

TITOLO INDAGINE

**ANALISI DELLE PROPRIETÀ DI UN IMMOBILE E DELLA
LORO INFLUENZA SUL PREZZO**

Marco Speciale

1. Introduzione

Nel biennio 2007- 2008 è scoppiata una grave crisi economica di portata nazionale e mondiale, che ha destabilizzato i mercati fino al 2013, in seguito allo scoppio di una bolla immobiliare negli Stati Uniti D'America. Gli effetti di quella che è stata chiamata "La grande recessione" non sono stati arginati, nonostante un saliscendi di crolli e riprese, fino ad intaccare l'economia dei nostri giorni. La situazione è stata ulteriormente aggravata dalla Pandemia globale di Sars-Cov-2, che, dalle strade di Wuhan, si è diffusa fino in Occidente. Molte famiglie, già provate dal rincaro delle materie prime e da una crisi creditizio-bancaria, hanno assistito ad un netto taglio del loro reddito, riuscendo con difficoltà ad acquistare immobili. In un periodo storico in cui un'eccessiva offerta coesiste con una contrazione della domanda, la figura professionale dello Statistico acquisisce sempre più rilevanza, poiché la sua conoscenza tecnica permette di analizzare e trarre risposte da innumerevoli quantità di dati utili nel guidare il cliente verso una scelta di acquisto consapevole.

I dati di cui il team di lavoro è in possesso, e che fornisce questo studio, sono relativi al comune di Palermo; si dimostrano, dunque, non generalizzabili per l'intero territorio nazionale, soprattutto se si tiene conto della grande varietà socioeconomica della Penisola.

Tale analisi si concentra sull'incidenza che i costituenti di un immobile esercitano sulla determinazione di un prezzo. Oggetto di particolare attenzione gli elementi maggiormente presi in considerazione da un ipotetico acquirente, quali zona di locazione dell'appartamento, piano all'interno del sistema edificio, quantità di vani e bagni a disposizione.

Attraverso il reperimento e l'osservazione di una quantità di annunci immobiliari sono state estratte delle informazioni che fanno riferimento al nome e la descrizione dell'immobile.

2. Estrazione e pulizia dei dataset

I nostri elementi di indagine sono stati ottenuti dal sito web Immobiliare.it, utilizzando l'estensione **selector gadget** di Google Chrome.

Per ricavare i dati ed analizzarli è stato utilizzato il software di programmazione "R".

I dataset estratti sono i seguenti:

- Dataset sugli appartamenti in vendita per intero in ottimo stato nel centro storico di Palermo
- Dataset sugli attici in vendita nel comune di Palermo
- Dataset dell'intersezione tra i due dataset sopracitati
- Dataset di un appartamento scelto tra quelli comuni

Dopo aver tratto le singole variabili ed averle unite in un dataframe, si è provveduto a verificare la presenza di eventuali ripetizioni nelle osservazioni.

Nel primo dataset sono state riscontrate 6 ripetizioni a fronte di 164 osservazioni. In prima analisi si è deciso di sfruttare la funzione **unique**, che elimina le ripetizioni. Tuttavia, riesaminando il dataset, si è notato che le ripetizioni non sono state rimosse del tutto, perché due di queste racchiudevano una **Descrizione** leggermente diversa dalle altre a livello testuale. Dunque, si è proceduto a rimuovere le suddette ripetizioni in eccesso tramite un'assegnazione di righe al dataset differente dal formato originale. Inoltre, la funzione **unique** ha modificato la numerazione delle osservazioni non in maniera immediatamente visibile, restituendo un'osservazione diversa rispetto a quella immessa nel codice; per rendere chiaro il concetto se ne riporta un esempio:

	Nome	Prezzo	NumeroLocali	NumeroBagni	Piano
51	Bilocale via del Celso 86, Tribunale, Palermo	100000	2	1	1
52	Bilocale piazza dei Tedeschi, Cassaro, Palermo	75000	2	1	1

Si è ritenuto opportuno riordinare le unità statistiche.

Il procedimento riportato è stato utilizzato pedissequamente per la creazione del **dataframe** degli attici. In questo dataset sono state riscontrate 3 ripetizioni a fronte di 116 osservazioni.

