificare l'effetto di ciascun fattore, controllando per le altre variabili incluse nel modello. Dunque questo approccio consente di quantificare l'impatto di ciascuna variabile esplicativa sulla variabile dipendente, controllando simultaneamente per altri fattori, e permette di ottenere stime più accurate rispetto a un'analisi univariata.

Formalmente, un modello di regressione lineare multipla può essere espresso come:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

dove:

- $Y$  è la variabile dipendente,
- $X_1, X_2 \dots X_n$  sono le variabili indipendenti,
- $\beta_0$  è l'intercetta,
- $\beta_1, \beta_2 \dots \beta_n$  sono i coefficienti che misurano l'effetto di ciascuna variabile indipendente su  $Y$ ,
- $\varepsilon$  è il termine di errore, che rappresenta la parte della variabilità di  $Y$  non spiegata dal modello.

La significatività dei coefficienti che misurano l'effetto di ciascuna variabile indipendente su  $Y$  viene determinata valutando il valore di  $p$  in una soglia di confidenza. Nello

specifico, un valore di  $p < 0.10$  (soglia convenzionale) indica che l'effetto della variabile indipendente sulla variabile dipendente è statisticamente significativo.

La capacità di questo modello di fornire risultati validi viene determinata misurando la percentuale di variabilità della variabile dipendente indicata dal modello (statistica R-quadrato).

La statistica R-quadrato indica che, all'aumentare della variabile dipendente, aumenta la capacità predittiva del modello anche se un valore molto altro di tale variabile potrebbe indicare over-fitting cioè un modello che si adatta troppo ai dati quindi perdendo capacità di generalizzazione.

### 4.6.3 Output delle regressioni

Nei modelli di regressione analizzati, come già anticipato, la Variabile Indipendente è la dummy Genere (Donna/Uomo), mentre come Variabile Dipendente vengono utilizzate variabili esprimenti il livello di performance dell'Host sia nella forma lineare che logaritmica. In particolare, sono state scelte:

- Il **numero di prenotazioni** sia in forma lineare (NumberofReservations) che logaritmica (lnNumberofReservations);
- I **ricavi per notte** sia in forma lineare (RevPAN) che logaritmica (lnRevPAN);
- Il **punteggio medio complessivo** attribuito dalle recensioni in forma lineare (Overallrating);
- Il **prezzo per notte** sia in forma lineare (ADR) che logaritmica (lnADR);
- il **Tasso di occupazione** (OCC) in forma lineare e logaritmica (lnOCC).

Come detto precedentemente i modelli utilizzano una regressione a effetti casuali (RE), appropriata quando si ipotizza che le caratteristiche non osservate specifiche degli host non siano correlate con le variabili esplicative incluse nel modello.

Le variabili di controllo scelte sono:

- Host\_Type: variabile categoriale ordinale che rappresenta la classificazione dell'host in base al numero di unità gestite (small, mid, large);
- Price\_Tier: variabile categoriale che indica la fascia di prezzo a cui appartiene la proprietà, suddivisa in budget, economy, midscale, upscale e luxury;

- HasPool: variabile dummy che indica se l'alloggio tra quelli inseriti dall'Host su Airbnb dispone di una piscina ("1 = vero" o "0 = falso");
- HasHotTub: dummy che indica gli alloggi tra quelli inseriti dall'Host su Airbnb che dispongono di vasca idromassaggio ("1 = vero" o "0 = falso");
- HasAirConditio: variabile dummy che indica gli alloggi tra quelli pubblicati dall'Host su Airbnb che dispongono di aria condizionata ("1 = vero" o "0 = falso");
- HasGym: variabile dummy che indica gli alloggi tra quelli inseriti dall'Host su Airbnb che dispongono di una palestra ("1 = vero" o "0= falso");
- HasKitchen: variabile dummy che indica se gli alloggi tra quelli pubblicati dall'Host su Airbnb dispongono di una cucina ("1 = vero" o "0 = falso");
- HasParking: variabile numerica che indica gli alloggi tra quelli inseriti dall'Host su Airbnb che dispongono di un parcheggio ("1 = vero" o "0= falso");
- Bedrooms: variabile numerica discreta che rappresenta il numero di posti letto presenti nell'alloggio;
- Listing\_Type: variabile categoriale che indica la tipologia di sistemazione, che può includere appartamento, stanza condivisa, stanza privata o stanza singola (Entire home/apt, Hotel Room, Private Room, Shared Room);
- AirbnbSuperhost: variabile categoriale che indica se l'host è un Superhost su Airbnb ("1 = vero" o "0= falso");
- Year: variabile categoriale introdotta per controllare il modello per gli effetti di tempo.

Per cominciare analizziamo i modelli lineari i cui risultati sono riportati in Tabella 2.

Il primo modello econometrico stima il numero di prenotazioni ricevute da un annuncio (variabile dipendente) in relazione al genere dell'host (variabile indipendente) considerando altre caratteristiche dell'annuncio (variabili di controllo).

L'obiettivo di questa specificazione è isolare l'impatto del genere dell'host sulle prenotazioni.

**Modello 1:**

$$\begin{aligned} \text{Number of Reservations}_{it} = & \beta_0 + \beta_1 \cdot \text{DGender}_{it} + \beta_2 \cdot \text{DListingType}_{it} + \beta_3 \cdot \\ & \text{DHostType}_{it} + \beta_4 \cdot \text{Bedrooms}_{it} + \beta_5 \cdot \text{DYear}_t + \beta_6 \cdot \text{DPriceTier}_{it} + \beta_7 \cdot \\ & \text{DSuperhost}_{it} + \beta_8 \cdot \text{DHasPool}_{it} + \beta_9 \cdot \text{DHasHotTub}_{it} + \beta_{10} \cdot \text{DHasAirCon}_{it} + \beta_{11} \cdot \\ & \text{DHasGym}_{it} + \beta_{12} \cdot \text{DHasKitchen}_{it} + \beta_{13} \cdot \text{DHasParking}_{it} + u_i + \epsilon_{it} \end{aligned}$$

Il modello successivo analizza l'effetto del genere dell'host sul ricavo per notte e ha l'obiettivo di verificare se esistano differenze, controllando per altre caratteristiche dell'annuncio (variabili di controllo).

**Modello 2:**

$$\begin{aligned} \text{RevPAN}_{it} = & \beta_0 + \beta_1 \cdot \text{DGender}_{it} + \beta_2 \cdot \text{DListingType}_{it} + \beta_3 \cdot \text{DHostType}_{it} + \beta_4 \cdot \\ & \text{Bedrooms}_{it} + \beta_5 \cdot \text{DYear}_t + \beta_6 \cdot \text{DPriceTier}_{it} + \beta_7 \cdot \text{DSuperhost}_{it} + \beta_8 \cdot \\ & \text{DHasPool}_{it} + \beta_9 \cdot \text{DHasHotTub}_{it} + \beta_{10} \cdot \text{DHasAirCon}_{it} + \beta_{11} \cdot \text{DHasGym}_{it} + \beta_{12} \cdot \\ & \text{DHasKitchen}_{it} + \beta_{13} \cdot \text{DHasParking}_{it} + u_i + \epsilon_{it} \end{aligned}$$

Il terzo modello econometrico ha come variabile dipendente il prezzo medio giornaliero (ADR) degli annunci su Airbnb. L'obiettivo di questo modello è verificare se il genere dell'host influisce sui prezzi medi giornalieri.

**Modello 3:**

$$\begin{aligned} \text{ADR}_{it} = & \beta_0 + \beta_1 \cdot \text{DGender}_{it} + \beta_2 \cdot \text{DListingType}_{it} + \beta_3 \cdot \text{DHostType}_{it} + \beta_4 \cdot \\ & \text{Bedrooms}_{it} + \beta_5 \cdot \text{DYear}_t + \beta_6 \cdot \text{DPriceTier}_{it} + \beta_7 \cdot \text{DSuperhost}_{it} + \beta_8 \cdot \\ & \text{DHasPool}_{it} + \beta_9 \cdot \text{DHasHotTub}_{it} + \beta_{10} \cdot \text{DHasAirCon}_{it} + \beta_{11} \cdot \text{DHasGym}_{it} + \beta_{12} \cdot \\ & \text{DHasKitchen}_{it} + \beta_{13} \cdot \text{DHasParking}_{it} + u_i + \epsilon_{it} \end{aligned}$$

Il prossimo modello analizza l'effetto del genere dell'host sulla variabile dipendente rappresentata dalle recensioni ottenute sugli annunci.

**Modello 4:**

$$\begin{aligned} \text{OverallRating}_i = & \beta_0 + \beta_1 \cdot \text{DGender}_i + \beta_2 \cdot \text{DListingType}_i + \beta_3 \cdot \text{DHostType}_i + \beta_4 \cdot \\ & \text{Bedrooms}_i + \beta_5 \cdot \text{DPriceTier}_i + \beta_6 \cdot \text{DSuperhost}_i + \beta_7 \cdot \text{DHasPool}_i + \beta_8 \cdot \\ & \text{DHasHotTub}_i + \beta_9 \cdot \text{DHasAirCon}_i + \beta_{10} \cdot \text{DHasGym}_i + \beta_{11} \cdot \text{DHasKitchen}_i + \beta_{12} \cdot \\ & \text{DHasParking}_i + \epsilon_i \end{aligned}$$

Il quinto modello studia l'impatto del genere sulla variabile dipendente rappresentata dal tasso di occupazione:

**Modello 5:**

$$\begin{aligned} \text{Occupation}_{it} = & \beta_0 + \beta_1 \cdot \text{DGender}_{it} + \beta_2 \cdot \text{DListingType}_{it} + \beta_3 \cdot \text{DHostType}_{it} + \beta_4 \cdot \\ & \text{Bedrooms}_{it} + \beta_5 \cdot \text{DYear}_t + \beta_6 \cdot \text{DPriceTier}_{it} + \beta_7 \cdot \text{DSuperhost}_{it} + \beta_8 \cdot \\ & \text{DHasPool}_{it} + \beta_9 \cdot \text{DHasHotTub}_{it} + \beta_{10} \cdot \text{DHasAirCon}_{it} + \beta_{11} \cdot \text{DHasGym}_{it} + \beta_{12} \cdot \\ & \text{DHasKitchen}_{it} + \beta_{13} \cdot \text{DHasParking}_{it} + u_i + \epsilon_{it} \end{aligned}$$

I modelli stimati riportati nelle Tabelle 2 e 3 analizzano la relazione tra la variabile indipendente "Genere" e le variabili dipendenti sia in forma lineare sia in forma logaritmica controllando per le variabili di controllo.

#### 4.6.3.1 Regressione multivariata lineare

Di seguito si propongono i risultati ottenuti dall'applicazione dei cinque modelli sui dati forniti. L'elaborazione dei dati è stata eseguita tramite il software di calcolo STATA.

