



**Politecnico
di Torino**

POLITECNICO DI TORINO

Collegio di ingegneria Gestionale e della Produzione – Classe LM/31

Corso di Laurea Magistrale in Ingegneria Gestionale

A.a. 2024/2025

Sessione di Laurea Aprile 2025

Dinamiche di prezzo e competitività territoriale nel mercato degli affitti di breve termine in Val Pellice

Relatori:

Dott. Luigi Buzzacchi
Dott. Francesco Luigi Milone

Candidato:

Simone Leo

TITOLO

Dinamiche di prezzo e competitività territoriale nel mercato
degli affitti di breve termine in Val Pellice

INDICE

TITOLO	2
INDICE.....	3
INDICE DELLE FIGURE	5
INDICE DELLE TABELLE.....	6
INDICE DEI GRAFICI.....	7
1. INTRODUZIONE.....	8
2. STORIA AIRBNB.....	10
3. LETTERATURA	18
Effects of location on Airbnb apartment pricing in Málaga (2020)	18
Housing market segmentation and hedonic prediction accuracy (2002).....	19
Real Estate Market Segmentation: Hotels as Exemplar (2018)	20
Hedonic pricing and the sharing economy: how profile characteristics affect Airbnb accommodation prices in Barcelona, Madrid, and Seville.....	22
Exploring changes in guest preferences for Airbnb accommodation with different levels of sharing and prices: Using structural topic model.....	22
A Review of the Problems and Some Proposals Tested on Bordeaux City Data	24
Comparing Airbnb and traditional accommodation experiences using text-mining methods – the hedonic and utilitarian values framework	25
Price determinants of Airbnb accommodation in the French cities.....	26
Airbnb’s disruption of the housing structure in London	27
4. APPROFONDIMENTO SULLE VARIABILI.....	28
Rappresentazione database	28
Data exploration.....	33
Pulizia database.....	39
5. ANALISI VALUTAZIONI	47
Punteggio valutazioni	47
6. ANALISI REGRESSIONE	50
Principi teorici.....	50
Variabili.....	51
Regressione	53
Calcolo residui	68

Statistiche descrittive:	69
7. MERCATO D'INFLUENZA DI UNA PROPRIETÀ	74
Dati spaziali	74
Pesi spaziali.....	79
Autocorrelazione spaziale	82
8. CONCLUSIONI	98
9. VINCOLI DELLO STUDIO	101
Bibliografia.....	102
Sitografia	104

INDICE DELLE FIGURE

Figura 1 Analisi AC e PAC dei 4 lag.....	71
Figura 2 Rappresentazione geografica delle località presenti in Val Pellice.....	74
Figura 3 Rappresentazione delle località presenti a febbraio 2023.	75
Figura 4 Matrice dei pesi basata sulla distanza (Caso 1).....	81
Figura 5 Matrice dei pesi KNN (Caso 2).....	82
Figura 6 Grafico di dispersione di Moran per i prezzi (Caso 1).....	84
Figura 7 Grafico di dispersione di Moran dei residui (Caso 1).	84
Figura 8 Grafico di dispersione di Moran dei prezzi (Caso 2).	84
Figura 9 Grafico di dispersione di Moran dei residui (Caso 2).	85
Figura 10 Mappa del cluster e di significatività di PriceUSD, Caso 1.	92
Figura 11 Cluster and spatial outliers high-high, Caso 1.....	93
Figura 12 Cluster and spatial outliers low-low, Caso 1.....	94
Figura 13. Outliers spaziali.....	94
Figura 14 Applicazione LISA di PriceUSD, con $p=0,01$	95
Figura 15 Mappa dei cluster, Bonferroni ($p<0.000119$).	95
Figura 16 Mappa dei cluster, $p<0.001$	96

INDICE DELLE TABELLE

Tabella 1 Descrizione variabili.	34
Tabella 2 Tipologie di proprietà.	45
Tabella 3 Statistiche recensioni.	49
Tabella 4 Overview variabili.	52
Tabella 5 Risultati regressione con PriceUSD.....	61
Tabella 6 Risultati regressione con lnPriceUSD.....	65
Tabella 7 Sintesi numerica dei residui.	69
Tabella 8 Rappresentazione property apposita per GeoDa.....	76

INDICE DEI GRAFICI

Grafico 1 Distribuzione delle strutture disponibili per ciascun comune in Val Pellice.....	41
Grafico 2 Suddivisione percentuale della variabile Price Tier.	42
Grafico 3 Boxplot del prezzo pre e post pulizia del database.....	43
Grafico 4 Distribuzione del prezzo a notte.	43
Grafico 5 Tipologie di Airbnb presenti nel territorio.	44
Grafico 6 Andamenti delle recensioni nelle proprietà di Airbnb.....	48
Grafico 7 Matrice di correlazione delle variabili.....	54
Grafico 8 Correlazione specifica delle variabili della proprietà.	55
Grafico 9 Scatter plot rappresentante la relazione tra PriceUSD e OverallRating. .	56
Grafico 10 Distribuzioni dei residui rispetto alle curve delle normali.	70
Grafico 11 Boxplot dei residui per comune.....	71
Grafico 12 Boxplot dei residui per gruppi di comuni.	72
Grafico 13 Grafico di connettività, Caso 1.	81
Grafico 14 Distribuzione di riferimento del prezzo, Caso 1.....	86
Grafico 15 Distribuzione di riferimento di e_{it} ed u_i , Caso 1.	86
Grafico 16 Distribuzione di riferimento del prezzo, Caso 2.....	87
Grafico 17 Distribuzione di riferimento di e_{it} ed u_i , Caso 2.	87
Grafico 18 Dispersione di Moran di PriceUSD, Caso 1.	88
Grafico 19 Visuale Linear Smoother dei residui e_{it} , Caso 1.	89
Grafico 20 Dynamic brushing PriceUSD di un primo campione, Caso2.	90
Grafico 21 Dynamic brushing PriceUSD di un secondo campione, Caso2.....	90
Grafico 22 Correlogramma spaziale PriceUSD, Caso 1.....	91

1. INTRODUZIONE

La presente proposta di tesi elabora e analizza la formazione dei prezzi nel mercato degli affitti di breve termine delle località presenti all'interno dei comuni della Val Pellice, mettendo in evidenza le dinamiche spaziali ed economiche che la caratterizzano. Analizzare a monte le variabili che determinano il prezzo finale è fondamentale, vista la continua evoluzione della piattaforma.

L'elaborato racconta inizialmente l'origine di Airbnb nel 2008, piattaforma considerata protagonista dell'economia circolare internazionale. Vengono spiegati gli ostacoli affrontati, il metodo attraverso cui prende forma la credibilità degli host con le recensioni e il modo in cui viene analizzata nel suo complesso l'intera crescita della piattaforma.

Nel capitolo successivo è stata svolta una ricerca di articoli e studi aventi obiettivi simili a quelli trattati nella parte di analisi successiva, costituendo la letteratura esistente del lavoro. Attraverso questo approfondimento, è stato possibile comprendere in maniera più efficiente le metodologie da utilizzare e in un secondo momento confrontare i risultati ricavati con quelli ottenuti dagli studi già esistenti. L'interpretazione dei dati è stata quindi contestualizzata in maniera più solida.

Lo scopo del lavoro è duplice. Il primo è identificare tutte le variabili indipendenti che condizionano significativamente la variabile dipendente, ovvero il prezzo. Il secondo è costituito dall'utilizzo di due strumenti, l'indice di Moran e LISA (Local Indicators of Spatial Association), i quali permettono di rilevare la presenza di autocorrelazione spaziale tra le osservazioni in questione. Il fine di questi indici è di individuare aree in cui si manifestano dei comportamenti simili di proprietà vicine.

Il primo obiettivo è stato conseguito mediante una regressione edonica. Il software Stata ha permesso lo svolgimento di questa fase, dove si è ricavata la relazione tra il prezzo degli affitti e i fattori indipendenti appartenenti al database. Il secondo obiettivo è stato raggiunto con il software GeoDa, il quale ha fornito una risposta sul quesito costituito dalla presenza dei cluster spaziali.

Da come si evince dai risultati, il valore della variabilità spiegata (R-Squared) aumenta al crescere dell'inclusione di variabili indipendenti significative nelle successive iterazioni di regressione, le quali vengono analizzate nel dettaglio. Per quanto riguarda l'analisi condotta sui residui del prezzo, è stata fondamentale per raggiungere l'obiettivo dei cluster spaziali. In particolare, i prezzi evidenziano una buona autocorrelazione spaziali e sottolineano la presenza di osservazioni con caratteristiche simili localizzate in aree vicine. I residui invece mostrano un'autocorrelazione spaziale assente, dove la variabilità residua è dipendente da decisione prese direttamente dagli host in maniera casuale nel processo di identificazione del prezzo degli affitti.

Le conclusioni propongono, qualora gli host vogliano definire strategie di prezzo adatte al mercato in cui risiedono, uno studio in profondità delle dinamiche di prezzo come soluzione al fine situarsi una posizione strategica all'interno del mercato di riferimento.

2. STORIA AIRBNB

1) Origine

Airbnb nasce da due ragazzi, Brian Chesky e Joe Gebbia, i quali nel 2007, appena laureati e vivendo a San Francisco, una delle città più costose della California, ebbero un'idea per guadagnare dei soldi approfittando di un evento che si sarebbe tenuto in quei giorni, proprio nella loro città, il quale aveva attirato moltissime persone che facevano fatica a trovare stanze, alberghi, posti dove stare.

I due giovani decisero quindi di destinare parte del proprio appartamento per accogliere coloro che avevano necessità di trovare da dormire in quei giorni, facendo pagare un affitto, e inoltre decisero di aggiungere un altro servizio compreso nel prezzo, ovvero quello della colazione. Per pubblicizzare questa loro proposta, chiamata "AirBed & Breakfast" (utilizzavano dei materassi ad aria per far dormire le persone), crearono un sito web apposito che portò loro i primi clienti, i quali usufruirono del loro servizio.

Il fatto che, inizialmente, non avessero investito in letti veri e propri evidenzia che questa fu un'idea concepita da due persone, le quali non avevano immaginato che questa loro proposta potesse fruttare loro un guadagno stabile e che soprattutto non avevano l'intenzione di creare un'azienda a livello globale.

Successivamente, notando la grande richiesta di sistemazioni dove dormire a poco prezzo e il successo della loro iniziativa, nel 2008, collaborarono con un altro giovane ragazzo, Nathan Blecharczyk, fondando ufficialmente l'azienda "AirBed & Breakfast".

Crearono una piattaforma digitale sulla quale ognuno poteva decidere di caricare una sistemazione da affittare a coloro, che viaggiando, avevano la necessità di trovare un luogo nel quale soggiornare per brevi periodi ad un prezzo più conveniente rispetto ai già noti hotel.

I primi tentativi di sponsorizzare la propria iniziativa attraverso delle partnership con degli eventi, che sapevano avrebbero attirato una mole di visitatori utile alla crescita della loro piattaforma, non andarono a buon fine. Questo poiché gli investitori non erano convinti che queste sistemazioni alternative, apparentemente poco accoglienti e di qualità, potessero funzionare.

Il vero cambiamento per la startup arrivò dopo una trovata geniale di pubblicizzazione per la campagna presidenziale che si è tenuta nello stesso anno, pensata per la raccolta di fondi per la loro piattaforma.

Un'altra mossa giusta fatta dai fondatori, che fu cruciale per la crescita dell'azienda, fu quella di unirsi al programma Y Combinator, che si occupava di accelerare il processo di crescita delle startup, attraverso dei fondi e soprattutto un supporto a livello di strategie di miglioramento del servizio proposto.

Venne cambiato il nome della piattaforma in "Airbnb", si focalizzarono su New York ed ebbero degli incontri con chi usufruiva del sito per ricevere dei feedback affinché potessero continuare a potenziare e perfezionare la loro proposta sul mercato.

Grazie a questi cambiamenti la piattaforma cominciò a crescere in maniera esponenziale, dal punto di vista delle prenotazioni e di conseguenza del ricavo.

Con il tempo "Airbnb" è cambiata molto, ha anche aggiunto possibilità di sistemazioni diverse, come l'affitto di intere case. Ad oggi però possiamo dire che ha portato ad una rivoluzione del modo di concepire il soggiorno, offrendo a chi viaggia l'opportunità di farlo in maniera più economica e rendendo lo stesso viaggiare accessibile a più persone.

2) Crescita

Come menzionato in precedenza ciò che ha permesso la crescita di Airbnb è stato un riadattamento continuo della piattaforma sulla base delle esigenze e dei riscontri avuti da chi ne usufruiva, per i cofondatori era infatti importante che gli utenti riuscissero ad utilizzare tale servizio senza avere difficoltà e proprio l'intuitività e la semplicità nel farne uso portarono ad un incremento dei fruitori.

Inoltre, per promuovere la loro azienda hanno sfruttato i social e la viralità che questi offrono, chiedendo ai loro utenti di raccontare le proprie esperienze positive, abbinato poi alla creazione di campagne pubblicitarie accattivanti e creative che furono molto efficaci.

Airbnb non si è limitata al territorio degli Stati Uniti, è riuscita ad espandersi a livello internazionale, adattando la piattaforma alle esigenze dei vari territori, traducendola nelle diverse lingue e tenendo in considerazione gli aspetti culturali di ogni paese, come le festività. Questa diffusione è stata resa possibile dai rapporti che l'azienda ha creato con i territori locali, presentandosi come supporto per la loro economia e allineandosi dal punto di vista normativo.

Per riuscire ad allargare la base d'utenza del servizio e adattarsi alle tendenze dei viaggi che cambiano nel tempo, Airbnb ha inserito nuove funzionalità come la possibilità di soggiornare per un periodo più lungo di tempo e l'inserimento di sistemazioni cosiddette di "lusso".

Inoltre, ha puntato sul concetto di comunità e di unione tra coloro che usufruiscono della piattaforma, pensando a degli eventi che permettessero agli host di conoscersi e confrontarsi tra di loro. Riuscì così a dare molta importanza alle recensioni e alle esperienze degli ospiti, creando un senso di affidabilità della piattaforma sia per chi ne fruisce quotidianamente che per chi se ne serve per la prima volta.

Inoltre, ha introdotto la possibilità di ricevere premi, quali sconti e benefici, per gli utenti già esistenti che invitavano altri membri ad iscriversi alla piattaforma e per i

nuovi registrati, così da incentivare ancora di più la conoscenza e l'utilizzo di tale servizio.

Tutti questi cambiamenti apportati sono basati su una rilevazione di dati concreti avvenuta nel tempo.

Un ultimo aspetto chiave sul quale la piattaforma ha dovuto lavorare in maniera molto efficace è stata la gestione delle difficoltà a livello normativo riscontrate in diverse città, le quali avevano introdotto delle regolamentazioni per controllare gli affitti a breve termine che risultavano avere un forte impatto sulla disponibilità di locazioni a lungo termine. Airbnb ha collaborato con le amministrazioni locali, cercando di porre delle limitazioni al numero di notti possibili da affittare e introducendo delle tassazioni, andando così incontro alle esigenze dei vari territori e permettendo di mantenere la loro influenza su quelle aree.

3) Modello di commissioni

Airbnb richiede che vengano pagate delle commissioni per l'utilizzo della piattaforma sia da parte degli utenti che mettono a disposizione i propri alloggi e sistemazioni da affittare agli altri utenti, che a coloro che ne usufruiscono ed effettuano le prenotazioni.

Gli host pagano una commissione che è standard e varia tra il 3% e il 5% del totale del prezzo della prenotazione. Questa minima variazione dipende dalle condizioni che l'host offre. Per esempio coloro che permettono la prenotazione mediante la cancellazione gratuita, pagheranno una commissione leggermente maggiore.

Tale entrata permette all'azienda di gestire i costi legati al mantenimento e riparazione del sito e l'assimilazione dei pagamenti.

Gli ospiti invece pagano al servizio una commissione variabile, in base al costo della prenotazione e alla durata del soggiorno, questa può andare da un minimo del 5% del totale della riservazione fatta ad un massimo del 20%.

Tali percentuali sono inversamente proporzionali rispetto al costo e alla durata del soggiorno, ovvero più il soggiorno è breve e il prezzo da pagare è basso, più la commissione è alta, viceversa più il soggiorno è di lunga durata e il prezzo è alto, più la quota trattenuta dalla piattaforma scende.

4) Difficoltà affrontate

Durante questi anni di crescita dell'azienda vi sono stati anche diversi momenti di difficoltà nei quali Airbnb ha dovuto affrontare situazioni problematiche sotto numerosi aspetti.

A livello normativo abbiamo già visto le restrizioni imposte sui limiti di tempo di affitto delle locazioni da parte di una serie di città, di cui le più esemplari furono Parigi, New York e Barcellona.

Un altro tipo di obblighi che gli host hanno in alcune città è quello di richiedere e possedere delle licenze particolari, oppure registrarsi alle autorità locali affinché il servizio che offrono sia legale. Questo naturalmente crea dei processi lunghi a livello amministrativo che è necessario percorrere, per i quali i proprietari di immobili non vogliono perdere del tempo limitando così la possibilità per l'azienda di ottenere nuovi host.

Un ulteriore problema riscontrato è quello del pagamento delle tasse, che nonostante sia stato regolamentato dalla piattaforma, anche attraverso sistemi di pagamento automatico, non sempre viene sostenuto in maniera regolare.

La nascita di tale proposta ha avuto un forte impatto sul mercato immobiliare andando ad aumentare di conseguenza la crisi abitativa.

La presenza di quest'azienda ha diminuito il numero di locazioni disponibili per affitti di lungo termine, con un conseguente aumento della domanda rispetto all'offerta di queste sistemazioni, portando ad un incremento dei prezzi.

Inoltre, la possibilità di affittare a breve termine è una fonte di guadagno a tutti gli effetti per coloro che possiedono degli immobili, per questo motivo molte persone hanno deciso di investire sull'acquisto di proprietà che verranno poi destinate solo ed esclusivamente a tale scopo, andando ad accrescere ancora di più la crisi abitativa già presente.

Connesso a ciò, nei luoghi e nelle città con maggiore affluenza di visitatori, si vengono a creare delle vere e proprie aree turistiche dove prima vivevano gli abitanti locali, andando a perdere e distruggere il senso di comunità di quella zona, oltre al disturbo causato dagli ospiti e ai comportamenti irresponsabili che vengono segnalati dai residenti. Si tratta del fenomeno del "overtourism", al quale Airbnb ha dato un forte contributo.

Queste situazioni hanno ridotto la qualità della vita degli abitanti del posto, portando a tensioni e dissenso contro l'azienda e più in generale a tale sistema degli affitti a breve termine.

Vi sono poi state problematiche connesse alla concorrenza con il settore alberghiero, che Airbnb ha dovuto affrontare in passato e affronta ancora tutt'ora. Da un lato ha sicuramente causato una diminuzione delle entrate degli hotel che si sono dunque dovuti riorganizzare per cercare di competere sul mercato, dall'altro è stato in grado di mediare tra gli host e gli ospiti

Proprio per tale non conformità delle norme di sicurezza da parte degli host si sono trovati a dover affrontare lamentele degli ospiti a causa di incidenti, che li hanno portati in situazioni di pericolo senza avere i giusti mezzi o strumenti utili.

In quanto piattaforma di intermediazione, non è in grado di controllare che ogni proprietà rispetti tutte le regole. La dimostrazione sono delle circostanze nelle quali gli ospiti hanno segnalato casi di violazione della privacy, come la presenza di telecamere in casa. Ciò va a influenzare negativamente la fiducia di coloro che

usufruiscono del servizio, nonostante Airbnb sia sempre intervenuta a supporto dei suoi utenti.

Infine, l'ultima grande difficoltà che Airbnb si è trovata ad affrontare è stato il periodo di pandemia da COVID-19 che ha in qualche modo rilevato uno dei punti di debolezza di tale azienda, poiché naturalmente se viene a mancare la possibilità di viaggiare, il business degli affitti a breve termine si riduce drasticamente. La piattaforma per mantenere soprattutto il supporto degli host ha cercato di cambiare il tipo di offerta, andando a proporre anche la possibilità di soggiorno a lungo termine e inserendo nuove norme sanitarie.

5) Generazione della credibilità dell'host

Per aumentare la credibilità dei proprietari che affittano le proprie case, affinché gli ospiti possano prenotare conoscendo la realtà e avendo maggiori certezze, Airbnb si basa su un sistema di recensioni, scritte da parte di entrambi i tipi di utenti sull'esperienza avuta.

Tali recensioni vengono pubblicate solo successivamente al completamento di esse, da parte di entrambi i soggetti coinvolti, affinché venga assicurata imparzialità e autenticità dei commenti.

Ogni giudizio dato include dettagli su diversi aspetti del soggiorno: pulizia, comunicazione e disponibilità dell'host, precisione dell'annuncio, posizione dell'alloggio.

Tutto ciò è utile ovviamente agli ospiti per prenotare in base alle proprie aspettative e preferenze, ma è uno stimolo anche per gli host ad offrire sistemazioni di sempre maggior qualità.

Airbnb, per migliorare ulteriormente e rendere più affidabile l'esperienza sia di host che di ospiti, ha previsto dei programmi che garantiscano loro una certa sicurezza e un supporto assicurativo.

Gli host possono affidarsi all'aiuto dell'azienda in caso di danni alle proprietà, la quale offre loro una copertura assicurativa in tali casi, gli ospiti possono ricevere rimborsi o alternative nel caso di problemi con le sistemazioni prenotate, come alloggi diversi da quella che era la descrizione o la presentazione tramite foto nell'annuncio o condizioni di inabitabilità.

Tale sistema permette di aumentare l'affidabilità di Airbnb e la fiducia degli utenti nell'utilizzo della piattaforma.

Infine, la piattaforma si assicura della veridicità delle proprietà proposte grazie alla possibilità d'identificazione degli utenti tramite documenti ufficiali, collegamento con i profili social, descrizioni personali e foto veritiere. In generale i profili degli host e gli annunci pubblicati prevedono l'inserimento di diverse informazioni, foto, descrizioni dettagliate dell'alloggio, al fine di aumentare la trasparenza e la realtà di ciò che viene proposto per rendere l'esperienza complessivamente più positiva e prevedibile.

3. LETTERATURA

In questo capitolo vengono raccontati alcuni articoli aventi come focus il prezzo e come fine ultimo dei target simili a quelli trattati nella parte analitica successiva, in modo da essere confrontati con gli esiti ottenuti.

Effects of location on Airbnb apartment pricing in Málaga (2020)

Il modello edonico spaziale è in grado di mostrare il ragionamento di fissaggio del prezzo effettuato a monte dagli host che possiedono le proprietà. Questo tema è ciò che l'articolo evidenzia per gli appartamenti turistici (Airbnb) di Malaga, mediante amenities di vario tipo (localizzazione, ospiti e alcune caratteristiche della struttura).

L'obiettivo principale dichiarato nello studio è l'analisi della dipendenza diretta tra prezzo e localizzazione. Altri scopi di carattere inferiore sono la quantificazione dell'effetto causato da determinate variabili di carattere ambientale, quali la camminabilità, l'accessibilità ai servizi della zona e il rumore del traffico. Per finire, capire, qualora esistessero degli host aventi appartamenti in zone adiacenti, se si possa verificare un effetto di ricaduta.

La metodologia utilizzata è quindi costituita dal modello edonico, il quale permette agli host e gli ospiti la determinazione del valore del prezzo insito alle amenities che condizionano il prezzo finale. Il numero di appartamenti utilizzati in questo studio è di 2967, i quali risalgono all'anno 2017.

I fattori selezionati sono stati suddivisi in più classi, come "Servizi", "Regole", "Ambiente", "Accessibilità", "Dimensioni", "Host" e "Sociologia".

Inoltre, in questo caso di studio, è stata svolta una correlazione spaziale tra le osservazioni, mediante un modello edonico spaziale. I dati ricavati dalla piattaforma e dalle variabili di localizzazione sono stati il mezzo attraverso cui il modello econometrico spaziale è riuscito a fornire una motivazione valida dei prezzi delle località. Per comprendere al meglio gli effetti spaziali sono state

utilizzate due specifiche standard: SAR (metodo di ritardo spaziale) e SEM (metodo di errore spaziale). Nonostante la specificazione fornita dalla prima sia maggiormente dettagliata, gli esiti ricavati sono simili per entrambi i metodi.

Da come si evince dai risultati esposti dallo studio, le variabili di carattere geografico che identificano ciascuna osservazione causano effetti positivi e/o negativi significativi sul prezzo degli appartamenti. In particolare, i fattori che possiedono un effetto positivo significativo sul prezzo sono la camminabilità nella zona circostante alla struttura, la vicinanza al centro di Malaga e alla costa. È stato riscontrato un effetto negativo significativo sul prezzo, ovvero il rumore. Come ci si poteva aspettare, gli ospiti preferiscono localizzarsi in aree più tranquille della città.

Housing market segmentation and hedonic prediction accuracy (2002)

Lo studio su cui hanno lavorato Goodman e Thibodeau tratta le stime edoniche appartenenti al valore della proprietà. Nello specifico, gli studiosi francesi vogliono verificare l'ipotesi secondo cui definire i sottomercati seguendo una metodologia da loro inventata, evidenzia un miglioramento delle stime edoniche. I sottomercati in questione sono stati formati mediante due vie. La prima è costituita dai tratti di censimento, la seconda dai distretti dei codici postali. Per questo test si sono usufruite 28000 transazioni di località unifamiliari nel periodo compreso tra il 1995 e il 1997. Le stime edoniche comprendono la maggior parte dell'area occupata dalla contea di Dallas. Gli esiti empirici mostrano significativi guadagni, e di conseguenza ottime previsioni edoniche, ricavati tramite disaggregazione spaziale.