Il dataset, composto dall'intersezione dei due precedenti, deriva dall'impiego della funzione **semi join** che prende in considerazione tutte le osservazioni che nel primo dataset (appartamenti) hanno una corrispondenza nel secondo (attici).

In virtù di ciò è stata scelta una tipologia di appartamento abbastanza comune per il cliente medio.

Per effettuare questo studio è stato scelto come oggetto d'indagine l'attico sito in Via Giuseppe Garibaldi, del quartiere Kalsa di Palermo.

Al fine di verificare che i dataset avessero valori validi, sono state applicate delle regole, con particolare riferimento alle variabili **piano – numero locali – numero bagni – prezzo**; per tutte le variabili numeriche è stata posta la condizione che fossero strettamente positive, che il prezzo fosse maggiore del numero di bagni e locali, e che il piano rispecchiasse gli standard del sito. (grafico regole → figura 1 in appendice)

Nel dataset appartamenti sono stati riscontrati 7 errori (come si evince dal grafico). Tali valori risultano plausibili, ne consegue che gli errori evidenziati siano dovuti alla presenza di NA; In virtù di tale constatazione il dataset non è stato ulteriormente modificato.

(grafico violazione regole → figura 2 in appendice)

Nel dataset attici non sono state riscontrate violazioni alle regole.

(grafico violazione regole → figura 3 in appendice)

3. Analisi esplorativa

Al fine di verificare quali fattori abbiano inciso maggiormente sulla determinazione del prezzo, si è proceduto con svariate rappresentazioni grafiche, dalle quali si evince che nella distribuzione del prezzo di entrambi i dataset (appartamenti e attici) vi è la presenza di un valore anomalo per ciascuno di essi.

Nel dataset appartamenti il valore anomalo corrisponde a 3.7 € milioni (MLN) .

Nel dataset attici il valore anomalo corrisponde invece a 2.2 € milioni (MLN).

Per ogni aspetto degli immobili sono state calcolate misure sintetiche con presenza e assenza di valore anomalo.

Per iniziare si è svolta l'analisi della variabile **Piano**. Le unità sono state raggruppate per ogni modalità di Piano, in seguito sono state calcolate le medie dei prezzi per ciascun gruppo. Non è stata evidenziata una sostanziale differenza tra le medie dei gruppi. Nonostante ci si aspettasse che ad un piano superiore corrispondesse un prezzo più elevato, ciò non si è verificato. Si è deciso di non rimuovere NA per non perdere ulteriormente informazione. (Tabella 1 in appendice)

In merito alla variabile **Numero di Locali**, il procedimento è stato analogo al precedente. In questo caso, però, si è notato dalla tabella di riferimento che il prezzo fosse direttamente proporzionale al numero di locali presenti nell'appartamento. (Tabella 5 in appendice)

In riferimento alla variabile **Numero di Bagni**, sono stati rilevati gli stessi risultati della precedente variabile Numero di locali. Si è evinto sia dalla tabella che dal grafico che il prezzo aumentasse al crescere del numero di bagni a disposizione. (Tabella 3 e Figura 6 in appendice)

Al fine di analizzare la variabile **Zona**, le unità sono state raggruppate per ogni modalità della variabile stessa e sono state, in seguito, calcolate le medie dei prezzi per ogni gruppo, come fatto per le variabili precedenti.

Il quartiere di S. Erasmo risulta in media quello con un prezzo maggiore, probabilmente per via della vicinanza al mare. Si è notato come il prezzo medio della **Zona** Kalsa sia stato fortemente influenzato dalla presenza del valore anomalo; infatti, dalla seconda posizione per prezzo medio più alto retrocedeva di un posto facendo sì che la **Zona** Castellammare si posizionasse al secondo posto. (Tabelle 9 – 10 in appendice).