	<b>M1</b>	<b>M2</b>	<b>M3</b>	<b>M4</b>	<b>M5</b>
	<b>Numero di Prenotazioni</b>	<b>RevPAN</b>	<b>ADR</b>	<b>Tasso di occupazione</b>	<b>Valutazione Complessiva</b>
Uomo	1.988 (0.640)***	-0.026 (0.626)	-0.108 (1.285)	0.006 (0.006)	-0.482 (0.222)**
Hotel Room	40.515 (13.564)***	-5.935 (9.781)	-41.899 (14.685)***	0.165 (0.099)*	2.769 (3.196)
Private Room	-4.014 (0.826)***	-14.171 (0.627)***	-34.684 (1.435)***	-0.043 (0.008)***	-0.056 (0.409)
Shared Room	-10.466 (1.736)***	-16.237 (1.087)***	-30.089 (2.064)***	-0.167 (0.018)***	-1.650 (1.309)
Host Mid	0.711 (1.006)	-1.281 (0.985)	-3.483 (2.354)	0.004 (0.009)	0.683 (0.575)
Host Small	-4.712 (0.914)***	-2.306 (1.006)**	-1.505 (2.382)	0.009 (0.008)	3.304 (0.541)***
Camera da letto	-1.240 (0.514)**	11.806 (0.932)***	36.905 (2.017)***	-0.022 (0.004)***	0.464 (0.177)***
Anno 2018	1.985 (0.615)***	3.071 (0.504)***	-0.054 (1.207)	0.040 (0.006)***	YES
Anno 2019	5.874 (0.736)***	6.103 (0.584)***	-1.725 (1.157)	0.093 (0.007)***	YES
Anno 2020	-11.051 (0.715)***	-8.515 (0.600)***	-3.645 (1.265)***	-0.117 (0.007)***	YES
Anno 2021	-6.484 (0.763)***	2.947 (0.733)***	4.596 (1.518)***	-0.008 (0.008)	YES
Anno 2022	4.248 (0.805)***	18.196 (0.775)***	23.023 (1.672)***	0.132 (0.007)***	YES
Anno 2023	6.995 (0.797)***	14.952 (0.744)***	16.257 (1.458)***	0.080 (0.007)***	YES
Fascia di Prezzo: Economy	-1.521 (1.015)	2.585 (0.630)***	11.466 (0.937)***	-0.065 (0.009)***	1.236 (0.337)***
Fascia di Prezzo: Luxury	-7.277 (1.011)***	16.997 (1.086)***	99.923 (2.704)***	-0.212 (0.009)***	2.540 (0.402)***
Fascia di Prezzo: Midscale	-3.063 (1.028)***	4.149 (0.655)***	23.193 (1.005)***	-0.115 (0.009)***	1.937 (0.321)***
Fascia di Prezzo: Upscale	-4.607 (1.016)***	9.355 (0.861)***	42.116 (1.324)***	-0.149 (0.009)***	1.965 (0.350)***
Superhost	16.595 (0.830)***	12.177 (0.682)***	6.348 (1.453)***	0.144 (0.006)***	3.667 (0.186)***
Ha la Piscina	3.780 (4.797)	6.236 (6.190)	0.243 (9.292)	0.029 (0.029)	0.912 (0.849)
Ha la Cucina	-11.385 (1.466)***	-0.415 (0.965)	-1.820 (2.153)	-0.017 (0.011)*	0.224 (0.518)

Ha la Vasca Idromassaggio	1.175 (1.857)	3.803 (3.001)	14.843 (7.763)*	-0.009 (0.016)	0.596 (0.409)
Ha l'Aria Condizionata	8.586 (0.684)***	7.534 (0.655)***	8.402 (1.409)***	0.058 (0.006)***	0.969 (0.219)***
Ha la Palestra	0.077 (0.718)	0.467 (3.217)	15.063 (13.483)	-0.012 (0.023)	-0.311 (0.701)
Ha il Parcheggio	0.077 (0.718)	-0.952 (0.721)	-0.463 (1.607)	0.001 (0.006)	-0.488 (0.267)*
Cons_	33.020 (2.008)***	1.795 (1.967)	-0.492 (4.054)	0.427 (0.017)***	89.008 (0.833)***
N	19,620	19,620	19,620	19,620	19,620
R <sup>2</sup>	0.1363	0.2565	0.3635	0.1964	0.1207

\*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

Tabella 2 : Regressione lineare

Il primo modello (M1) utilizza come variabile dipendente il numero di prenotazioni in forma lineare. I risultati mostrano che il coefficiente di determinazione complessivo (R-squared overall) è pari al 13.63%.

Il coefficiente della variabile principale, ovvero il genere dell'host, è significativo al 1% con un valore di 1.988, suggerendo che gli host uomini ricevono in media due prenotazioni in più rispetto alle host donne. In questa regressione, emerge dunque un'evidenza di gender gap.

Analizzando le variabili di controllo, in particolare focalizzandoci sulla tipologia di alloggio, emerge una significativa differenza nelle prenotazioni:

- La categoria "hotel" risulta  $\beta = 40.515$ ,  $p < 0.01$ . Questo indica che gli hotel ricevono un numero significativamente maggiore di prenotazioni rispetto agli appartamenti interi che sono utilizzati come categoria di riferimento;
- La categoria "camere private" risulta  $\beta = -4.014$ ,  $p < 0.01$ . Il coefficiente negativo indica che ricevono meno prenotazioni rispetto agli interi appartamenti;
- La categoria "stanze condivise" risulta  $\beta = -10.466$ ,  $p < 0.01$  suggerendo che la condivisione degli spazi comporta una riduzione sostanziale delle prenotazioni.

In aggiunta, si evince che un aumento dei posti letto è associato a una leggera riduzione del numero di prenotazioni suggerendo che gli spazi più grandi riducono le prenotazioni:  $\beta = -1,240$  e  $p < 0,10$ .

Anche la tipologia di HOST impatta sul numero di prenotazioni infatti la categoria di host "SMALL" risulta  $\beta = -4.712$ ,  $p = 0.001$  cioè si verifica un numero di prenotazioni inferiore rispetto alla categoria di base rappresentata dall' "Host Large".

Proseguendo l'analisi valutando le fasce di prezzo degli alloggi risulta che le proprietà classificate come "luxury", "midscale" e "upscale" mostrano tutti coefficienti significativi e negativi rispetto alla categoria di riferimento "budget" e come ci si poteva aspettare il segmento luxury ha un impatto fortemente negativo nel numero di prenotazioni ( $\beta = -7.277$ ,  $p < 0.001$ ).

È significativo l'incremento delle prenotazioni per gli host classificati come Superhost che si attesta a 16,59 prenotazioni in più ( $p < 0,001$ ), confermando il valore percepito dall'etichetta Superhost.

La presenza di aria condizionata nei locali mostra un effetto positivo e significativo ( $\beta = 8.58$ ,  $p = 0.001$ ). Mentre, contro le aspettative, la presenza di una cucina sembra avere un effetto negativo significativo ( $\beta = -11.38$ ,  $p = 0.01$ ).

Risultano poco impattanti le restanti variabili nella scelta dell'appartamento quali ad esempio palestra e parcheggio.

Si osserva inoltre che nel numero di prenotazioni il 2020 evidenzia un effetto negativo e significativo ( $\beta = -11.05$ ,  $p < 0.01$ ) dovuto all'impatto del Covid-19.

In conclusione: il numero di prenotazioni è molto influenzato dalla tipologia dell'alloggio e dalla reputazione dell'host. Gli hotel e i Superhost ricevono molte più prenotazioni, mentre le stanze condivise e gli host piccoli ne ricevono di meno. Si riscontra inoltre un gender gap a favore degli host uomini per questa variabile di performance.

Il secondo modello (M2) ha come variabile dipendente il ricavo per notte (RevPAN) e mostra che il coefficiente di determinazione complessivo (R-squared overall) è pari a 0.2565 ovvero, il 25.65% della variabilità totale nei ricavi per notte è spiegata dalle covariate.

Analizzando la variabile principale "Genere" si osserva che il genere dell'host non risulta avere un effetto statisticamente significativo sul RevPAN. ( $\beta = -0.026$ ,  $p = 0.626$ )

Esaminando le variabili di controllo presenti nel modello si osserva che:

- Per la tipologia di alloggio: gli annunci con stanze private generano un ricavo inferiore di -14.17 unità ( $p < 0,01$ ), mentre quelli con stanze condivise registrano

una riduzione di  $-16.23$  unità ( $p > 0,01$ ) rispetto alla categoria di riferimento rappresentata dall'intero appartamento. In generale gli ospiti preferiscono spazi privati rispetto agli spazi condivisi.

- Il numero di posti letto ha un impatto fortemente positivo e altamente significativo ( $\beta = 11.806$  e  $p < 0,01$ ): un aumento dei posti letto è associato a un incremento sostanziale nei ricavi, suggerendo che gli spazi più grandi attraggono più prenotazioni o prenotazioni più costose.
- La tipologia di host "SMALL" indica che gli host con un numero ridotto di annunci hanno ricavi per notte inferiori rispetto a quelli che gestiscono un numero maggiore di proprietà.
- Per la tipologia di proprietà: le proprietà classificate nelle fasce di prezzo luxury, midscale, upscale ed economy mostrano tutti coefficienti positivi e significativi rispetto alla categoria budget. In particolare, il segmento luxury genera l'aumento maggiore nei ricavi ( $\beta = 16.997$ ,  $p < 0,01$ ), evidenziando come il posizionamento premium possa massimizzare i guadagni.
- I Superhost registrano un incremento significativo di  $12,177$  unità di ricavo ( $p < 0,01$ ) indicando come l'ottenimento del badge, possa portare a ottenere ricavi maggiori.
- Dotazioni: tra le dotazioni che caratterizzano la proprietà, gli annunci con aria condizionata ( $\beta = 7.534$ ,  $p < 0,01$ ) hanno un effetto positivo sull'aumento dei ricavi. Le altre variabili legate alle caratteristiche degli annunci (es. cucina, palestra, parcheggio) non mostrano risultati significativi, suggerendo che il loro effetto sui ricavi è meno pronunciato.
- Year: gli anni 2018 e 2019 mostrano un effetto positivo e significativo sul ricavo per notte, con coefficienti pari a  $\beta = 3.071$  e  $\beta = 6.103$  rispettivamente. Nel 2020 si è registrato un calo significativo a causa dell'impatto negativo della pandemia sul fatturato. A partire dal 2022 ( $+18.196$ ,  $p < 0,01$ ) e nel 2023 ( $+14.952$ ,  $p < 0,01$ ), si osserva una forte ripresa dei ricavi per notte, segnalando un recupero significativo nel periodo post-pandemia.

Il terzo modello, indicato come M3, utilizza la tariffa media giornaliera (ADR) come variabile dipendente e restituisce un coefficiente di determinazione (R-quadro) pari a

0.3635, indicando che il 36.35% della variazione del ADR è spiegata dalle variabili indipendenti considerate.

Per quanto riguarda la variabile principale *Genere*, non emerge una relazione significativa tra il genere e l'ADR, dunque, si conclude che il genere non rappresenta un fattore determinante nel contesto delle tariffe giornaliere. In questa regressione non emerge alcuna evidenza del gender gap.

Proseguendo l'analisi valutando il contributo apportato dalle variabili di controllo, si rileva che la tipologia di alloggio ha un impatto significativo sul ADR.

Emerge infatti che:

- La categoria “Hotel”, il coefficiente è negativo e significativo ( $\beta = -41.899$ ,  $p < 0.01$ ), indicando che questi alloggi registrano un ADR significativamente inferiore rispetto agli appartamenti interi, che fungono da categoria di riferimento;
- Le “stanze private” presentano un effetto negativo e significativo ( $\beta = -34.6$ ,  $p < 0.01$ ), confermando che il prezzo per notte per questa tipologia di alloggio è generalmente inferiore rispetto a quello degli appartamenti interi.
- Le “stanze condivise” si è calcolato un coefficiente negativo e significativo ( $\beta = -30.089$ ,  $p < 0.01$ ), riflettendo il minor valore economico associato a una sistemazione meno privata.

La variabile di controllo “Camera da letto” ( $\beta = 36.905$ ,  $p < 0.01$ ) indica che l'aggiunta di una camera da letto incide sull'ADR comportandone un importante incremento. Si può asserire che le proprietà più ampie tendono ad avere tariffe giornaliere più elevate, riflettendo una più elevata capacità ricettiva e un maggior valore percepito da parte degli ospiti.

Si nota inoltre che tutte le categorie di prezzo mostrano un effetto positivo e significativo, ciò conferma che il livello di prezzo incide direttamente sulla tariffa giornaliera. Nello specifico, la categoria luxury vanta un coefficiente altamente positivo ( $\beta = 99.92$ ,  $p < 0.01$ ), evidenziando un impatto significativo sull'ADR, con tariffe nettamente superiori rispetto alla categoria budget.

Analizzando l'effetto dei servizi offerti in relazione a ADR, risulta che:

- Ha la vasca idromassaggio:  $\beta = 14.843$ ,  $p < 0.1$

- Ha l'aria condizionata:  $\beta = 8.402$ ,  $p < 0.01$

Tale risultato evidenzia che la presenza di servizi come vasca idromassaggio oppure aria condizionata impatta sulla tariffa media giornaliera (ADR), evidenziando il valore aggiunto di queste dotazioni per gli ospiti. Al contrario, la presenza del servizio cucina non sembra avere un impatto rilevante sulla variabile dipendente.

Il fattore reputazione host si è rilevato particolarmente impattante infatti il coefficiente della variabile Superhost:  $\beta = 6.348$ ,  $p < 0.01$  è positivo e significativo. Da ciò si evince che gli ospiti sono disposti a pagare un premio di fiducia per soggiornare presso host considerati più affidabili e con una reputazione consolidata.

Valutando gli anni:

- 2022:  $\beta = 23.023$ ,  $p < 0.01$
- 2023:  $\beta = 16.257$ ,  $p < 0.01$

I prezzi sono aumentati, indicando una ripresa del settore turistico e una crescente domanda di alloggi rispetto agli anni 2019 e 2020 dove si riscontra un effetto negativo delle tariffe giustificabile con la pandemia Covid.

