

Data Mining - Modello Robusto

Anno accademico 2022/2023

A. Davitti (857132), N. Mauri (856896), M. Rivolta (856863), A. Suardi (846627)

Indice

1	Introduzione: target e covariate	2
2	Gestione dei valori mancanti	5
3	Primo modello	7
4	Analisi della Multicollinearità e Associazione	7
5	Linearità	9
5.1	Trasformazione Box-Cox del target	9
5.2	Trasformazioni delle covariate quantitative	10
5.3	Optimal grouping delle covariate qualitative	10
6	Model selection	11
7	Punti influenti	12
8	Controllo Eteroschedasticità	13
9	Controllo inferenza con Bootstrap	14
10	Conclusioni	14
11	Modello logistico	15

1 Introduzione: target e covariate

Le osservazioni utilizzate nel modello sono le inserzioni presenti su Airbnb nella città di Amsterdam il 7 e 8 dicembre 2018. La variabile target scelta è il prezzo di affitto giornaliero di ciascuna struttura e come covariate le seguenti 49 variabili di cui 35 quantitative (*Q*), 7 dummy (*D*) e 11 variabili factor con più di 6 livelli (*F*). Le ultime cinque covariate sono state da noi costruite a partire da altri dati resi disponibili da Airbnb.

- **host_is_superhost** (*D*): host è/non è superhost
- **host_response_rate** (*Q*): tasso di risposta dell'host ai messaggi
- **host_acceptance_rate** (*Q*): tasso di accettazione delle richieste di prenotazione
- **host_neighbourhood** (*F*): quartiere di residenza dell'host
- **host_listings_count** (*Q*): numero di inserzioni totali dell'host presenti su Airbnb
- **host_has_profile_pic** (*D*): host ha/non ha foto profilo
- **host_identity_verified** (*D*): host ha/non ha identità verificata
- **neighbourhood** (*F*): quartiere indicato dall'host nell'inserzione
- **neighbourhood_cleansed** (*F*): quartiere identificato tramite le coordinate della posizione geografica
- **latitude** (*Q*): latitudine
- **longitude** (*Q*): longitudine
- **property_type** (*F*): tipo di struttura
- **room_type** (*F*): tipo di alloggio
- **accommodates** (*Q*): capienza massima
- **bathrooms** (*Q*): numero di bagni
- **bedrooms** (*Q*): numero di stanze da letto
- **beds** (*Q*): numero di letti
- **bed_type** (*F*): tipo di letto
- **square_feet** (*Q*): superficie in piedi quadri della struttura
- **security_deposit** (*Q*): costo cauzione
- **cleaning_fee** (*Q*): spese di pulizia
- **guests_included** (*Q*): numero massimo di ospiti inclusi nel prezzo
- **extra_people** (*Q*): numero massimo di ospiti extra
- **minimum_nights** (*D*): numero minimo di notti richieste agli ospiti per soggiornare
- **maximum_nights** (*Q*): numero massimo di notti che gli ospiti possono prenotare
- **has_availability** (*D*): disponibile/non disponibile nel momento dello scraping dal sito

- **availability_30** (*D*): numero di giorni di disponibilità nel successivo mese
- **availability_60** (*D*): numero di giorni di disponibilità nei successivi due mesi
- **availability_90** (*D*): numero di giorni di disponibilità nei successivi tre mesi
- **availability_365** (*D*): numero di giorni di disponibilità nel successivo anno
- **number_of_reviews** (*D*): numero di recensioni
- **review_scores_rating** (*Q*): voto medio complessivo
- **review_scores_accuracy** (*Q*): voto medio dell'accuratezza dell'inserzione
- **review_scores_cleanliness** (*Q*): voto medio della pulizia
- **review_scores_checkin** (*Q*): voto medio del check-in
- **review_scores_communication** (*Q*): voto medio della comunicazione dell'host
- **review_scores_location** (*Q*): voto medio della location
- **review_scores_value** (*Q*): voto medio del rapporto qualità/prezzo
- **instant_bookable** (*D*): la struttura può/non può essere prenotata automaticamente
- **cancellation_policy** (*F*): politica di cancellazione della prenotazione
- **require_guest_profile_picture** (*D*): necessità/non necessità di avere foto profilo per prenotare
- **require_guest_phone_verification** (*D*): necessità/non necessità di avere numero di telefono verificato per prenotare
- **calculated_host_listings_coun** (*Q*): numero di inserzioni totali dell'host presenti su Airbnb ad Amsterdam nel momento dello scraping
- **reviews_per_month** (*Q*): numero di recensioni per mese
- **host_for_date** (*Q*): numero di giorni passati dalla pubblicazione dell'inserzione
- **days_since_last_review** (*Q*): numero di giorni passati dall'ultima recensione
- **days_since_first_review** (*Q*): numero di giorni passati dalla prima recensione
- **host_verifications_num** (*Q*): numero di contatti social dell'host
- **amenities_num** (*Q*): numero di servizi disponibili