Al fine di verificare la presenza di eventuali affinità tra le descrizioni degli annunci si è sviluppato un **grafico wordcloud**, da cui si è desunto che la tipologia di annuncio più frequente riguardi trilocali e bilocali siti in Via Roma e nel quartiere Tribunale. (Figura 4 in appendice)

Le medesime considerazioni fatte per le variabili Numero Locali (tabella 6 in appendice) e Piano (Tabella 2 vedi in appendice) si dimostrano ancora più marcate se si analizza il dataset **Attici** ad eccezione della variabile Numero Bagni (Tabella 4 e Figura 7 in appendice) dove il prezzo medio per numero di bagni triplica passando da 2 a 3 bagni nel dataset appartamenti, mentre nell'altro dataset tale prezzo raddoppia.

Per quanto concerne invece la variabile **Zona**, sono state rilevate delle differenze rispetto al dataset precedente. Ciò è avvenuto per due motivazioni:

- È presente un numero maggiore di zone perché il dataset in questione tiene in considerazione il comune di Palermo nella sua interezza, al contrario del dataset appartamenti (che considera solo ed esclusivamente il centro storico di Palermo).
- È presente un minor numero di osservazioni per ciascun gruppo, dato che la numerosità campionaria del dataset attici è minore a quello degli appartamenti. (Figure 8 - 9 in appendice)

Partendo da queste considerazioni si deduce che il valore anomalo abbia un peso maggiore sulla media di prezzo del suo gruppo di riferimento poiché non può essere assorbito dagli altri valori.

Si è notato come il prezzo medio della **Zona** Borgo Vecchio sia fortemente influenzato dalla presenza del valore anomalo, infatti, da **Zona** con il prezzo medio più elevato retrocede poi di ben 15 posizioni; è emerso che le prime tre zone con il prezzo medio più alto, senza il valore anomalo, siano in ordine: Politeama, Strasburgo e S. Erasmo. (Tabelle 7 – 8 e Figura 10 in appendice).

La **Zona** S. Erasmo si riconferma tra le zone con il prezzo medio più elevato, anche tenendo conto di ogni **Zona** del comune di Palermo e non solo nel centro storico.

Al fine di verificare la presenza di eventuali similarità tra le descrizioni degli annunci è stato disegnato un **grafico wordcloud**, dal quale, infine, si evince che la tipologia di annuncio più frequente descriva un attico luminoso con terrazza con visione panoramica nelle zone di Via Lanza di Scalea e Corso Calatafimi.

(Figura 5 in appendice)

4. Conclusioni

Dalla seguente analisi si mette in luce come gli elementi che maggiormente incidano sul prezzo siano:

- Numero di locali
- Numero di bagni
- Zona

Il piano non è particolarmente rilevante per la determinazione del prezzo di un immobile. Si denota, inoltre, come le tipologie di immobile maggiormente messe in vendita siano bilocali e trilocali. Inoltre, le zone in cui risiedono immobili con valore più alto sono rispettivamente Politeama, Strasburgo e S. Erasmo.

I risultati potrebbero non essere del tutto attendibili, in quanto le unità statistiche rilevate sono esigue.