In conclusione, il prezzo giornaliero degli alloggi in questo modello non sembra dipendere direttamente dal genere dell'host ma è influenzato da tipologia, numero di camere, categoria di prezzo, anno e dotazioni oltre che dalla reputazione dell'host.

Il Modello 4 valuta il tasso di occupazione (OCC) come variabile dipendente: per esso il coefficiente di determinazione generale (R-squared overall) è pari a 0.1964.

Analizzando il coefficiente della variabile "Genere" non emerge una relazione significativa, indicando che il genere non rappresenta un fattore determinante in questo contesto.

Passando all'analisi delle variabili di controllo si ha:

- Per la tipologia abitativa: sia le stanze private che le stanze condivise hanno un effetto negativo e significativo ( $\beta = -0.043$ ,  $\beta = -0.167$ ) sul tasso di occupazione rispetto agli appartamenti interi.
- Numero di camere da letto:  $\beta = -0.022$ ,  $p = 0.01$  indicano un effetto negativo significativo sul tasso di occupazione cioè, all'aumentare del numero di camere da letto, il tasso di occupazione tende a diminuire.

- Prezzo: tutte le fasce mostrano un effetto negativo e significativo sul tasso di occupazione, in particolare le categorie luxury ( $\beta = -0.212$ ,  $p < 0.01$ ) e upscale ( $\beta = -0.149$ ,  $p < 0.01$ ). Dunque, le proprietà più costose che tendono ad avere una frequenza di occupazione inferiore.
- Il badge di Superhost ha un effetto positivo e significativo ( $\beta = 0.144$ ,  $p < 0.01$ ), indicando che gli alloggi di Superhost registrano una percentuale di occupazione maggiore rispetto agli altri.
- Dotazioni: la disponibilità dell'aria condizionata è associata a un aumento significativo del tasso di occupazione ( $\beta=0.058$  e  $p<0.01$ ), ma indicatori, come HasHotTub e HasGym, non mostrano un impatto statisticamente significativo sul tasso di occupazione.
- Years: In accordo con i modelli precedenti si rileva un calo nel biennio 2020-2021, con una ricrescita negli anni 2022-2023.

In conclusione questa regressione non restituisce alcuna prova della presenza di gender gap.

Il modello 5 (M5) esamina se esiste una relazione tra genere dell' host e punteggio ottenuto nelle recensioni a parità di caratteristiche dell' alloggio.

L'Analisi mostra un R-squared overall pari a 0.1207, indicando che circa il 12% della varianza totale di OverallRating è spiegata dalle variabili indipendenti incluse nel modello.

La variabile principale "Genere dell' host" in questo caso ha un impatto negativo e significativo ( $\beta=-0.482$ ,  $p=0.05$ ) indicando che gli host uomini ricevono in media una valutazione complessiva degli annunci peggiore. Questo risultato potrebbe suggerire che le host donne dedicano maggiore attenzione alla cura degli alloggi e offrono un'esperienza di ospitalità più accogliente per gli ospiti.

Analizzando l'impatto delle variabili di controllo sul punteggio delle recensioni si ha che:

- Numero di posti letto:  $\beta= 0.464$ ,  $p<0.01$  mostra che un aumento del numero di posti letto è associato a un miglioramento significativo delle valutazioni complessive;

- Fascia di prezzo: le sistemazioni di fascia più alta tendono a ricevere valutazioni migliori: la categoria luxury ha un coefficiente significativo ( $\beta=2.540$ ,  $p<0.01$ ), suggerendo un aumento nelle valutazioni complessive rispetto alla categoria budget.
- Caratteristiche della proprietà: la presenza dell'aria condizionata ha un impatto positivo e significativo sul punteggio delle recensioni ( $\beta=0.969$ ,  $p<0.01$ ). Al contrario, in modo controintuitivo, la disponibilità di parcheggio sembra avere un effetto negativo e significativo ( $\beta=-0.488$ ,  $p<0.01$ ) sul punteggio complessivo delle recensioni.
- Avere il badge di Superhost aiuta ad ottenere un punteggio migliore nelle recensioni.

In conclusione in seguito all'analisi dei cinque modelli di regressione lineare si può affermare che il genere dell'host ha un effetto significativo sul numero di prenotazioni e sul punteggio delle recensioni. Gli host uomini ottengono maggiori prenotazioni rispetto alle host donne a parità di caratteristiche dell'alloggio, mentre le host donne riescono ad ottenere valutazioni migliori.

#### 4.6.3.2 *Regressione multivariata logaritmica*

Per migliorare l'interpretazione dei coefficienti e ridurre l'eteroschedasticità, si è applicata una trasformazione logaritmica a ADR, Number of Reservations, RevPAN e Tasso di Occupazione. In questo modo, i coefficienti possono essere interpretati come elasticità, indicando la variazione percentuale della variabile dipendente a seguito di una variazione percentuale della variabile indipendente.

	<b>M1</b>	<b>M2</b>	<b>M3</b>	<b>M4</b>
	<b>(ln) Numero di Prenotazioni</b>	<b>(ln) RevPAN</b>	<b>(ln) ADR</b>	<b>(ln)Tasso di occupazione</b>
<b>Uomo</b>	<b>0.053</b> <b>(0.027)**</b>	-0.001 (0.024)	-0.012 (0.021)	0.006 (0.006)

Camera di albergo	YES	YES	YES	YES
Camera privata	YES	YES	YES	YES
Camera condivisa	YES	YES	YES	YES
Host Medio	YES	YES	YES	YES
Host Piccolo	YES	YES	YES	YES
Camera da letto	YES	YES	YES	YES
Anno 2018	YES	YES	YES	YES
Anno 2019	YES	YES	YES	YES
Anno 2020	YES	YES	YES	YES
Anno 2021	YES	YES	YES	YES
Anno 2022	YES	YES	YES	YES
Anno 2023	YES	YES	YES	YES
Fascia di Prezzo: Economica	YES	YES	YES	YES
Fascia di Prezzo: Lussuosa	YES	YES	YES	YES
Fascia di Prezzo: Media	YES	YES	YES	YES
Fascia di Prezzo: Esclusiva	YES	YES	YES	YES
Superhost	YES	YES	YES	YES
Ha la Piscina	YES	YES	YES	YES
Ha la Cucina	YES	YES	YES	YES
Ha la Vasca Idromassaggio	YES	YES	YES	YES
Ha l'Aria Condizionata	YES	YES	YES	YES
Ha la Palestra	YES	YES	YES	YES
Ha il Parcheggio	YES	YES	YES	YES
N	19,620	19,620	19,620	19,620
R <sup>2</sup>	0.1363	0.2565	0.3635	0.1964

Tabella 3: Regressione logaritmica

\*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

L'uso della trasformazione logaritmica ha confermato che la variabile indipendente rappresentata dal genere dell'host ha un impatto significativo sul numero di prenotazioni. Inoltre, la coerenza dei risultati ottenuti con le regressioni in forma lineare e quelle in forma logaritmica indica che le nostre stime non sono distorte da problemi di eteroschedasticità.

#### 4.6.3.3 *Effetto di moderazione*

Un effetto di moderazione si verifica quando la relazione tra una variabile indipendente e una variabile dipendente viene influenzata dalla presenza di una terza variabile, detta moderatrice.

L'analisi di moderazione coinvolge tre variabili:

- Variabile indipendente  $\rightarrow X$
- Variabile dipendente  $\rightarrow Y$
- Moderatore  $\rightarrow M$

Il moderatore è un fattore che cambia l'intensità o la direzione del legame tra la variabile dipendente e la variabile indipendente. Questo significa che la presenza o assenza di un moderatore può amplificare, attenuare o persino invertire la relazione tra le due variabili. L'analisi dell'effetto di moderazione è cruciale per comprendere come diversi fattori possano influenzare il risultato di un fenomeno e per identificare le condizioni sotto le quali un certo effetto si verifica con maggiore o minore intensità.

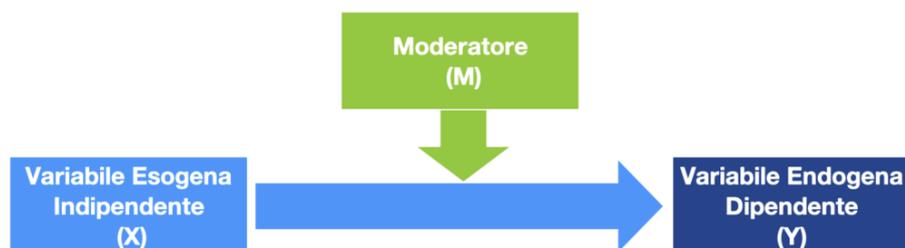


Figura 12: Modello concettuale della moderazione semplice

#### 4.6.3.4 *Regressioni multivariata con interazioni*

Sono state effettuate delle regressioni introducendo le interazioni per analizzare come l'effetto del genere dell'host sulle variabili dipendenti varia in funzione di altre caratteristiche, dette variabili di moderazione.

La variabile moderatrice non ha solo un effetto diretto sugli esiti, ma influenza anche il modo in cui il genere dell'host impatta tali esiti, permettendo di analizzare se il gender gap nelle performance è costante o dipende da alcune caratteristiche dell'host e dell'annuncio e se esistano situazioni in cui l'effetto del genere è amplificato o attenuato.

Questo approccio, dunque, permette di andare oltre una semplice misurazione del gender gap e di identificare le condizioni specifiche in cui il divario si manifesta, contribuendo a una lettura più accurata delle discriminazioni nel contesto della sharing economy.

Nello specifico sono state scelte le seguenti interazioni:

- *Gender × Superhost*: questa interazione permette di analizzare se lo status di Superhost modera l'effetto del genere sulle performance. Se lo status di Superhost riducesse o amplificasse il gender gap, ciò potrebbe indicare che il riconoscimento della qualità da parte della piattaforma influisce sulla percezione degli ospiti rispetto al genere dell'host.
- *Gender × Listing Type*: questa interazione indaga se l'impatto del genere dell'host varia in funzione del tipo di alloggio offerto. È possibile che gli ospiti reagiscano in modo diverso a host uomini e donne a seconda della tipologia di sistemazione.
- *Gender × Price Tier*: questa interazione verifica se l'effetto del genere è differenziato in base alla fascia di prezzo dell'annuncio. L'effetto del genere potrebbe variare tra fasce di prezzo.
- *Gender × Year*: questa interazione consente di valutare se il gender gap è cambiato nel tempo, osservando se l'effetto del genere sulle performance varia tra gli anni considerati (2017-2022). Questa analisi è utile per comprendere eventuali evoluzioni nel comportamento degli ospiti o nelle dinamiche di mercato su Airbnb.