Gli autori dell'articolo intendono calcolare i parametri di un modello gerarchico per i prezzi delle case e di individuare i perimetri dei sottomercati in questione. Propongono quattro soluzioni differenti per definire i sottomercati immobiliari della zona interessata: usufruire dei tratti di censimento, dei distretti dei codici postali, non utilizzare nessuna disaggregazione spaziale e infine la loro metodologia.

Il dettaglio di ciascun sottomercato è stato eseguito appositamente. In prima battuta gli studiosi costruiscono un campione fondandosi su aspettative esistenti. Successivamente ogni sottosettore viene trattato individualmente per lo svolgimento di una regressione edonica, mentre la verifica della riduzione della somma dei residui quadratici significativa è svolta mediante i test F. Se la riduzione si dimostra significativa allora i sottosettori vengono considerati appropriati. Qualora il prezzo per unità degli appartamenti evidenzia cambiamenti spaziali, la presenza di sottosettori è confermata.

I test eseguiti al fine di inquadrare la qualità dei sottosettori sono i seguenti: la riduzione dell'errore quadratico, il test F per i sottosettori e un test non annidato utilizzando il test J.

La procedura di definizione dei sottomercati esposta nello studio ha previsto 55 sottomercati dai distretti postali e 82 sottomercati dalle sezioni censuarie.

Dagli esiti si evince un'accuratezza maggiore delle stime del valore di mercato tra i sottomercati ricavati dal metodo utilizzato da Goodman e Thibodeau, a discapito dei sottogruppi delle sezioni censuarie. Di conseguenza, l'accuratezza delle stime edoniche dei prezzi s'incrementa e si migliora in presenza di disaggregazione spaziale. Inoltre, l'efficienza delle aspettative del valore di mercato può essere influenzata dal metodo con cui si sono formati i sottomercati.

Infine, si può concludere affermando che più ci si trova in un'area limitata geograficamente, più le prestazioni si dimostrano efficienti rispetto alle stime aggregate.

Real Estate Market Segmentation: Hotels as Exemplar (2018)

I due temi principali di questo articolo sono le caratteristiche specifiche delle abitazioni e il reddito da locazione, in una situazione dove l'utente finale della piattaforma si aspetta una segmentazione precisa ed in linea alle sue preferenze.

Dopodichè è stata effettuata una valutazione sistematica delle principali variabili indipendenti che condizionano il prezzo degli hotel, quali il tasso di occupazione, metriche di reddito, caratteristiche della proprietà e tariffe di locazione.

Nel caso in cui si considerasse un mercato unico aggregato per l'intera industria, si potrebbero riscontrare stime distorte. Le variabili indipendenti costituite dal tasso di occupazione, dalla tariffa giornaliera di ciascuna camera e dal ricavo di quest'ultima sono le fondamenta su cui viene costruita la forza economica del mercato in cui la stessa proprietà. Ovviamente, per ciascun segmento di proprietà l'influenza di queste variabili indipendenti muta.

Solamente nella condizione in cui i mercati delle proprietà alberghiere sono suddivisi per classe allora il rapporto tra il prezzo e il numero di camere è significativo. È una relazione monotona. Per questi motivi la presente consiglia, qualora fossero presenti grandi categorie di proprietà, di disaggregare a seconda dei profili degli utenti finali e per qualità.

Utilizzando i dataset forniti da STR Global, sono stati individuati i 25 mercati alberghieri più rilevanti degli Stati Uniti. Ciò comprende 2257 transazioni dal gennaio 2000 al luglio 2012.

Il focus risiede nella realizzazione di un paragone tra la segmentazione o disaggregazione del mercato e le classi di hotel tramite un modello di determinazione dei prezzi. Il logaritmo del prezzo della proprietà è la principale variabile dipendente.

Gli esiti dello studio affermano una correlazione diretta tra il ricavo operativo e il prezzo della proprietà nel settore alberghiero. In particolare, mostra una segmentazione insita appartenente a ciascuna classe all'interno del settore in considerazione.

In conclusione, questo caso di studio mostra l'importanza di modelli che comprendano sia caratteristiche di stampo edonico sia di mercato. Propone quindi una segmentazione caratteristica di ciascuna tipologia di proprietà che dipende

maggiormente dalla qualità dell'utente finale e della proprietà. Inoltre, lo studio sottolinea che affidarsi esclusivamente alle qualità edonistiche non è sufficiente per procurarsi un disegno chiaro e completo del settore immobiliare. I modelli ibridi invece assegnano più indicazioni sulle variabili indipendenti che condizionano i valori della proprietà

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

La concentrazione di questo studio risiede sul modo in cui i prezzi degli affitti vengono influenzati da determinate genericità. Il campione preso in considerazione è composto da Barcellona, Siviglia e Madrid negli anni 2015 e 2016. Le caratteristiche interessate sono quelle inerenti alla tipologia di host, il numero massimo di ospiti, la distanza dal centro, il numero di camere da letto e di bagni, il punteggio complessivo e il numero di recensioni.

Gli esiti della regressione sono simili per tutte le città spagnole considerate. La distanza dal centro, il numero di recensioni e il numero di annunci per host sono le uniche variabili che possiedono una correlazione negativa ai prezzi. Tutte le altre caratteristiche elencate precedentemente sono correlate in modo positivo al prezzo degli affitti.

Al fine di evitare la presenza di multicollinearità, il VIF deve possedere valori inferiori a 10 per ognuna delle variabili considerate.

La variazione del fissaggio dei prezzi delle proprietà che un host possiede, al fine di farsi riconoscere tra gli alloggi dei competitor, è la conseguenza di ambiente di concorrenza monopolistica; infatti, questo approccio è dovuto al fatto che gli alloggi sono non perfettamente sostituibili.

Exploring changes in guest preferences for Airbnb accommodation with different levels of sharing and prices: Using structural topic model

I target di questo articolo sono la comprensione del modo attraverso cui la condivisione e il prezzo condizionino la scelta degli ospiti in fase di prenotazione,

specialmente per alcune caratteristiche di servizio. Il numero di osservazioni interessate nello studio è di 181.190, localizzate in Malesia e Kuala Lumpur. Si differenziano per fasce di prezzo e per tipologie di alloggio.

Il metodo di risoluzione in questione è STM e si fonda sull'analisi delle recensioni. In particolare, individua gli argomenti rilevanti citati nei commenti e la quantità di volte che sono stati individuati. Sono state identificate cinque categorie, corrispondenti a posizione, valore, esperienza generale, strutture e servizio.

Le preferenze identificate per la proprietà coincidono in prima battuta ai valori edonici, nello specifico quest'ultimi sono sottolineati degli stessi ospiti delle proprietà. I turisti reputano la property essenziale nella loro vacanza.

Un secondo valore è rappresentato dal rapporto col proprietario dell'alloggio. Il sostegno che fornisce l'host nei confronti dell'ospite è un aspetto cruciale, soprattutto per la risoluzione degli imprevisti che si presentano nel corso dell'esperienza. Infine, gli ospiti cercano degli ingredienti utili a permettere un'esperienza gradita e congrua ai desideri dei bambini. Un secondo fattore è rappresentato dalla proprietà condivisa, nella quale si localizzano i valori utilitari, l'efficienza nell'interazione con l'host e l'accesso.

L'ultima preferenza individuata è il condizionamento del prezzo sulle decisioni effettuate dall'ospite. Un prezzo più basso della struttura significa un maggiore interessamento all'accessibilità dei servizi, quali i mezzi di trasporto, i ristoranti e i centri commerciali. La localizzazione si dimostra quindi cruciale per questa tipologia di clienti. Un prezzo più alto indica una maggiore attenzione all'estetica e al comfort della proprietà, a discapito della posizione.

Dalle conclusioni si evince come la fascia di prezzo e la tipologia di abitazione condizionino le scelte degli utenti. Le esperienze più adattabili alle loro esigenze e i valori edonici rappresentano le maggiori richieste degli ospiti di proprietà intere. Per quanto riguarda coloro che prediligono esperienze più economiche la scelta ricade sulle proprietà condivise. I risultati di questo articolo mostrano soluzioni in

vista di un miglioramento delle proposte degli host, incrementando la fiducia degli utenti nei loro confronti.

A Review of the Problems and Some Proposals Tested on Bordeaux City Data

L'obiettivo di tale articolo esamina i metodi edonistici utilizzati per la valutazione dei prezzi delle proprietà. La città in questione è Bordeaux, nella quale vengono effettuati dei test al fine di individuare dei miglioramenti nel modello. Le soluzioni fornite da questo studio sono cinque. La prima è la specificazione degli attributi; infatti, creando sei categorie di variabili e aggiungendo maggiori dettagli si è raggiunta una maggiore precisione dei modelli. Tuttavia, questa proposta ha causato multicollinearità. La seconda è la segmentazione del mercato, causata da una diversificazione di domanda e offerta presente sulla piattaforma. Ad ogni segmento è associato un numero differente di variabili costituenti il prezzo, sottolineando l'importanza di suddividere le categorie al fine di ricavare stime di prezzo più realistiche. Il metodo edonistico successivo è la forma funzionale, qualora fosse scelta correttamente incrementerebbe la precisione della stima dei prezzi. La penultima proposta si basa sull'eterogeneità temporale, la quale dimostra avere un'influenza determinante sul miglioramento della parte spiegata dal modello, aumentando notevolmente l'accuratezza. Infine, la dipendenza spaziale ha un impatto positivo sulla stima.

Concludendo, tutte le proposte sottolineano che è fondamentale l'uso di modelli maggiormente dettagliati e diversificati nelle sue variabili al fine di migliorare la previsione dei prezzi. Nonostante ciò, sono presenti dei punti deboli all'interno di questa metodologia, poiché in caso di raffigurazione di segmenti più specifici potrebbe essere necessario l'utilizzo di approcci qualitativi.

Comparing Airbnb and traditional accommodation experiences using text-mining methods – the hedonic and utilitarian values framework

Questo articolo esamina le diversità tra i giudizi espressi dagli ospiti in alcune scelte tradizionali, come gli hotel, e in Airbnb mediante un procedimento di text-mining per approfondire 437.820 recensioni di viaggiatori. Questa metodologia è utile per estrapolare informazioni rilevanti da una grande quantità di dati; infatti, in questo caso si evidenziano dei valori edonici che danno luce alle preferenze degli ospiti. Gli output principali dello studio sono la ricerca di fattori che sono di gradimento o di delusione nei viaggiatori nelle due casistiche, e quali invece sono differenti tra le esperienze.

Tra le varie tecniche utilizzate di text-mining sono state scelte la Structural Topic Modelling (l'analisi dei temi) e l'analisi dei sentimenti, attraverso le quali sono stati definiti i temi e i sentimenti maggiormente presenti. Il luogo deciso per l'analisi è Praga, nel periodo che parte da novembre 2017 e si estende fino a novembre 2019.

Vi sono vari fattori chiave che influenzano positivamente la piattaforma di Airbnb:

- Il rapporto con l'host. L'importanza dell'host espressa tramite l'accoglienza, la gentilezza e la sua disponibilità. Questo è spesso considerato il principale aspetto positivo. Inoltre, gli ospiti spesso menzionano il nome dell'host nella loro recensione, sottolineando una relazione amichevole.
- Atmosfera del quartiere. I punti di interesse principali, quali negozi o ristoranti, sono spesso apprezzati dagli ospiti fornendo loro un senso di appartenenza al contesto urbano.
- Locali ed esperienze consigliate. L'host gioca un ruolo fondamentale sull'esperienza vissuta dall'ospite, in quanto gli raccomanda locali ed esperienze indirizzandolo verso i percorsi turistici principali.

Tuttavia, Airbnb possiede anche delle criticità. Esse sono caratterizzate principalmente dalla qualità delle camere e degli arredi. In particolare, il comfort del letto, il bagno sono il motivo di insoddisfazione in diverse strutture. Ciò è quello che differenzia maggiormente questa piattaforma dagli hotel, nei quali è presente un focus sul comfort, efficienza e servizi standardizzati.

Dagli esiti dello studio, si evince che le aspettative degli ospiti sono l'ago della bilancia della scelta tra gli hotel ed Airbnb. Quest'ultimo attira chi preferisce vivere un'esperienza unica, coinvolgente e arricchente mentre gli hotel sono scelti da chi desidera standardizzazione, comodità ed efficienza.

Price determinants of Airbnb accommodation in the French cities

L'articolo esamina le caratteristiche che contribuiscono al cambiamento dei prezzi degli Airbnb in 93 città francesi, studiando un totale di 140.000 annunci tra ottobre 2016 e marzo 2017. I fattori presi in considerazione sono sia interni all'alloggio, le sue caratteristiche intrinseche, che esterni, come la localizzazione, la tipologia di città e il suo stato rispetto ad altre forme di alloggio.

Viene studiato il prezzo implicito di ciascuna caratteristica della proprietà, in mercato concorrenziale. Le dinamiche urbane sono rispettate dai prezzi degli alloggi; infatti, le preferenze degli ospiti confermano maggior affluenza nelle proprietà in posizioni centrali e dotate di caratteristiche specifiche in base alla loro tipologia.

I principali fattori su cui si è basato lo studio sono l'influenza della centralità e la dimensione della città, l'effetto delle caratteristiche interne, l'effetto della concorrenza con altri alloggi e variabili urbane ed ambientali.

I risultati suggeriscono come gli host possano ottimizzare i costi e migliorare l'attrattiva della propria proprietà in base alla dimensione della città in cui si trovano. Nello specifico, nelle grandi città occorre offrire servizi di alta qualità ed un'alto servizio di accoglienza, mentre nelle medie e piccole città è consigliato

puntare su servizi ed esperienze che valorizzino la natura e le attrazioni locali, oltre che tenere i prezzi competitivi.

Airbnb's disruption of the housing structure in London

L'articolo in questione elabora come la disponibilità e l'accessibilità degli alloggi a Londra siano state influite dalla crescita degli affitti brevi mediante Airbnb. I punti principali sono i seguenti:

- Calo degli affitti a lungo termine: la piattaforma di Airbnb, espandendosi velocemente specialmente in aree urbane, ha causato la riduzione di opzioni abitative a lungo termine e all'incremento dei prezzi di affitto.
- Densità geografica e distribuzione di Airbnb: Le aree in cui sono presenti più proprietà appartenenti alla piattaforma sono i quartieri centrali delle città, oltre che alcune zone periferiche. In generale, queste aree sono caratterizzate da una forte presenza di appartamenti e flat.
- Relazione tra tipologia di Airbnb e unità abitativa: Le aree con maggiore eterogeneità abitativa, che includono case singole o bifamiliari, sono poco interessate alla diffusione di Airbnb.
- Impatto sul mercato degli affitti e il settore privato: Airbnb ha trovato una base solida nelle aree con una grande percentuale di case in affitto privato, a scapito delle abitazioni di proprietà. Le proprietà destinate agli affitti sono per lo più appartamenti.

In sintesi, il documento mette in luce come Airbnb stia cambiando profondamente la struttura abitativa di Londra, contribuendo alla gentrificazione e riducendo la disponibilità di alloggi a lungo termine.

4. APPROFONDIMENTO SULLE VARIABILI

Lo scopo di questi capitoli è analizzare le variabili indipendenti che influenzano il prezzo di una proprietà, ovvero la variabile dipendente. Le variabili indipendenti possono essere sia interne che esterne. Per variabili interne s'intendono il numero di stanze, la metratura o il numero di ospiti, mentre le variabili esterne sono rappresentate dal quartiere a cui appartiene la proprietà, le recensioni o l'etnia.

Questa operazione si svolge tramite l'utilizzo di un database che raccoglie tutte le informazioni relative alle variabili in questione. Per agevolare l'analisi si è utilizzato un software statistico, chiamato "Stata", largamente utilizzato in ambito accademico per l'analisi di dati e in grado di svolgere una molteplicità di funzioni quali gestione di database, analisi statistico-econometriche e analisi grafiche.

Rappresentazione database

Il database fornisce i valori di numerose variabili associate a ciascuna proprietà. Queste variabili sono principalmente di cinque tipologie: VARCHAR (abbreviazione di CHARACTER VARYING, permette di definire stringhe di caratteri a lunghezza variabile con lunghezza massima predefinita), NUMBER (variabile numerica), BOOLEAN (0 se è falsa, 1 se è vera), TIMESTAMP_NTZ (numero che esprime una grandezza temporale) e ARRAY (struttura di dati formata da più componenti ordinate dello stesso tipo).

Di seguito l'elenco delle variabili e il loro significato.

- Status: Stato della proprietà. Prenotato (Status = R), Sfitto (Status = A) o Bloccato (Status = B)
- Property ID: Identificatore univoco AirDna per la proprietà.
- Airbnb Property ID: Identificatore univoco Airbnb per la proprietà.
- Airbnb HOST ID: ID univoco dell'host Airbnb.
- VRBO Property ID: ID univoco della proprietà VRBO.
- VRBO Host ID: ID host di VRBO.

- Listing Title: Titolo dell'annuncio dell'affitto per le vacanze. Se la proprietà è elencata sia su Airbnb che su Vrbo, si utilizza l'annuncio Airbnb.
- Property Type: Tipo di sistemazione, come casa, appartamento o pensione.
- Real Estate Property Type: Tipo di proprietà immobiliare, come casa/villa o appartamento/condominio/Loft.
- Listing Type: Tre tipi di annunci di affitto per le vacanze: casa intera, stanza privata e stanza condivisa.
- Property Manager: Nome del Property Manager nel caso in cui sia specificato per la proprietà.
- Host Type: Tipo di host della proprietà in base al numero di unità che possiede. Classificato in 1 unità, 2-5 unità, 6-20 unità, o più di 20 unità.
- Country: Paese in cui si trova l'immobile in affitto per le vacanze.
- State: Stato in cui si trova l'immobile in affitto per le vacanze.
- City: Città in cui si trova l'immobile in affitto per le vacanze.
- Zipcode: Codice postale in cui si trova l'immobile in affitto per le vacanze.
- Neighborhood: Quartiere in cui si trova la proprietà in affitto per le vacanze, ove disponibile.
- Latitude: Latitudine della proprietà in affitto per le vacanze.
- Longitude: Longitudine della proprietà in affitto per le vacanze.
- Exact Location: TRUE = Latitudine e Longitudine sono le coordinate esatte della proprietà. FALSE = Latitudine e Longitudine sono coordinate confuse entro un raggio di 500 metri per la privacy dell'host.
- Location Type: Tipo di posizione dell'annuncio. Classificato in Destinazione/Resort - Montagne/Lago, Destinazione/Resort - Costiera, Piccola città/Rurale, Grande città - Periferia, Grande città - Urbana o Città di medie dimensioni. Può anche essere nullo.
- Price Tier: I livelli di prezzo della tariffa giornaliera media in MarketMinder segmentano gli annunci all'interno di un mercato in diverse fasce di prezzo: Budget, Economy, Midscale, Upscale e Luxury.
- Bedrooms: Numero di camere da letto in un annuncio di affitto per le vacanze.
- Bathrooms: Numero di bagni in un annuncio di affitto per le vacanze.

- Max Guests: Numero di ospiti che l'annuncio può ospitare.
- Cancellation Policy: Politica di cancellazione per la perdita dell'affitto di una casa vacanze.
- Minimum Stay: Il soggiorno minimo predefinito richiesto dall'host.
- Created Date: La data in cui è stata creata l'inserzione dell'affitto per le vacanze.
- Last Scraped Date: L'ultima data in cui gli scraper hanno trovato l'annuncio di affitto per le vacanze. Ogni annuncio viene estratto ogni tre giorni.
- Last Calendar Update: L'ultima data in cui gli scraper hanno catturato le informazioni del calendario per l'annuncio di affitto per le vacanze. Ogni giorno si esegue lo scraping dei calendari attivi.
- Response Rate: La percentuale di nuove richieste e richieste di prenotazione a cui un host risponde (accettando/pre-approvando o rifiutando) entro 24 ore.
- Response Time: Il tasso medio di risposta dell'host alle nuove richieste e alle richieste di prenotazione da parte di un ospite.
- Currency (USD): USD (dollari USA (\$)).
- ADR (USD): Tariffa media giornaliera delle notti prenotate in dollari USA (\$). $ADR = \text{Ricavi totali} / \text{Notti prenotate}$. Include le spese di pulizia.
- Revenue LTM (USD): Fatturato degli ultimi dodici mesi (LTM), in dollari USA (\$). Include le spese di pulizia ma non altre spese aggiuntive.
- Revenue Potential LTM (USD): Ultimi dodici mesi (LTM) di potenziali entrate, in dollari USA (\$). Include le spese di pulizia ma non altre spese aggiuntive.
- Occupancy Rate LTM: Tasso di occupazione = $\text{Conteggio dei giorni di prenotazione} / \text{Notti in prenotazioni attive}$.
- Number of Bookings LTM: Numero di prenotazioni negli ultimi dodici mesi (LTM). Se la proprietà ha annunci sia su Airbnb che su Vrbo, sarà l'aggiunta di prenotazioni su ciascuna piattaforma.
- Number of Reviews: Numero totale di recensioni per la proprietà. Se la proprietà ha annunci sia su Airbnb che su Vrbo, saranno aggiunte le recensioni su ciascuna piattaforma.

- Security Deposit (USD): Deposito cauzionale in dollari USA (\$).
- Cleaning Fee (USD): Spese di pulizia addebitate per prenotazione in dollari USA (\$).
- Extra People Fee: Supplemento per persone extra in dollari USA (\$).
- Published Nightly Rate (USD): Tariffa notturna predefinita per un annuncio di affitto per le vacanze in dollari USA (\$).
- Published Monthly Rate (USD): Tariffa mensile predefinita per un annuncio di affitto per le vacanze in dollari USA (\$).
- Published Weekly Rate (USD): Tariffa settimanale predefinita per un annuncio di affitto per le vacanze in dollari USA (\$).
- Count Reservation Days LTM: Numero totale di giorni di calendario classificati come riservati negli ultimi dodici mesi.
- Count Available Days LTM: Numero totale di giorni di calendario che sono stati classificati come disponibili per la prenotazione, ma non effettivamente prenotati negli ultimi dodici mesi.
- Count Blocked Days LTM: Numero totale di giorni di calendario classificati come bloccati dalla ricezione di una prenotazione negli ultimi dodici mesi.
- Number of Photos: Numero di foto in un annuncio di affitto per le vacanze.
- Instantbook: Vero = la proprietà in affitto per le vacanze può essere prenotata senza alcuna comunicazione tra host e ospite.
- Airbnb Superhost: Vero o falso a seconda che l'host sia un Superhost su Airbnb.
- Listing URL: Link all'annuncio dell'affitto per le vacanze su Airbnb o Vrbo.
- Listing Main Image URL: Collegamento all'immagine principale dell'annuncio di affitto per le vacanze.
- Listing Images: Elenco dei link delle immagini per l'annuncio di affitto per le vacanze.
- License: Numero di licenza Airbnb / Vrbo.
- Overall Rating: Valutazione media degli ospiti della struttura su 100.
- Airbnb Communication Rating: Valutazione media della comunicazione della proprietà su 10.

- Airbnb Accuracy Rating: Valutazione media della precisione della proprietà su 10.
- Airbnb Cleanliness Rating: Valutazione media della pulizia della struttura su 10.
- Airbnb Checkin Rating: Valutazione media del check-in della struttura su 10.
- Airbnb Location Rating: Valutazione media della posizione della struttura su 10.
- Airbnb Value Rating: Valutazione media del valore della struttura su 10.
- Has Pool: Se è vero, l'annuncio ha una piscina.
- Has Hot Tub: Se è vero, l'annuncio ha una vasca idromassaggio.
- Has Air Con: Se è vero, l'annuncio ha l'aria condizionata.
- Has Gym: Se è vero, l'annuncio ha una palestra.
- Pets Allowed: Se è vero, l'annuncio consente animali domestici.
- Has Kitchen: Se è vero, l'annuncio ha una cucina.
- Has Parking: Se è vero, l'annuncio ha un parcheggio.
- Amenities: Elenco dei servizi pubblicati nell'annuncio - ogni servizio sarà contrassegnato come vero o falso.
- Currency (Native): Valuta del paese in cui è elencata la proprietà.
- ADR (Native): Tariffa media giornaliera (ADR) delle notti prenotate nella valuta del paese in cui è elencata la proprietà. $ADR = \text{Ricavi totali} / \text{Notti prenotate}$. Include le spese di pulizia.
- Revenue LTM (Native): Ultimi dodici mesi (LTM) o entrate totali mensili delle inserzioni, a seconda del report visualizzato, nella valuta del paese in cui è elencata la proprietà. Include le spese di pulizia, ma non altre spese aggiuntive.
- Revenue Potential LTM (Native): Ultimi dodici mesi (LTM) di potenziali entrate, nella valuta del paese in cui è quotata la proprietà. Include le spese di pulizia, ma non altre spese aggiuntive.
- Cleaning Fee (Native): Spese di pulizia addebitate per prenotazione, nella valuta del paese in cui è elencata la proprietà.

- Published Nightly Rate (Native): Tariffa notturna predefinita per un annuncio di affitto per le vacanze, nella valuta del paese in cui è elencata la proprietà.
- Published Monthly Rate (Native): Tariffa mensile predefinita per un annuncio di affitto per le vacanze, nella valuta del paese in cui è elencata la proprietà.
- Published Weekly Rate (Native): Tariffa settimanale predefinita per un annuncio di affitto per le vacanze, nella valuta del paese in cui è elencata la proprietà.

Data exploration

L'obiettivo è l'esplorazione dei dati relativi agli Airbnb presenti in Val Pellice, una valle alpina della Città Metropolitana di Torino, posta a sud della Val Chisone e a nord della Valle Po. Essa è composta dai comuni di Angrogna, Bibiana, Bobbio Pellice, Bricherasio, Luserna San Giovanni, Lusernetta, Rorà, Torre Pellice e Villar Pellice. Lo scopo deciso in una prima fase è di prendere in considerazione due giorni singoli dell'anno 2023, ma ciò ha registrato pochissime osservazioni (circa 200), spiegate dal fatto che gli Airbnb attivi nei giorni stabiliti erano solamente 101. Il numero di proprietà osservate si è dimostrato troppo basso per ottenere inferenza statistica, di conseguenza dedurre le caratteristiche della popolazione dall'osservazione di una parte di essa risulta complesso.