Di seguito una rappresentazione della distribuzione delle strutture in studio in una mappa geografica (*Fig.1*)

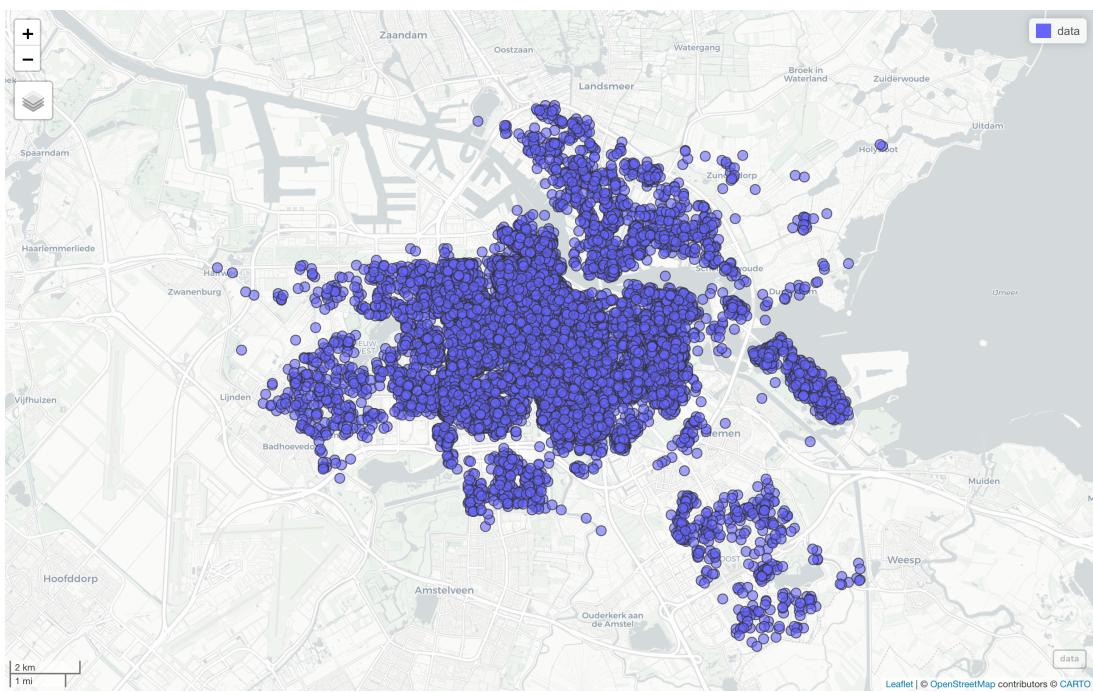


Fig.1 - mappa geografica delle strutture analizzate

e la distribuzione di densità della variabile target (*Fig.2*).