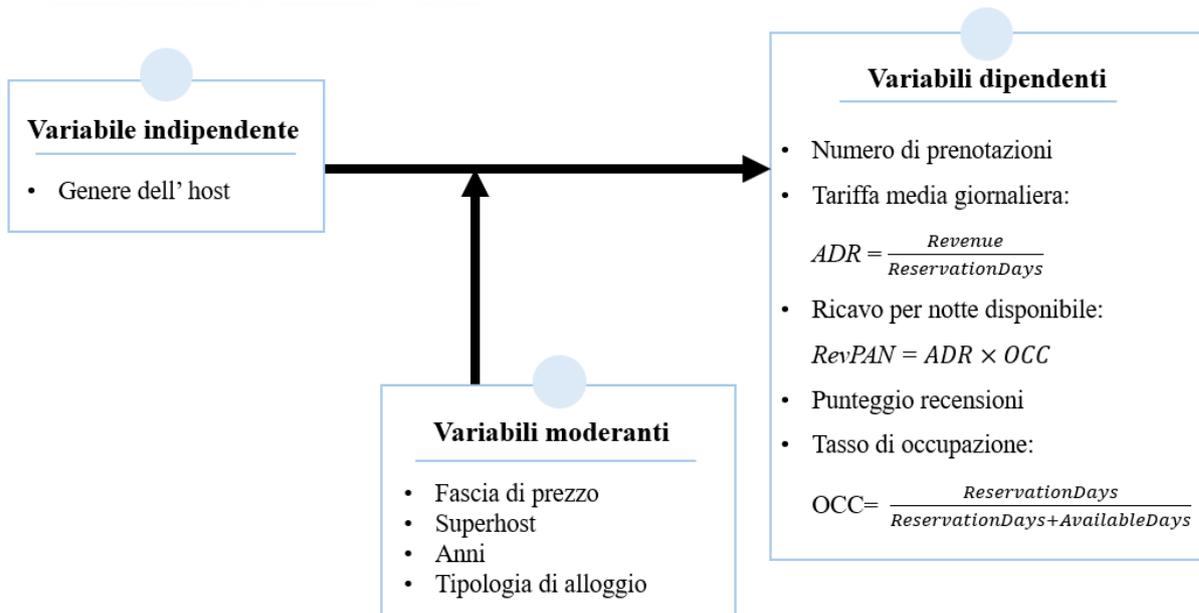


Figura 13: Effetto di interazione

Il seguente modello econometrico analizza il numero di prenotazioni ricevute da un annuncio su Airbnb, considerando il genere dell'host e la sua interazione con altre variabili chiave. La specificazione del modello è la seguente:

**Modello 1:**

$$\begin{aligned} \text{Number of Reservations}_{it} = & \beta_0 + \beta_1 D\text{Gender}_{it} + \beta_2 D\text{Superhost}_{it} + \beta_3 (D\text{Gender}_{it} \cdot \\ & D\text{Superhost}_{it}) + \beta_4 D\text{ListingType}_{it} + \beta_5 D\text{HostType}_{it} + \beta_6 \text{Bedrooms}_{it} + \\ & \beta_7 \text{Bedrooms}_{it}^2 + \beta_8 D\text{Year}_t + \beta_9 D\text{PriceTier}_{it} + \beta_{10} D\text{HasPool}_{it} + \beta_{11} D\text{HasHotTub}_{it} + \\ & \beta_{12} D\text{HasAirCon}_{it} + \beta_{13} D\text{HasGym}_{it} + \beta_{14} D\text{HasKitchen}_{it} + \beta_{15} D\text{HasParking}_{it} + \\ & \beta_{16} (D\text{Gender}_{it} \cdot D\text{Year}_t) + \beta_{17} (D\text{Gender}_{it} \cdot D\text{ListingType}_{it}) + \beta_{18} (D\text{Gender}_{it} \cdot \\ & D\text{PriceTier}_{it}) + u_i + \epsilon_{it} \end{aligned}$$

L'obiettivo del secondo modello è analizzare se l'effetto del genere dell'host sul RevPAN varia in base alle variabili moderatrici.

**Modello 2:**

$$\begin{aligned}
\text{RevPAN} = & \beta_0 + \beta_1 \text{DGender}_{\{it\}} + \beta_2 \text{DSuperhost}_{\{it\}} + \beta_3 (\text{DGender}_{\{it\}} \cdot \text{DSuperhost}_{\{it\}}) + \\
& \beta_4 \text{DListingType}_{\{it\}} + \beta_5 \text{DHostType}_{\{it\}} + \beta_6 \text{Bedrooms}_{\{it\}} + \beta_7 \text{Bedrooms}_{\{it\}}^{\{2\}} + \\
& \beta_8 \text{DYear}_{\{it\}} + \beta_9 \text{DPriceTier}_{\{it\}} + \beta_{10} \text{DHasPool}_{\{it\}} + \beta_{11} \text{DHasHotTub}_{\{it\}} + \\
& \beta_{12} \text{DHasAirCon}_{\{it\}} + \beta_{13} \text{DHasGym}_{\{it\}} + \beta_{14} \text{DHasKitchen}_{\{it\}} + \beta_{15} \text{DHasParking}_{\{it\}} + \\
& \beta_{16} (\text{DGender}_{it} \cdot \text{DYear}_t) + \beta_{17} (\text{DGender}_{it} \cdot \text{DListingType}_{it}) \\
& + \beta_{18} (\text{DGender}_{it} \cdot \text{DPriceTier}_{it}) + u_i + \epsilon_{it}
\end{aligned}$$

Il terzo modello analizza l'effetto del genere dell'host in relazione alla tariffa giornaliera (ADR). Oltre alla variabile principale del genere, il modello include diverse interazioni che permettono di valutare se e come l'impatto del genere varia in funzione di fattori chiave, come lo status di Superhost, il tipo di annuncio, la fascia di prezzo e l'anno di osservazione. Queste interazioni consentono di verificare se il divario di genere nei ricavi sia uniforme o se emerga in specifici contesti, offrendo un'analisi più dettagliata delle dinamiche della piattaforma.

**Modello 3:**

$$\begin{aligned}
\text{ADR} = & \beta_0 + \beta_1 \text{DGender}_{it} + \beta_2 \text{DSuperhost}_{it} + \beta_3 (\text{DGender}_{it} \cdot \text{DSuperhost}_{it}) + \\
& \beta_4 \text{DListingType}_{it} + \beta_5 \text{DHostType}_{it} + \beta_6 \text{Bedrooms}_{it} + \beta_7 \text{Bedrooms}_{it}^2 + \beta_8 \text{DYear}_t + \\
& \beta_9 \text{DPriceTier}_{it} + \beta_{10} \text{DHasPool}_{it} + \beta_{11} \text{DHasHotTub}_{it} + \beta_{12} \text{DHasAirCon}_{it} + \\
& \beta_{13} \text{DHasGym}_{it} + \beta_{14} \text{DHasKitchen}_{it} + \beta_{15} \text{DHasParking}_{it} + \beta_{16} (\text{DGender}_{it} \cdot \\
& \text{DYear}_t) + \beta_{17} (\text{DGender}_{it} \cdot \text{DListingType}_{it}) + \beta_{18} (\text{DGender}_{it} \cdot \text{DPriceTier}_{it}) + u_i + \\
& \epsilon_{it}
\end{aligned}$$

Il quarto modello analizza la correlazione tra il genere dell'host e il tasso di occupazione includendo variabili di moderazione.

In particolare, vengono esplorate interazioni con lo status di Superhost, il tipo di annuncio, la fascia di prezzo e l'anno di osservazione, permettendo di individuare eventuali differenze nell'impatto del genere a seconda del contesto. Questo approccio consente di comprendere se il divario di genere nei ricavi sia uniforme o se emerga solo in determinate situazioni.

**Modello 4:**

$$\begin{aligned} \text{Occupation} = & \beta_0 + \beta_1 \text{DGender}_{it} + \beta_2 \text{DSuperhost}_{it} + \beta_3 (\text{DGender}_{it} \cdot \\ & \text{DSuperhost}_{it}) + \beta_4 \text{DListingType}_{it} + \beta_5 \text{DHostType}_{it} + \beta_6 \text{Bedrooms}_{it} + \\ & \beta_7 \text{Bedrooms}_{it}^2 + \beta_8 \text{DYear}_t + \beta_9 \text{DPriceTier}_{it} + \beta_{10} \text{DHasPool}_{it} + \beta_{11} \text{DHasHotTub}_{it} + \\ & \beta_{12} \text{DHasAirCon}_{it} + \beta_{13} \text{DHasGym}_{it} + \beta_{14} \text{DHasKitchen}_{it} + \beta_{15} \text{DHasParking}_{it} + \\ & \beta_{16} (\text{DGender}_{it} \cdot \text{DYear}_t) + \beta_{17} (\text{DGender}_{it} \cdot \text{DListingType}_{it}) + \beta_{18} (\text{DGender}_{it} \cdot \\ & \text{DPriceTier}_{it}) + u_i + \epsilon_{it} \end{aligned}$$

Il seguente modello econometrico analizza come il genere dell' host influenzi il punteggio delle recensioni e se questo effetto varia in maniera diversa in base alle interazioni con le altre variabili.

**Modello 5:**

$$\begin{aligned} \text{OverallRating} = & \beta_0 + \beta_1 \text{DGender}_{it} + \beta_2 \text{DSuperhost}_{it} + \beta_3 (\text{DGender}_{it} \cdot \\ & \text{DSuperhost}_{it}) + \beta_4 \text{DListingType}_{it} + \beta_5 \text{DHostType}_{it} + \beta_6 \text{Bedrooms}_{it} + \\ & \beta_7 \text{Bedrooms}_{it}^2 + \beta_8 \text{DYear}_t + \beta_9 \text{DPriceTier}_{it} + \beta_{10} \text{DHasPool}_{it} + \beta_{11} \text{DHasHotTub}_{it} + \\ & \beta_{12} \text{DHasAirCon}_{it} + \beta_{13} \text{DHasGym}_{it} + \beta_{14} \text{DHasKitchen}_{it} + \beta_{15} \text{DHasParking}_{it} + \\ & \beta_{16} (\text{DGender}_{it} \cdot \text{DYear}_t) + \beta_{17} (\text{DGender}_{it} \cdot \text{DListingType}_{it}) + \beta_{18} (\text{DGender}_{it} \cdot \\ & \text{DPriceTier}_{it}) + u_i + \epsilon_{it} \end{aligned}$$

L'implementazione dei precedenti modelli nel software di calcolo STATA restituisce i seguenti risultati:

	<b>M1</b>	<b>M2</b>	<b>M3</b>	<b>M4</b>	<b>M5</b>
	<b>NumberofReservations</b>	<b>RevPan</b>	<b>ADR</b>	<b>Occupation</b>	<b>OverallRating</b>
<b>Uomo</b>	<b>4.261</b> (1.967)**	<b>2.700</b> (1.317)**	2.211 (2.480)	<b>0.034</b> (0.017)**	-0.516 (0.382)
Hotel Room	40.460 (13.666)***	-5.210 (9.995)	-40.449 (15.386)***	0.163 (0.099)*	4.267 (0.550)***
Private Room	-4.559 (1.035)***	-14.595 (0.869)***	-36.043 (2.026)***	-0.049 (0.010)***	0.510 (0.205)**
Shared Room	-13.968 (2.000)***	-14.250 (2.157)***	-28.205 (4.319)***	-0.111 (0.043)**	-5.809 (1.910)***
Year 2018	2.891 (0.810)***	3.604 (0.695)***	1.432 (1.793)	0.042 (0.008)***	0.150 (0.286)

Year 2019	6.576 (0.936)***	7.358 (0.803)***	-0.768 (1.629)	0.099 (0.009)***	0.461 (0.276)*
Year 2020	-10.169 (0.864)***	-7.543 (0.799)***	-2.391 (1.715)	-0.110 (0.009)***	0.390 (0.276)
Year 2021	-6.359 (0.969)**	3.079 (0.901)***	6.472 (1.936)***	-0.013 (0.010)	0.297 (0.268)
Year 2022	4.239 (1.007)***	19.723 (1.025)***	26.522 (2.457)***	0.137 (0.010)***	0.319 (0.257)
Year 2023	6.455 (0.994)**	15.865 (0.966)***	19.733 (2.099)***	0.081 (0.009)***	0.569 (0.249)**
Economy	0.443 (1.389)	3.295 (0.870)***	12.040 (1.253)***	-0.051 (0.012)***	0.713 (0.197)***
Luxury	-6.248 (1.317)***	17.845 (1.452)***	100.101 (3.614)***	-0.197 (0.012)***	2.266 (0.239)***
Midscale	-1.757 (1.387)	4.457 (0.871)***	23.000 (1.273)***	-0.104 (0.012)***	1.486 (0.192)***
Upscale	-3.813 (1.303)***	9.279 (1.020)***	39.788 (1.523)***	-0.133 (0.012)***	2.041 (0.188)***
Superhost	17.491 (1.116)***	13.105 (0.909)***	4.348 (1.459)***	0.152 (0.008)***	3.296 (0.106)***
Private room * Uomo	1.142 (1.723)	0.985 (1.170)	2.803 (2.570)	0.013 (0.014)	-0.077 (0.318)
Shared room * Uomo	4.395 (2.921)	-2.220 (2.454)	-0.888 (4.764)	-0.072 (0.048)	5.034 (2.107)**
Year 2018 * Uomo	-2.099 (1.244)*	-1.279 (1.002)	-3.517 (2.324)	-0.005 (0.013)	0.117 (0.462)
Year 2019 * Uomo	-1.641 (1.504)	-2.966 (1.154)**	-2.422 (2.262)	-0.013 (0.014)	-0.015 (0.447)
Year 2020 * Uomo	-2.009 (1.478)	-2.273 (1.198)*	-3.045 (2.497)	-0.016 (0.014)	0.077 (0.461)
Year 2021 * Uomo	-0.350 (1.558)	-0.427 (1.492)	-4.426 (3.051)	0.010 (0.015)	-0.093 (0.460)
Year 2022 * Uomo	0.003 (1.651)	-3.437 (1.555)**	-7.867 (3.254)**	-0.010 (0.015)	0.156 (0.424)
Year 2023 * Uomo	1.213 (1.640)	-2.020 (1.482)	-7.626 (2.795)***	-0.002 (0.014)	0.006 (0.402)
Economy * Uomo	-3.930 (2.018)*	-1.068 (1.224)	-0.560 (1.725)	-0.026 (0.018)	-0.161 (0.335)
Luxury * Uomo	-1.835 (2.020)	-1.471 (2.173)	0.047 (5.678)	-0.029 (0.017)*	-0.216 (0.375)
Midscale * Uomo	-2.342 (2.047)	-0.027 (1.288)	1.548 (1.889)	-0.018 (0.017)	0.080 (0.290)
Upscale * Uomo	-1.220 (2.051)	0.907 (1.705)	6.249 (2.505)**	-0.030 (0.017)*	-0.679 (0.334)**
Superhost * Uomo	-2.029 (1.650)	-1.964 (1.321)	4.292 (2.826)	-0.016 (0.012)	0.479 (0.180)***
Has Kitchen	YES	YES	YES	YES	YES
Has Hot Tub	YES	YES	YES	YES	YES