Al fine di risolvere questo tema, sono state analizzate due possibilità:

- Considerare tutti i giorni del 2023 con lo stesso focus geografico.
- Cambiare focus geografico.

Si è scelta la prima opzione, coinvolgendo i dati di tutti i giorni del 2023 anziché di soli due giorni. Le osservazioni registrate in questo caso sono chiaramente maggiori.

Usfruendo di Stata, si è utilizzata la funzione `import excel` `"/"fonte_del_file"/"nome_del_file".xlsx"`, sheet ("Sheet1") firstrow per importare il file su Stata. Il database completo possiede 33695 osservazioni e 28 variabili. Il periodo interessato va dal giorno 1/1/2023 al giorno 31/12/2023.

Analizzando il database fornito, si sono elencate tutte le variabili che potrebbero avere un'influenza sul prezzo. Esse sono rappresentate sia da valori numerici che testuali (come "Status", "PropertyType", "HostType" e "City"). Per quest'ultime sono state create delle dummy, in grado di evidenziare la presenza di ciascuna specifica casistica testuale. Lo stesso discorso è valido per la voce "Amenities", la quale è composta da diverse voci al suo interno. Si procede quindi con la creazione di alcune variabili dummy, le quali indicano la presenza di uno specifico servizio all'interno della struttura. Per evitare di trascurare delle covariate, che possono rivelarsi determinanti quando si dovrà affrontare la regressione, si sono inserite tutte le caratteristiche possibili osservabili dall'annuncio (al netto del fatto che abbiano un VIF ≤ 10 , verificato nella fase successiva). La tabella seguente è stata ricavata attraverso il comando *sum* di Stata, il quale fornisce per ogni variabile il numero di osservazioni, la media, la deviazione standard, il valore minimo e il valore massimo. È stata aggiunta una breve descrizione per le dummy. Non tutte le variabili possiedono la totalità delle osservazioni.

Tabella 1 Descrizione variabili.

Variabile	Osservazioni	Media	Deviazione Standard	Minimo	Massimo	Note
PropertyID	0					
Date	0					
PriceUSD	33,695	77.51346	192.4299	0	3116	
BookedDate	0					
Status	0					
YEAR	0					
PropertyType	0					
PropertyMa~r	0					
HostType	0					
City	0					
Neighborhood	0					
Latitude	33,695	44.8191	.0070124	44.79741	44.83699	
Longitude	33,695	7.217736	.036873	7.137733	7.32946	
PriceTier	0					
Bedrooms	33,33	1.423792	.8466529	0	6	
Bathrooms	33,695	1.216694	.7012814	0	6	
MaxGuests	33,695	3.702923	1.547225	1	12	
ResponseRate	32,323	95.06667	17.92743	0	100	

ResponseTime	7,752	25757.67	35896.48	295	86400	
AirbnbSupert	30,206	.2287956	.4200643	0	1	
OverallRating	25,415	94.76313	10.56629	20	100	
AirbnbCommng	25,415	9.865827	.5371414	6	10	
AirbnbAccung	25,415	9.496911	1.203002	2	10	
AirbnbCleaning	25,415	9.511155	1.03975	4	10	
AirbnbLocation	25,415	9.780877	.5860551	6	10	
AirbnbValue	25,415	9.476412	1.174705	2	10	
wireless	33,695	.6098531	.4877903	0	1	= 1 se tra i servizi c'è il wi-fi
parking	33,695	.7566998	.4290812	0	1	= 1 se tra i servizi c'è il parcheggio
tv	33,695	.6700994	.4701837	0	1	= 1 se tra i servizi c'è la tv
washer	33,695	.6494732	.477142	0	1	= 1 se tra i servizi c'è la doccia
conditioner	33,695	.1399614	.3469521	0	1	= 1 se tra i servizi c'è il conditioner
elevator	33,695	.0541623	.2263411	0	1	= 1 se tra i servizi c'è l'ascensore
coffee_maker	33,695	.3839739	.4863589	0	1	= 1 se tra i servizi c'è la macchina del caffè
fireplace	33,695	.1924915	.3942627	0	1	= 1 se tra i servizi c'è il camino
hangers	33,695	.7560172	.4294888	0	1	= 1 se tra i servizi ci sono le grucce
hairdryer	33,695	.783321	.4119882	0	1	= 1 se tra i servizi c'è phon
heating	33,695	.718801	.4495911	0	1	= 1 se tra i servizi c'è il riscaldamento
essentials	33,695	.8463867	.360583	0	1	= 1 se tra i servizi ci sono le cose essenziali

first_aid_~t	33,695	.3866152	.4869814	0	1	= 1 se tra i servizi c'è il kit di primo soccorso
lockbox	33,695	.0912598	.2879825	0	1	= 1 se tra i servizi c'è la cassaforte
stair_gates	33,695	.0478112	.2133698	0	1	= 1 se tra i servizi c'è lo scorrimento
hot_water	33,695	.7895237	.407653	0	1	= 1 se tra i servizi c'è l'acqua calda
bed_linens	33,695	.6127022	.4871399	0	1	= 1 se tra i servizi c'è la biancheria da letto
extra_pill~s	33,695	.5347974	.4987951	0	1	= 1 se tra i servizi ci sono cuscini e coperte extra
refrigerator	33,695	.6875501	.463499	0	1	= 1 se tra i servizi c'è il frigorifero
dishes_and~e	33,695	.7146164	.4516036	0	1	= 1 se tra i servizi ci sono piatti e argenteria
cooking_ba~s	33,695	.6468912	.4779432	0	1	= 1 se tra i servizi ci sono le cose essenziali per cucinare
stove	33,695	.4652619	.4987992	0	1	= 1 se tra i servizi c'è la stufa
patio_or_b~y	33,695	.4795073	.4995873	0	1	= 1 se tra i servizi c'è il patio o il balcone
garden_or_~d	33,695	.5348865	.4987889	0	1	= 1 se tra i servizi c'è il giardino
long_term_~d	33,695	.3052678	.4605276	0	1	= 1 se tra i servizi c'è il permesso di lungo termine

lake_access	33,695	.0216649	.1455891	0	1	= 1 se tra i servizi c'è l'accesso al lago
kitchen	33,695	.8753524	.3303238	0	1	= 1 se tra i servizi c'è la cucina
laptopfrie~y	33,695	.4429441	.4967413	0	1	= 1 se tra i servizi c'è un laptop
iron	33,695	.4688529	.4990363	0	1	= 1 se tra i servizi c'è il ferro da stiro
shampoo	33,695	.603502	.4891773	0	1	= 1 se tra i servizi c'è lo shampoo
bathhtub	33,695	.2231785	.4163832	0	1	= 1 se tra i servizi c'è la vasca da bagno
body_soap	33,695	.3443538	.4751641	0	1	= 1 se tra i servizi c'è il bagnoschiuma
microwave	33,695	.3667903	.4819358	0	1	= 1 se tra i servizi c'è il microonde
luggage_dr~d	33,695	.2722362	.4451175	0	1	= 1 se tra i servizi c'è il permesso di lasciare i bagagli
hot_water_~e	33,695	.2783202	.4481786	0	1	= 1 se tra i servizi c'è il bollitore
alfresco_d~g	33,695	.3786022	.4850459	0	1	= 1 se tra i servizi c'è la possibilità di mangiare all'aperto
dining_table	33,695	.4040362	.4907118	0	1	= 1 se tra i servizi c'è il tavolo da cucina
toaster	33,695	.1490726	.3561653	0	1	= 1 se tra i servizi c'è il tosta pane
outdoor_se~g	33,695	.2836326	.4507673	0	1	= 1 se tra i servizi ci sono

						le sedie da esterno
freezer	33,695	.3045853	.4602384	0	1	= 1 se tra i servizi c'è il freezer
nespresso_~e	33,695	.0324974	.1773196	0	1	= 1 se tra i servizi c'è la macchina del caffè nespresso
laundromat~y	33,695	.1590444	.3657229	0	1	= 1 se tra i servizi c'è una lavanderia nelle vicinanze
cleaning_p~s	33,695	.4075679	.4913893	0	1	= 1 se tra i servizi ci sono dei prodotti di pulizia
clothes_dr~k	33,695	.3967057	.4892212	0	1	= 1 se tra i servizi c'è lo stendino
wardrobe_o~t	33,695	.4528565	.4977799	0	1	= 1 se tra i servizi c'è un armadio
wine_glasses	33,695	.2270367	.4189227	0	1	= 1 se tra i servizi ci sono dei bicchieri di vino
bbq_area	33,695	.1372014	.3440649	0	1	= 1 se tra i servizi c'è un'area barbeque
bidet	33,695	.4193797	.4934649	0	1	= 1 se tra i servizi c'è il bidet
piano	33,695	.0324974	.1773196	0	1	= 1 se tra i servizi c'è il piano
fire_extin~r	33,695	.2347826	.4238692	0	1	= 1 se tra i servizi c'è l'estintore
gym	33,695	.0324974	.1773196	0	1	= 1 se tra i servizi c'è la palestra
high_chair	33,695	.1944502	.3957827	0	1	= 1 se tra i servizi c'è il seggiolone

crib	33,695	.1383588	.3452813	0	1	= 1 se tra i servizi c'è la culla
libero	33,695	.4582876	.4982644	0	1	
apartment	33,695	.3642083	.4812146	0	1	
cabin	33,695	.0433299	.2036016	0	1	
chalet	33,695	.0108325	.1035154	0	1	
cottage	33,695	.0108325	.1035154	0	1	
guest_house	33,695	.0099124	.0990681	0	1	
house	33,695	.1525746	.3595823	0	1	
tiny_house	33,695	.0216649	.1455891	0	1	
villa	33,695	.0623238	.2417463	0	1	
bed_and_br~t	33,695	.0587624	.235183	0	1	
campsite	33,695	.0072118	.0846165	0	1	
condo	33,695	.1508236	.3578821	0	1	
farm_stay	33,695	.0298857	.1702746	0	1	
guest_suite	33,695	.0378691	.1908825	0	1	
tent	33,695	.0072711	.0849615	0	1	
vacation_h~e	33,695	.0324974	.1773196	0	1	
small_host	33,695	.5130732	.4998365	0	1	
medium_host	33,695	.3838255	.4863235	0	1	
big_host	33,695	.1031014	.3040957	0	1	
villar_pel~e	33,695	.1569669	.363775	0	1	
torre_pell~e	33,695	.5558095	.4968829	0	1	
luserna	33,695	.2112776	.4082209	0	1	
bricherasio	33,695	.075946	.2649156	0	1	
economy	33,695	.2158777	.411436	0	1	
midscale	33,695	.1707078	.376259	0	1	
luxury	33,695	.1140525	.3178798	0	1	
upscale	33,695	.1453628	.3524715	0	1	
budget	33,695	.31429	.4642394	0	1	

Pulizia database

Per ottenere degli esiti veritieri, le analisi e i test devono essere sostenuti da dati riconoscibili e rappresentativi della popolazione. Per questo motivo, sono state prese le seguenti decisioni:

- 1) Le strutture aventi PriceUSD=0 sono state eliminate. Ciascuna proprietà deve avere un valore economico equivalente ad una prenotazione giornaliera. Se il prezzo è zero la struttura non può essere presa in considerazione. Per rispettare questo criterio sono state eliminate 411 osservazioni.
- 2) Siccome Airbnb è una piattaforma online in cui i proprietari di immobili possono inserire i loro alloggi e metterli a disposizione dei viaggiatori, ciascuno di essi deve almeno possedere una camera da letto per dormire. Quindi non è ammissibile avere Bedrooms=0. Per questa motivazione, 1087 osservazioni sono state eliminate.
- 3) Per un motivo pressoché identico al punto precedente la variabile “Bathrooms” deve essere necessariamente presente per motivi fisiologici e legali (requisito da soddisfare per far parte della piattaforma). Sono state eliminate 365 osservazioni.
- 4) Molte osservazioni risultano non possedere alcun valore di PriceTier. Per ottenere una situazione di partenza più equilibrata, dove il numero di informazioni è il medesimo, si sono eliminate le osservazioni aventi valore “PriceTier” pari a zero (1341).
- 5) I giorni in cui la proprietà risulta bloccata, quindi con Status=0, sono stati eliminati, in quanto non rappresentativi di giorni effettivi presenti sul mercato (pari a 11986). Il numero di osservazioni è calato, ma non crollato in maniera considerevole.
- 6) Le osservazioni che possiedono, in corrispondenza della variabile “Airbnb Superhost”, un valore diverso da uno o da zero (rispettivamente VERO o FALSO, essendo una variabile booleana) vengono cancellate. Questo a causa di numerose registrazioni con la cella vuota, quindi non di aiuto nell’analisi generale del lavoro. Le osservazioni cancellate sono 1241.

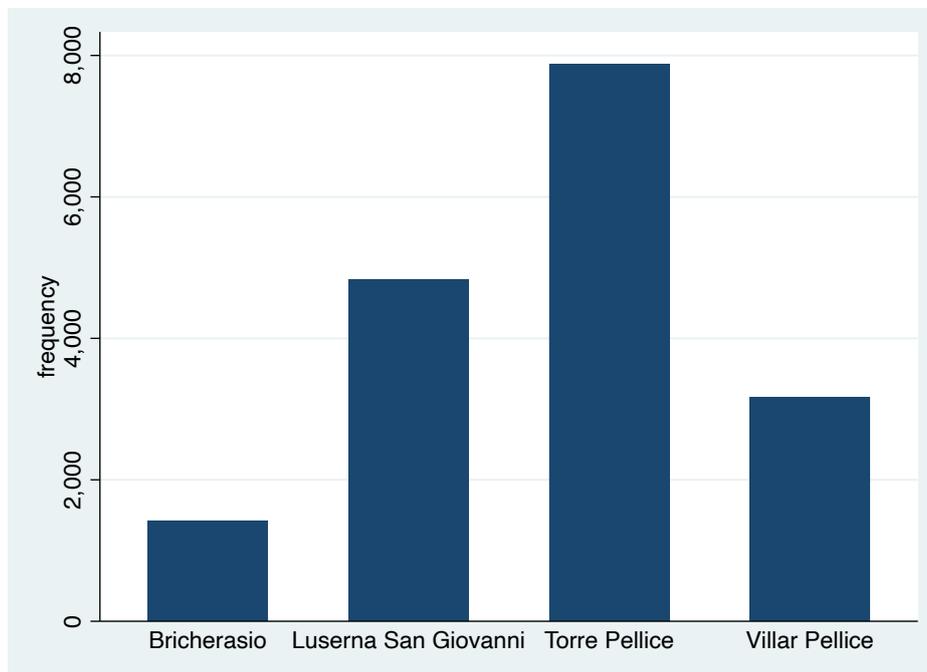
Sommando tutte le osservazioni eliminate si ottiene un totale di 14875. Le osservazioni su cui si andranno compiere le analisi saranno quindi 17281.

Nell'esplorazione del database è importante suddividere la quantità di strutture in ciascuno dei suoi comuni di appartenenza, per capire la densità di proprietà appartenenti a ciascuno di essi.

Per compiere questa azione si è utilizzata la funzione *count if City=="nome_del_comune"*. Ad ogni comune è stata assegnata una percentuale della quantità totale di osservazioni presenti nell'intero campione.

- Torre Pellice: $(7878/17281) = 45,59\%$
- Villar Pellice: $(3167/17281) = 18,33\%$
- Bobbio Pellice: 0 %
- Bricherasio: $(1411/17281) = 8,16\%$
- Luserna San Giovanni: $(4825/17281) = 27,92\%$
- Rorà: 0%
- Angrogna: 0%
- Bibiana: 0%
- Lusernetta: 0%

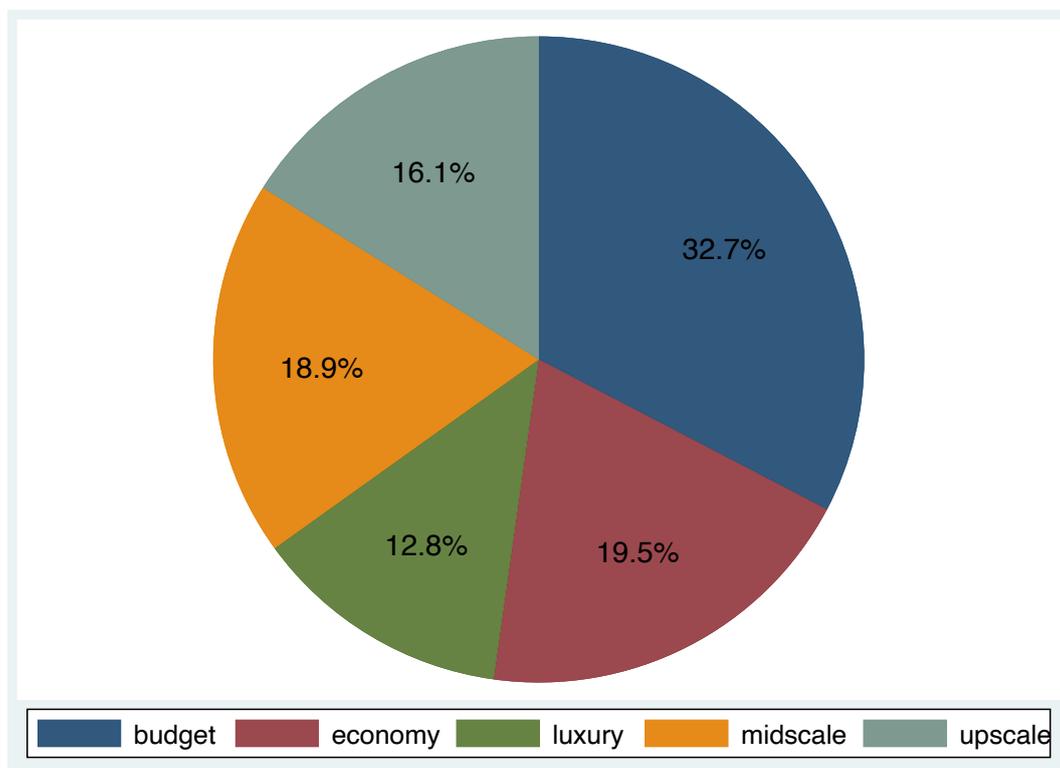
Grafico 1 Distribuzione delle strutture disponibili per ciascun comune in Val Pellice.



Ci sono varie fasce di prezzo su cui si possono suddividere le proprietà. I livelli di prezzo della tariffa giornaliera media in MarketMinder segmentano gli annunci all'interno di un mercato in cinque categorie, rappresentate nel Grafico 2.

Da come si evince dal grafico, il 52,2% degli Airbnb presenti nell'area descritta appartiene alle categorie budget ed economy, contro un 35% offerto dalla coppia midscale ed upscale. La qualità offerta, in termini di prezzo, dalle strutture presenti in Val Pellice non risulta essere elevatissima. Essendo un'area distante dalla città ed economicamente modesta, oltre che prevalentemente industriale piuttosto che turistica, si dimostra essere in linea con le sue aspettative.

Grafico 2 Suddivisione percentuale della variabile Price Tier.



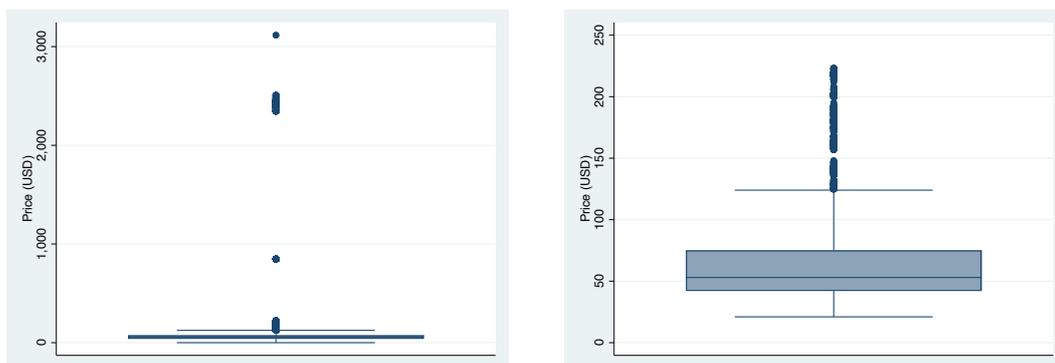
Uno dei punti di forza della piattaforma è il prezzo. Esso si dimostra più basso rispetto ad altre opportunità che offre il mercato, a pari condizioni. Un aspetto interessante è proprio l'analisi della variabile PriceUSD, la quale è maggiore di 300 \$ per 298 proprietà sul totale. Di queste, ben 208 possiedono un prezzo maggiore di 2000 \$. Assumendo la presenza di errori di battitura o di altro genere, si può notare come la maggior parte degli Airbnb (99,064 %) siano al di sotto di una soglia

coincidente a 300\$. Per questo motivo, oltre al fatto che quelle proprietà nel mondo reale non possono valere quel prezzo (quindi non veritiere), si è proceduto con l'eliminazione delle strutture aventi $PriceUSD > 2000$ \$. Il numero di osservazioni totale è rimasto invariato, in quanto le strutture aventi il prezzo superiore a tale soglia erano già state tutte eliminate dai vincoli precedenti.

Dal grafico sottostante si evince come nella situazione originaria fossero presenti degli outlier lungo la verticale del grafico, rappresentanti osservazioni che si discostavano totalmente dalla maggioranza. I motivi potrebbero essere errori di battitura in fase trascrizione o di altro genere.

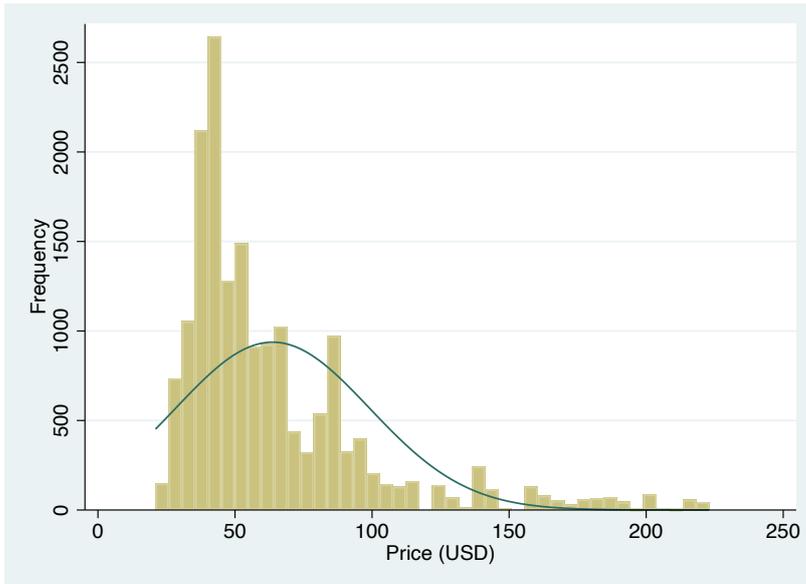
La situazione a seguito della pulizia del database si mostra più chiara e definita. I valori limite della "box" coincidono a circa 40\$ (primo quartile) e a 75\$ (terzo quartile), compresi in questo range vi sono il 50% dei valori del database. Il minimo e il massimo, non considerando gli outlier, sono 21\$ e 125\$, mentre la mediana risiede a 53\$. Sapendo che ciascun quartile occupa il 25% dei dati, si può affermare come nella fascia di prezzo compresa nel secondo quartile (40\$-53\$) vi sia la maggior densità di osservazioni.

Grafico 3 Boxplot del prezzo pre e post pulizia del database.



Il grafico della densità normale aggiunta all'istogramma mostra come il picco si raggiunga in corrispondenza di un prezzo di circa 60 \$; infatti, grazie al comando *mean PriceUSD*, si può risalire effettivamente al suo valore di 63,71981 \$.

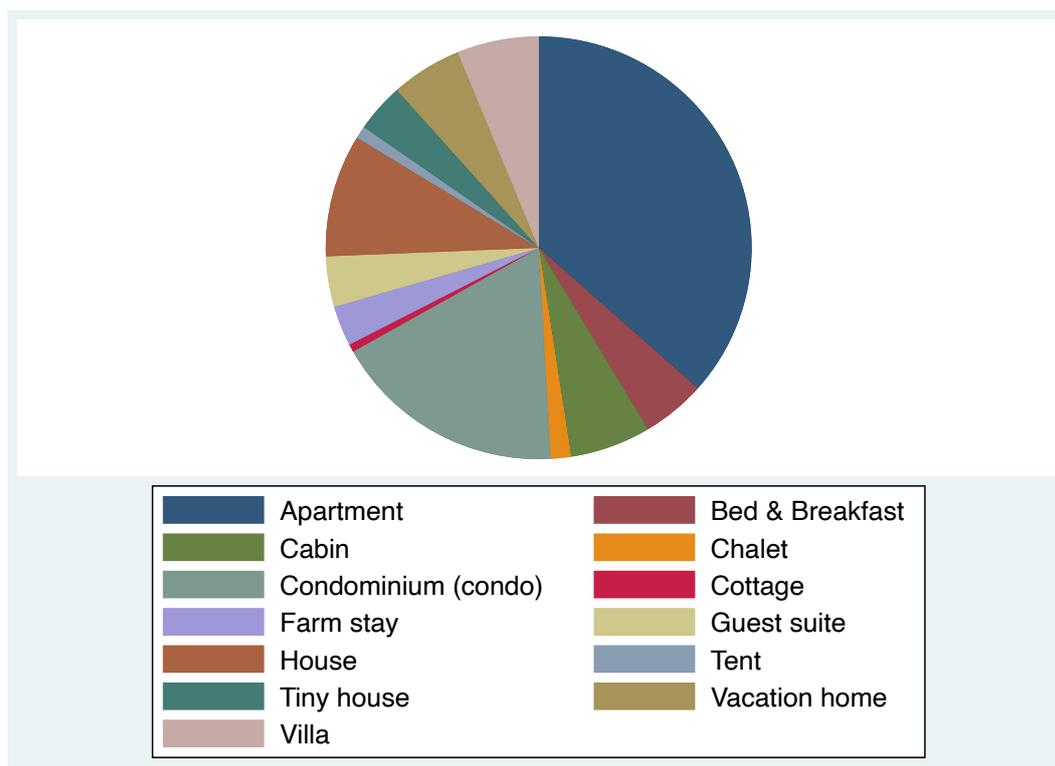
Grafico 4 Distribuzione del prezzo a notte.