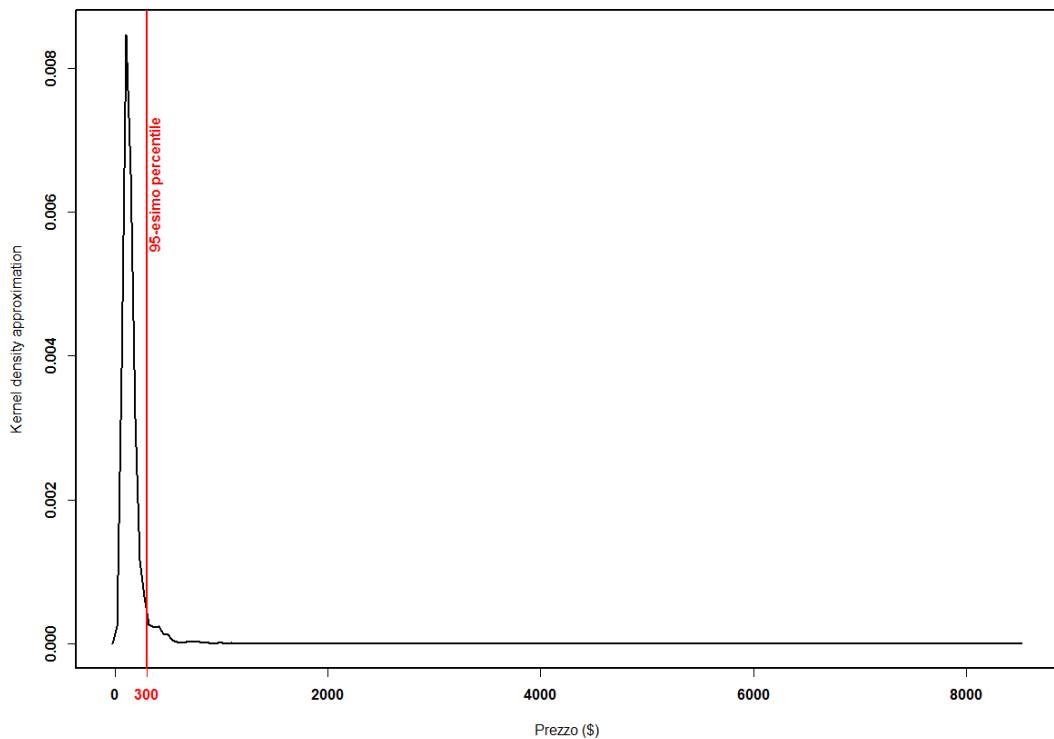


Fig.2 - distribuzione di densità del prezzo delle strutture analizzate

2 Gestione dei valori mancanti

La situazione iniziale dei dati mancanti è rappresentata in *Fig.3*.

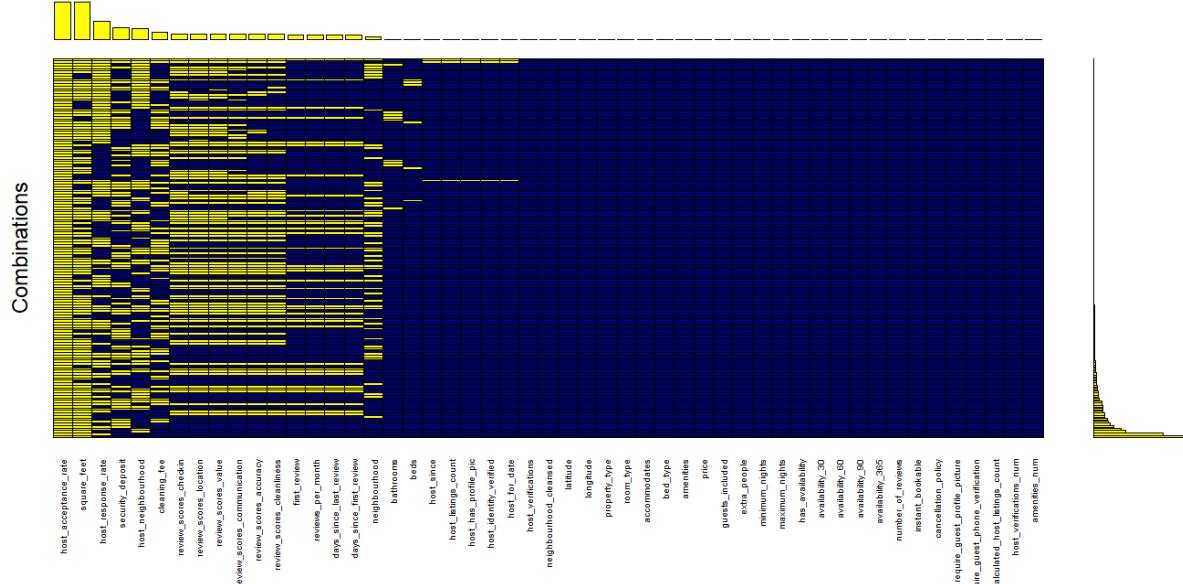


Fig.3 - profili dei dati mancanti iniziali

La gestione dei missing values è stata svolta nei seguenti tre passaggi:

1. eliminazione delle variabili che presentano una percentuale di dati mancanti superiore al 20% delle osservazioni totali (*Fig.4*):

host_since "0.02 %"	host_response_rate "47.34 %"	host_acceptance_rate "100 %"
host_neighbourhood "29 %"	host_listings_count "0.02 %"	host_verifications "0 %"
host_has_profile_pic "0.02 %"	host_identity_verified "0.02 %"	neighbourhood "8.25 %"
neighbourhood_cleansed "0 %"	latitude "0 %"	longitude "0 %"
property_type "0 %"	room_type "0 %"	accommodates "0 %"
bathrooms "0.05 %"	beds "0.03 %"	bed_type "0 %"
amenities "0 %"	square_feet "97.97 %"	price "0 %"
security_deposit "30.78 %"	cleaning_fee "18.12 %"	guests_included "0 %"
extra_people "0 %"	minimum_nights "0 %"	maximum_nights "0 %"
has_availability "0 %"	availability_30 "0 %"	availability_60 "0 %"
availability_90 "0 %"	availability_365 "0 %"	number_of_reviews "0 %"
first_review "12.01 %"	review_scores_accuracy "13.23 %"	review_scores_cleanliness "13.22 %"
review_scores_checkin "13.29 %"	review_scores_communication "13.24 %"	review_scores_location "13.28 %"
review_scores_value "13.28 %"	instant_bookable "0 %"	cancellation_policy "0 %"
require_guest_profile_picture "0 %"	host_for_date "0.02 %"	calculated_host_listings_count "0 %"
reviews_per_month "12.01 %"	host_verifications_num "0 %"	days_since_last_review "12.01 %"
days_since_first_review "12.01 %"		amenities_num "0 %"