Has Air Condition	YES	YES	YES	YES	YES
Has Gym	YES	YES	YES	YES	YES
Has Parking	YES	YES	YES	YES	YES
Host Pool	YES	YES	YES	YES	YES
Host Mid	YES	YES	YES	YES	YES
Host Small	YES	YES	YES	YES	YES
Bedroom	YES	YES	YES	YES	YES
Bedroom^2	YES	YES	YES	YES	YES
Cons	33.414 (2.280)***	4.486 (2.085)**	6.712 (4.813)	0.427 (0.019)***	87.014 (0.471)***
N	19,620	19,620	19,620	19,620	19,620
R <sup>2</sup>	0.1378	0.2633	0.3690	0.1977	0.12

Tabella 4 : Regressione lineare con interazione

\* p<0.1; \*\* p<0.05; \*\*\* p<0.01

Il primo modello esamina il numero di prenotazioni ricevute dall'host, valutando l'impatto del genere su questa variabile dipendente e analizzando se tale effetto subisce variazioni in relazione alle interazioni incluse nel modello. Il coefficiente di determinazione complessivo (R-squared overall) indica che circa il 13,78% della variabilità totale nel numero di prenotazioni è spiegata dalle variabili indipendenti incluse nel modello.

Il coefficiente associato al genere dell' host mostra un valore positivo e significativo ( $\beta=4.261$ ,  $p=5\%$ ) indicando che gli host uomini ricevono in media 4.261 prenotazioni in più rispetto alle donne, a parità di altre condizioni. Ciò indica la presenza di un vantaggio competitivo per gli host uomini nel mercato di Airbnb.

Da un'analisi dell'effetto diretto delle variabili moderatrici, emerge un impatto significativo della tipologia di alloggio:

- Gli hotel ricevono molte più prenotazioni rispetto agli appartamenti interi ( $\beta = 40.460$ ,  $p<0.01$ );

- Le stanze private hanno un effetto negativo e significativo ( $\beta = -4.559$ ,  $p < 0.01$ ), indicando che ricevono meno prenotazioni rispetto agli appartamenti interi;
- Le stanze condivise sono molto meno richieste rispetto agli appartamenti interi ( $\beta = -13.968$ ,  $p < 0.01$ );
- Soluzioni Luxury, midscale e upscale indicano un effetto negativo sul numero di prenotazioni per gli annunci di fascia medio-alta;
- Il badge di Superhost ha un effetto positivo sul numero di prenotazioni in quanto è un indicatore di qualità.

Le variabili temporali hanno un impatto significativo infatti, analogamente alle analisi precedenti nel 2019, si registra un aumento delle prenotazioni ( $\beta = 6.576$ ,  $p < 0.01$ ), seguito da un brusco calo nel 2020 ( $\beta = -10.169$ ,  $p < 0.01$ ). Successivamente, si osserva una ripresa con il numero di prenotazioni che torna ai livelli pre-COVID ( $\beta = 6.45$ ,  $p = 0.01$ ).

Esaminando le interazioni tra la variabile indipendente “Genere dell’ host” e le variabili moderatrici, emergono effetti congiunti che arricchiscono l’interpretazione dei risultati. L’interazione significativa “ Uomo  $\times$  Economy ” ( $\beta = -3.930$ ,  $p < 0.1$ ) mostra che, nella categoria economy, gli host uomini perdono gran parte del loro vantaggio, ottenendo risultati simili a quelli delle host donne. Dunque in questa categoria il gap si riduce dimostrando che l’effetto del genere dell’ host sul numero di prenotazioni è influenzato dalla fascia di prezzo dell’alloggio.

Questo suggerisce che il gender gap nelle prenotazioni dipende dal segmento di mercato e che le preferenze degli ospiti possono variare in base alla fascia di prezzo dell’ alloggio. Le altre interazioni tra il genere e le variabili moderatrici non risultano essere significative.

Analizzando il secondo modello (M2) si ha come variabile dipendente il RevPAN.

L’R-squared overall è pari a 0,2633, indicando che il modello spiega circa il 26.3% della variabilità nei ricavi per notte è spiegata dal modello.

Il genere dell’host ha un impatto positivo e significativo sul RevPAN ( $\beta = 2.700$ ,  $p < 0.05$ ), indicando che, a parità di caratteristiche dell’alloggio, gli uomini ottengono un ricavo per

notte disponibile maggiore rispetto alle donne. Si evince dunque la presenza di un gender gap a favore degli uomini.

Passando all'analisi dell'effetto diretto delle variabili moderatrici incluse nel modello si osserva che il tipo di alloggio ha un effetto significativo, con le stanze private e condivise che registrano una riduzione ( $\beta = -14.595$  e  $\beta = -14.250$ ,  $p < 0.01$ ) del ricavo medio per notte rispetto agli interi appartamenti.

Le categorie di prezzo esercitano tutte un impatto importante sul RevPAN, si nota infatti che per la categoria Luxury si ottiene un incremento del ricavo medio per notte pari a +17.89 rispetto alla categoria Budget.

Si analizza l'effetto che le variabili moderatrici hanno nella relazione tra la variabile indipendente "Genere" e una variabile dipendente RevPAN:

- L'interazione " Uomo  $\times$  Year 2019 " ( $\beta = -2.966$ ,  $p < 0.05$ ) suggerisce che, nonostante un aumento generale del fatturato nel 2019 ( $\beta = 7.358$ ,  $p < 0.01$ ) rispetto all'anno di riferimento gli host uomini hanno registrato un calo significativo nei ricavi per notte. Gli ospiti potrebbero aver iniziato a favorire più spesso le host donne nel 2019.
- L'interazione " Uomo  $\times$  Year 2020 " ( $\beta = -2.273$ ,  $p < 0.1$ ) mostra che nel 2020, gli host uomini hanno guadagnato mediamente circa 2.3 unità di ricavo in meno rispetto agli anni precedenti, a parità di altre condizioni.
- L'interazione " Uomo  $\times$  Year 2022 " ( $\beta = -3.437$ ,  $p < 0.05$ ) evidenzia che, sebbene ci sia stato un aumento del RevPAN per entrambi i generi nel 2022, gli uomini hanno registrato un incremento inferiore rispetto alle donne.

Analizzando l'interazione " Uomo  $\times$  Price " si osserva che l'interazione non è significativa per nessuna delle categorie di prezzo, dunque non si può concludere che la fascia di prezzo influenzi la relazione tra genere dell'host e RevPAN.

L'interazione " Uomo  $\times$  Superhost " ha un effetto negativo ma non significativo dunque anche in questo caso non si può concludere che il badge di Superhost ha un impatto rilevante nella relazione tra il genere dell'host e RevPAN. In altre parole l'effetto del

genere dell' host sul RevPAN non cambia in modo rilevante in base alla variabile moderatrice Superhost.

Il terzo modello (M3) analizza l'ADR, ovvero la tariffa media giornaliera. L'R-squared overall è pari al 36,90%, indicando che una parte significativa della variabilità dell'ADR è spiegata dal modello.

Si denota in primis che il genere dell'host ha un effetto positivo, ma non significativo ( $\beta = 2.211$ ), indicando dunque che non ci sono differenze significative tra uomini e donne nel prezzo medio per notte.

Analizzando l'effetto diretto della tipologia di alloggio sulle tariffe per notte si osserva:

- Gli hotel tendono a offrire prezzi mediamente più bassi rispetto agli appartamenti interi, come evidenziato dal coefficiente negativo ( $\beta = -40.449$ ,  $p < 0.01$ ).
- Le stanze private e condivise mostrano una riduzione significativa del prezzo medio rispetto agli appartamenti interi, con coefficienti rispettivamente pari a  $\beta = -36.043$  e  $\beta = -28.205$  ( $p < 0.01$ ).

Dunque, le strutture con spazi condivisi tendono ad avere tariffe inferiori rispetto alle soluzioni abitative interamente dedicate agli ospiti.

Come era ragionevole aspettarsi, le fasce di prezzo degli alloggi influenzano significativamente l'ADR: gli alloggi luxury applicano tariffe giornaliere molto più alte rispetto alla categoria budget ( $\beta = 100.326$ ,  $p < 0.01$ ), mentre le proprietà midscale e upscale hanno prezzi intermedi.

Lo status di Superhost ha un impatto positivo e significativo ( $\beta=4.34$ ,  $p<0.01$ ) sull'ADR, delineando la tendenza degli ospiti a spendere di più per annunci gestiti da host con questo riconoscimento.

La variabile temporale Year conferma che mentre nel periodo 2020-2021 le tariffe sono diminuite (così come anche il numero di richieste) nel biennio successivo i coefficienti positivi e significativi evidenziano un aumento rilevante delle tariffe. Questo suggerisce una crescita significativa dell'ADR in questi anni, influenzata da cambiamenti nella domanda di mercato.

L'analisi delle interazioni mostra che l'effetto del genere dell'host sull'ADR non è costante, ma varia in funzione delle diverse condizioni:

- L'interazione “ Uomo × Year 2022 ” ( $\beta = -7.86$ ,  $p < 0.05$ ) mostra che l'effetto positivo di Year 2022 sull'ADR è ridotto di 7.86 per gli uomini rispetto alle donne, indicando potenziali differenze di strategia o di domanda rispetto alle host donne.
- L'interazione “ Uomo × Year 2023 ” ( $\beta = -7.626$ ,  $p < 0.01$ ) evidenzia che sebbene vi sia un incremento significativo dell'ADR nel 2023 per entrambi i generi, gli host uomini applicano un ADR inferiore rispetto alle donne;
- L'interazione “ Uomo × Upscale ” ( $\beta = 6.249$ ) indica che gli uomini traggono un maggiore vantaggio in termini di ADR rispetto alle donne quando appartengono alla categoria upscale, suggerendo che per gli host uomini questa categoria ha un impatto più significativo sulle tariffe.

Dunque, mentre negli alloggi di fascia bassa non c'è una differenza significativa tra host uomo e donna, negli alloggi upscale gli uomini ottengono risultati migliori.

Da ciò si evince che il gender gap esiste solo in determinati segmenti di mercato.

Il quarto modello (M4) ha come variabile dipendente il tasso di occupazione e mostra un coefficiente di determinazione R-squared overall = 0.1977

Il coefficiente associato al genere “Uomo” dell'host è positivo e significativo ( $\beta = 0.034$ ,  $p < 0.05$ ) dunque, gli host uomini risultano avere un tasso di occupazione leggermente superiore rispetto alle host donne.

Analizzando l'effetto delle variabili moderanti la categoria dell'alloggio ha un impatto rilevante:

- Gli hotel mostrano un coefficiente positivo e significativo ( $\beta = 0.163$ ,  $p = 0.10$ ), dimostrando che il loro tasso di occupazione è sensibilmente superiore a quello degli appartamenti interi;
- Le stanze private ( $\beta = -0.049$ ,  $p < 0.01$ ) denotano meno prenotazioni in proporzione alle notti disponibili rispetto agli appartamenti interi;
- Le stanze condivise hanno un effetto ancora più negativo ( $\beta = -0.111$ ,  $p < 0.01$ ), confermando che questa tipologia di alloggio è meno richiesta e ha un minore tasso di occupazione rispetto agli appartamenti interi.

Analizzando le categorie di fascia di prezzo, si determina che tutte le categorie superiori alla categoria budget presentano coefficienti negativi e significativi, delineando una relazione inversa tra aumento della fascia di prezzo e il tasso di occupazione.