Un'ultima esplorazione sul database fornito riguarda il tipo di proprietà. Per dare una motivazione ed un senso a ciascun prezzo, è bene associarlo ad una tipologia di struttura. Le varianti disponibili sono molteplici, in quantità maggiore sono presenti appartamenti, condomini e case indipendenti. Insieme occupano circa il 65% dell'intera torta. Le tipologie rimanenti sono presenti in maniera più o meno equa nel mercato.

Come da previsione, essendo una realtà piccola distante dalla città, i condomini non rappresentano la maggioranza delle proprietà presenti. Come mostra il Grafico 4 la Val Pellice, essendo una zona abbastanza montuosa, non dispone ovviamente di numerosi palazzi, ma di una discreta quantità di strutture indipendenti.

Grafico 5 Tipologie di Airbnb presenti nel territorio.



Al fine di non sovraccaricare ulteriormente il database non sono state inserite tutte le tipologie di proprietà, bensì solo “Apartment” e “Condominium (Condo)”, le quali sono le uniche che si distaccano nettamente dalle restanti per numero di osservazioni.

Le frequenze registrate di “Apartment” coincidono a 6321, quelle di “Condominium (Condo)” a 3064. Tutte le altre si aggirano intorno a 1000 osservazioni, oppure al di sotto di questo numero. Di conseguenza si è scelto di mantenere le dummy rappresentative delle due tipologie di proprietà più numerose e di crearne una terza che inglobi le altre tipologie restanti, chiamata “Proprietà restanti”.

Tabella 2 Tipologie di proprietà.

Proprietà	Frekuensi	Percentuale
Apartment	6,321	36.58
Bed & Breakfast	837	4.84
Cabin	1,062	6.15
Chalet	267	1.55
Condominium (condo)	3,064	17.73

Cottage	102	0.59
Farm stay	531	3.07
Guest suite	666	3.85
House	1,621	9.38
Tent	163	0.94
Tiny house	645	3.73
Vacation home	930	5.38
Villa	1,072	6.20
Total	17,281	100.00

5. ANALISI VALUTAZIONI

Punteggio valutazioni

Nel database disponibile vi sono sette diverse valutazioni che si possono assegnare alla struttura:

- Overall Rating: livello di soddisfazione della struttura nel suo complesso.
- Airbnb Communication Rating: valutazione media della comunicazione.
- Airbnb Accuracy Rating: valutazione media della puntualità e precisione.
- Airbnb Cleanliness Rating: valutazione media della pulizia nella struttura.
- Airbnb Location Rating: valutazione media della posizione.
- Airbnb Checkin Rating: valutazione media del check-in.
- Airbnb Value Rating: valutazione media del valore dell'immobile.

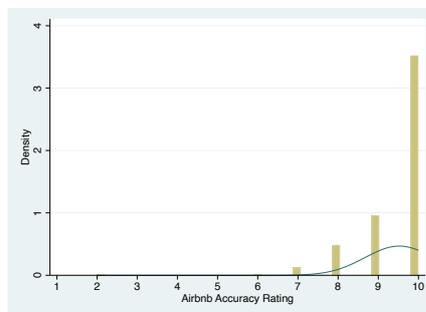
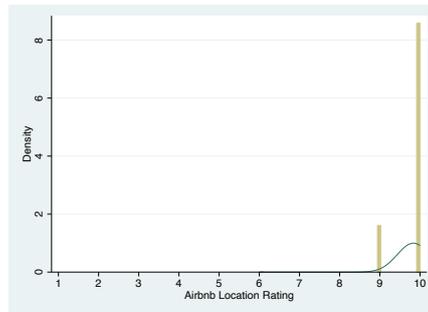
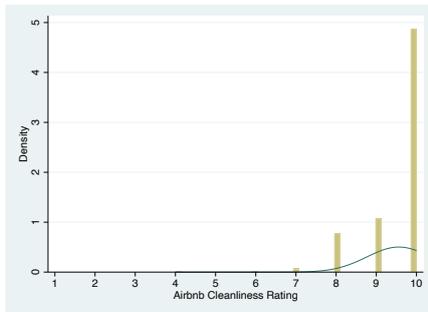
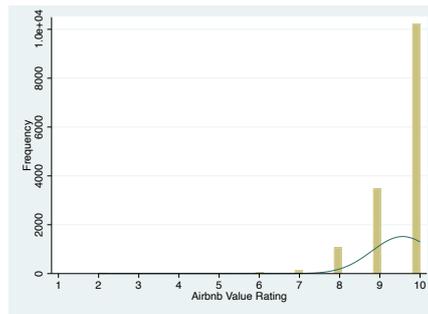
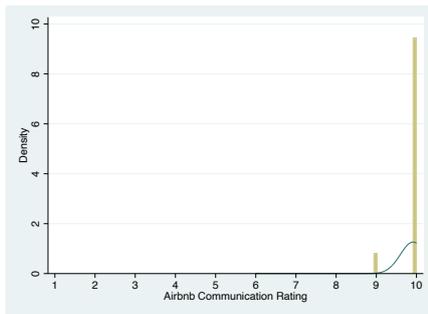
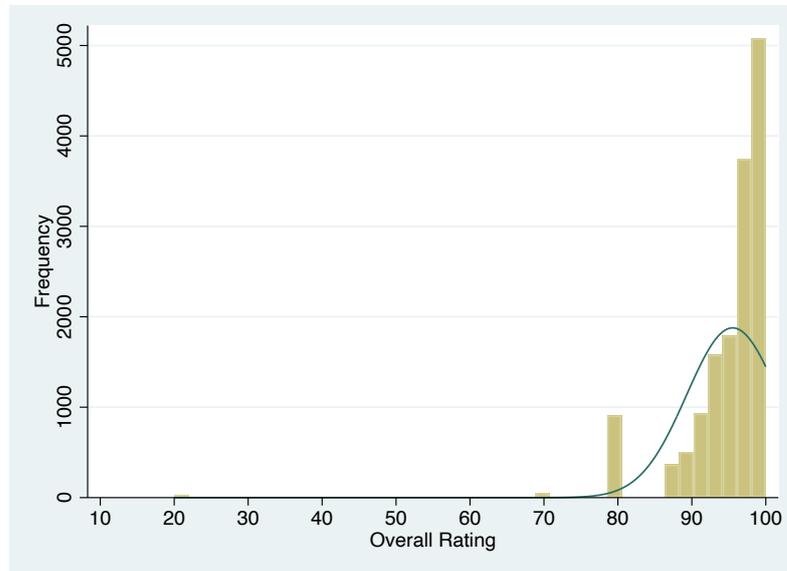
Tutti i voti sono distribuiti in un intervallo da 0 a 10, ad esclusione di “Overall Rating” che va fino ad un massimo di 100.

Affinché l'overall rating (valutazione complessiva) di un annuncio sia visibile, sono necessarie almeno tre recensioni. Una volta raggiunto questo numero, Airbnb calcolerà la media delle valutazioni e la mostrerà agli utenti sulla pagina dell'annuncio.

Le recensioni su questa piattaforma tendono a essere molto positive per una serie di motivi. Gli ospiti, in alcune occasioni, non scrivono appositamente recensioni negative per non compromettere il rapporto con l'host, come è stato confermato nell'articolo *Comparing Airbnb and traditional accommodation experiences using text-mining methods – the hedonic and utilitarian values framework* (2019).

Il risultato è che la media dei punteggi assegnati alle proprietà non coincide con la media matematica dell'intervallo dei valori disponibili, ma è notevolmente spostata verso l'alto. In ogni grafico il picco della normale è posizionato in prossimità nel valore massimo.

Grafico 6 Andamenti delle recensioni nelle proprietà di Airbnb.



Non si possiedono informazioni testuali sulle recensioni avvenute, ma per quanto riguarda le valutazioni numeriche si possono riassumere nella Tabella 3.

Tabella 3 Statistiche recensioni.

Tipologia valutazione	Osservazioni	Media	Deviazione Standard	Minimo	Massimo
Overall Rating	14,985	95.51992	6.212427	20	100
Airbnb Communication Rating	14,985	9.914648	.3157457	6	10
Airbnb Accuracy Rating	14,985	9.521655	.8523241	2	10
Airbnb Cleanliness Rating	14,985	9.559626	.79461	4	10
Airbnb Location Rating	14,985	9.834101	.4000112	6	10
Airbnb Value Rating	14,985	9.567501	.7689575	2	10

Tutte le valutazioni si aggirano intorno al 95% del valore massimo.

Nonostante questi valori elevati, il numero di persone che effettua una recensione alla fine del periodo di pernottamento non combacia con la totalità degli ospiti.

Ricordando che il numero di osservazioni a monte della pulizia del database è 33695, solamente 25415 possiedono delle valutazioni, ovvero il 75,43%.

6. ANALISI REGRESSIONE

Principi teorici

La regressione multipla è un'estensione della regressione lineare semplice, utilizzata per analizzare la relazione tra una variabile dipendente, in questo studio il prezzo, e due o più variabili indipendenti. Permette di determinare come la variabile dipendente viene condizionata da un insieme di variabili indipendenti contemporaneamente.

Gli obiettivi principali sono l'analisi della variabile dipendente a seguito dell'inserimento nel sistema di variabili indipendenti, prevedere i valori del prezzo degli affitti sapendo le variabili indipendenti e individuare quali di esse sono significative per dare una motivazione alla variazione della variabile dipendente.

La forma generale del modello di regressione multipla è:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

Fornendo una spiegazione per ciascuna variabile rappresentata, Y è la variabile dipendente, cioè il prezzo degli affitti da spiegare. X_1, X_2, \dots, X_k rappresentano le variabili indipendenti, le quali potrebbero condizionare la Y . Il valore che ci si aspetta possieda Y quando le variabili indipendenti sono zero è rappresentato da β_0 , l'intercetta. β_1 è l'effetto di un'unità aggiuntiva di X_1 su Y , quando tutte le altre variabili indipendenti rimangono costanti (ragionamento identico per β_2, \dots, β_k). Infine, ε è l'errore residuo, che rappresenta la porzione di variabilità del prezzo non spiegata dal modello.

Uno dei metodi più comuni per stimare i parametri in un modello di regressione lineare è il Metodo dei Minimi Quadrati Ordinari (OLS, Ordinary Least Squares). L'obiettivo dell'OLS è trovare i coefficienti di regressione che minimizzano la somma dei quadrati dei residui (o errori), cioè le differenze tra i valori osservati della variabile dipendente e i valori previsti dal modello di regressione.

L'errore residuo è la differenza tra il valore osservato di Y e il valore previsto \hat{Y} :

$$\text{Errore residuo} = e_i = Y_i - \hat{Y}_i$$

$$\text{Somma dei quadrati degli errori: } \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Il metodo OLS minimizza questa somma, che rappresenta la discrepanza tra il modello e i dati reali.

Per contrastare l'eteroschedasticità e ottenere stime corrette in Stata, si può utilizzare il comando *regress* con l'opzione *robust*, che calcola gli errori standard robusti all'eteroschedasticità.

La regressione fornisce anche il valore di R^2 , conosciuto anche come variabilità spiegata. È una misura che indica quanto bene il modello di regressione lineare riesce a spiegare la variabilità della variabile dipendente in base alle variabili indipendenti. La formula è la seguente.

$$R^2 = 1 - \frac{SSR}{SST}$$

Dove SSR rappresenta la somma dei quadrati dei residui, cioè la parte di variabilità non spiegata dal modello della variabile dipendente. SST è la somma totale dei quadrati, ossia la variabilità del prezzo a confronto della sua media.

Variabili

Successivamente all'operazione di pulizia del database si può eseguire un overview delle variabili rimanenti e dei loro valori statistici. Si sono stilate tutte le variabili interessate nell'analisi attraverso il comando *sum* di Stata, escludendo quelle testuali come "PropertyID", "Date", "BookedDate", "Status", "YEAR", "PropertyType", "PropertyManager", "HostType", "City", "Neighborhood", "PriceTier" e "Amenities".

Come emerge dalla Tabella 2 l'eliminazione di osservazioni, effettuata nei passaggi precedenti per ottenere un database più pulito possibile, ha causato l'annullamento totale delle variabili "campsite" e "guest_house". Il numero di osservazioni che

possiede un valore del “Response Time” è solamente 3814, si registra quindi il 77,93% di missing sul totale. Ci potrebbero essere degli errori nella selezione delle osservazioni, introducendo dei bias. Per questo motivo questa variabile è stata eliminata.

Tabella 4 Overview variabili.

Variabile	Osservazioni	Media	Deviazione Standard	Minimo	Massimo
PriceUSD	17,281	63.71981	35.36431	21	223
Latitude	17,281	44.81926	.0065192	44.80187	44.83699
Longitude	17,281	7.218089	.0383398	7.137733	7.299938
Bedrooms	16,967	1.439677	.7925151	1	6
Bathrooms	17,281	1.22924	.6824145	1	6
MaxGuests	17,281	3.757884	1.514303	1	9
ResponseRate	17,125	96.40368	12.30283	0	100
ResponseTime	3,814	14601.1	24927.93	295	86400
AirbnbSupe~t	17,281	.2775881	.447822	0	1
OverallRat~g	14,985	95.51992	6.212427	20	100
AirbnbComm~g	14,985	9.914648	.3157457	6	10
AirbnbAccu~g	14,985	9.521655	.8523241	2	10
AirbnbClea~g	14,985	9.559626	.79461	4	10
AirbnbLoca~g	14,985	9.834101	.4000112	6	10
AirbnbValu~g	14,985	9.567501	.7689575	2	10
wirless	17,281	1	0	1	1
parking	17,281	.6667438	.4713909	0	1
tv	17,281	.7694578	.4211921	0	1
washer	17,281	.7187663	.4496143	0	1
conditioner	17,281	.6305191	.4826782	0	1
elevator	17,281	.1442046	.3513073	0	1
coffee maker	17,281	.0510966	.2202011	0	1
fireplace	17,281	.4162954	.492958	0	1
libero	17,281	.2356924	.4244431	0	1
apartment	17,281	.7185927	.4496987	0	1
cabin	17,281	.3657774	.4816614	0	1
chalet	17,281	.0614548	.2401696	0	1
cottage	17,281	.0154505	.1233396	0	1
house	17,281	.0059024	.0766025	0	1
tiny_house	17,281	0	0	0	0
villa	17,281	.0938024	.2915621	0	1

bed_and_br~t	17,281	.0373242	.1895606	0	1
condo	17,281	.0620334	.2412233	0	1
farm_stay	17,281	.0484347	.2146892	0	1
guest_suite	17,281	0	0	0	0
tent	17,281	.1773046	.3819373	0	1
vacation_h~e	17,281	.0307274	.1725831	0	1
small_host	17,281	.0385394	.1925001	0	1
medium_host	17,281	.0094323	.0966638	0	1
big_host	17,281	.0538163	.2256614	0	1
villar_pel~e	17,281	.4965569	.5000026	0	1
torre_pell~e	17,281	.4097564	.4918029	0	1
luserna	17,281	.0936867	.2914008	0	1
bricherasio	17,281	.1832649	.3868947	0	1
economy	17,281	.4558764	.4980637	0	1
midscale	17,281	.2792084	.4486231	0	1
luxury	17,281	.0816504	.2738392	0	1
upscale	17,281	.1953012	.3964439	0	1
budget	17,281	.1885308	.3911467	0	1

Tutte le variabili presenti nella tabella sono state utilizzate nelle regressioni successive.

Regressione

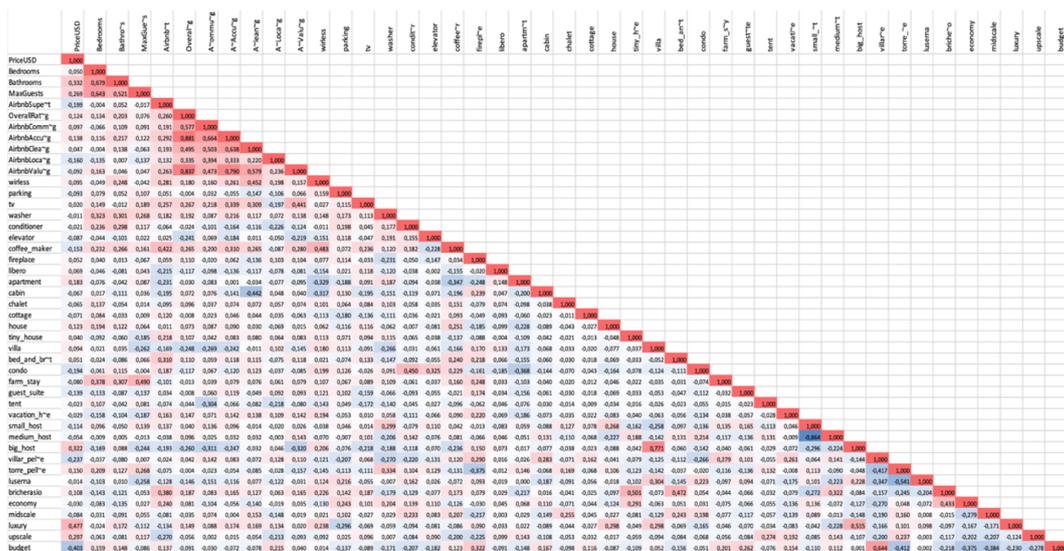
La prima fase di analisi consiste in una regressione lineare in cui il prezzo è la variabile dipendente mentre le restanti, comprese quelle generate in un secondo momento, sono variabili indipendenti. L'elevata numerosità di caratteristiche in gioco ha portato ad un approfondimento necessario sugli indici di correlazione. Perciò, attraverso il comando *correlate* di Stata, si è creata una matrice di correlazione, dove i numeri al suo interno possono variare da -1 ad 1, nello specifico:

- 1 indica una correlazione perfetta positiva.
- 0 indica nessuna correlazione lineare.
- -1 indica una correlazione perfetta negativa.

Nelle regressioni, se due variabili sono altamente correlate, possono causare problemi di multicollinearità. Essa può rappresentare un problema perché può rendere i coefficienti di regressione instabili e di difficile interpretazione.

Con una rappresentazione a colori si riescono a distinguere le variabili altamente correlate (colorate di rosso) da quelle scarsamente correlate (colorate di blu). Da questa situazione (Grafico 7) si esaminano gli elementi strutturali degli Airbnb; infatti, un indice di correlazione alto tra il prezzo ed una variabile può indicare una maggiore influenza sulla variabile dipendente. In una prima fase non sono state inserite tutte le variabili create. Nel grafico sottostante, sulla totalità di dummy appartenenti alle Amenities, vi sono rappresentate solo le più rilevanti, considerando sia elementi interni all'alloggio che esterni, come affermato dall'articolo *Determinants of Airbnb prices in European cities: A spatial econometrics approach (2021)*.

Grafico 7 Matrice di correlazione delle variabili.



In questa matrice è possibile identificare gli elementi strutturali maggiormente correlati con il prezzo. Questi coincidono con “luxury”, “Bathrooms”, “big_host”, “upscale”, “MaxGuests” e “apartment”. Quattro di queste (“luxury”, “big_host”, “upscale” e “apartment”) sono delle dummy che assumono valore 1 quando si presenta la rispettiva situazione. Bisogna prestare attenzione all’indice di correlazione tra le variabili stesse, qualora superasse il valore di 0,8 occorre evidenziare il dato e procedere con un’ulteriore verifica attraverso il VIF. Tuttavia, in questo caso, la correlazione equivalente a 0,5008 presente tra “big_host” e “luxury” rappresenta quella più elevata; quindi, non vi è la necessità di eliminare alcuna coppia di variabili dalla matrice.

Grafico 8 Correlazione specifica delle variabili della proprietà.

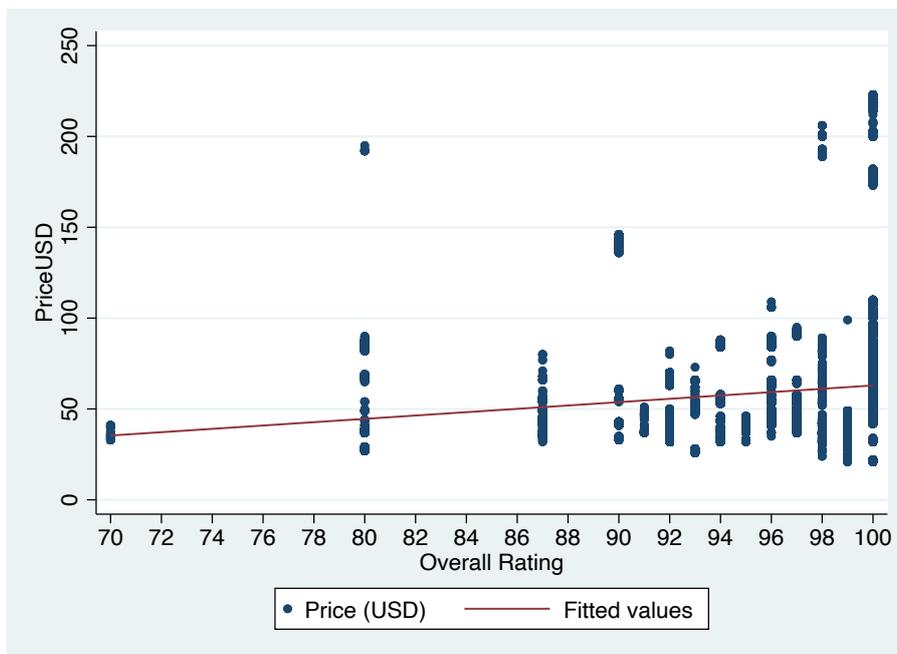
	luxury	Bathrooms	big_host	upscale	MaxGuests	apartment
luxury	1					
Bathrooms	0,0078	1				
big_host	0,5008	0,0048	1			
upscale	-0,164	0,0871	-0,137	1		
MaxGuests	-0,093	0,267	-0,226	-0,042	1	
apartment	0,0953	-0,109	0,0417	0,1351	0,0459	1

Da come si evince dal Grafico 8 sono state inserite ulteriori dummy per includere tutte le specifiche possibili, assimilabili dall’annuncio, relative ad un’osservazione. Tra queste si può trovare il comune di appartenenza (“City”: Torre Pellice, Villar

Pellice, Luserna San Giovanni e Bricherasio), la tipologia di struttura (“PropertyType”: Appartamento, Casa vacanze, Condominio...), il tipo di host (piccolo, medio o grande), lo stato (“Status”: A, B o R) e la tipologia per fascia di prezzo (“Price Tier”: budget, luxury, upscale, midscale ed economy).

Un ulteriore modo per verificare l’indice di correlazione è a livello grafico, attraverso uno scatter plot. Per esaminare la relazione tra il “PriceUSD” e una delle variabili con cui ha maggior correlazione, come ad esempio “OverallRating”, si utilizza uno scatter plot. Da come si evince dal grafico i punti tendono a disporsi lungo una linea ascendente; quindi, la relazione è considerata positiva. All’aumentare dell’overall rating assegnata alla struttura il prezzo incrementa.

Grafico 9 Scatter plot rappresentante la relazione tra PriceUSD e OverallRating.



Si sono esaminati tutti i servizi presenti nel database. Una volta elencati, si è svolta una prima regressione con quelli considerati più influenti sul prezzo: “Bathrooms”, “Bedrooms”, “elevator”, “parking”, “fireplace”, “wirless”, “coffee_maker” e “OverallRating”. L’esito della regressione è raffigurato nella Tabella 5, dove si può notare come la variabilità spiegata in questo primo modello è pari a 23,3% (più precisamente $Adj\ R\text{-squared} = 0.2326$). Sia per questa regressione che per quelle successive si è sempre verificato il VIF, il quale per ogni variabile deve essere

minore o uguale a 10 per evitare un'alta multicollinearità. Come da previsione le variabili “Bathrooms”, “fireplace”, “wirless” e “OverallRating” hanno un segno positivo, mentre “elevator”, “Bedrooms”, “parking” e “coffee_maker” influenzerebbero negativamente il prezzo. Questo comportamento si spiega momentaneamente con l'esclusione di variabili determinanti.

- *Posizione*

Si valutata l'influenza della posizione dell'alloggio sul prezzo. La Val Pellice è composta da dieci comuni, solamente in quattro di essi (Torre Pellice, Villar Pellice, Luserna San Giovanni, Bricherasio) si registrano degli Airbnb attivi. Ciascuna osservazione appartenente al database si trova quindi in uno di questi comuni.

Si sono quindi aggiunte le dummy relative a questi paesi nella regressione. I risultati mostrano come la dummy rispettiva a “Villar Pellice” sia stata omessa a causa di una collinearità troppo elevata, mostrando solo i dati degli altri tre comuni rimanenti. Quest'ultimi possiedono tutti un coefficiente altamente significativo e sono positivamente correlati al prezzo.