Fig.4 - percentuali di dati mancanti per ogni variabile

2. eliminazione delle osservazioni con un numero di review pari a zero e che di conseguenza hanno valori mancanti su tutte le variabili relative ai voti delle recensioni (*Fig.5*):

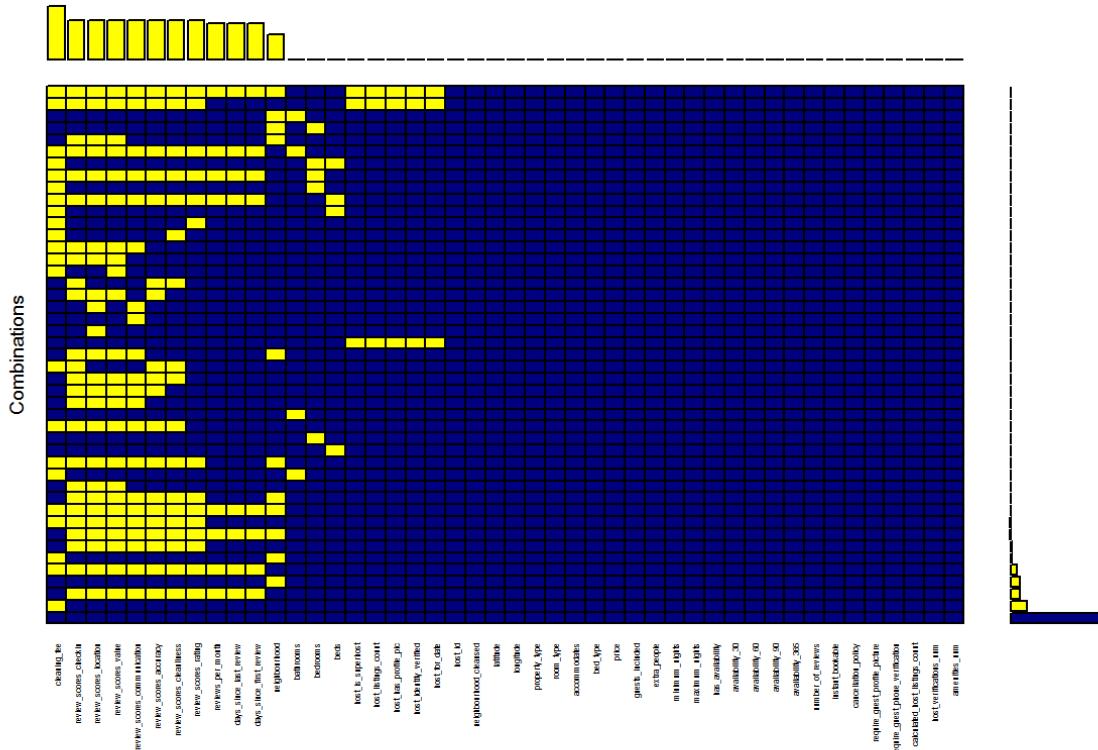


Fig.5 - profili dei dati mancanti dopo l'eliminazione delle variabili con valori mancanti superiori al 20%

3. imputazione dei valori mancanti rimanenti utilizzando la metodologia *cart* nella procedura mice ipotizzando quindi che si tratti di Missing At Random (*MAR*) (*Fig.6*):

```
library (mice)
tempData <- mice (covariate, m = 3, maxit = 10, meth = 'cart', seed = 500)
```

Fig.6 - imputazione dei valori mancanti utilizzando R

3 Primo modello

Il modello iniziale presenta tutte le covariate rimanenti a seguito dello step precedente ad esclusione di *has_availability* che è degenere. La bontà di adattamento del modello ai dati è pari al 21% ($R_{adj}^2 = 21.26\%$).

I grafici delle diagnostiche sui residui sono contenuti in *Fig.22* nelle Conclusioni.

4 Analisi della Multicollinearità e Associazione

Si analizza la multicollinearità delle covariate quantitative calcolando per ognuna il *VIF* ed eliminando una variabile alla volta tra quelle che ne presentano un valore superiore alla soglia pari a 5 (*Fig.7*). La variabile *inter* non viene eliminata perché presenta multicollinearità per costruzione.