Lo status di Superhost ha un impatto positivo e significativo sul tasso di occupazione ( $\beta = 0.152$ ,  $p < 0.01$ ), ottenendo un numero significativamente maggiore di prenotazioni in proporzione alla disponibilità.

L'anno 2019 ha visto un incremento significativo del tasso di occupazione ( $\beta = 0.099$ ,  $p < 0.01$ ), indicando un picco nella domanda prima dell'inizio della pandemia di COVID-19. Il 2020 segna una riduzione ( $\beta = -0.110$ ,  $p < 0.01$ ) e successivamente, a partire dal 2021, si osserva una ripresa progressiva.

Passando all'analisi delle variabili di interazione si osserva che la relazione tra genere dell'host e tasso di occupazione non è uniforme ma varia a seconda della fascia di prezzo:

- La variabile di interazione “ Uomo  $\times$  Luxury ” ( $\beta=-0.029$ ,  $p<0.05$ ) suggerisce che in questa fascia di prezzo, il tasso di occupazione risulta sostanzialmente simile tra uomini e donne dunque, il vantaggio degli uomini nel tasso di occupazione si riduce.
- La variabile di interazione “ Uomo  $\times$  Upscale ” ( $\beta=-0.030$ ,  $p=0.10$ ) indica che sebbene gli alloggi upscale tendano ad avere un tasso di occupazione inferiore indipendentemente dal genere dell'host, gli host uomini tendono a ottenere risultati inferiori rispetto alle donne anche se l'effetto non è fortemente significativo.

Questi risultati suggeriscono che il gender gap su Airbnb non è solo una questione di discriminazione diretta, ma può essere influenzato da fattori strutturali, strategie di gestione e percezioni dei clienti.

Il quinto modello (M5) ha come variabile dipendente il punteggio medio delle recensioni che rappresenta un indicatore chiave della qualità percepita dell'host e della struttura.

Il modello ha un R-squared overall = 0.12.

Dai risultati ottenuti si riporta che il genere dell'host non ha un impatto significativo sul punteggio medio delle recensioni ( $\beta = -0.516$ ) dunque gli host uomini non ricevono valutazioni significativamente diverse dalle host donne.

Analizzando l'effetto diretto delle variabili moderatrici si ha:

- La tipologia di alloggio: gli hotel raccolgono più recensioni positive rispetto agli interi appartamenti, in controtendenza troviamo le camere condivise registrano un effetto negativo e significativo ( $\beta = -5.809$ ,  $p < 0.01$ ), indicando che gli ospiti tendono a dare recensioni decisamente più basse per questo tipo di alloggio.
- Nelle fasce di prezzo la categoria luxury si distingue per avere un effetto molto positivo e significativo ( $\beta=2.26$ ,  $p<0.01$ ): gli alloggi di fascia alta tendono ad ottenere valutazioni particolarmente positive.
- Gli host con status di Superhost ricevono in media valutazioni più alte ( $\beta = 3.296$ ,  $p < 0.01$ ), confermando che questa etichetta è associata a una migliore esperienza per gli ospiti e a un miglioramento della reputazione complessiva.

L'analisi delle interazioni tra il genere dell'host e alcune caratteristiche dell'alloggio evidenzia che:

- La variabile di interazione “ Uomo  $\times$  Upscale ” ( $\beta=-0.679$ ,  $p<0.05$ ) indica che per gli uomini, l'effetto positivo di avere un annuncio upscale è più debole rispetto alle donne. Mentre gli annunci di fascia alta ottengono valutazioni più alte, gli uomini che gestiscono queste strutture ricevono valutazioni più basse rispetto alle donne;
- La variabile di interazione “ Uomo  $\times$  Shared room ” ( $\beta=5.034$ ,  $p<0.05$ ) mostra che sebbene gli spazi condivisi hanno un impatto molto negativo sull'Overall Rating in generale, gli host maschi subiscono una penalizzazione significativamente minore rispetto alle donne;
- L'effetto positivo e significativo della variabile di interazione “ Uomo  $\times$  Superhost” ( $\beta=0.479$ ,  $p<0.01$ ) permette agli host maschi che possiedono il badge di Superhost di recuperare parte dello svantaggio sull'Overall Rating.

In conclusione, il genere gap non risulta essere uniforme, ma varia in base al segmento di mercato e dallo status di Superhost.

Di seguito è applicata una trasformazione logaritmica alle variabili Number of Reservations, RevPAN, ADR e Tasso di Occupazione. In questo modo, i coefficienti possono essere interpretati come elasticità, indicando la variazione percentuale della variabile dipendente a seguito di una variazione percentuale della variabile indipendente.

	M1	M2	M3	M4
	logNumberofReservations	logRevPan	logADR	logOccupation
<b>Uomo</b>	<b>0.161</b> (0.086)*	0.087 (0.076)	0.061 (0.082)	<b>0.023</b> (0.012)*
Hotel Room	1.050 (0.377)***	0.116 (0.397)	-0.239 (0.166)	0.116 (0.070)*
Private Room	-0.253 (0.048)***	-0.669 (0.044)***	-0.558 (0.041)***	-0.038 (0.007)***
Shared Room	-0.726 (0.173)***	-1.004 (0.188)***	-0.646 (0.184)***	-0.083 (0.032)**
Year 2018	0.154 (0.043)***	0.199 (0.044)***	0.177 (0.060)***	0.031 (0.006)***
Year 2019	0.290 (0.048)***	0.362 (0.047)***	0.208 (0.059)***	0.070 (0.007)***
Year 2020	-0.400 (0.047)***	-0.424 (0.048)***	0.058 (0.061)	-0.083 (0.007)***
Year 2021	-0.292 (0.051)***	-0.035 (0.052)	0.169 (0.063)***	-0.013 (0.008)*
Year 2022	0.142 (0.049)***	0.628 (0.048)***	0.507 (0.057)***	0.093 (0.007)***
Year 2023	0.155 (0.049)***	0.443 (0.048)***	0.408 (0.057)***	0.054 (0.007)***
Economy	0.046 (0.059)	0.082 (0.047)*	0.171 (0.040)***	-0.033 (0.009)***
Luxury	-0.295 (0.059)***	0.203 (0.053)***	0.681 (0.049)***	-0.138 (0.008)***
Midscale	-0.084 (0.058)	0.052 (0.047)	0.256 (0.042)***	-0.071 (0.008)***
Upscale	-0.136 (0.056)**	0.167 (0.048)***	0.440 (0.044)***	-0.090 (0.008)***
Superhost	0.922 (0.042)**	0.706 (0.031)***	0.386 (0.026)***	0.113 (0.006)***
Private room * Uomo	-0.039 (0.068)	0.066 (0.061)	-0.053 (0.055)	0.012 (0.010)
Shared room * Uomo	0.260 (0.204)	-0.138 (0.206)	-0.219 (0.200)	-0.048 (0.036)
Year 2018 * Uomo	-0.143 (0.065)**	-0.132 (0.065)**	-0.171 (0.090)*	-0.007 (0.009)
Year 2019 * Uomo	-0.077 (0.072)	-0.097 (0.068)	-0.114 (0.086)	-0.011 (0.010)

Year 2020 * Uomo	-0.031 (0.074)	-0.050 (0.072)	-0.089 (0.091)	-0.010 (0.010)
Year 2021 * Uomo	0.023 (0.078)	0.042 (0.078)	-0.104 (0.092)	0.008 (0.011)
Year 2022 * Uomo	-0.030 (0.075)	-0.071 (0.071)	-0.143 (0.083)*	-0.007 (0.011)
Year 2023 * Uomo	0.004 (0.075)	-0.052 (0.071)	-0.115 (0.083)	-0.002 (0.010)
Economy * Uomo	-0.115 (0.084)	-0.072 (0.067)	0.006 (0.058)	-0.019 (0.013)
Luxury * Uomo	-0.095 (0.085)	-0.055 (0.077)	0.026 (0.072)	-0.020 (0.012)*
Midscale * Uomo	-0.020 (0.083)	-0.003 (0.067)	0.111 (0.059)*	-0.012 (0.012)
Upscale * Uomo	-0.026 (0.082)	-0.034 (0.071)	0.162 (0.061)***	-0.021 (0.012)*
Superhost * Uomo	-0.108 (0.058)*	-0.054 (0.044)	-0.029 (0.037)	-0.011 (0.009)
Has Kitchen	YES	YES	YES	YES
Has Hot Tub	YES	YES	YES	YES
Has Air Condition	YES	YES	YES	YES
Has Gym	YES	YES	YES	YES
Has Parking	YES	YES	YES	YES
Host Pool	YES	YES	YES	YES
Host Mid	YES	YES	YES	YES
Host Small	YES	YES	YES	YES
Bedroom	YES	YES	YES	YES
Bedroom^2	YES	YES	YES	YES
Cons	2.760 (0.095)***	2.294 (0.087)***	3.171 (0.083)***	0.332 (0.014)***
N	19,620	19,620	19,620	19,620
R <sup>2</sup>	0.1467	0.2163	0.1330	0.2001

Tabella 5: Regressioni logaritmiche con interazione

\*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

Il modello in forma logaritmica conferma i risultati ottenuti nel modello lineare per le variabili dipendenti, mantenendo la coerenza nelle relazioni tra le variabili. Tuttavia,

poiché il modello logaritmico fornisce una rappresentazione più appropriata per analizzare le variazioni relative piuttosto che quelle assolute, ci concentreremo sull'interpretazione delle interazioni significative che emergono in questa forma.

Il primo modello (M1) utilizza come variabile dipendente il logaritmo del numero di prenotazioni e presenta un coefficiente di determinazione R-squared overall = 0.1467 dunque, circa il 14,67% della variabilità nel numero di prenotazioni è spiegato dalle variabili indipendenti incluse nel modello.

La variabile che rappresenta il genere dell'host risulta con un coefficiente positivo e significativo ( $\beta = 0.161$ ,  $p < 0.10$ ), delineando che il genere maschile ha un impatto significativo e positivo sul numero di prenotazioni ricevute dagli host. Nello specifico, gli uomini risultano avere un numero di prenotazioni del 16.1% maggiore rispetto a un host donna a parità di altre condizioni e ciò conferma un evidente presenza di gender gap.

Analizzando le interazioni emerge che:

- L'interazione “Uomo  $\times$  Superhost” ( $\beta = -0.108$ ,  $p < 0.10$ ) suggerisce che, sebbene il badge di Superhost abbia un effetto positivo sul numero di prenotazioni ( $\beta = 0.992$ ,  $p < 0.01$ ) per tutti gli host, per gli host maschi il vantaggio derivante dal possesso di questo riconoscimento si riduce rispetto alle donne;
- L'interazione “Uomo  $\times$  Year 2018” ( $\beta = -0.143$ ,  $p < 0.05$ ) mostra che, nonostante nel 2018 ci sia stato un incremento generale del 15,4% nel numero di prenotazioni, per gli host uomini questo aumento è stato più contenuto.

Il Modello 2 (M2) utilizza come variabile dipendente il logaritmo del RevPAN. Il valore di  $R^2$  è pari al 21,63%, indicando una discreta capacità del modello di spiegare la variabilità del fatturato. La variabile indipendente che rappresenta il genere dell'host presenta un coefficiente positivo, ma il risultato non è statisticamente significativo, suggerendo che, in questo caso, il genere non ha un impatto rilevante sul ricavo per notte.

Analizzando le interazioni emerge che:

- L'anno 2018 risulta significativo e positivo ( $\beta = 0.199$ ,  $p < 0.01$ ), indicando un aumento generale dei ricavi nel tempo per tutti gli host. Tuttavia, l'interazione “Uomo  $\times$  Year 2018” ( $\beta = -0.132$ ,  $p < 0.05$ ) suggerisce che l'incremento dei ricavi nel 2018 è stato più contenuto per gli uomini rispetto alle donne.

Il terzo modello (M3) ha come variabile dipendente il logaritmo della tariffa giornaliera (ADR) e mostra un coefficiente di determinazione R-squared overall= 0.1330 indicando che circa il 13.30% della variabilità è spiegata dalle variabili indipendenti incluse. Anche in questo caso la variabile indipendente del genere dell'host risulta non avere un effetto significativo sulla tariffa giornaliera.