- *Tipologia alloggio*

Le strutture presenti nella valle sono caratterizzate da una discreta diversificazione. Per questo motivo sono state create numerose dummy, quindici in totale, per coinvolgere ciascuna tipologia presente e rendere l'analisi più completa possibile. Due di queste sono state omesse per motivi di alta collinearità, “cottage” e “campsite”. Tra le variabili non statisticamente significative “apartment”, “chalet” e “tent” hanno un segno positivo, mentre “house” e “vacantion_home” un segno negativo. Quelle statisticamente significative invece hanno tutte un'influenza negativa sul prezzo e coincidono con “guest_house”, “tiny_house”, “villa”, “bed&breakfast”, “condo”, “farm_stay” e “guest_suite”. Con l'aggiunta di queste dummy la variabilità spiegata segna un discreto incremento rispetto alla regressione precedente, arrivando ad un Adj R-squared = 0.4011.

Tuttavia, come spiegato nella sezione dell'analisi dei dati, si è deciso di sintetizzare questa moltitudine di variabili in 3 dummy: “apartment”, “condo” e “Proprietà_restanti”. Per questo motivo, nella tabella riassuntiva dei risultati appaiono solamente queste tre voci relative nel tipo di proprietà.

- *Host*

A seconda delle unità possedute si possono suddividere gli host in: small (1 Unit), medium (2-5 Units) e big (6-20 Units). Le dummy, per analizzare queste casistiche, sono state create ed aggiunte alla regressione per valutarne l'influenza. Gli utenti della piattaforma si dimostrano interessati a questi dettagli nel processo di prenotazione della proprietà; infatti, potrebbero essere condizionati da un host avente ulteriori alloggi.

I risultati mostrano come i “medium_host” siano stati omessi dalla regressione per collinearità troppo elevata, mentre “small_host” e “big_host” sono entrambi statisticamente significativi e con segno positivo. Sembrerebbe quindi che gli host con più proprietà abbiano maggiori possibilità di scegliere un prezzo maggiore rispetto ad host più piccoli, a causa di maggiore affidabilità garantita agli utenti.

- *Fasce di prezzo*

Per valutare l'impatto che hanno le diverse fasce di prezzo si sono inserite cinque dummy coincidenti a “Budget”, “Economy”, “Midscale”, “Upscale” e “Luxury”. Quando il loro valore è pari a 1 la struttura in questione si trova all'interno di quella rispettiva fascia di prezzo.

Come si poteva prevedere, le tre fasce più basse possiedono un coefficiente negativo, che s'incrementa man mano che la fascia di prezzo si abbassa. Quindi più sarà alto il prezzo di queste strutture meno sarà la richiesta di questi utenti, i quali si aspettano delle offerte che siano pari ai servizi offerti. “Upscale” è invece stata omessa per alta collinearità, mentre “Luxury” è l'unica ad avere un coefficiente positivo. Questo siccome strutture considerate di lusso vengono giustamente affiancate ad un prezzo elevato che spiega il loro status.

- *Cattura degli effetti fissi di mese*

Un aspetto da analizzare ed approfondire nel database sono gli effetti fissi. Questi rappresentano fattori sistematici in grado di introdurre variazioni in modo prevedibile, quindi non casuali ed invariabili nel tempo.

Tra le principali tipologie di effetti fissi, vi sono gli effetti temporali, geografici, individuali ed esterni. I primi, a seconda del periodo di riferimento, condizionano i dati (giornalieri, settimanali o mensili). I secondi, hanno caratteristiche differenti per quanto riguarda la localizzazione (città, regioni o paesi). I terzi, la variabilità è correlata a tipologie di categorie ben definite. Infine, gli ultimi sono caratterizzati da eventi economici o politici esterni.

La Val Pellice, situata ad un'altitudine media di 600 metri sopra il livello del mare, si aspetta accogliere turisti non solo nel periodo invernale, ma soprattutto in quello estivo, disponendo di diversi sentieri montani e panorami che potrebbero attirare l'attenzione di turisti e forestieri appassionati nel settore.

Al fine di ripulire il database da eventuali effetti stagionali, i quali potrebbero manifestarsi con un incremento del prezzo nei mesi estivi o invernali e con una diminuzione del prezzo nei periodi di bassa stagione, si è deciso di adottare la strategia della regressione. Si sono create undici dummy, una per ogni mese (non dodici perché una dummy viene omessa a causa di alta collinearità), dove il valore della variabile è 1 se il mese è gennaio, 2 a febbraio etcc. Così facendo si può notare come varia il prezzo in ciascun mese dell'anno. Il modello di regressione in questione è rappresentato nella colonna 6 della Tabella 2.

- *Cattura degli effetti fissi di giorno della settimana*

Un'altra alternativa a quella proposta precedentemente riguarda la cattura degli effetti fissi per giorno della settimana. Questo a causa della possibile presenza di differenze dei dati in occasione di festività nazionali o religiose. Nello specifico, il centro della comunità valdese in Italia è situato a Torre Pellice. In occasione della

festa dei valdesi, celebrata il 17 febbraio, è solito notare l'affluenza di turisti che giungono in Val Pellice per commemorare questo evento.

Per svolgere questa operazione, si è utilizzato il comando di Stata *dow ()*, il quale ottiene un giorno della settimana per ciascuna osservazione espressa in YYYY-MM-DD. Successivamente si sono create sette dummy, una per ciascun giorno della settimana ("1.weekday" corrisponde a lunedì, "2.weekday" a martedì ecc.). Infine, si è aggiunto in fondo alla regressione "i.weekday" in modo tale da escludere completamente una dummy tra le sette create come base di riferimento. I risultati di questo modello appartengono alla colonna 7 del Tabella 2.

- *Cattura degli effetti fissi di mese e di giorno della settimana*

L'ultima regressione riguarda il coinvolgimento di entrambi gli effetti fissi, sia di mese sia di giorno della settimana. Il risultato è una variabilità spiegata che si dimostra maggiore di tutti gli altri modelli precedenti, dovuto anche ad un numero di variabili maggiore.

Questo modello porta a diversi vantaggi rispetto all'analisi di due modelli separati, come il 6 e il 7. Modellare separatamente queste due casistiche potrebbe lasciare effetti residui non catturati.

- *Risultati*

Nella seguente tabella vi sono i risultati associati a ciascuna regressione descritta. Ogni colonna rappresenta una di queste, in cui per tutte le variabili utilizzate vi è il beta corrispondente, la varianza (standard error) e il p-value indicato con un numero di asterischi a seconda del livello di significatività: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Il significato di queste tre soglie si può riassumere così:

- $p < 0,01$: Evidenza molto forte contro l'ipotesi nulla.
- $p < 0,05$: Evidenza moderata contro l'ipotesi nulla.
- $p < 0,1$: Evidenza debole contro l'ipotesi nulla.

L'ipotesi nulla è la condizione in cui il coefficiente è uguale a zero con il rispettivo livello di significatività. Inoltre, al fondo di ogni colonna, ci sono il numero di osservazioni utilizzate per la rispettiva regressione e R-squared.

Tabella 5 Risultati regressione con PriceUSD.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABILI	Struttura	Posizione	Alloggio	Host	Fasce di prezzo	Effetti fissi mensili	Effetti fissi giorno della settimana	Effetti fissi di mese e di giorno della settimana
Bathrooms	44.06***	46.41***	51.59***	47.42***	46.12***	45.91***	46.11***	45.90***
	-1.416	-1.280	-1.214	-1.090	-1.035	-1.021	-1.035	-1.021
elevator	-8.323***	-10.12***	4.810***	9.880***	-8.164***	-7.576***	-8.159***	-7.570***
	(0.864)	(0.863)	(0.733)	(0.747)	(0.785)	(0.776)	(0.777)	(0.769)
Bedrooms	-13.08***	-15.02***	-18.59***	-16.92***	-8.465***	-8.316***	-8.460***	-8.311***
	(0.696)	(0.653)	(0.660)	(0.594)	(0.475)	(0.466)	(0.475)	(0.466)
parking	-7.550***	-11.44***	-10.97***	-10.11***	-4.353***	-3.898***	-4.358***	-3.899***
	(0.762)	(0.663)	(0.597)	(0.528)	(0.573)	(0.571)	(0.573)	(0.571)
fireplace	3.725***	10.78***	10.59***	10.16***	13.18***	13.03***	13.18***	13.03***
	(0.558)	(0.635)	(0.605)	(0.583)	(0.510)	(0.505)	(0.510)	(0.505)
wirless	5.551***	0.901	1.329**	-0.631	-5.087***	-4.546***	-5.087***	-4.549***
	(0.599)	(0.636)	(0.599)	(0.580)	(0.553)	(0.558)	(0.552)	(0.557)
coffee_maker	-20.25***	-17.08***	-9.089***	-1.107***	-6.056***	-6.097***	-6.051***	-6.086***
	(0.475)	(0.411)	(0.473)	(0.429)	(0.407)	(0.405)	(0.406)	(0.404)
OverallRating	0.485***	0.279***	0.0879***	0.505***	-0.193***	-0.170***	-0.193***	-0.170***
	(0.0491)	(0.0379)	(0.0311)	(0.0365)	(0.0363)	(0.0361)	(0.0362)	(0.0360)
o.villar_pellic e		-	-	-	-	-		
torre_pellice		21.70***	25.56***	23.58***	-12.26***	-13.04***	-30.85***	-30.74***
		(0.654)	(0.642)	(0.674)	(0.688)	(0.682)	(0.998)	(0.997)
luserna		13.44***	20.12***	15.60***	-11.87***	-12.29***	-30.46***	-29.98***
		(0.625)	(0.615)	(0.689)	(0.613)	(0.611)	-1.064	-1.067
bricherasio		33.61***	39.84***	48.93***	18.59***	17.69***		
		(0.692)	(0.746)	(0.670)	-1.101	-1.101		
o.apartment			-	-	-	-		
condo			-26.81***	-22.39***	-10.61***	-10.91***	-3.770***	-3.853***
			(0.654)	(0.608)	(0.652)	(0.641)	(0.736)	(0.728)

Proprietà restanti			4.640***	6.342***	6.841***	7.065***		
			(0.488)	(0.504)	(0.459)	(0.459)		
small_host				2.787***	3.718***	3.629***	-14.12***	-14.13***
				(0.501)	(0.599)	(0.600)	-1.025	-1.004
o.medium_host				-	-	-		
big_host				18.59***	17.83***	17.74***		
				-1.378	-1.212	-1.192		
AirbnbSuperhost				-18.65***	-0.904**	-0.774**	-0.907**	-0.782**
				(0.420)	(0.369)	(0.368)	(0.369)	(0.368)
economy					-33.89***	-33.46***	22.97***	23.02***
					-1.297	-1.284	-1.105	-1.098
midscale					-26.73***	-26.36***	30.14***	30.12***
					(0.916)	(0.908)	(0.680)	(0.679)
o.luxury					-	-		
upscale					1.931*	2.399**	58.79***	58.87***
					(0.995)	(0.989)	(0.981)	(0.982)
budget					-56.86***	-56.47***		
					(0.768)	(0.760)		
o.mese_1						-		-
mese_2						0.414		-4.305***
						(0.755)		(0.799)
mese_3						1.222*		-3.494***
						(0.735)		(0.780)
mese_4						1.269*		-3.496***
						(0.735)		(0.776)
mese_5						0.810		-3.852***
						(0.704)		(0.750)
mese_6						2.579***		-2.174***
						(0.732)		(0.774)
mese_7						5.844***		1.106
						(0.752)		(0.794)
mese_8						9.007***		4.337***
						(0.794)		(0.835)
mese_9						5.119***		0.316
						(0.777)		(0.822)
mese_10						1.417*		-3.253***
						(0.752)		(0.805)

mese_11						2.853***		-1.833**
						(0.767)		(0.821)
mese_12						4.817***		
						(0.807)		
villar_pellice							-18.59***	-17.70***
							-1.102	-1.101
o.bricherasio							-	-
apartment							6.835***	7.060***
							(0.459)	(0.459)
o.Proprietà_r estanti							-	
medium_host							-17.84***	-17.76***
							-1.211	-1.191
o.big_host							-	
luxury							56.84***	56.46***
							(0.768)	(0.760)
o.budget							-	
1.weekday							-0.405	-0.339
							(0.575)	(0.570)
2.weekday							-0.340	-0.345
							(0.574)	(0.570)
3.weekday							-0.360	-0.415
							(0.576)	(0.572)
4.weekday							-0.332	-0.447
							(0.576)	(0.572)
5.weekday							1.497**	1.428**
							(0.594)	(0.590)
6.weekday							1.714***	1.586***
							(0.594)	(0.589)
mese_1								-4.683***
								(0.808)
o.mese_12								-
Constant	-8.474*	-1.283	9.752***	-27.99***	77.84***	71.95***	57.16***	55.50***
	-4.642	-3.572	-2.907	-3.396	-3.563	-3.584	-2.861	-2.851
Observations	14,671	14,671	14,671	14,671	14,671	14,671	14,671	14,671
R-squared	0.252	0.329	0.394	0.454	0.651	0.657	0.652	0.658

La significatività rimane nel complesso relativamente costante. Il numero di osservazioni dipende dalle variabili che non possiedono un valore. Tuttavia, le osservazioni sono rimaste costanti per tutte e sette le regressioni, come da previsione. Siccome la pulizia del database è stata svolta al fine di eliminare tutte le strutture che non possedevano valori anche solo per una variabile. Inoltre, si può evincere come la variabilità spiegata cresca di regressione in regressione, come indica il valore espresso da R-squared. Solamente la penultima regressione effettuata registra un lieve calo in questo indice.

Mantenendo gli stessi parametri si è eseguita una regressione utilizzando la trasformata del prezzo come variabile dipendente, ovvero $\ln(\text{PriceUSD})$, esattamente come svolto negli articoli *Effects of location on Airbnb apartment pricing in Málaga (2020)* e *Real Estate Market Segmentation: Hotels as Exemplar (2018)*. Usare il prezzo direttamente può portare a problemi di eteroschedasticità, cioè una varianza non costante degli errori. La trasformazione logaritmica tende a comprimere le differenze tra valori molto alti e molto bassi, riducendo l'eteroschedasticità e migliorando la qualità delle stime.

Effettuando un confronto tra le due variabili dipendenti si può notare come siano presenti delle differenze di significatività. In particolare le variabili “small_host”, “AirbnbSuperhost” e “medium_host” diventano poco significative mentre “mese_10” diventa significativa e “5.weekday” altamente significativa. Per quanto riguarda “Bathrooms”, “fireplace”, “economy”, “midscale”, “upscale”, “apartment”, “luxury”, tutti i mesi ad eccezione di “mese_3” e i due giorni della settimana “5.weekday” e “6.weekday” continuano a contribuire positivamente al prezzo, esattamente come “elevator”, “Bedrooms”, “wirless”, “coffee_maker”, “Overallrating”, “torre pellice”, “villar pellice”, “luserna”, “condo” e tutti i giorni della settimana a meno dei due citati precedentemente hanno sempre un’influenza negativa sulla variabile dipendente. Inoltre, vi sono delle variabili che non hanno confermato la loro tipologia di influenza. In particolare, “parking”, “small_host”, “AirbnbSuperhost” e “medium_host”, oltre che perdere la loro significatività, acquistano un’influenza positiva. La situazione opposta non è stata registrata. Infine, si evince come la variabilità spiegata dal modello con la trasformata

logaritmica come variabile dipendente sia più elevata di quella avente il prezzo standard, di conseguenza “lnPriceUSD” fornisce una rappresentazione della realtà più veritiera rispetto a “PriceUSD”.

Tabella 6 Risultati regressione con lnPriceUSD.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABILI	Struttura	Posizione	Alloggio	Host	Fasce di prezzo	Effetti fissi mensili	Effetti fissi giorno della settimana	Effetti fissi di mese e di giorno della settimana
Bathrooms	0.374*** (0.0153)	0.416*** (0.0135)	0.477*** (0.0128)	0.463*** (0.0126)	0.441*** (0.0104)	0.435*** (0.0102)	0.440*** (0.0104)	0.435*** (0.0102)
elevator	-0.137*** (0.0134)	-0.157*** (0.0134)	0.0279** (0.0114)	0.114*** (0.0108)	-0.169*** (0.0106)	-0.159*** (0.0105)	-0.169*** (0.0104)	-0.158*** (0.0103)
Bedrooms	-0.0757*** (0.00825)	- 0.0957*** (0.00754)	-0.138*** (0.00795)	-0.144*** (0.00822)	-0.0152** (0.00615)	-0.0120** (0.00602)	-0.0151** (0.00615)	-0.0119** (0.00602)
parking	-0.0131 (0.00966)	- 0.0796*** (0.00855)	- 0.0726*** (0.00782)	- 0.0809*** (0.00758)	0.0317*** (0.00625)	0.0381*** (0.00622)	0.0316*** (0.00624)	0.0381*** (0.00621)
fireplace	0.00484 (0.00749)	0.0944*** (0.00792)	0.0953*** (0.00785)	0.111*** (0.00765)	0.152*** (0.00560)	0.150*** (0.00551)	0.152*** (0.00560)	0.150*** (0.00551)
wirless	0.0996*** (0.00890)	0.0250*** (0.00921)	0.0331*** (0.00887)	0.0357*** (0.00844)	- 0.0431*** (0.00648)	- 0.0335*** (0.00658)	- 0.0431*** (0.00646)	- 0.0335*** (0.00655)
coffee_maker	-0.263*** (0.00721)	-0.225*** (0.00627)	-0.124*** (0.00685)	- 0.0249*** (0.00624)	-0.108*** (0.00545)	-0.110*** (0.00540)	-0.107*** (0.00544)	-0.110*** (0.00539)
OverallRating	0.0102*** (0.000924)	0.00695** * (0.000752)	0.00455** * (0.000587)	0.00824** * (0.000635)	- 0.00337** * (0.000375)	- 0.00289** * (0.000365)	- 0.00337** * (0.000372)	- 0.00289** * (0.000363)
o.villar_pellice		-	-	-	-	-		
torre_pellice		0.288*** (0.00889)	0.335*** (0.00867)	0.343*** (0.00921)	-0.208*** (0.00994)	-0.220*** (0.00982)	-0.518*** (0.0104)	-0.516*** (0.0104)
luserna		0.201*** (0.00913)	0.282*** (0.00901)	0.267*** (0.00942)	-0.157*** (0.00864)	-0.164*** (0.00856)	-0.467*** (0.0109)	-0.459*** (0.0108)
bricherasio		0.552*** (0.00981)	0.630*** (0.0102)	0.762*** (0.00911)	0.310*** (0.0127)	0.296*** (0.0126)		-

villar pellice								-0.296***
								(0.0126)
o.apartment			-	-	-	-		
condo			-0.338***	-0.312***	-0.132***	-0.137***	-	-
			(0.00829)	(0.00802)	(0.00764)	(0.00743)	0.0479***	0.0492***
Proprietà_restanti			0.0692***	0.0733***	0.0846***	0.0883***		-
			(0.00777)	(0.00746)	(0.00605)	(0.00597)		
apartment								0.0882***
								(0.00597)
small_host				-0.00204	0.00867	0.00728	0.0190*	0.0188*
				(0.00654)	(0.00618)	(0.00617)	(0.0108)	(0.0105)
o.medium_host				-	-	-		
big_host				0.0497***	-0.0107	-0.0118		-
				(0.0179)	(0.0125)	(0.0123)		
medium_host								0.0114
								(0.0123)
AirbnbSuperhost				-0.276***	0.00335	0.00585	0.00329	0.00571
				(0.00597)	(0.00492)	(0.00491)	(0.00493)	(0.00491)
economy					-0.620***	-0.613***	0.327***	0.327***
					(0.0132)	(0.0130)	(0.0111)	(0.0110)
midscale					-0.477***	-0.471***	0.470***	0.470***
					(0.0102)	(0.0101)	(0.00865)	(0.00849)
o.luxury					-	-		
upscale					-	-	0.874***	0.874***
					0.0738***	0.0669***	(0.0114)	(0.0114)
					(0.0113)	(0.0113)		
budget					-0.947***	-0.941***		
					(0.00956)	(0.00948)		
luxury								0.941***
								(0.00950)
o.budget								-
o.mese_1						-		
mese_1								-
								0.0863***
								(0.0100)
mese_2						0.000624		-
						(0.00984)		0.0864***
								(0.0102)

mese_3						-0.00444		-	0.0914***
						(0.0102)			(0.0106)
mese_4						0.00500		-	0.0829***
						(0.00992)			(0.0103)
mese_5						0.00932		-	0.0766***
						(0.00935)			(0.00975)
mese_6						0.0464***		-	0.0412***
						(0.00939)			(0.00971)
mese_7						0.0940***			0.00665
						(0.00935)			(0.00960)
mese_8						0.129***			0.0429***
						(0.00961)			(0.00988)
mese_9						0.0665***			-0.0220**
						(0.00958)			(0.00984)
mese_10						0.0203**		-	0.0657***
						(0.00936)			(0.00967)
mese_11						0.0417***		-	0.0447***
						(0.00940)			(0.00971)
mese_12						0.0888***			
o.mese_12									-
						(0.0100)			
villar_pellice								-0.310***	
								(0.0127)	
o.bricherasio								-	
apartment								0.0845***	
								(0.00605)	
o.Proprietà_r estanti								-	
medium_hos t								0.0103	
								(0.0125)	
o.big_host								-	
luxury								0.947***	
								(0.00957)	
o.budget								-	-0.00823
1.weekday								-0.00935	(0.00712)
								(0.00722)	-0.00685

2.weekday							-0.00709	(0.00713)
							(0.00724)	-0.00751
3.weekday							-0.00742	(0.00715)
							(0.00725)	-0.00832
4.weekday							-0.00720	(0.00716)
							(0.00726)	0.0269***
5.weekday							0.0278***	(0.00737)
							(0.00747)	0.0288***
6.weekday							0.0305***	(0.00734)
							(0.00745)	3.824***
Constant	2.752***	2.886***	3.019***	2.711***	4.494***	4.396***	3.843***	(0.0304)
	(0.0866)	(0.0702)	(0.0548)	(0.0590)	(0.0387)	(0.0385)	(0.0310)	
Observations	14,671	14,671	14,671	14,671	14,671	14,671	14,671	14,671
R-squared	0.168	0.272	0.331	0.397	0.679	0.689	0.681	0.690

Calcolo residui

In una regressione lineare, i residui rappresentano le differenze tra i valori osservati della variabile dipendente (Y) e i valori predetti dal modello (\hat{Y}). I residui aiutano a comprendere dove e come il modello non riesce a catturare la relazione tra le variabili.

Il comando “*predict e_it, e*” in Stata viene utilizzato dopo una regressione panel per calcolare i residui idiosincratici del modello (e_{it}), che rappresentano la parte dell'errore che varia sia tra le unità (i) che nel tempo (t). Sono ciò che rimane dopo aver considerato sia gli effetti fissi o casuali (α_i) sia l'effetto delle variabili esplicative (X_{it}).

In un modello panel, la struttura dell'errore totale può essere decomposta come:

$$u_{it} = \alpha_i + e_{it}$$

Dove u_{it} rappresenta l'errore totale per l'unità i e il tempo t , α_i è l'effetto specifico dell'unità i ed e_{it} è il residuo idiosincratico. Quest'ultimo è la parte non spiegata del

modello dopo aver considerato gli effetti fissi o casuali α_i e le variabili esplicative X_{it} .

In sostanza questo errore varia sia tra la property che nel tempo. È una variazione casuale che non è spiegata dal modello. Sono quindi fondamentali per diagnosticare un modello e comprendere la variabilità non spiegata restante dopo aver considerato gli effetti principali del modello panel.

Per quanto riguarda il comando “*predict u_i, u*” viene utilizzato per calcolare gli effetti specifici stimati per ciascuna unità del pannello, cioè gli effetti fissi o gli effetti casuali a seconda del modello stimato.

In essenza, e_{it} e u_{it} rappresentano ciò che rimane del prezzo (che non si può osservare) dopo aver depurato di tutte le caratteristiche inserite nel modello 8. Rappresenta quindi, con buona probabilità, parte della componente del prezzo formatasi a causa di interazione strategica con annunci vicini.

Statistiche descrittive:

Per capire la variabilità dei residui e degli effetti specifici è utile fare una sintesi numerica che fornisca media, deviazione standard, minimo, massimo e percentili sia per e_{it} che per α_i .

Tabella 7 Sintesi numerica dei residui.

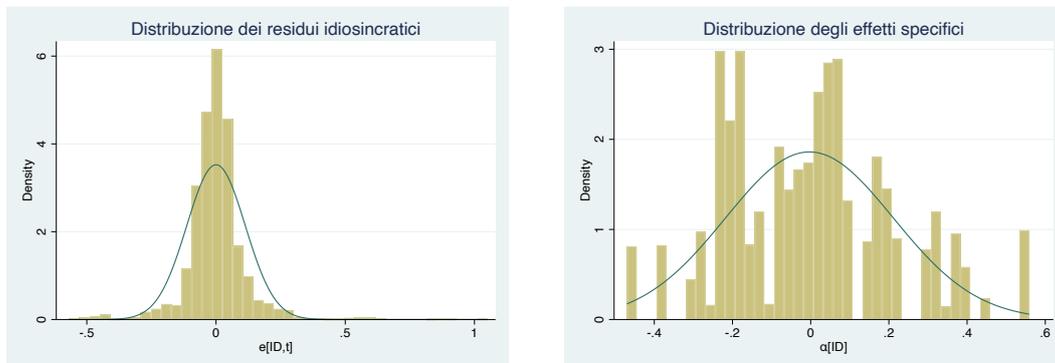
α_i				
Percentiles	Smallest			
1%	-.4697048	-.4697048		
5%	-.3061712	-.4697048		
10%	-.2365065	-.4697048	Obs	14,671
25%	-.1797969	-.4697048		
50%	-.0137625		Mean	-.003234
	Largest	Std. Dev.		.2144293
75%	.1057549	.5587688		
90%	.3277942	.5587688	Variance	.0459799
95%	.3635044	.5587688	Skewness	.3693845
99%	.5587688	.5587688	Kurtosis	2.982567

e_{it}				
Percentiles	Smallest			
1%	-.4272634	-.5673236		
5%	-.1514232	-.5673236		
10%	-.0948437	-.5582018	Obs	14,671
25%	-.0466727	-.5582018		
50%	.000324		Mean	-4.19e-11
	Largest	Std. Dev.		.1131855
75%	.043235	.9039142		
90%	.1031228	.9094551	Variance	.012811
95%	.1623925	.9194185	Skewness	.4718811
99%	.3134395	1.050.357	Kurtosis	12.47166

La kurtosis indica quanto una distribuzione è “appuntita” o “piatta” rispetto alla normale, mentre la skewness misura quanto una distribuzione è asimmetrica rispetto alla media.