	VIF
host_listings_count	7.3450
latitude	1.0297
longitude	1.0424
accommodates	3.4452
bathrooms	1.0382
bedrooms	2.8296
beds	3.8216
cleaning_fee	1.3484
guests_included	1.5519
extra_people	1.2302
minimum_nights	1.0103
maximum_nights	1.0462
availability_30	5.6997
availability_60	34.3806
availability_90	24.2035
availability_365	2.5216
number_of_reviews	3.0679
review_scores_rating	3.4487
review_scores_accuracy	2.1220
review_scores_cleanliness	1.9291
review_scores_checkin	1.9520
review_scores_communication	2.0873
review_scores_location	1.2280
review_scores_value	2.0633
calculated_host_listings_count	7.3685
reviews_per_month	2.7499
host_for_date	1.3936
days_since_last_review	1.5018
days_since_first_review	2.3434
host_verifications_num	1.1181
amenities_num	1.2187

Fig.7 - valori di *VIF* dopo aver eliminato le covariate multicollineari

L'associazione tra le covariate qualitative viene invece controllata calcolando il *chi-quadro* normalizzato (*Fig.8*). Nessuna variabile viene eliminata poiché nessuna coppia di covariate risulta particolarmente associata.

Row	Column	Chi.Square.norm
host_is_superhost	host_has_profile_pic	0.00
host_is_superhost	host_identity_verified	0.00
host_is_superhost	neighbourhood	0.01
host_is_superhost	neighbourhood_cleansed	0.01
host_is_superhost	property_type	0.03
host_is_superhost	room_type	0.03
host_is_superhost	bed_type	0.00
host_is_superhost	instant_bookable	0.00
host_is_superhost	cancellation_policy	0.01
host_is_superhost	require_guest_profile_picture	0.00
host_is_superhost	require_guest_phone_verification	0.00
host_has_profile_pic	host_identity_verified	0.00
host_has_profile_pic	neighbourhood	0.00
host_has_profile_pic	neighbourhood_cleansed	0.00
host_has_profile_pic	property_type	0.00
host_has_profile_pic	room_type	0.00
host_has_profile_pic	bed_type	0.00
host_has_profile_pic	instant_bookable	0.00
host_has_profile_pic	cancellation_policy	0.00
host_has_profile_pic	require_guest_profile_picture	0.00
host_has_profile_pic	require_guest_phone_verification	0.00
host_identity_verified	neighbourhood	0.01
host_identity_verified	neighbourhood_cleansed	0.00
host_identity_verified	property_type	0.01
host_identity_verified	room_type	0.00
host_identity_verified	bed_type	0.00
host_identity_verified	instant_bookable	0.02
host_identity_verified	cancellation_policy	0.01
host_identity_verified	require_guest_profile_picture	0.00
host_identity_verified	require_guest_phone_verification	0.00
neighbourhood	neighbourhood_cleansed	0.71
neighbourhood	property_type	0.03
neighbourhood	room_type	0.03
neighbourhood	bed_type	0.00
neighbourhood	instant_bookable	0.02
neighbourhood	cancellation_policy	0.01
neighbourhood	require_guest_profile_picture	0.00
neighbourhood	require_guest_phone_verification	0.00
neighbourhood_cleansed	property_type	0.02
neighbourhood_cleansed	room_type	0.02
neighbourhood_cleansed	bed_type	0.00
neighbourhood_cleansed	instant_bookable	0.02
neighbourhood_cleansed	cancellation_policy	0.00
neighbourhood_cleansed	require_guest_profile_picture	0.00
neighbourhood_cleansed	require_guest_phone_verification	0.00
neighbourhood_cleansed	property_type	0.06
neighbourhood_cleansed	room_type	0.00
neighbourhood_cleansed	bed_type	0.00
neighbourhood_cleansed	instant_bookable	0.03
neighbourhood_cleansed	cancellation_policy	0.00
neighbourhood_cleansed	require_guest_profile_picture	0.00
neighbourhood_cleansed	require_guest_phone_verification	0.00
property_type	room_type	0.02
property_type	bed_type	0.00
property_type	instant_bookable	0.03
property_type	cancellation_policy	0.00
property_type	require_guest_profile_picture	0.00
property_type	require_guest_phone_verification	0.00
room_type	bed_type	0.00
room_type	instant_bookable	0.03
room_type	cancellation_policy	0.00
room_type	require_guest_profile_picture	0.00
room_type	require_guest_phone_verification	0.00
bed_type	instant_bookable	0.00
bed_type	cancellation_policy	0.00
bed_type	require_guest_profile_picture	0.00
bed_type	require_guest_phone_verification	0.00
instant_bookable	cancellation_policy	0.01
instant_bookable	require_guest_profile_picture	0.00
instant_bookable	require_guest_phone_verification	0.00
cancellation_policy	require_guest_profile_picture	0.00
cancellation_policy	require_guest_phone_verification	0.00
require_guest_profile_picture	require_guest_phone_verification	0.59

Fig.8 - valori di chi-quadro normalizzati per ogni coppia delle covariate qualitative

Al termine di questo step viene stimato nuovamente il modello a cui viene poi aggiunta la covariata *inter*, l'interazione tra le due covariate quantitative più significative (*bedrooms* e *cleaning_fee*).