Le interazioni che risultano significative sono:

- La variabile di interazione “ Year2018 × Uomo ” ( $\beta = -0.171$ ,  $p < 0.10$ ) indica che l'effetto positivo dell'aumento della tariffa media giornaliera nel 2018 è meno marcato per gli host uomini rispetto alle donne;
- L'interazione “ Year 2022 × Uomo ” ( $\beta = -0.143$ ,  $p < 0.10$ ) è negativa e significativa, indicando che, sebbene la tariffa media giornaliera sia aumentata nel 2022 per tutti gli host, gli uomini hanno registrato una crescita inferiore rispetto alle donne;
- L'interazione “ Midscale × Uomo ” ( $\beta = 0.111$ ,  $p < 0.10$ ) suggerisce che, quando l'host è un uomo e l'alloggio appartiene alla fascia midscale, il prezzo medio per notte aumenta del 11,1% rispetto alle donne con alloggi della stessa categoria;
- L'interazione “ Upscale × Uomo ” ( $\beta = 0.162$ ,  $p < 0.01$ ) indica che, negli alloggi upscale, gli host uomini riescono a fissare una tariffa media giornaliera più elevata rispetto alle host donne.

Il quarto modello utilizza come variabile dipendente il logaritmo del tasso di occupazione e presenta un coefficiente di determinazione R-squared overall = 0.2001.

La variabile “ Uomo ” ha un effetto positivo e significativo ( $\beta = 0.023$ ,  $p < 0.10$ ) sul logaritmo del tasso di occupazione, evidenziando la tendenza per gli host uomini ad avere un tasso di occupazione maggiore rispetto agli host donne.

Valutando le variabili di interazione, si osserva che queste giocano un ruolo significativo nel modellare l'effetto della variabile principale sul tasso di occupazione:

- La variabile di interazione “ Uomo × Luxury ” ( $\beta = -0.020$ ,  $p = 0.10$ ) mostra che per gli host uomini all'aumentare del prezzo dell'appartamento diminuisce leggermente il tasso di occupazione rispetto alle host donne;

- La variabile di interazione “ Uomo × Upscale” ( $\beta = -0.020, p = 0.10$ ) indica che gli alloggi di fascia medio-alta hanno un effetto negativo ma, questa penalizzazione risulta meno accentuata per gli uomini rispetto alle donne.

## 4.7 Interaction plot

Di seguito sono presentati gli interaction plot, grafici che rappresentano le interazioni tra variabili e consentono di visualizzare in che modo la relazione tra una variabile indipendente e quella dipendente cambia in funzione dei valori assunti da una variabile moderatrice.

In particolare, si esamina l'effetto di due variabili moderatrici: la fascia di prezzo dell'alloggio e il possesso del badge di Superhost.

Le seguenti Figure forniscono dunque una rappresentazione grafica degli effetti moderatori significativi riscontrati in questo studio.

Il gender gap nelle prenotazioni è più evidente per gli alloggi di fascia Budget, Upscale e Midscale dove gli uomini ricevono più prenotazioni rispetto alle donne mentre, per fasce di prezzo Luxury, il genere non sembra influenzare le prenotazioni in modo significativo. Questo suggerisce che il genere può avere un impatto sulle performance, ma in modo differenziato a seconda della fascia di prezzo dell'alloggio.

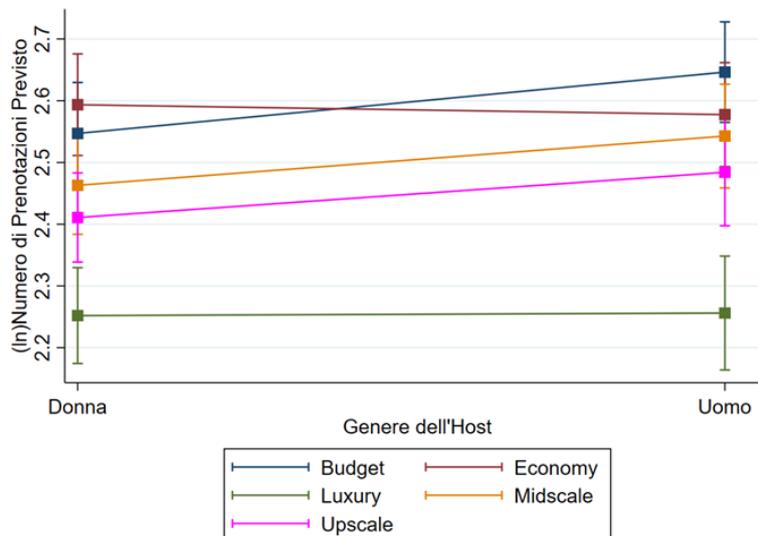


Figura 14: Diagramma di interazione per l'effetto moderatore del prezzo nella relazione tra genere dell'host e numero di prenotazioni

Il gender gap nei ricavi per notte tende a ridursi per le fasce di prezzo più alte, dove le differenze tra uomini e donne sono minime.

Negli alloggi Luxury e Upscale, il genere dell'host sembra avere un impatto minimo sui ricavi mentre per Budget e Midscale, invece, gli uomini tendono a ottenere ricavi più elevati rispetto alle donne.

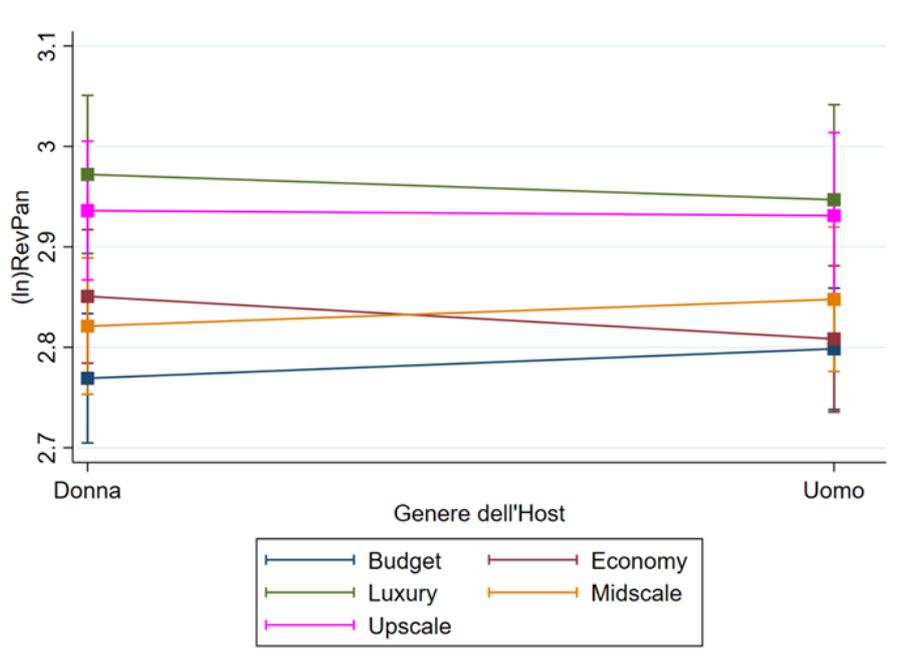


Figura 15: Diagramma di interazione per l'effetto moderatore del prezzo nella relazione tra genere dell'host e RevPan

La fascia di prezzo non modifica la relazione tra il genere dell'host e il tasso di occupazione, fatta eccezione per la categoria Budget, dove gli host uomini registrano un tasso di occupazione più elevato.

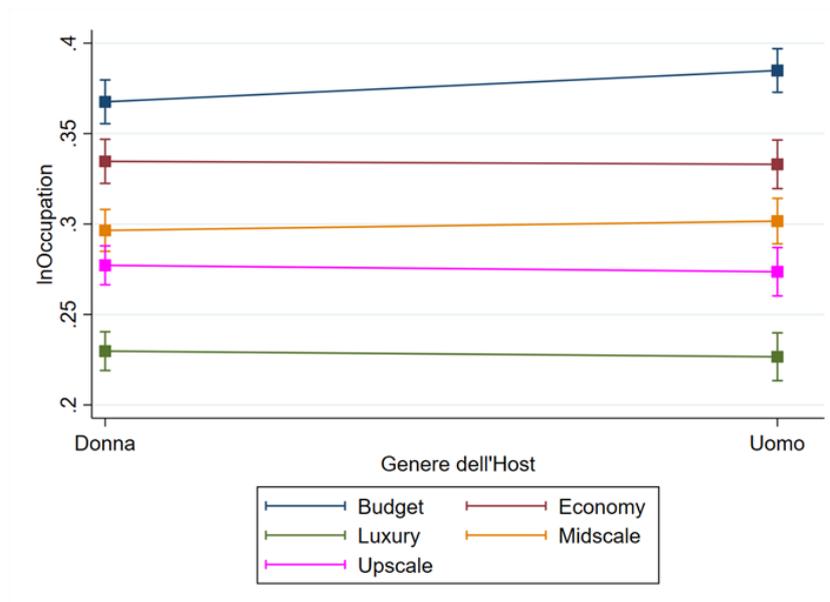


Figura 16: Diagramma di interazione per l'effetto moderatore del prezzo nella relazione tra genere dell'host e Tasso di occupazione

L'analisi dell'effetto moderatore del badge di Superhost suggerisce che esso contribuisce a ridurre le differenze di genere, portando le host donne a raggiungere livelli di prenotazione, RevPAR e tasso di occupazione simili a quelli degli uomini.

In assenza del badge di Superhost, gli host uomini registrano performance superiori rispetto alle donne. Tuttavia, con il badge, il divario di genere si attenua significativamente o si inverte leggermente a favore delle donne.

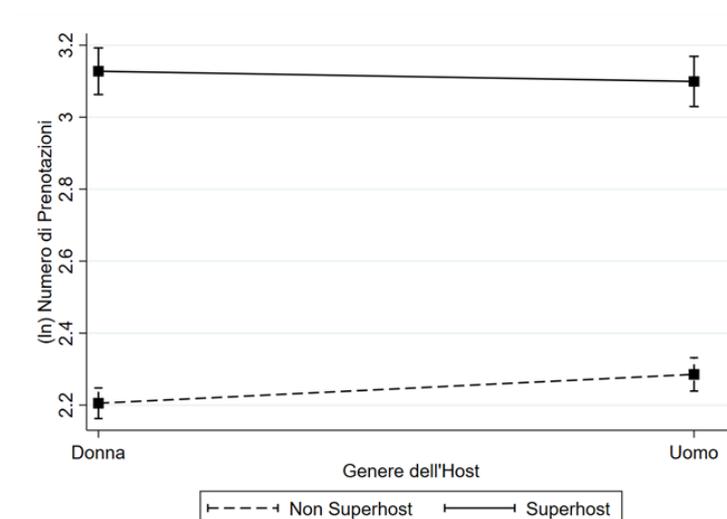


Figura 17: Diagramma di interazione per l'effetto moderatore del Superhost nella relazione tra Genere dell'host e Numero di prenotazioni

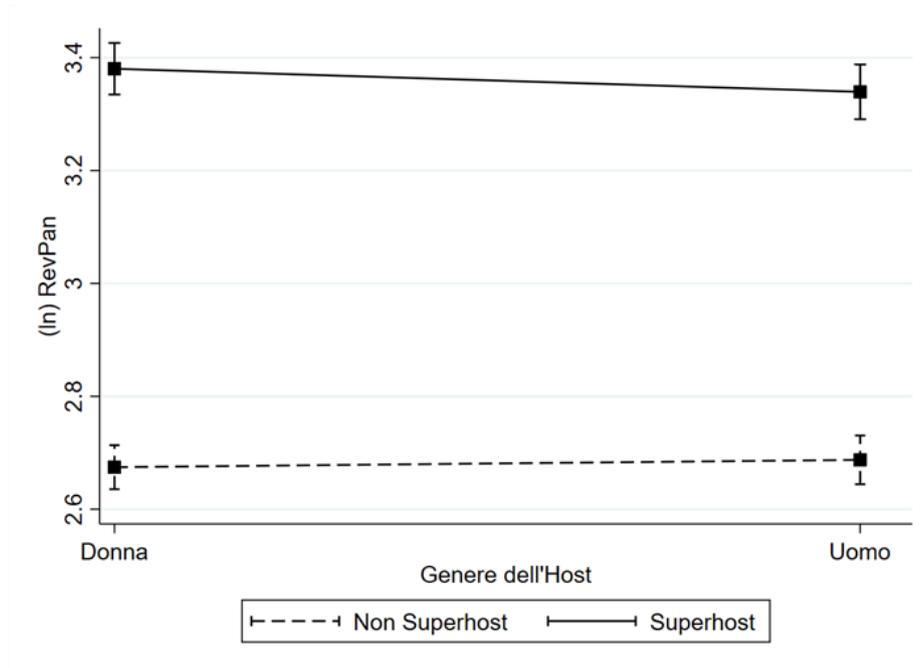


Figura 18: Diagramma di interazione per l'effetto moderatore del Superhost nella relazione tra genere dell' host e RevPan