Appendice

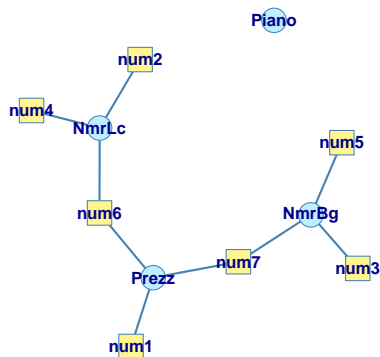


Figura 1 regole

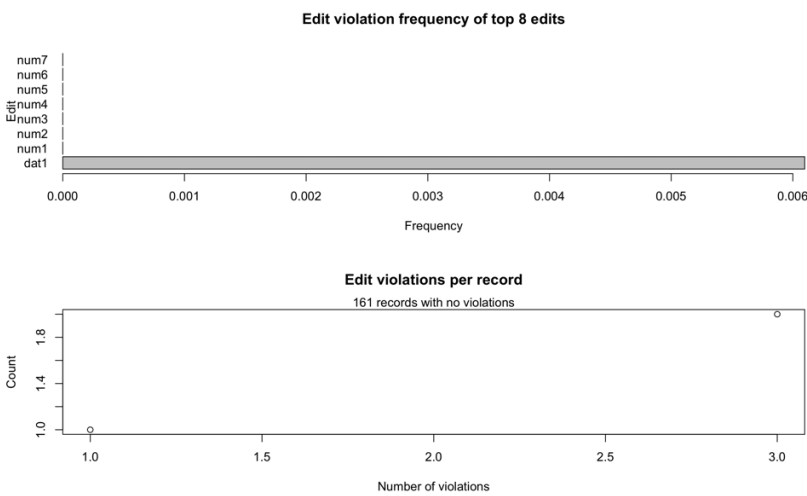


Figura 2 violazione regole df appartamenti

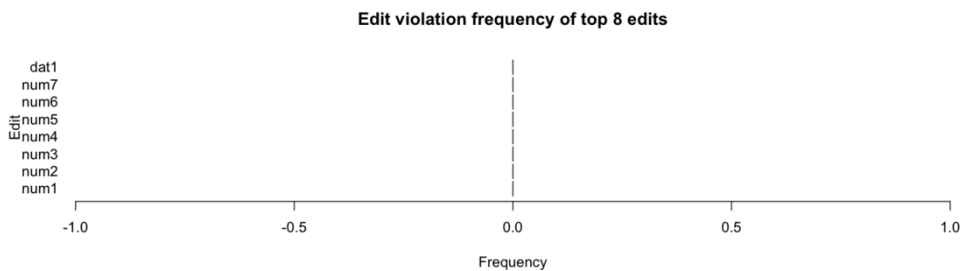


Figura 3 violazione regole df attici

	Piano	count	Price_group
1	6	1	89.0000
2	R	5	109.9800
3	T - 3	1	170.0000
4	1	67	176.2836
5	7	1	220.0000
6	2	29	224.9643
7	4	3	238.3333
8	7 - 8	1	250.0000
9	3	30	265.8333
10	NA	1	340.0000
11	2 - 3	2	387.5000
12	3 - 4	3	395.0000
13	R - 1	1	545.0000
14	T	11	561.4545
15	T - 2	2	695.0000

Tabella 1 medie di Prezzo per Piano df appartamenti

	Piano	count	Price_group
1	1	1	100.0000
2	A	1	125.0000
3	5	7	185.7143
4	10	4	199.7500
5	7	10	217.8000
6	9	16	218.5267
7	6	16	257.6250
8	11	4	277.5000
9	2	6	289.0000
10	3 - 4	4	346.2500
11	8	8	350.2500
12	4	5	350.8000
13	15	1	359.0000
14	3	16	407.2606
15	14	2	447.5000
16	13	3	476.6667
17	2 - 3	5	493.8000
18	5 - 7	1	650.0000
19	5 - 6	1	680.0000
20	5 - 5	1	684.8000
21	T - 2	1	1200.0000

Tabella 1 medie di Prezzo per Piano df attici

	NumeroBagni	count	Price_group
1	1	79	118.8468
2	2	54	239.3704
3	3	25	673.9583

Tabella 3 medie di Prezzo per Numero di bagni df appartamenti

	NumeroBagni	count	Price_group
1	1	50	143.0382
2	2	41	332.5290
3	3	22	681.9909

Tabella 4 medie di Prezzo per Numero di bagni df attici

	NumeroLocali	count	Price_group
1	2	48	103.9354
2	1	6	151.8333
3	3	46	179.9348
4	4	16	215.3125
5	NA	1	450.0000
6	5	41	510.4500

Tabella 5 medie di Prezzo per Numero di locali df appartamenti

	NumeroLocali	count	Price_group
1	1	1	95.0000
2	2	10	132.3000
3	4	32	174.5446
4	3	16	181.8363
5	5	54	479.1961