5 Linearità

5.1 Trasformazione Box-Cox del target

Per migliorare la linearità del modello e diminuire l'eteroschedasticità dei residui si cerca la migliore trasformazione della variabile target secondo *Box-Cox*. Lambda (λ) risulta pari a -0.06 e quindi il target viene trasformato applicando il logaritmo (*Fig.9, Fig.10*).

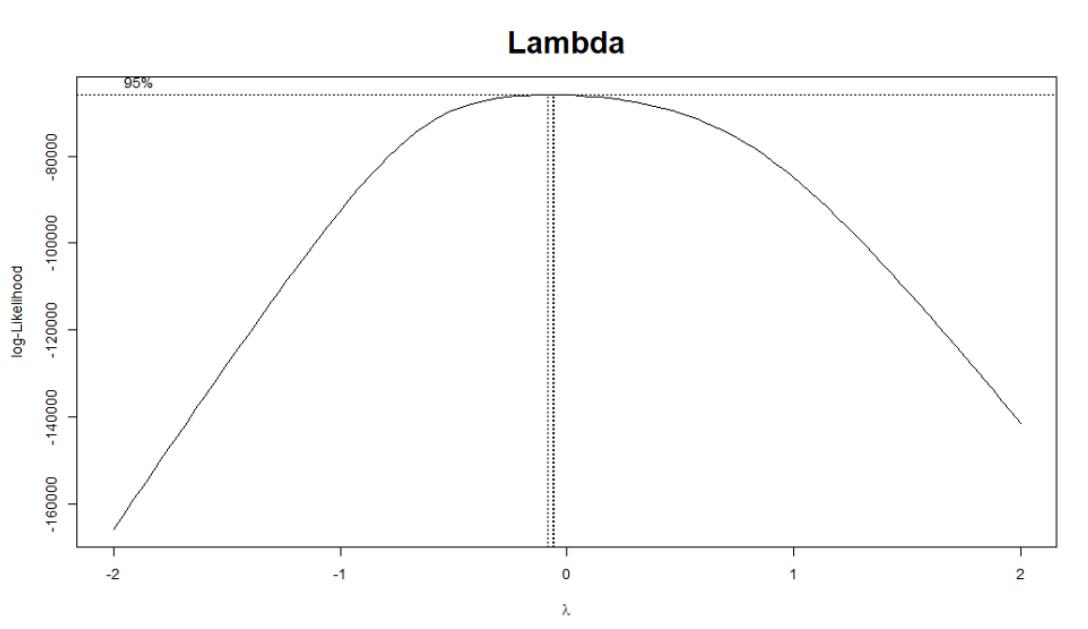


Fig.9 - Lambda restituito dalla procedura Box-Cox applicata sul target

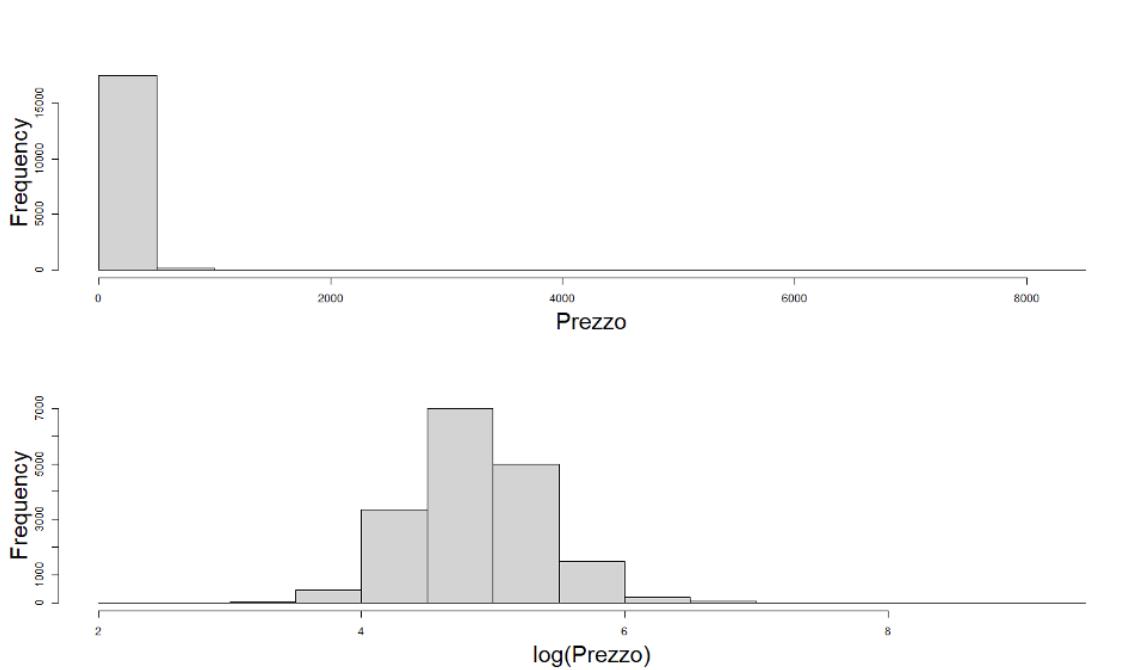


Fig.10 - distribuzione delle frequenze assolute del target prima e dopo la trasformazione

La bontà di adattamento del modello con il target trasformato è pari al 57% ($R^2adj = 56.88\%$).