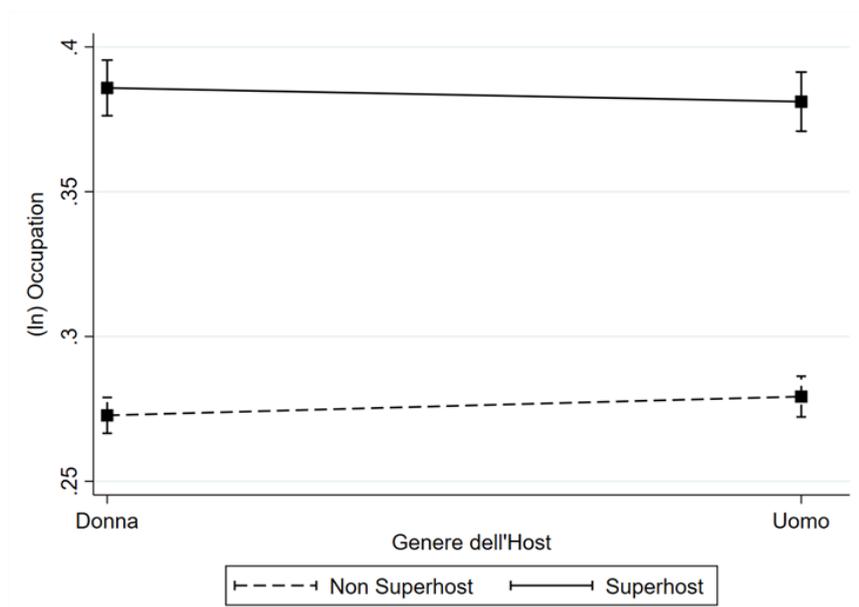


Figura 19: Diagramma di interazione per l'effetto moderatore del Superhost nella relazione tra genere dell' host e Occupation

# Conclusioni

In conclusione, il lavoro presentato, considerando i dati elaborati dal database di Airbnb a Torino, conferma la domanda di ricerca iniziale rilevando un significativo divario di genere nelle performance degli host.

In particolare, gli host uomini riescono ad ottenere mediamente un numero maggiore di prenotazioni, un ricavo per notte più elevato, un più alto tasso di occupazione ed in definitiva migliori guadagni, nonostante le host donne riescano ad ottenere un punteggio maggiore nelle recensioni.

Benché una parte della letteratura proponga che nelle piattaforme digitali il gender gap si riduca, in questo lavoro si trova evidenza della persistenza del divario di genere nei mercati peer-to-peer digitali.

E' però importante sottolineare che tenendo in considerazione le caratteristiche specifiche dei singoli alloggi e dei singoli host, la dimensione di questo gender gap si riduce.

Le interazioni tra le variabili infatti, hanno rivelato che il gender gap si manifesta in maniera diversa in alcune categorie di alloggio e in specifiche fasce di prezzo. Inoltre, il divario di genere tende a ridursi in contesti particolari, come nel caso dello stato di Superhost, che favorisce una significativa riduzione del gap grazie all'introduzione di un sistema di valutazione oggettivo, piuttosto che basato su percezioni soggettive.

Questo evidenzia che, implementando politiche attive in merito al contrasto della discriminazione, le piattaforme hanno la possibilità di ridurre il divario di genere.

In conclusione, il divario di genere non si manifesta in maniera uniforme, ma varia in base alla combinazione di caratteristiche specifiche dell'host e del tipo di alloggio.

# Implicazioni future

I risultati di questa ricerca offrono spunti significativi per un approfondimento del divario di genere nell'economia della condivisione.

Un possibile sviluppo futuro potrebbe riguardare l'estensione dell'analisi a un campione geografico più ampio, includendo diverse città o Paesi, al fine di comprendere se e in che modo il contesto locale influenzi il gender gap. Le differenze normative, culturali ed economiche potrebbero infatti incidere sulle dinamiche di discriminazione di genere nei mercati digitali.

Un altro aspetto da approfondire potrebbe essere quello di analizzare le strategie di pricing adottate dagli host. L'analisi del pricing è fondamentale per comprendere il gender gap sulle piattaforme digitali come Airbnb perché aiuta a distinguere tra scelte strategiche dell'host e pregiudizi della domanda, fornendo una visione più chiara delle cause della disparità di performance tra uomini e donne.

Studi passati hanno infatti evidenziato che i guadagni più bassi su piattaforme digitali come Airbnb non sono necessariamente attribuibili a una discriminazione di genere, ma potrebbero derivare da una maggiore avversione al rischio da parte delle donne, che tende a portarle a fissare prezzi più bassi.

Inoltre, sarebbe interessante approfondire il funzionamento degli algoritmi di visibilità, esplorando se Airbnb valuti la possibilità che i suoi algoritmi di ranking possano favorire involontariamente un genere rispetto all'altro, e prendere provvedimenti per correggere eventuali distorsioni, rendendo la piattaforma più equa e neutrale.

Infine, sarebbe utile esaminare più a fondo l'impatto delle politiche specifiche implementate da Airbnb per ridurre le disparità di genere. A livello normativo, i risultati di questo studio potrebbero sostenere l'adozione di politiche mirate a promuovere l'uguaglianza di genere all'interno della sharing economy.

In conclusione, sebbene questo studio abbia dimostrato la persistenza del gender gap su Airbnb, ulteriori ricerche potranno ampliare la comprensione del fenomeno e fornire indicazioni utili per l'implementazione di politiche più efficaci nel promuovere l'equità di genere nei mercati digitali.

# Bibliografia

Agovino, M., Bevilacqua, M., & Cerciello, M. (2023). Measuring female discrimination through language: a novel indicator and its effect on production efficiency in Italy. *International Journal of Manpower*, 44(9), 128-155.

Belleflamme, P., & Peitz, M. (2021). *The Economics of Platforms*. Cambridge University Press.

Carter, R. T., Johnson, V. E., Kirkinis, K., Roberson, K., Muchow, C., & Galgay, C. (2019). A meta-analytic review of racial discrimination: Relationships to health and culture. *Race and Social Problems*, 11, 15-32.

Cutillo, A., & Centra, M. (2017). Gender-based occupational choices and family responsibilities: The gender wage gap in Italy. *Feminist Economics*, 23(4), 1-31.

Davidson, A., & Gleim, M. R. (2022). The gender earnings gap in sharing economy services: The role of price, number of stays, and guests accommodated on Airbnb. *Journal of Marketing Theory and Practice*, 31(4), 490–501.

Ert, E., & Fleischer, A. (2020). What do Airbnb hosts reveal by posting photographs online and how does it affect their perceived trustworthiness?. *Psychology & Marketing*, 37(5), 630-640.

Ert, E., Fleischer, A., & Kopolovich, D. (2024). Gender earning gap on digital platforms: The Airbnb case. *Tourism Economics*, 13548166241263877.

European Institute for Gender Equality (2024). European Institute for Gender Equality.

Eurostat (2024) Gender pay gap statistics

Furlotti, K., Mazza, T., Tibiletti, V., & Triani, S. (2019). Women in top positions on boards of directors: Gender policies disclosed in Italian sustainability reporting. *Corporate Social Responsibility and Environmental Management*, 26(1), 57-70.

Istat. (2017). Piano di Uguaglianza di Genere 2024-2026. Istituto Nazionale di Statistica.

Marchenko, A. (2019). The impact of host race and gender on prices on Airbnb. *Journal of Housing Economics*, 46, 101635.

- Möhlmann, M. (2015). Collaborative consumption: determinants of satisfaction and the likelihood of using a sharing economy option again. *Journal of consumer behaviour*, 14(3), 193-207.
- Pena-Boquete, Y., De Stefanis, S., & Fernandez-Grela, M. (2010). The distribution of gender wage discrimination in Italy and Spain: a comparison using the ECHP. *International Journal of Manpower*, 31(2), 109-137.
- Quintano, C., Castellano, R., & Rocca, A. (2010). Male-female discrimination: an analysis of gender gap and its determinants. *Statistica*, 70(2), 171-190.
- Ritter, M., & Schanz, H. (2019). The sharing economy: A comprehensive business model framework. *Journal of cleaner production*, 213, 320-331.
- Severini, F., Felici, F., Ferracuti, N., Pretaroli, R., & Socci, C. (2019). Gender policy and female employment: a CGE model for Italy. *Economic Systems Research*, 31(1), 92-113.
- World Economic Forum. (2024). Global Gender Gap Report 2024.
- Worthen, M. G., Lingiardi, V., & Caristo, C. (2017). The roles of politics, feminism, and religion in attitudes toward LGBT individuals: A cross-cultural study of college students in the USA, Italy, and Spain. *Sexuality Research and Social Policy*, 14, 241-258.
- Zhu, X., & Liu, K. (2021). A systematic review and future directions of the sharing economy: business models, operational insights and environment-based utilities. *Journal of cleaner production*, 290, 125209.
- Zervas, G., Proserpio, D., & Byers, J. W. (2017). The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry.

## Indice delle Figure

Figura 1: World Economic Forum, Global Gender Gap 2024, Gender Equality Index by subindex.....	12
Figura 2: World Economic Forum,Global Gender Gap 2024, Evolution of the Global Gender Gap Index and subindexes over time.....	12
Figura 3: World Economic Forum, Global Gender Gap 2024,Regional result for subindexes .....	13
Figura 4: World Economic Forum, Global Gender Gap 2024, Italy .....	14
Figura 5: Global Gender Gap Index Indicators,2024, Italia .....	16
Figura 6: Eurostat, Gender Pay gap based on hourly gross salary .....	17
Figura 7: Eurostat, Salary gap man vs woman .....	18
Figura 8: Codice Python Estrazione Nomi parte 1 .....	30
Figura 9: Codice Python Estrazione Nomi parte 2 .....	30
Figura 10: Codice Python per ricavare il genere dal nome parte 1 .....	31
Figura 11: Codice Python per ricavare il genere dal nome parte 2 .....	31
Figura 12: Modello concettuale della moderazione semplice .....	64
Figura 13:Effetto di interazione.....	66
Figura 14: Diagramma di interazione per l'effetto moderatore del prezzo nella relazione tra genere dell' host e numero di prenotazioni.....	81
Figura 15: Diagramma di interazione per l'effetto moderatore del prezzo nella relazione tra genere dell'host e RevPan.....	82
Figura 16:Diagramma di interazione per l'effetto moderatore del prezzo nella relazione tra genere dell'host e Tasso di occupazione .....	83
Figura 17: Diagramma di interazione per l'effetto moderatore del Superhost nella relazione tra Genere dell'host e Numero di prenotazioni .....	83
Figura 18: Diagramma di interazione per l'effetto moderatore del Superhost nella relazione tra genere dell' host e RevPan .....	84
Figura 19: Diagramma di interazione per l'effetto moderatore del Superhost nella relazione tra genere dell' host e Occupation .....	84

## **Indice delle Tabelle**

Tabella 1: Frequenza assoluta, relativa e cumulata del numero di proprietà suddivise per Listing Tipe.....	42
Tabella 2 : Regressione lineare .....	56
Tabella 3: Regressione logaritmica.....	63
Tabella 4 : Regressione lineare con interazione .....	70
Tabella 5: Regressioni logaritmiche con interazione.....	78

## **Indice dei Grafici**

Grafico 1: Distribuzione degli host per genere.....	33
Grafico 2: Distribuzione degli annunci per genere.....	34
Grafico 3: Prenotazione medie annuali per genere.....	35
Grafico 4: Andamento della Revenue annuale in base al genere dell' host.....	36
Grafico 5: Overall Rating per genere .....	37
Grafico 6: Distribuzione dei Superhost per genere .....	38
Grafico 7:Distribuzione dei servizi offerti per genere dell'host.....	39
Grafico 8: Divisione per fascia di prezzo in base al genere dell'host .....	40
Grafico 9: Andamento dell'ADR per genere .....	41
Grafico 10: Tipologia di annuncio per genere .....	42
Grafico 11: Media dei giorni prenotati per genere .....	43
Grafico 12: Media dei giorni disponibili per genere .....	44
Grafico 13: Tasso di Occupazione per Genere dell' host.....	45
Grafico 14: Distribuzione del numero di camere per host suddivisa per genere.....	46
Grafico 15: Distribuzione del ricavo medio per notte suddiviso per genere .....	47