Tabella 6 medie di Prezzo per Numero di localii df attici

	Zona	count	Price_group
1	San Filippo Neri	7	85.0000
2	Oreto – Perez	6	110.6667
3	Montegrappa	7	139.0000
4	Calatafimi Bassa – Indipendenza	4	140.0000
5	Brancaccio	2	144.5000
6	Sperone	3	146.3333
7	Zisa	4	147.2500
8	CEP – Michelangelo Alta	1	149.0000
9	Cruillas	2	158.5000
10	Acqua dei Corsari	1	165.0000
11	Cardillo	1	175.0000
12	Uditore – Leonardo Da Vinci Alta	6	181.6667
13	Altarello	1	195.0000
14	Palermo	1	226.5300
15	Noce	4	234.5000
16	Cassaro	3	239.3900
17	Pallavicino – Villaggio Ruffini	2	257.0000
18	Resuttana	2	257.5000
19	Montepellegrino	6	295.1667
20	Giotto Galilei – Palagonia	7	308.8424
21	Sferracavallo – Barcarello	1	320.0000
22	Tribunale	3	320.0000
23	Porto	1	359.0000
24	Castellammare	4	363.7500
25	Malaspina	2	379.5000
26	Kalsa	5	390.8000
27	San Lorenzo	1	395.0000
28	Università	2	395.0000
29	De Gasperi – Croce Rossa	5	437.0000
30	Notarbartolo – Sciuti	4	442.2500
31	Libertà – Villabianca	2	495.0000
32	Arenella	1	684.8000
33	Sant'Erasmo	2	750.0000
34	Strasburgo – Belgio	2	765.0000
35	Politeama – Ruggiero Settimo	5	852.0000
36	Borgo Vecchio	3	944.6667

Tabella 7 medie di Prezzo per Zona df attici
compreso valore anomalo

	Zona	count	Price_group
1	San Filippo Neri	7	85.0000
2	Oreto – Perez	6	110.6667
3	Montegrappa	7	139.0000
4	Calatafimi Bassa – Indipendenza	4	140.0000
5	Brancaccio	2	144.5000
6	Sperone	3	146.3333
7	Zisa	4	147.2500
8	CEP – Michelangelo Alta	1	149.0000
9	Cruillas	2	158.5000
10	Acqua dei Corsari	1	165.0000
11	Cardillo	1	175.0000
12	Uditore – Leonardo Da Vinci Alta	6	181.6667
13	Altarello	1	195.0000
14	Palermo	1	226.5300
15	Noce	4	234.5000
16	Cassaro	3	239.3900
17	Pallavicino – Villaggio Ruffini	2	257.0000
18	Resuttana	2	257.5000
19	Montepellegrino	6	295.1667
20	Giotto Galilei – Palagonia	7	308.8424
21	Borgo Vecchio	2	317.0000
22	Sferracavallo – Barcarello	1	320.0000
23	Tribunale	3	320.0000
24	Porto	1	359.0000
25	Castellammare	4	363.7500
26	Malaspina	2	379.5000
27	Kalsa	5	390.8000
28	San Lorenzo	1	395.0000
29	Università	2	395.0000
30	De Gasperi – Croce Rossa	5	437.0000
31	Notarbartolo – Sciuti	4	442.2500
32	Libertà – Villabianca	2	495.0000
33	Arenella	1	684.8000
34	Sant'Erasmo	2	750.0000
35	Strasburgo – Belgio	2	765.0000
36	Politeama – Ruggiero Settimo	5	852.0000

Tabella 8 medie di Prezzo per Zona df attici
escluso valore anomalo

	Zona	count	Price_group
1	Tribunale	33	138.5455
2	Zisa	3	165.0000
3	Cassaro	41	169.2171
4	Oreto – Perez	1	215.0000
5	Castellammare	31	250.5806
6	Kalsa	42	322.2439
7	Sant'Erasmo	7	755.7143

Tabella 9 medie di Prezzo per Zona df appartamenti
compreso valore anomalo

	Zona	count	Price_group
1	Tribunale	33	138.5455
2	Zisa	3	165.0000
3	Cassaro	41	169.2171
4	Oreto – Perez	1	215.0000
5	Kalsa	41	237.8000
6	Castellammare	31	250.5806
7	Sant'Erasmo	7	755.7143