5.2 Trasformazioni delle covariate quantitative

Osservando i grafici delle *spline*, si cercano delle trasformazioni possibili per le variabili quantitative.

Le variabili che vengono trasformate sono le seguenti:

- trasformazione quadratica: *host_listings_count*, *guests_included*, *extra_people*, *inter*;
- trasformazione cubica: *cleaning_fee*, *minimum_nights*, *maximum_nights*, *review_scores_rating*, *review_scores_checkin*, *review_scores_communication*, *host_for_date*, *days_since_last_review*, *host_verifications_num*;
- trasformazione logaritmica: *availability_365*, *days_since_first_review*, *amenities_num*.

Le successive tre figure (*Fig.11*, *Fig.12*, *Fig.13*) mostrano le spline di covariate di cui poi è stata aggiunta una trasformazione.

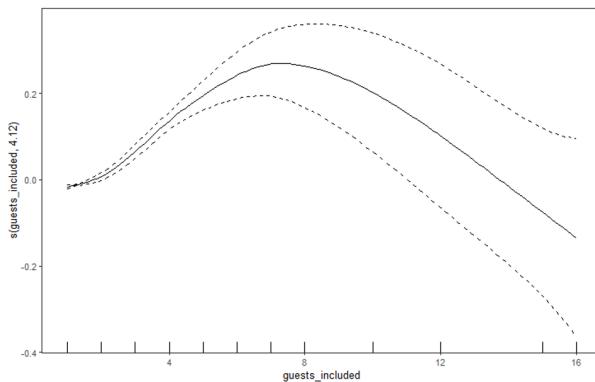


Fig.11 - spline di guests_included - trasf. quadratica

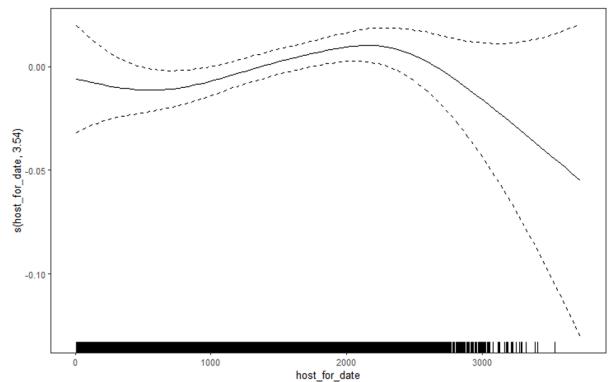


Fig.12 - spline di host_for_date - trasf. cubica

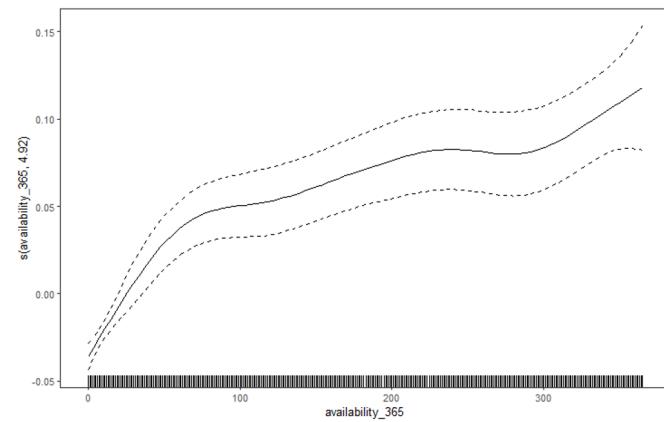


Fig.13 - spline di availability_365 - trasf. logaritmica

5.3 Optimal grouping delle covariate qualitative

Le covariate qualitative con un numero di livelli elevato vengono sottoposte a *optimal grouping*. I risultati finali sono i seguenti:

- la variabile *property_type* passa da 31 a 6 livelli
- la variabile *neighbourhood* passa da 44 a 10 livelli
- la variabile *neighbourhood_cleansed* passa da 22 a 8 livelli

Si riporta di seguito il grafico (*Fig.14*) del risultato dell'optimal grouping relativo alla variabile *neighbourhood*.