Entrambi gli istogrammi dei residui, rappresentati nel Grafico 4, possiedono una distribuzione normale simmetrica rispetto alla rispettiva media; quindi, si esclude la possibilità di residui non normali. Le code delle distribuzioni appaiono coerenti con quelle della curva normale, questa situazione nel caso della rappresentazione degli effetti specifici non si presenta in maniera del tutto chiara. La coda in corrispondenza dei valori negativi risulta più spessa rispetto alla curva normale, indicando problemi di specificazione. Per quanto riguarda invece il picco registrato dalla distribuzione dei residui idiosincratici coincide con la normale, a differenza della distribuzione degli effetti specifici dove il picco si registra intorno al valore -2 anziché al valore zero come dimostra la curva normale. Un picco troppo basso potrebbe indicare dei residui sotto-dispersi.

Grafico 10 Distribuzioni dei residui rispetto alle curve delle normali.



Per controllare la possibile presenza di correlazioni temporali si è scelto di aggregare i residui a livello mensile, convertendo la data standard in un formato mensile e analizzando l’autocorrelazione sui residui mensili aggregati. Si è quindi calcolata la media mensile dei residui e_{it} per ciascun mese. Con il comando “*corrgram e_it*” si può visualizzare la correlazione con i lag, cioè i ritardi temporali. Occorre verificare i valori di autocorrelazione (AC) e della funzione di autocorrelazione parziale (PAC). L’esito di 4 lag aventi tutti valori negativi sia per AC che per PAC dimostra come i residui abbiano un comportamento anticorrelato

su più periodi temporali. Nello specifico un'AC negativa su più lag suggerisce un comportamento oscillatorio o ciclico nei residui, con un'alternanza sistematica tra valori alti e bassi. Inoltre, essendo le PAC negative su tutti e 4 i lag, potrebbe esserci un ciclo temporale più lungo non catturato dal modello. Concludendo, se i lag negativi non sono significativi è probabile che il pattern sia casuale e non richieda interventi specifici.

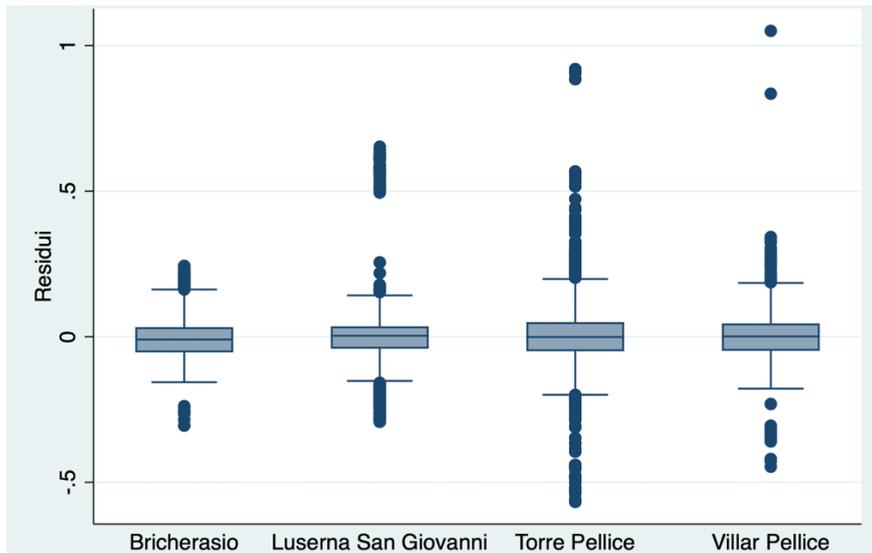
Figura 1 Analisi AC e PAC dei 4 lag.

LAG	AC	PAC	Q	Prob>Q	-1	0	1	-1	0	1
					[Autocorrelation]			[Partial Autocor]		
1	-0.0772	-0.1309	.09097	0.7629						
2	-0.1639	-0.2914	.54252	0.7624						
3	-0.2333	-0.4024	1.5588	0.6688						
4	-0.2608	-0.5281	2.9872	0.5600						

Si potrebbe applicare il test di Durbin-Watson per verificare ulteriormente l'autocorrelazione nei residui medi mensili. L'esito di questo test fornisce un valore pari a 1.67923, il quale essendo vicino a 2 dimostra l'assenza di autocorrelazione.

Un altro aspetto da considerare è la correlazione di residui che si può creare tra gruppi di comuni. In questi termini è stata svolta in prima battuta una rappresentazione di boxplot dei residui per ciascun singolo comune, mentre in un secondo step si è svolto un confronto tra coppie di comuni al fine di evincere dai grafici se i residui siano correlati tra i comuni o a gruppi di 2/3 comuni.

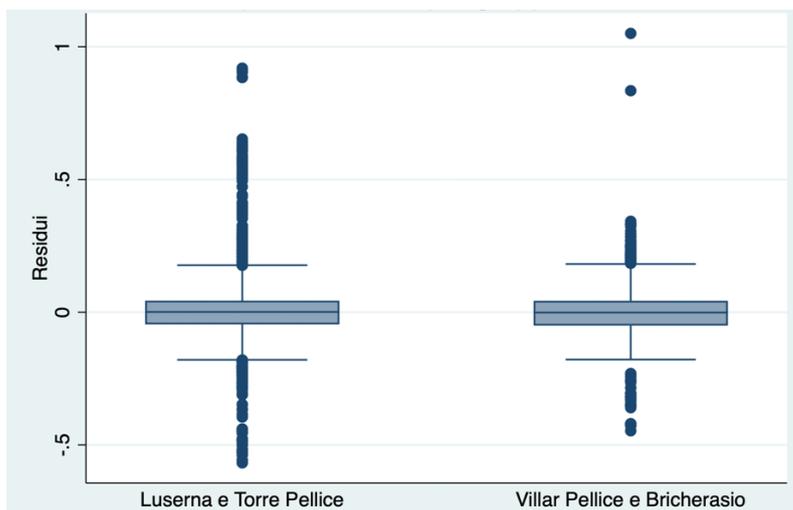
Grafico 11 Boxplot dei residui per comune.



Le caratteristiche grafiche da tenere sotto controllo per quest'analisi sono le differenze nelle medie dei residui, le differenze nella variabilità, gli outlier e la sovrapposizione delle distribuzioni.

Tra tutte le possibili coppie di comuni si è scelto di mettere a confronto quelle che possedevano più differenze a livello grafico.

Grafico 12 Boxplot dei residui per gruppi di comuni.



Come si può notare dai boxplot presenti nel Grafico 5 e nel Grafico 6 nessuna rappresentazione dei residui presenta una mediana, un intervallo interquartile o delle lunghezze delle linee che indicano il range significativamente diverso dalle

altre. Queste caratteristiche suggeriscono l'assenza di correlazione tra i residui sia tra singoli comuni sia tra coppie di comuni. Un'ultima osservazione riguarda gli outlier, ovvero dati che si discostano significativamente dalla maggior parte degli altri valori. Analizzando tutti i boxplot proposti si evince come nei comuni di Torre Pellice e di Villar Pellice ci sia maggiore dispersione sui prezzi, oltre che ad essere presenti outlier positivi che si discostano maggiormente dalle altre rappresentazioni. Ciò potrebbe essere spiegato dalla presenza di fattori non modellati e che influenzano il comune rispetto agli altri oppure da una sottostima da parte del modello dei valori in quel comune.

7. MERCATO D'INFLUENZA DI UNA PROPRIETÀ

Dati spaziali

L'ultimo passo determinante di questo elaborato si basa sull'influenza posseduta da un singolo Airbnb e dal conseguente effetto che può scaturire quest'ultimo sulle altre property circostanti.

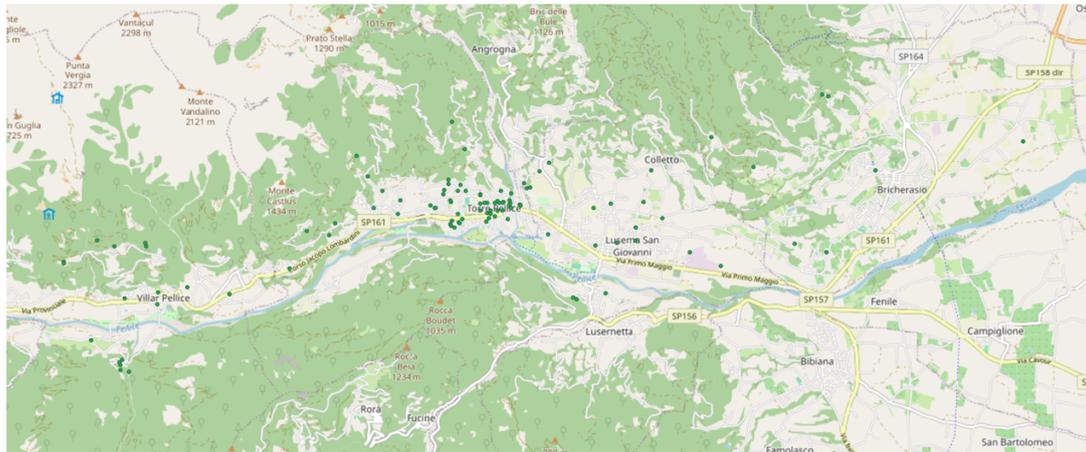
I dati spaziali prendono forma dalla combinazione di due fattori importanti. Il primo è costituito dalle informazioni sulle variabili tradizionali, esattamente come in qualsiasi altra analisi statistica. Quando si parla di analisi spaziale, questo concetto è insito nelle informazioni sugli attributi. In genere, questo è contenuto in una tabella in cui sulle righe vi sono le osservazioni e sulle colonne gli attributi. Il dataset di questo caso di studio è strutturato esattamente così. Il secondo aspetto è distinto e riguarda la localizzazione. Ciò consiste nella definizione di oggetti spaziali, classificati come punti, linee o rette. In sostanza, la caratterizzazione formale di qualsiasi oggetto spaziale si riduce alla descrizione delle coordinate x e y dei punti nello spazio. In questo elaborato, essendo coinvolte delle osservazioni, questo coincide semplicemente nelle loro coordinate (latitudine e longitudine).

Le tabelle di dati tradizionali non hanno problemi a includere le coordinate x e y come colonne, ma come tali non possono gestire la definizione del confine di unità spaziali irregolari. Di conseguenza, è necessaria una struttura dati specializzata, tipicamente contenuta in un sistema informativo geografico.

Nella Figura 2 sono rappresentate tutte le proprietà appartenenti al database completo. Attraverso il software GeoDa è stato possibile utilizzare la longitudine e la latitudine di ciascuna osservazione e rappresentarla sulla cartina utilizzando l'opzione BaseMap.

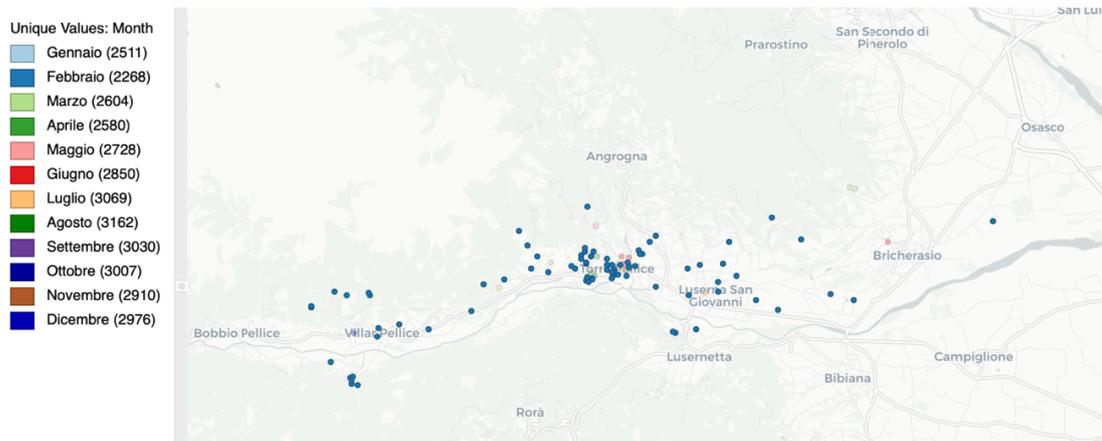
Dall'immagine si intuisce perfettamente dove ci sia la maggiore concentrazione di proprietà (prevalentemente a Torre Pellice e Luserna San Giovanni) e dove invece è più scarsa (Villar Pellice e Bricherasio).

Figura 2 Rappresentazione geografica delle località presenti in Val Pellice.



Siccome sono stati presi in considerazione tutti gli Airbnb nel corso dell'intero anno 2023 per raggiungere un numero di località adeguato alla ricerca, è normale notare come la stessa osservazione sia presente in più mesi. Nella Figura 3 vi è una rappresentazione temporale delle property, dove per ciascun mese è affidato un colore e un preciso numero di osservazioni.

Figura 3 Rappresentazione delle località presenti a febbraio 2023.



L'analisi eseguita sui residui sarà fondamentale per affrontare due questioni:

- Errori sistematici simili per annunci più simili
- L'autocorrelazione spaziale in un errore idiosincratico

In prima battuta occorre quindi capire se il residuo dell'annuncio è autocorrelato spazialmente; quindi, se annunci geograficamente vicini possiedono degli errori

sistematici simili tra loro. L'obiettivo finale è determinare il mercato di influenza di un singolo Airbnb. Questo perché qualora i residui di una proprietà non fossero minimamente influenzati dai residui di un'altra proprietà localizzata in un comune differente, significherebbe che i due comuni in questione sarebbero completamente separati come mercati. Facendo un esempio tra due proprietà presenti rispettivamente nei comuni di Luserna e Bricherasio, se i residui dell'Airbnb di Luserna fossero influenzati dalla property di Bricherasio e quindi fosse presente una correlazione tra ciò che rimane dei residui di una proprietà rispetto a ciò che rimane dei residui dell'altra proprietà, i due comuni si muoverebbero assieme come mercati. Di conseguenza un Airbnb di Luserna potrebbe competere con un Airbnb di Bricherasio.

Quest'analisi potrebbe considerarsi un primo step per una ricerca dai volumi più rilevanti, per scoprire ad esempio quale potrebbe essere l'area di influenza media di un singolo Airbnb sul territorio nazionale. Questo caso di studio può considerarsi più agevole, rispetto a quello nazionale, in quanto il numero di annunci è controllato.

È stata svolta un'analisi per verificare se i residui della regressione, nel caso degli annunci Airbnb, siano autocorrelati spazialmente (ovvero se gli errori sistematici siano simili per annunci vicini o simili). Si è utilizzato il software GeoDa.

Tuttavia, le property aventi lo stesso ID vengono ripetute più volte nel database a causa dell'arco temporale considerato (l'intero anno 2023). Al fine di trovare una soluzione a questo problema, è stata creata una tabella idonea al software. Per ciascuna property è stata calcolata la media delle seguenti variabili: "PriceUSD", "e_it" e "u_i". Inoltre, sono state inserite le variabili maggiormente rilevanti per il raggiungimento dello scopo su GeoDa. "Latitude" e "Longitude" sono fondamentali per identificare la presenza di autocorrelazione spaziale tra le abitazioni. Ulteriori variabili, quali "Bedrooms", "HostType" e "PriceTier", sono state comprese in Tabella 8 per effettuare altre analisi circostanti. Le osservazioni totali sono quindi 84, di seguito la rappresentazione.

Tabella 8 Rappresentazione property apposta per GeoDa.

PropertyID	PriceUSD	e it	u i	Latitude	Longitude	Bedrooms	HostType	PriceTier
abnb_1006337222011929423	42.97015			44,807827	7,1499924	1	1 Unit	budget
abnb_12949167	83.578651	.0001403	.123458	44,82276	7,22108	2	1 Unit	luxury
abnb_13422674	28.142458	-.0001999	-.3538884	44,82355	7,23034	1	6-20 Units	budget
abnb_14548470	41.836121	-.0002107	-.3114568	44,812801	7,137733	3	1 Unit	budget
abnb_14850565	66.818962	.0002027	.348823	44,82663	7,27568	1	2-5 Units	economy
abnb_16489864	85.651428	.0001413	.1222817	44,82037	7,22106	1	1 Unit	upscale
abnb_18017313	44.41573	-.0001045	-.1380078	44,81747	7,18632	2	1 Unit	midscale
abnb_18987991	41.737705	-.0000594	-.071608	44,80802	7,23952	2	2-5 Units	budget
abnb_20510257	40.711655	.0000469	.0756221	44,80868	7,1567	1	2-5 Units	budget
abnb_20992463	36.303371	.0000788	.1040379	44,80187	7,14326	1	1 Unit	budget
abnb_21195168	64.6763			44,81566	7,28381	1	2-5 Units	upscale
abnb_21521565	36.05817	-.0000301	-.0537138	44,80705	7,15646	1	2-5 Units	budget
abnb_29348723	139.9726	.0004079	.7361574	44,82426	7,22977	1	1 Unit	upscale
abnb_2974895	194.83908	.0007517	.3233292	44,82185	7,22682	4	2-5 Units	economy
abnb_31682555	46.521614	-.0000818	-.1402922	44,82151	7,22232	2	1 Unit	midscale
abnb_33129868	56.12162	-.0002238	-.2455805	44,81439	7,2629	1	2-5 Units	upscale
abnb_33919693	27.098726			44,81991	7,22313	1	2-5 Units	budget
abnb_34293946	43.058823	-.0007264	-.0610528	44,81987	7,20447	2	1 Unit	economy
abnb_34856112	84	.0004692	.4987126	44,80847	7,17088	1	1 Unit	midscale
abnb_36974897	35	-.0319468	-.1579434	44,8226	7,22104	1	1 Unit	budget
abnb_37894427	43.849316	-.0001687	-.0608745	44,81955	7,22387	2	1 Unit	economy
abnb_38443100	43.10955	-.0001098	-.1933348	44,82715	7,23477	1	1 Unit	economy
abnb_4273467	86.068497	-.0000929	-.1676563	44,82367	7,23094	1	6-20 Units	luxury
abnb_43728628	36.415092	.000751	.1967793	44,82126	7,229	1	1 Unit	budget
abnb_43967532	48.453201	-.0000166	-.0166726	44,82315	7,20137	2	1 Unit	midscale
abnb_44028351	44.914288	-.0001536	-.1860634	44,80771	7,24031	3	2-5 Units	budget
abnb_44069251	42.419674	-.0002041	-.3077008	44,81916	7,2265	1	1 Unit	economy
abnb_44135650	54.396397	.0001545	.2543485	44,81576	7,15414	2	2-5 Units	budget
abnb_46445865	58	-.0001934	-.0258206	44,82188	7,21508	1	6-20 Units	economy
abnb_46446487	69			44,82267	7,2137	2	6-20 Units	economy
abnb_46530077	37			44,82331	7,21373	1	6-20 Units	economy
abnb_46541012	96.696968			44,82156	7,2152	3	6-20 Units	economy
abnb_49289810	50.176472	.0002443	.349089	44,80942	7,16245	1	2-5 Units	budget
abnb_49526919	113.94475			44,81885	7,22222	6	1 Unit	upscale
abnb_49568381	91.984848	.0004252	.2774888	44,8213	7,24703	2	1 Unit	midscale
abnb_50122340	66.926025	.00004	.0722355	44,82076	7,24369	2	1 Unit	economy
abnb_50367087	60.573425	-.0005536	-.3913605	44,82314	7,21652	2	2-5 Units	upscale
abnb_50619713	47.185715	-.0001459	-.1010012	44,81999	7,22236	1	2-5 Units	midscale
abnb_50892207	36.569832	.0000483	.0854142	44,81927	7,25736	1	1 Unit	budget

abnb_51732372	54.330303	.0001625	.2650798	44,8153	7,15424	1	2-5 Units	budget
abnb_53188548	108.74157			44,81995	7,21659	2	1 Unit	luxury
abnb_53490301	33.710915	-.0000915	-.1534218	44,813114	7,137756	1	1 Unit	budget
abnb_570300681354966205	40.988201	-.0000318	-.0532221	44,8216	7,25358	1	1 Unit	budget
abnb_585145275354233618	31.338913	-.0002174	-.2568801	44,81525	7,14785	1	1 Unit	budget
abnb_600497917802236852	43.099998	-.0000595	-.1029901	44,81253	7,2691	1	2-5 Units	midscale
abnb_604622011754394228	61.031872	-.000162	-.2010508	44,808559	7,24606	1	1 Unit	luxury
abnb_6049306	41.766666	-.0016707	-.4955935	44,8252	7,19849	1	2-5 Units	economy
abnb_609291357984952628	37.60606	-.0000141	-.022951	44,81789	7,21594	1	1 Unit	economy
abnb_619098465366870259	45.213699	-.000038	-.0686055	44,81863	7,21675	1	2-5 Units	midscale
abnb_637569777313836413	74.21875	-.0005523	-.1747568	44,82475	7,21475	2	2-5 Units	luxury
abnb_6513591	44.962566	-.000148	-.1368599	44,82044	7,22301	1	1 Unit	economy
abnb_651851708364509881	32.519608	.0000581	.0293077	44,81618	7,14451	2	1 Unit	budget
abnb_658486555425954358	50.047619	-.0000451	-.0327756	44,819161	7,2173954	1	2-5 Units	midscale
abnb_6610380	62.040909	-5.60e-06	-.0060886	44,81893	7,21542	2	2-5 Units	upscale
abnb_665268980263429351	94.494316	.0001128	.1962228	44,82074	7,21186	1	2-5 Units	upscale
abnb_679576900540037767	68.036255	.0000705	.1153281	44,830902	7,267163	1	2-5 Units	economy
abnb_680294711040489227	54.157448	-.0000617	-.071647	44,81792	7,252164	1	1 Unit	midscale
abnb_681070629088186414	69.252052			44,82073	7,21213	1	2-5 Units	upscale
abnb_794983317713539168	83.019997			44,818494	7,2149767	1	1 Unit	upscale
abnb_7951792	51.012657	.0000577	.0900924	44,81607	7,25209	1	1 Unit	midscale
abnb_800040754274010781	61.195572			44,82613	7,25525	3	1 Unit	budget
abnb_807714420743251577	52.705521	-.0004009	-.3230949	44,82913	7,21791	2	2-5 Units	upscale
abnb_807917041484238720	103.85168	-.0001233	-.1273594	44,82114	7,2110514	2	1 Unit	luxury
abnb_824508099509848994	60.940559	.0001285	.1816703	44,821903	7,2050564	3	1 Unit	budget
abnb_850669219942746632	80.811249	.0001666	.2051287	44,83699	7,28929	1	2-5 Units	midscale
abnb_851524134349447462	69.318626	.0000629	.0634123	44,836682	7,2905519	1	2-5 Units	economy
abnb_855028845586010385	53.732826	-9.42e-06	-.0060993	44,82146	7,22188	1	1 Unit	midscale
abnb_855792735053106091	51.546585	-.0004941	-.3933312	44,82149	7,22392	2	1 Unit	midscale
abnb_874320875560720072	127.41969			44,821091	7,2291721	1	2-5 Units	luxury
abnb_874434916848113102	164.34555			44,821091	7,2291721	2	2-5 Units	luxury
abnb_886865111234962177	39.951725	.0000642	.0460421	44,82035	7,22453	2	2-5 Units	budget
abnb_909346006980708370	200.8212	.0005064	.3780829	44,82316	7,2252	1	6-20 Units	luxury
abnb_909346071203748511	192.76712	.0002345	.1692824	44,82164	7,2256	1	6-20 Units	luxury
abnb_909346140003672220	170.649			44,82151	7,22696	1	6-20 Units	luxury
abnb_909346204425371070	97.670731			44,82278	7,22712	1	6-20 Units	luxury
abnb_909346270889722448	93.389832			44,8215	7,22552	1	6-20 Units	luxury
abnb_909346334625722436	111.14286			44,82126	7,22693	1	6-20 Units	luxury
abnb_909977441283196248	75.132072	.0006897	.3614315	44,826108	7,2999384	2	1 Unit	economy
abnb_917271120768959883	37.914677	-.0002564	-.3714644	44,82815	7,1963621	1	1 Unit	economy

abnb_925855446707691264	75.82119	.0001059	.0790846	44,816899	7,1908074	1	1 Unit	upscale
abnb_930345382215916832	77.07547	.0014926	.3910927	44,815772	7,2482936	2	1 Unit	economy
abnb_931925372188932287	40.012047	.0000263	.0108069	44,821744	7,2244201	2	2-5 Units	budget
abnb_938143129399677928	41.522728			44,820372	7,2255341	2	2-5 Units	economy
abnb_947768490181431945	56.638657	-.0002849	-.1676219	44,820781	7,2272117	1	1 Unit	upscale

Pesi spaziali

I pesi spaziali giocano un ruolo cruciale nell'analisi trasversale della dipendenza spaziale. Sono un punto cardine della costruzione delle statistiche di autocorrelazione spaziale, poiché definiscono la struttura delle relazioni spaziali tra le unità geografiche o i punti nello spazio. L'autocorrelazione spaziale misura il grado in cui un valore di una variabile osservata in una data località è simile (o diverso) ai valori osservati nelle località vicine.