Tabella 10 medie di Prezzo per Zona df appartamenti
escluso valore anomalo



Figura 4 wordcloud df appartamenti



Figura 5 wordcloud df attici

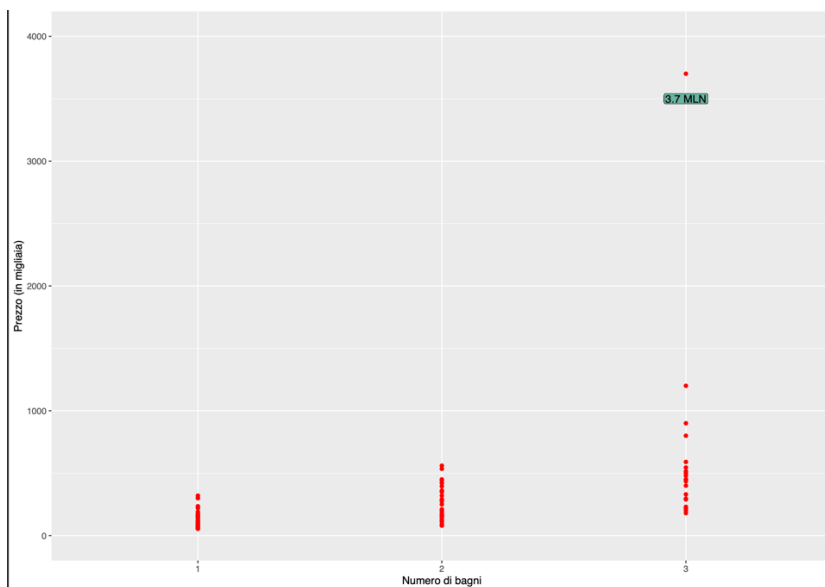


Figura 6 Prezzo condizionatamente al Numero di bagni df appartamenti