La definizione dei pesi spaziali avviene tramite una matrice (spatial weights matrix), comunemente indicata come W , che codifica le relazioni tra le unità geografiche. Ogni elemento w_{ij} della matrice specifica la "forza" della relazione spaziale tra l'unità i e l'unità j . Un esempio di matrice dei pesi per un sistema con quattro punti è il seguente.

$$W = \begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \\ w_{31} & w_{32} & w_{33} & w_{34} \\ w_{41} & w_{42} & w_{43} & w_{44} \end{bmatrix}$$

Il valore che assume l'elemento w_{ij} dipende dalla modalità con cui si vuole costruire la matrice, a seconda della natura del fenomeno studiato. In generale, i pesi spaziali sono diversi da zero quando i due punti i e j sono vicini, e zero altrimenti. L'intera diagonale della matrice è pari zero, siccome la relazione di auto-vicino è esclusa, quindi $w_{ii} = 0$.

La forma più basilare della matrice corrisponde a quella binaria, la quale esprime l'esistenza di una relazione vicina come una relazione binaria, con pesi 1 e 0. Nello

specifico, ciascuna unità spaziale w_{ij} è rappresentata da una riga I e i possibili vicini delle colonne J , con I diverso da J . La casistica in cui si presenta una relazione vicina tra l'unità spaziale corrispondente alla riga I e l'unica colonna corrispondente J segue poi come $w_{ji}=w_{ij}=1$. Con poche eccezioni, le analisi in GeoDa che impiegano pesi spaziali li utilizzano nella cosiddetta forma standardizzata per riga. La standardizzazione per riga assume i pesi ottenuti w_{ij} e li divide per la somma dei pesi della riga:

$$w_{ij(s)} = \frac{w_{Ij}}{\sum_j w_{Ij}}$$

Di conseguenza la somma di tutti i pesi standardizzati presenti nella riga è equivalente a uno. Inoltre, la somma S_0 di tutti i pesi all'interno della matrice corrisponde a n , il numero totale di osservazioni.

$$S_0 = \sum_I \sum_j w_{Ij}$$

I pesi all'interno della matrice possono essere basati su due logiche: sulla contiguità o sulla banda di distanza.

Il criterio di contiguità si fonda sulla definizione di regole aventi lo scopo di determinare quali elementi in una griglia o in una matrice sono considerati adiacenti (contigui).

La banda di distanza è la seconda logica presa in considerazione, sulla quale è stato scelto di applicare il caso di studio.

La matrice di distanza generale è fondata sulla differenza di posizione misurata tra gli annunci nello spazio multi-attributo, essendo presenti più criteri. A seconda del caso di studio in questione si può scegliere tra la distanza euclidea e la distanza di Manhattan. Siccome il database è già stato ripulito, non vi sono osservazioni che si distaccano in maniera anomala dalla media e i dati hanno una natura continua, oltre che degli attributi ben scalati. Di conseguenza è preferita la distanza euclidea, la quale rappresenta la distanza media "naturale". Nel caso le osservazioni avessero

avuto valori degli attributi molto diversi tra loro, la distanza di Manhattan sarebbe stata la scelta migliore.

Sono state individuate due casistiche, da cui derivano due matrici di pesi. Il Caso 1, la cui matrice dei pesi è presentata in Figura 4, è fondato sulla banda di distanza. In particolare, si sono selezionate nel database le variabili di Latitudine e Longitudine, la trasformazione di tipo *Raw*, la distanza euclidea come metrica e utilizzando una larghezza di banda minima.

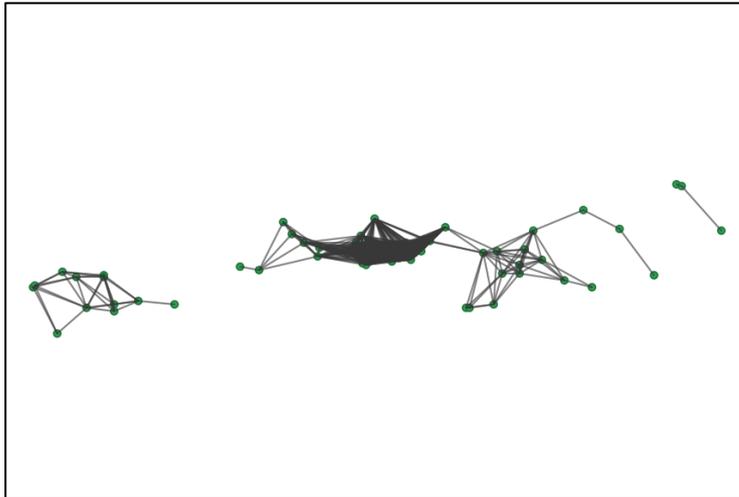
Figura 4 Matrice dei pesi basata sulla distanza (Caso 1).

Property	Value
type	threshold
symmetry	symmetric
file	GeoDa bsm.gwt
id variable	PropertyID
distance metric	Euclidean
distance vars	Latitude,Longitude
distance unit	unspecified
threshold value	0.0141392
# observations	84
min neighbors	1
max neighbors	45
mean neighbors	25.64
median neighbors	33.00
% non-zero	30.53%

Le proprietà dei pesi della banda di distanza possono essere ulteriormente studiate attraverso il grafico di connettività.

Il modello mostrato nel Grafico 13 evidenzia come la connettività suddivida i punti in tre sotto grafici interconnessi. Le diverse sottoreti non hanno alcuna connessione tra loro.

Grafico 13 Grafico di connettività, Caso 1.



La Figura 5 rappresenta invece il Caso 2, dove la metodologia scelta e la selezione delle variabili sono rimaste invariate. Tuttavia, il metodo utilizzato è variato in *K-Nearest*. Il numero di vicini considerato in questo modello è quattro.

Figura 5 Matrice dei pesi KNN (Caso 2).

Property	Value
type	k-NN
symmetry	asymmetric
file	GeoDa kn4.gwt
id variable	PropertyID
distance metric	Euclidean
distance vars	Latitude,Longitude
neighbors	4
# observations	84
min neighbors	4
max neighbors	4
mean neighbors	4.00
median neighbors	4.00
% non-zero	4.76%

Autocorrelazione spaziale

Al fine di raggiungere lo scopo di questo capitolo sono stati presi in considerazione 2 indici di correlazione spaziali: l'indice di Moran e LISA (Local Indicators of Special Association). Entrambi rappresentano degli strumenti utilizzati nell'analisi spaziale, ma hanno scopi diversi.

- L'indice di Moran:

Il suo obiettivo è misurare la correlazione spaziale globale, ossia quanto un fenomeno è distribuito in modo omogeneo o raggruppato nello spazio. Il valore può oscillare in un intervallo da -1 a +1. Valori positivi significano che aree con valori simili tendono ad essere vicine (aggregazione), mentre valori negativi indicano che aree con valori diversi tendono ad essere vicine (dispersione). Una distribuzione casuale si verifica in presenza di valori vicino a zero.

Si rivela utile quando bisogna conoscere la struttura spaziale significativa su tutto il territorio considerato di una variabile.

- LISA:

Il suo scopo è misurare la correlazione spaziale locale, individuando le specifiche aree che mostrano un cluster significativo. Questo indice è in grado di individuare quattro cluster differenti:

- High-High (HH): valori alti circondati da valori alti.
- Low-Low (LL): valori bassi circondati da valori bassi.
- High-Low (HL): valore alto circondato da valori bassi.
- Low-High (LH): valore basso circondato da valori alti.

Il suo uso si dimostra particolarmente efficiente, in quanto è in grado di rilevare aree con pattern particolari che potrebbero non essere evidenti con Moran's I globale.

La differenza sostanziale è che Moran's I fornisce una risposta sull'esistenza di una correlazione spaziale significativa in tutto il dataset, mentre LISA restituisce un chiarimento sulla presenza di aree con una correlazione spaziale significativa.

Moran's I:

La statistica I di Moran è con ogni probabilità l'indicatore più usato dell'autocorrelazione spaziale globale. Inizialmente scoperto da Moran (1948), fu reso popolare attraverso il lavoro sull'autocorrelazione spaziale di Cliff e Ord (1973). Si tratta di una statistica del prodotto trasversale tra una variabile e il suo ritardo spaziale, con la variabile espressa in deviazioni dalla sua media. Per

un'osservazione localizzato in I , questo è espresso come $z_i = x_i - \bar{x}$, dove \bar{x} è la media della variabile x .

La statistica I di Moran è data quindi da:

$$I = \frac{\sum_i \sum_j z_i w_{ij} * \frac{z_j}{S_0}}{\sum_i \frac{z_i^2}{n}}$$

Con w_{ij} elementi della matrice dei pesi spaziali, $S_0 = \sum_i \sum_j w_{ij}$ come la somma di tutti i pesi e n come il numero di osservazioni.

Di seguito sono rappresentati i grafici a dispersione del Caso 1 e del Caso 2.

Figura 6 Grafico di dispersione di Moran per i prezzi (Caso 1).

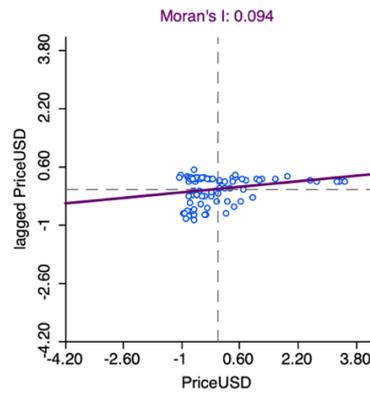


Figura 7 Grafico di dispersione di Moran dei residui (Caso 1).

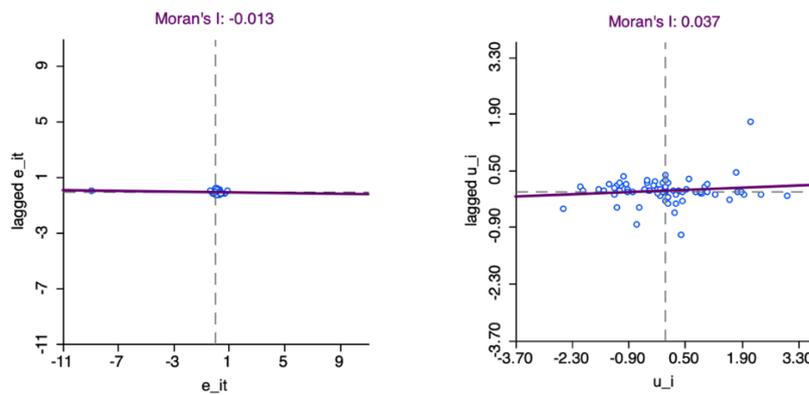


Figura 8 Grafico di dispersione di Moran dei prezzi (Caso 2).

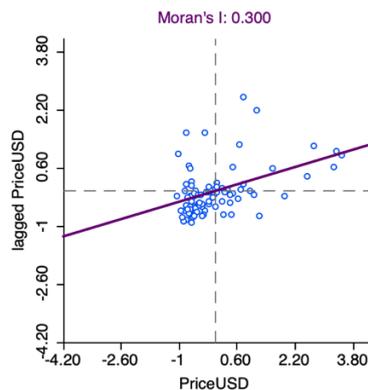
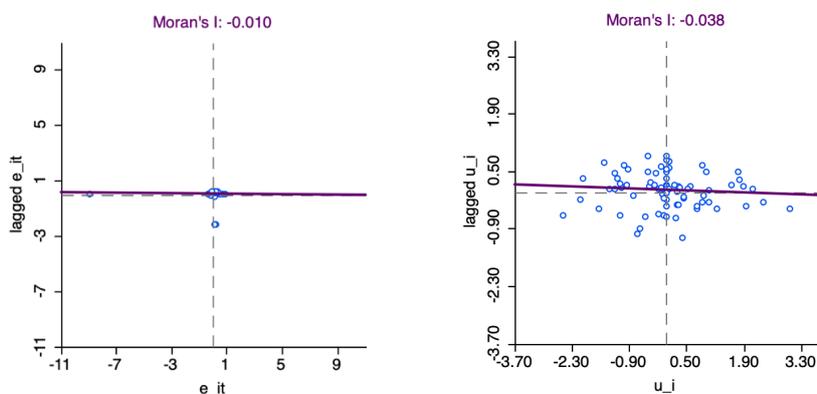


Figura 9 Grafico di dispersione di Moran dei residui (Caso 2).



Come mostrano i grafici, i prezzi in entrambe le casistiche possiedono una buona autocorrelazione spaziale; infatti, la retta di regressione assume una pendenza positiva indicando autocorrelazione positiva. I residui dei prezzi possiedono un'inclinazione pressoché piatta, evidenziando l'assenza di autocorrelazione spaziale.

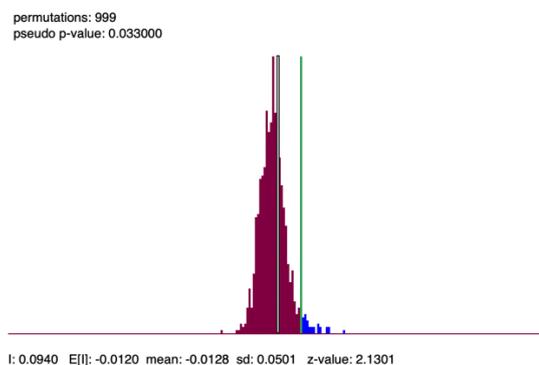
Questo significa che vi sono proprietà con caratteristiche simili collocate vicine; quindi, ciò che rimane è la decisione dell'host casuale del proprio prezzo.

Tuttavia, l'inclinazione della retta di regressione restituisce solo una stima dell'indice di Moran. Il significato della statistica non viene quindi colto con precisione, per questo motivo si utilizza l'opzione di randomizzazione di GeoDa. È una metodologia che consiste nel diminuire o eliminare i bias che potrebbero condizionare l'analisi.

Per considerare affidabile l'inferenza statistica si considerano almeno 999 permutazioni. L'esito di questa operazione è una distribuzione statistica, raffigurata da un istogramma.

L'ipotesi nulla in questa casistica consiste nell'assenza di un effetto o di una differenza significativa tra le osservazioni analizzate. Se il valore della statistica per i dati effettivi risulta diversa dalla distribuzione di riferimento, la situazione si dimostra a favore dell'ipotesi alternativa.

Grafico 14 Distribuzione di riferimento del prezzo, Caso 1.



Il valore di 0,0940 rappresenta la statistica per i dati effettivi, che si trova a destra della distribuzione di riferimento. Questo indica un rifiuto dell'ipotesi nulla.

Oltre al numero di permutazioni, in alto a sinistra, è anche fornito lo pseudo p-value. Quest'ultimo è il rapporto tra il numero di valori che sono maggiori o uguali del valore osservato. In questo caso le osservazioni che superano il riferimento sono 33, risultante del prodotto tra il numero di permutazioni e lo pseudo p-value (rispettivamente $999 \cdot 0,033$).

Grafico 15 Distribuzione di riferimento di e_{it} ed u_{it} , Caso 1.

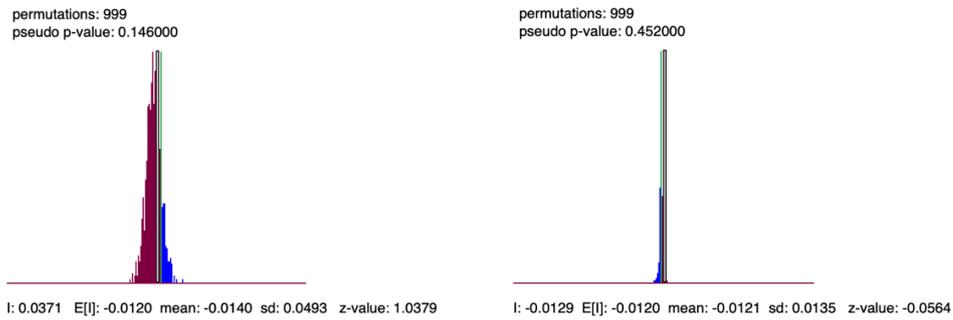
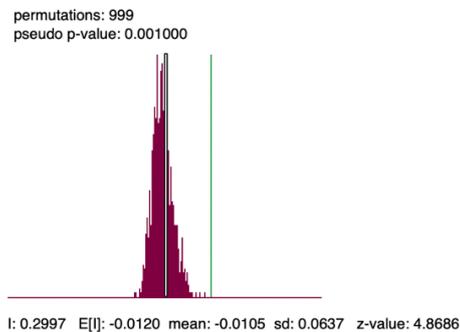


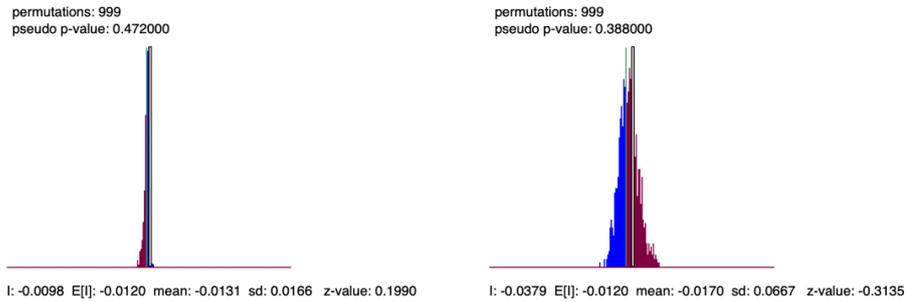
Grafico 16 Distribuzione di riferimento del prezzo, Caso 2.



I valori di riferimento sottostanti la figura possiedono un rispettivo significato. In questo caso:

- $I = 0,2997$ rappresenta il valore effettivamente osservato
- $E[I]$ è il valore teorico atteso. Si ricava mediante la formula $-I/(I-n)$, dove n è il numero di osservazioni, ovvero 84. Il valore -0.0120 è dato infatti da $-1/(84-1)$.
- Mean è la media della distribuzione di riferimento, la quale è estremamente simile al valore teorico atteso.
- Z-value è una misura statistica che misura quanto un valore si discosta dalla media di distribuzione. Uno valore positivo, come in questo caso, indica che i valori nelle vicinanze tendono ad essere simili.

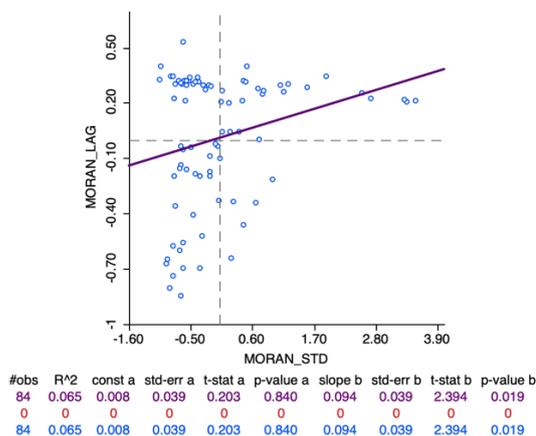
Grafico 17 Distribuzione di riferimento di e_{it} ed u_{it} , Caso 2.



I grafici sulle distribuzioni di riferimento confermano i risultati dei grafici di dispersione antecedenti; infatti, i prezzi dimostrano la presenza di autocorrelazione spaziale positiva. I residui, avendo un valore di z-value pressoché uguale a zero, evidenziano l'assenza di autocorrelazione spaziale significativa.

È possibile creare un grafico a dispersione di Moran come grafico a dispersione standard. A differenza del grafico specifico, le informazioni vengono visualizzate sotto il grafico. Considerando il Caso 1 di analisi del PriceUSD, Grafico 18, la pendenza è di 0,094 come mostrato già precedentemente.

Grafico 18 Dispersione di Moran di PriceUSD, Caso 1.



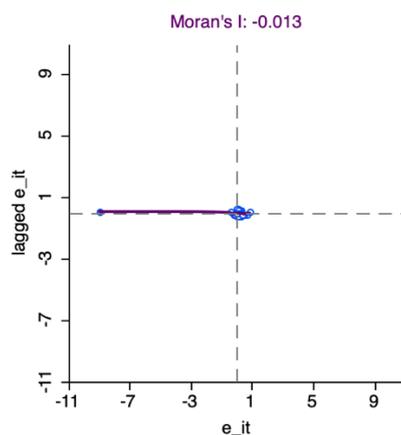
MORAN_STD rappresenta la variabile standardizzata, mentre MORAN_LAG il suo ritardo spaziale.

L'implementazione attuale della funzionalità del grafico di Moran in GeoDa ignora i valori specifici dei pesi e prende in considerazione solo la presenza di connettività

(cioè, pesi diversi da zero). Di conseguenza, nel calcolo del lag spaziale, tutti i vicini ricevono lo stesso peso.

Nel processo di costruzione dei grafici, fino ad ora, si è utilizzata una visuale “Linear Smoother”, in cui una retta di regressione interpola le osservazioni interessate restituendo un andamento del campione preso in considerazione. Esiste una visuale differente, chiamata “Lowess Smoother”, la quale permette di identificare potenziali rotture di struttura nei pattern di autocorrelazione spaziale. Se il grafico mostra un andamento inclinato e positivo, vi è una forte presenza di autocorrelazione spaziale. Altrimenti, se l’andamento si dimostrasse piatto, non ci sarebbe autocorrelazione. Di quest’ultimo caso ne sono una prova i residui di entrambi i casi citati precedentemente.

Grafico 19 Visuale Linear Smoother dei residui e_{it} , Caso 1.

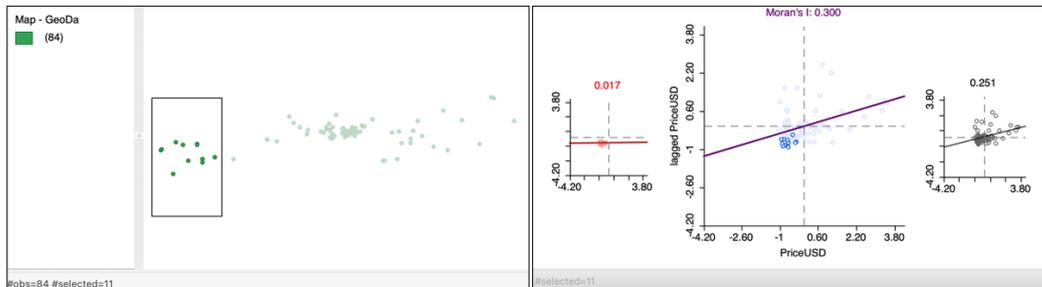


Da come si evince dal Grafico 19, il pattern mostrato da questa tipologia di residui è perfettamente piatto, marcando l’assenza di autocorrelazione spaziale dei residui e_{it} .

Un'opzione per valutare la stabilità dell'autocorrelazione spaziale in tutto l'insieme di dati, o per trovare una potenziale eterogeneità spaziale, è l'esplorazione del grafico a dispersione mediante *dynamic brushing*. Questa visualizzazione è in grado di rappresentare dei sotto-set di dati, raffigurandoli attraverso l'utilizzo di 3 grafici a dispersione di Moran (Grafico 20). Il grafico a sinistra evidenzia in rosso le osservazioni selezionate, il grafico a destra mostra in grigio quelle non selezionate,

mentre in centro vi è la rappresentazione comprendente la pendenza per tutte le osservazioni.

Grafico 20 Dynamic brushing PriceUSD di un primo campione, Caso2.

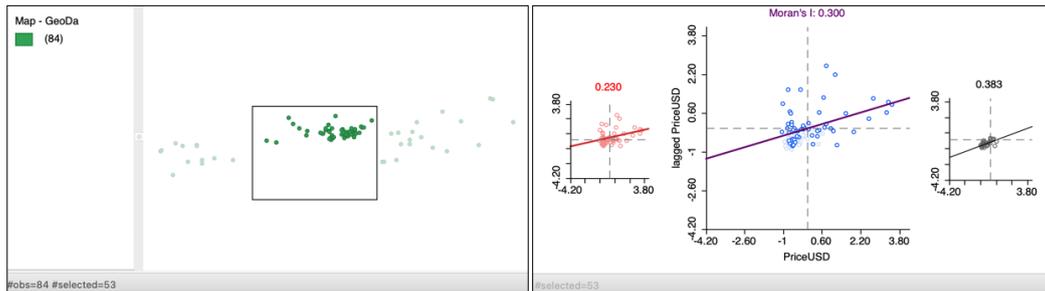


In prima battuta è stato selezionato un campione di 11 proprietà (incluso principalmente le proprietà di Villar Pellice), con il conseguente aggiornamento dei due grafici in rosso e in grigio, ai lati dello scatter plot di Moran. Per ciascuno dei sottoinsiemi di osservazioni, selezionati e non selezionati, i pesi spaziali sono corretti per rimuovere i collegamenti alle osservazioni appartenenti all'altro sottoinsieme. In altre parole, non esistono effetti marginali e ogni sottoinsieme è trattato come un'entità autonoma.

Nella nostra selezione iniziale, le osservazioni selezionate non mostrano alcuna autocorrelazione spaziale (un valore di 0,017), mentre il complemento (non selezionato) ottiene un I di Moran di 0,251, sostanzialmente lo stesso della statistica generale di 0,300. Ciò suggerirebbe una scarsissima autocorrelazione spaziale tra i prezzi di vendita nelle 11 osservazioni selezionate, mentre il resto dei dati mostra lo stesso modello dell'insieme.

In un secondo momento son state selezionate un numero di 53 proprietà (incluso osservazioni presenti a Torre Pellice e Luserna San Giovanni), ottenendo degli esiti differenti come mostra il Grafico 21.

Grafico 21 Dynamic brushing PriceUSD di un secondo campione, Caso2.