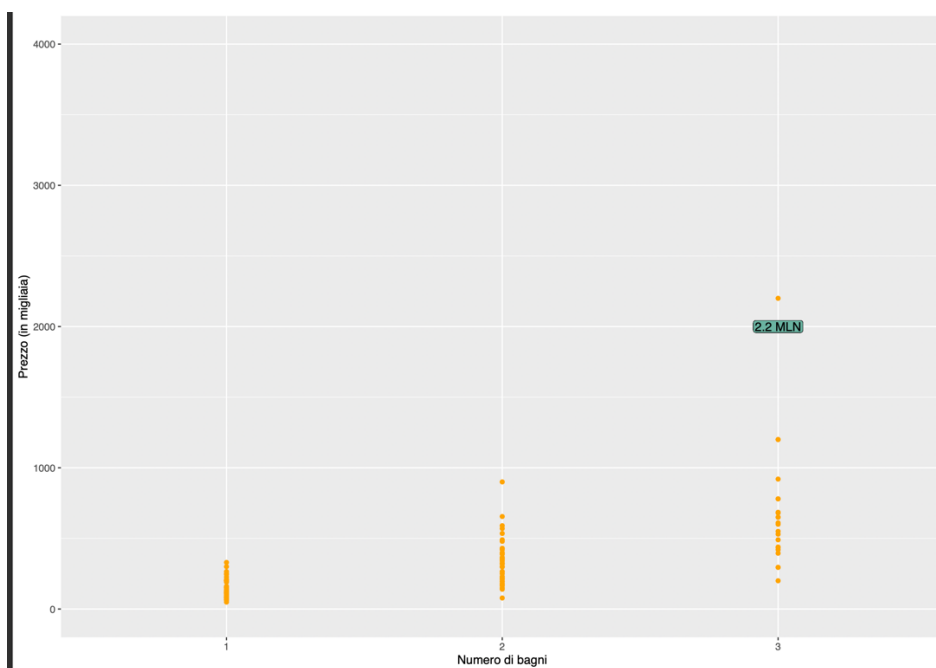


Figura 7 Prezzo condizionatamente al Numero di bagni df attici

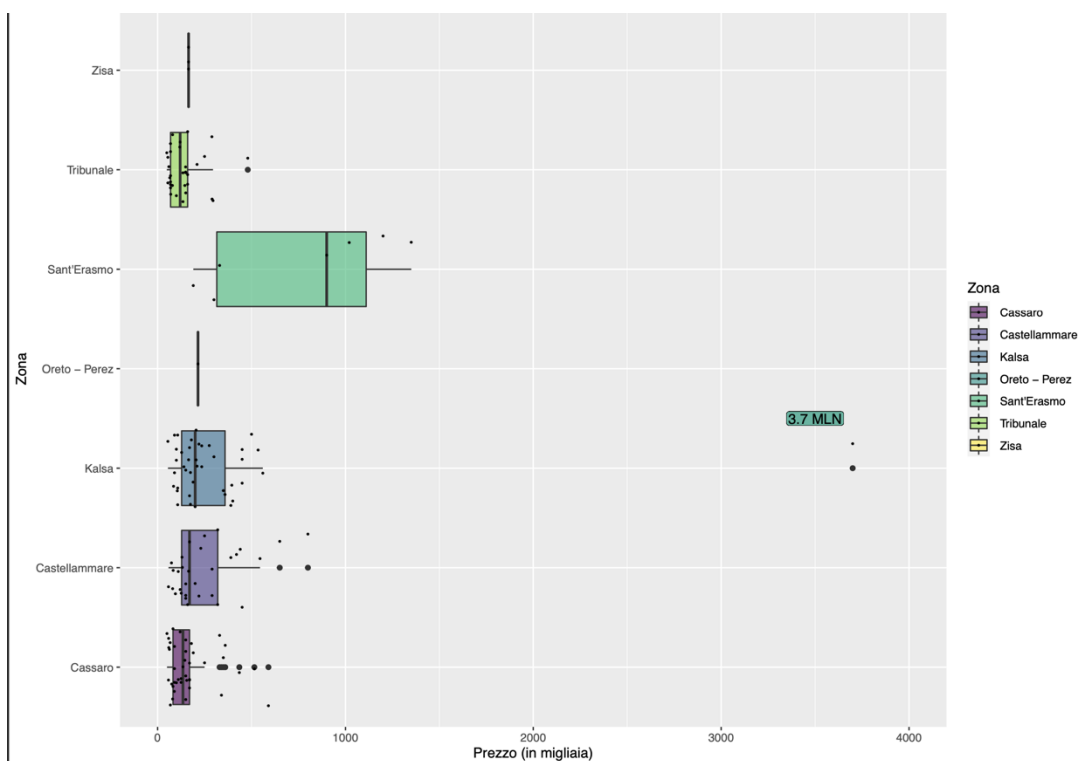


Figura 8 Prezzo condizionatamente alla Zona df appartamenti

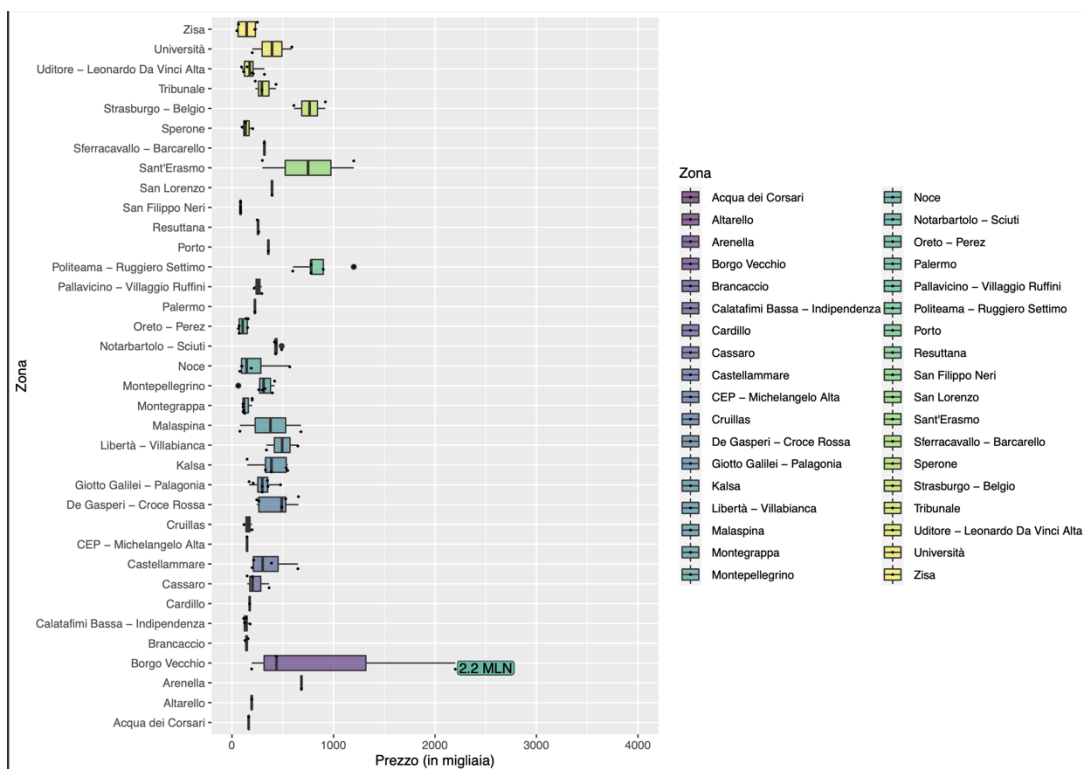


Figura 9 Prezzo condizionatamente alla Zona df atti compreso valore anomalo

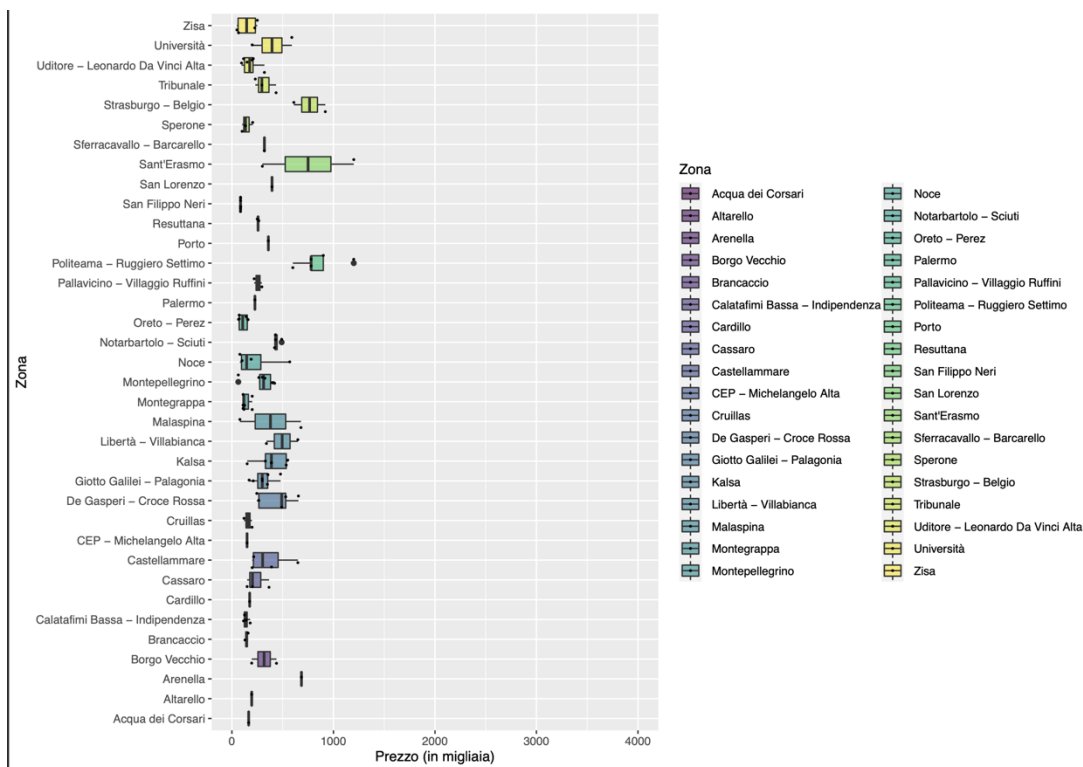


Figura 10 Prezzo condizionatamente alla Zona df attici escluso valore anomalo