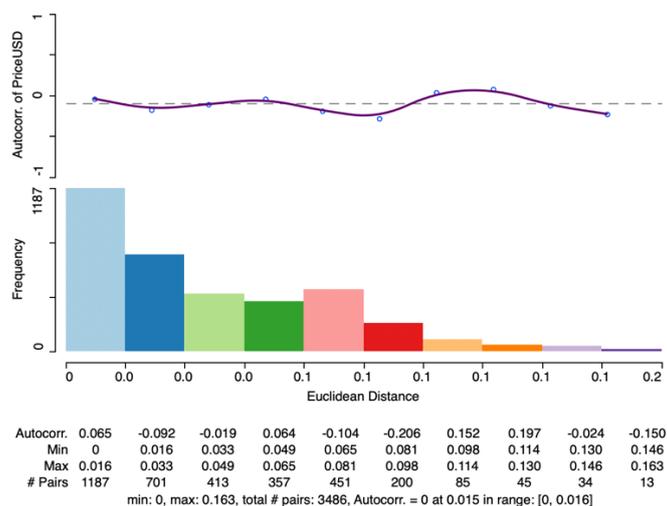


Questa seconda selezione mostra meno evidenza di eterogeneità spaziale, con un I di Moran di 0,230 per le osservazioni selezionate e 0,383 per il complemento.

Una rappresentazione geografica che consente di mettere in relazione l'autocorrelazione spaziale dei dati con la distanza spaziale tra le osservazioni è il correlogramma spaziale (Grafico 22).

A seconda del numero di “contenitori” scelto e la distanza massima dell’insieme, si ricava il numero di elementi in ogni contenitore. Vengono utilizzate tutte le coppie disponibili, le quali considerando le 84 osservazioni risultano essere $[84^2 - 84] / 2 = 3.486$ coppie.

Grafico 22 Correlogramma spaziale PriceUSD, Caso 1.



Il grafico più in alto nella figura mostra il correlogramma effettivo, il quale fornisce informazioni su come l'autocorrelazione spaziale cambia con la distanza. L'asse zero tratteggiato determina l'intervallo di autocorrelazione spaziale. L'intersezione

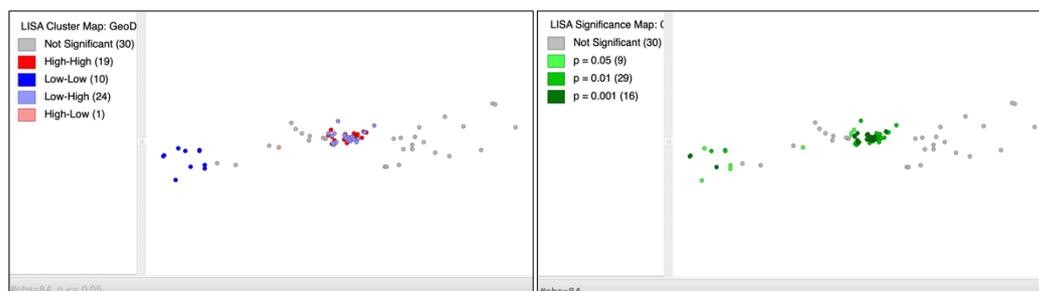
tra quest'ultimo e il correlogramma avviene inizialmente in corrispondenza del secondo intervallo. L'autocorrelazione fluttua attorno alla linea zero fino al quarto intervallo, dopodiché per i successivi due intervalli è negativa per poi assumere un andamento positivo, ripetendo questo comportamento successivamente. Vi è una regola empirica, la quale afferma che per ciascun contenitore devono essere presenti almeno 30 coppie, ciò non succede nell'ultimo contenitore, in cui sono presenti solamente 13 coppie.

Come discusso precedentemente, le statistiche sull'autocorrelazione spaziale sono studiate per rifiutare l'ipotesi nulla della casualità spaziale per favorire un'alternativa di *clustering*. Questo raggruppamento non fornisce un'indicazione sulla posizione dei cluster. Il concetto di un indicatore locale di associazione spaziale, o *LISA*, è progettato per rimediare a questa situazione.

Questo indicatore possiede due caratteristiche principali. La prima propone una statistica per ogni posizione con una valutazione del significato. La seconda restituisce una relazione proporzionale tra la somma delle statistiche locali e una statistica globale corrispondente.

La maggior parte delle statistiche di autocorrelazione spaziale globale può essere espressa come una doppia somma sugli indici i e j , come $\sum_i \sum_j g_{ij}$. La forma locale di tale statistica sarebbe quindi, per ogni osservazione (posizione) I , la somma dell'espressione rilevante sull'indice j , $\sum_j g_{ij}$.

Figura 10 Mappa del cluster e di significatività di PriceUSD, Caso 1.



La mappa dei cluster aumenta le località significative con un'indicazione del tipo di associazione spaziale, in base alla posizione del valore e al suo ritardo spaziale nel grafico a dispersione di Moran. In questo esempio, tutte e quattro le categorie sono

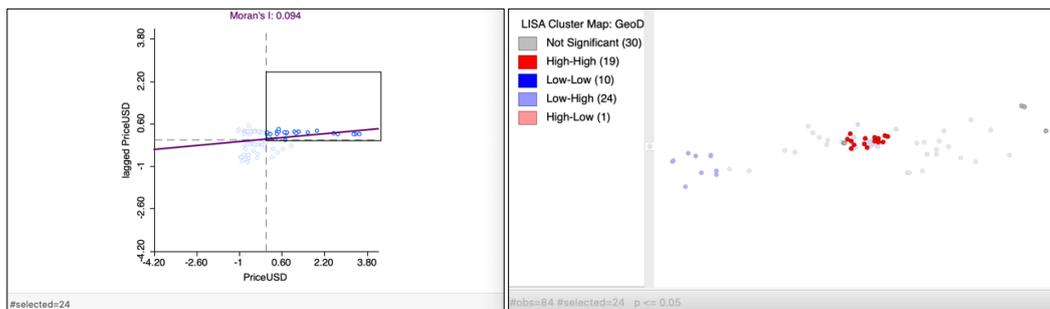
rappresentate, con il rosso scuro per i cluster alto-alto (diciannove in questo caso di studio), blu scuro per i cluster basso-basso (dieci località), blu chiaro per i valori anomali spaziali basso-alto (ventiquattro località) e rosso chiaro per i valori anomali spaziali alto-basso (una località). Si identifica quindi un'area centrale, costituita dai punti rosso scuro, dove è presente un cluster con valori elevati.

La mappa di significatività mostra le località con una statistica locale significativa, con il grado riflesso in tonalità di verde sempre più scure.

È inoltre possibile evidenziare la connessione tra il grafico a dispersione di Moran e la mappa del cluster. È importante tenere presente che vi è una differenza tra una proprietà (e il suo ritardo spaziale) che si trova in un dato quadrante della trama, e quella proprietà che è un cluster locale significativo.

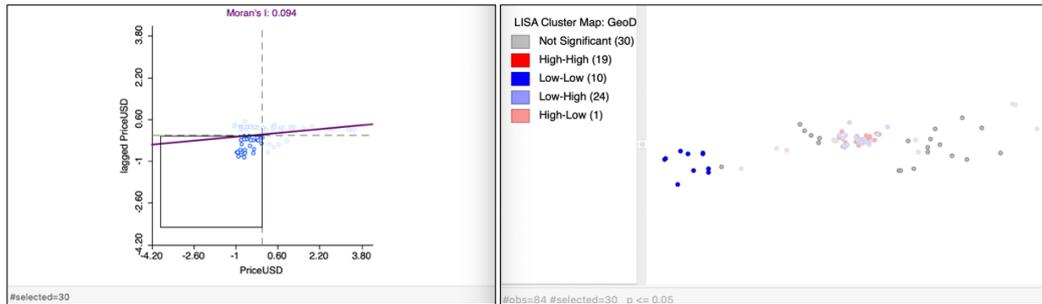
In Figura 11 si utilizza la funzione di collegamento, selezionando tutte le posizioni del quadrante superiore destro del grafico a dispersione di Moran. Esse sono immediatamente evidenziate nella corrispondente mappa del cluster nel riquadro a destra della figura. Mentre nel grafico a dispersione sono stati selezionati 24 punti, sulla mappa vi sono 19 località significative ($p < 0,05$).

Figura 11 Cluster and spatial outliers high-high, Caso 1.



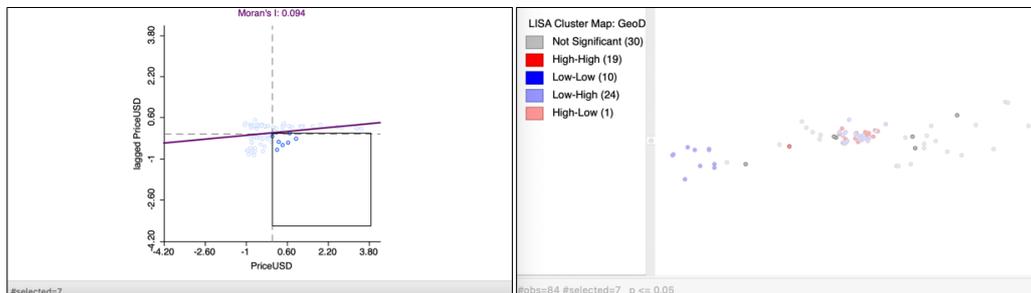
Si può applicare la stessa prospettiva ai punti nel quadrante inferiore sinistro del grafico a dispersione di Moran. In Figura 11, tutti i 30 punti selezionati sono mostrati nel pannello di sinistra. Nella mappa dei cluster a destra, solo 10 di queste osservazioni sono effettivamente significative, con un'ampia fascia di osservazioni nell'area centrale (Torre Pellice e Luserna San Giovanni) e ad est (Bricherasio) che non raggiunge quel livello.

Figura 12 Cluster and spatial outliers low-low, Caso 1.



Gli outliers spaziali sono osservazioni situate nei quadranti fuori diagonale del grafico a dispersione di Moran (Figura 13). I 7 outliers spaziali Low-High si trovano nel quadrante inferiore destro del grafico a dispersione di Moran. In questo caso, nessuno di questi valori è considerato significativo (tutti i punti grigi non sono significativi).

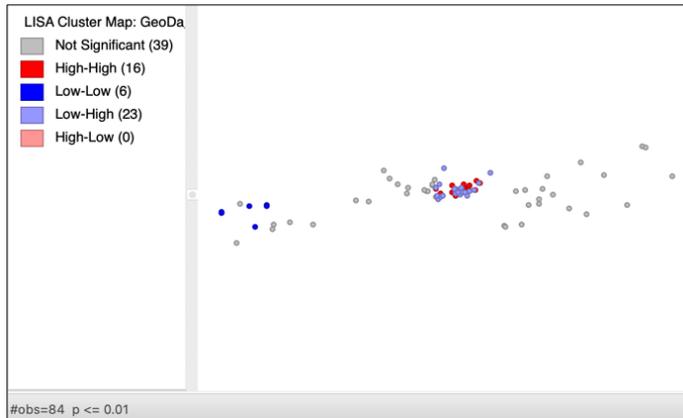
Figura 13. Outliers spaziali



Una conseguenza del calcolo degli pseudo valori p è il problema dei confronti multipli, come afferma l'articolo *Quantile local spatial autocorrelation (2019)*, i quali si presentano quando si effettuano più test nello stesso momento aumentando il rischio di falsi positivi. In questo caso di studio significherebbe riconoscere delle significatività che nella realtà non sono veritiere.

Non vi è tutt'ora una soluzione completamente efficace a questo problema, esistono solamente delle regolazioni. Ad esempio, si può cambiare il valore di p , adottando un filtro di significatività, fissandolo a 0.01, 0.001 oppure 0.0001. Adottando un valore di $p = 0.01$ il risultato è rappresentato dalla Figura 13.

Figura 14 Applicazione LISA di PriceUSD, con $p=0,01$.

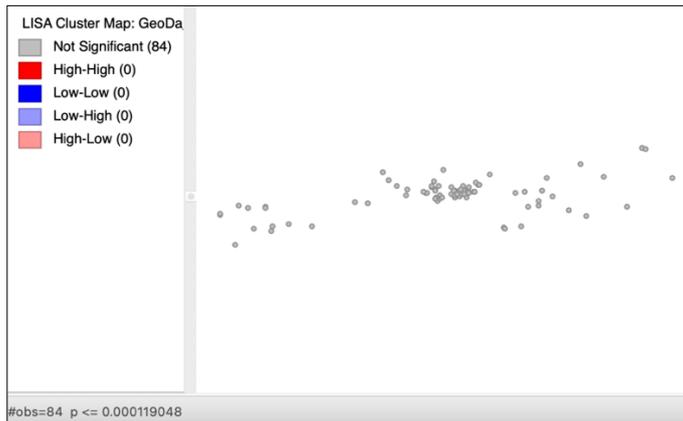


Si può notare come variando il valore di p a 0.01, il numero di punti significativi cala a 22. Precedentemente (Figura 13), con un pseudo valore- p pari a 0.05, il numero di osservazioni significative era pari a 29. Infine, abbassando notevolmente il valore di p a 0.001 e 0.00001, si ottengono rispettivamente significatività pari a 6 e 0.

Un metodo più accurato per la selezione del valore p è attraverso l'inferenza personalizzata del filtro di significatività. L'obiettivo è il fissaggio di un valore α come tasso di errore complessivo dell'indice I . Per l'analisi dei big data, questo dato potrebbe essere fissato su 0.1, come suggerito da Efron e Hastie (2016). In questo caso di studio, poiché le 84 osservazioni difficilmente costituiscono big data, si può mantenere il valore a 0.01. Questo perché avendo un numero ridotto di osservazioni, si sceglie un valore più basso per ridurre il rischio di falsi positivi.

Infine, un approccio più raffinato utilizzato per la scelta del corretto uso del valore p è legato a Bonferroni. Esso consiste nell'adottare come limite del valore p una soglia pari al rapporto tra α ed il numero di confronti multipli. Quest'ultimo, in questo caso, coincide al numero di osservazioni. A seguito di ciò, p è pari $0.01/84=0.000119$. Nessun'osservazione soddisfa questo criterio, per cui il suo pseudo valore p è inferiore al taglio (Figura 15), non risultando statisticamente significativa secondo questo criterio.

Figura 15 Mappa dei cluster, Bonferroni ($p<0.000119$).

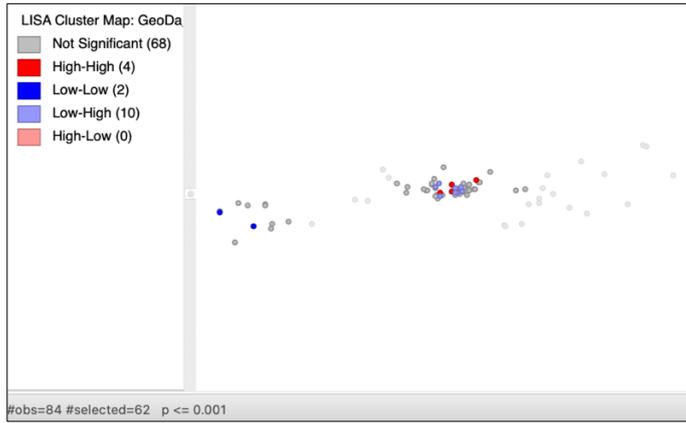


È successo ciò a causa dell'abbassamento della soglia di significatività, che permette di ridurre il rischio di commettere falsi positivi, ma di conseguenza alza il rischio di perdere rilevanza statistica.

Affinché la procedura di Bonferroni sia funzionale, occorre possedere un numero concreto di permutazioni in modo tale che il valore p sia inferiore ad α/n . Il software GeoDa è in grado di applicare un numero di permutazioni massimo di 99999; quindi, per raggiungere la significatività si deve verificare la condizione secondo cui $\alpha/n > 0,00001$. Considerando questo vincolo e le 84 osservazioni a disposizione, gli unici valori di α che permettono un risultato significativo sono 0.01 e 0.001. Inserendo $\alpha = 0.0001$ il vincolo non viene più rispettato. Tuttavia, considerando 0.01 e 0.001, il numero di osservazioni significative sulla mappa è pari zero. Di conseguenza, il metodo di Bonferroni in questo studio non permette di evidenziare nessun particolare interessante a causa della soglia di p troppo conservativa.

Escludendo il limite di Bonferroni, l'unico input di significatività che permette al tempo stesso di mostrare delle significatività e non avere un rischio di falsi positivi troppo elevato rimane $p=0.001$, come evidenzia la Figura 16.

Figura 16 Mappa dei cluster, $p < 0.001$.



8. CONCLUSIONI

Le conclusioni di questo lavoro di ricerca si concentrano sull'analisi dei risultati ottenuti, le metodologie adottate e le implicazioni degli esiti ricavati per la comprensione delle dinamiche dei prezzi e dei fattori che li influenzano.

Per ottenere risultati accurati e veritieri, è fondamentale che il database utilizzato per le analisi sia il più rappresentativo possibile dei servizi e della popolazione. A tal fine, è stata effettuata una pulizia approfondita del database, che ha comportato l'eliminazione del 46% delle osservazioni iniziali.

L'analisi delle valutazioni assegnate dagli ospiti alle strutture ha evidenziato che i punteggi ottenuti non corrispondono alla media matematica dell'intervallo disponibile, risultando notevolmente spostati verso l'alto. Ciò è probabilmente dovuto alla presenza di alcuni comportamenti che inducono gli ospiti ad evitare l'assegnazione di valutazioni negative, per non compromettere il rapporto con l'host.

Nel perseguire il primo obiettivo dell'elaborato, sono state selezionate le variabili indipendenti ritenute significative per la previsione del prezzo. Per iniziare, è stata realizzata una matrice di correlazione, utile a verificare la presenza di multicollinearità e per identificare le variabili coinvolte. Nessuna coppia di variabili è stata eliminata, in quanto nessuna ha superato la soglia critica di correlazione.

La fase successiva ha riguardato l'analisi di regressione, prima utilizzando la variabile dipendente "PriceUSD" e successivamente la sua trasformazione logaritmica, "lnPriceUSD". Le variabili indipendenti analizzate appartengono a diverse macroaree, tra cui la posizione, la tipologia di alloggio, le caratteristiche dell'host, le fasce di prezzo, e gli effetti fissi mensili e settimanali. Confrontando i modelli basati su queste due variabili dipendenti, è emerso che alcuni predittori, come il bagno, il camino, le fasce di prezzo (lusso, alta gamma, medio livello, economico), l'appartamento, e tutti i mesi eccetto marzo, continuano ad avere un impatto positivo sul prezzo. Al contrario, variabili come l'ascensore,

le camere da letto, il wi-fi, l'overallrating e il condominio hanno mostrato un effetto negativo sul prezzo. Infine, il parcheggio, l'host di piccole dimensioni, il super host e l'host medio, hanno mostrato variazioni di significatività rispetto ai modelli.

L'applicazione della trasformazione logaritmica ha portato a risultati significativamente migliori, con valori di R^2 e R^2 aggiustato più elevati, a indicare una maggiore capacità di spiegare la varianza della variabile dipendente. Inoltre, i risultati complessivi hanno mostrato percentuali più alte di spiegazione della varianza, confermando la validità dell'approccio adottato.

Nella fase successiva della regressione edonica, sono stati calcolati i residui per analizzare la presenza di autocorrelazioni temporali. L'analisi ha evidenziato una forte anti-correlazione dei residui su periodi temporali successivi, con un comportamento oscillatorio caratterizzato da un'alternanza sistematica tra valori alti e bassi. È emersa anche l'assenza di correlazione tra i residui, sia all'interno dei singoli comuni che tra le coppie di comuni. Per quanto riguarda gli outlier, l'analisi dei box plot ha rivelato che i comuni di Torre Pellice e Villar Pellice presentano una maggiore dispersione nei prezzi, con outlier positivi che si discostano significativamente dalle altre osservazioni.

Successivamente sono stati costruiti pesi spaziali per analizzare le statistiche di autocorrelazione spaziale. Nonostante siano state esaminate diverse configurazioni, come bande di distanza e k-nearest neighbors, i risultati sono stati coerenti. I grafici hanno mostrato un'autocorrelazione spaziale positiva per i prezzi, mentre i residui non hanno mostrato alcuna autocorrelazione spaziale significativa. Esaminando in dettaglio le singole aree all'interno della Val Pellice, utilizzando la tecnica di dynamic brushing, è emerso che gli alloggi di Villar Pellice presentano una scarsa autocorrelazione spaziale, mentre le proprietà di Torre Pellice e Luserna San Giovanni mostrano una minore eterogeneità spaziale.

Infine, è stato impiegato l'indicatore locale di associazione spaziale (LISA) per identificare più dettagliatamente la posizione dei cluster. La mappa dei cluster

ha evidenziato che le località di Torre Pellice hanno un livello di significatività significativamente più elevato rispetto a quelle di Villar Pellice, che invece formano un cluster con valori più bassi.

L'analisi ha fornito importanti indicazioni sul comportamento dei prezzi in relazione a vari fattori spaziali e temporali, evidenziando un'autocorrelazione spaziale positiva tra i prezzi delle strutture e la distribuzione geografica delle proprietà. L'assenza di autocorrelazione spaziale nei residui suggerisce che il modello ha efficacemente catturato la variabilità spaziale dei dati. Tuttavia, le differenze tra le aree della Val Pellice, con la bassa autocorrelazione spaziale a Villar Pellice, indicano che le dinamiche locali possono variare notevolmente. L'approccio utilizzato, potenziato dalla trasformazione logaritmica, ha dimostrato di essere più adatto a spiegare la varianza dei prezzi, con un miglioramento significativo della qualità predittiva. Nel complesso, i risultati sottolineano l'importanza di considerare variabili spaziali e temporali nell'analisi dei prezzi, nonché la necessità di affrontare la multicollinearità e le problematiche di autocorrelazione per ottenere modelli più robusti e affidabili.

9. VINCOLI DELLO STUDIO

Il presente elaborato ha esaminato un database composto da sole 84 osservazioni, ripetute nel tempo per l'intero anno 2023, con l'obiettivo di ottenere inferenze statistiche più robuste. Tuttavia, la riduzione del campione potrebbe comportare una limitata capacità di generalizzare i risultati a una popolazione più ampia, soprattutto considerando la dimensione relativamente ridotta del campione originale.

Le valutazioni assegnate dagli ospiti alle strutture sono risultate significativamente più alte rispetto alla media dell'intervallo di valori disponibile, suggerendo la presenza di un bias di risposta. Questo bias può aver distorto l'interpretazione delle reali preferenze degli utenti e, di conseguenza, aver influenzato i risultati delle analisi.

L'analisi dei residui ha rivelato la presenza di un comportamento oscillatorio tra i residui mensili, con un'anti-correlazione su più periodi temporali. Ciò suggerisce la necessità di un'ulteriore indagine sulle dinamiche temporali non catturate pienamente dal modello. Tuttavia, è emersa l'assenza di correlazione tra i residui dei vari comuni, un segnale positivo che indica che il modello ha adeguatamente considerato le differenze locali nella distribuzione dei prezzi.

Un aspetto cruciale emerso dal lavoro è stato l'uso dei pesi spaziali per analizzare l'autocorrelazione spaziale dei dati. Le differenze tra le aree, come la scarsa autocorrelazione spaziale a Villar Pellice rispetto a Torre Pellice e Luserna San Giovanni, suggeriscono che fattori locali, quali l'accessibilità o la presenza di attrazioni turistiche, possano influenzare in modo significativo la distribuzione dei prezzi.

Bibliografía

- Chica-Olmo, J., & González-Morales, J.G., JoséLuisZafra-Gómez J.L. (2019). Effects of location on Airbnb apartment pricing in Málaga. *Tourism Management*. <https://doi.org/10.1016/j.tourman.2019.103981>
- Goodman, A.C., & Thibodeau, T.G. (2002). Housing market segmentation and hedonic prediction accuracy. *Journal of Housing Economics*. [https://doi.org/10.1016/S1051-1377\(03\)00031-7](https://doi.org/10.1016/S1051-1377(03)00031-7)
- Beracha, E., & Hardin, W.G., Skiba, H.M (2017). Real Estate Market Segmentation: Hotels as Exemplar. *Journal of Real Estate Finance and Economics*. <https://doi.org/10.1007/s11146-017-9598-z>
- Tong, B., & Gunter, U. (2020). Hedonic pricing and the sharing economy: how profile characteristics affect Airbnb accommodation prices in Barcelona, Madrid, and Seville. *Current Issues in Tourism*, 25(20), 3309–3328. <https://doi.org/10.1080/13683500.2020.1718619>
- Ding K, Choo WC, Ng KY and Zhang Q (2023) Exploring changes in guest preferences for Airbnb accommodation with different levels of sharing and prices: Using structural topic model. *Front. Psychol*. <https://doi.org/10.3389/fpsyg.2023.1120845>
- Faye, B., (2021). A review of the problems and some proposal tested on Bordeaux city data. *Annals of Tourism Research*. <https://doi.org/10.1016/j.annals.2020.103079>
- Anselin, L., (2019). Quantile local spatial autocorrelation. *Lett Spat Resour Sci*. <https://doi.org/10.1007/s12076-019-00234-0>
- Klegr, T., (2024). Comparing Airbnb and traditional accommodation experiences using text-mining methods – the hedonic and utilitarian values framework. *European Journal of Tourism Research*. <https://doi.org/10.54055/ejtr.v38i.3299>
- Gyódi, K., Nawaro, L., (2021). Determinants of Airbnb prices in European cities: A spatial econometrics approach. *Tourism Management*. <https://doi.org/10.1016/j.tourman.2021.104319>
- Shabrina, Z., Arcaute, E., & Batty, M. (2021). Airbnb and its potential impact on the London housing market. *Urban Studies*. <https://doi.org/10.1177/0042098020970865>
- Hastie E., & Bradley and Trevor. (2016). Computer Age Statistical Inference. Algorithms, Evidence, and Data Science. *Cambridge University Press*.

<https://doi.org/10.1017/CBO9781316576533>

Sitografia

<https://www.airbnb.it>

<https://geodacenter.github.io/documentation.html>

<https://www.stata.com>

<https://www.researchgate.net/search/publication?q=Airbnb>

https://en.wikipedia.org/wiki/Airbnb?utm_source=chatgpt.com

https://buildd.co/marketing/airbnb-marketing-strategy?utm_source=chatgpt.com