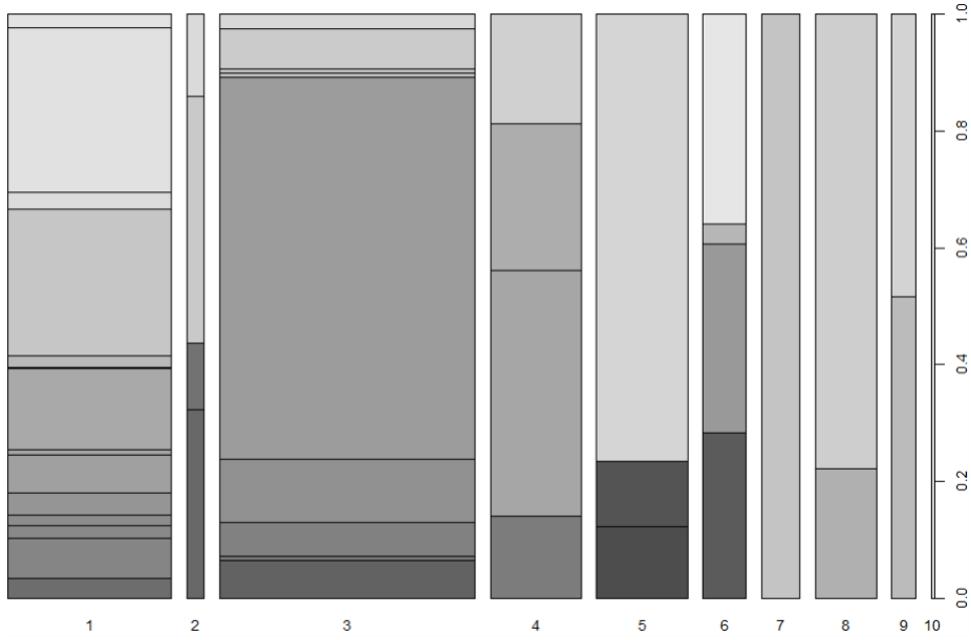


Fig.14 - risultato dell'*optimal grouping* della variabile *neighbourhood*

6 Model selection

Per ridurre il numero di covariate si utilizza un approccio ibrido diviso in due parti. Nella prima parte si svolge una best subset selection basata sull'*AIC* e nella seconda parte si eliminano tutte le variabili che non risultano significative. In questo modo il modello passa da avere 64 (*Fig.15*) variabili a 55 (*Fig.16*) tutte significative (*AIC* diminuito di 13.64) a cui però dobbiamo aggiungere nuovamente due covariate di primo grado per rispettare la gerarchia tra le diverse trasformazioni.

```
Start:  AIC=-40647.96
log(price + 1) ~ host_is_superhost + host_listings_count + I(host_listings_count^2) +
  host_has_profile_pic + host_identity_verified + og_neighbourhood +
  og_neighbourhood_cleansed + latitude + longitude + og_property_type +
  room_type + accommodates + bathrooms + bedrooms + beds +
  bed_type + cleaning_fee + I(cleaning_fee^2) + I(cleaning_fee^3) +
  guests_included + I(guests_included^2) + extra_people + I(extra_people^2) +
  minimum_nights + I(minimum_nights^2) + I(minimum_nights^3) +
  maximum_nights + I(maximum_nights^2) + I(maximum_nights^3) +
  availability_30 + availability_90 + log(availability_365 +
  1) + number_of_reviews + review_scores_rating + I(review_scores_rating^2) +
  I(review_scores_rating^3) + review_scores_accuracy + review_scores_cleanliness +
  review_scores_checkin + I(review_scores_checkin^2) + I(review_scores_checkin^3) +
  review_scores_communication + I(review_scores_communication^2) +
  I(review_scores_communication^3) + review_scores_location +
  review_scores_value + instant_bookable + cancellation_policy +
  require_guest_profile_picture + require_guest_phone_verification +
  reviews_per_month + host_for_date + I(host_for_date^2) +
  I(host_for_date^3) + days_since_last_review + I(days_since_last_review^2) +
  I(days_since_last_review^3) + log(days_since_first_review +
  1) + host_verifications_num + I(host_verifications_num^2) +
  I(host_verifications_num^3) + log(amenities_num + 1) + inter +
  I(inter^2)
```

Fig.15 - modello prima di eseguire la model selection con relativo *AIC*

```

Step: AIC=-40661.6
log(price + 1) ~ host_is_superhost + host_listings_count + I(host_listings_count^2) +
og_neighbourhood + og_neighbourhood_cleaned + longitude +
og_property_type + room_type + accommodates + bathrooms +
bedrooms + beds + bed_type + cleaning_fee + I(cleaning_fee^2) +
guests_included + I(guests_included^2) + extra_people + I(extra_people^2) +
minimum_nights + I(minimum_nights^2) + I(minimum_nights^3) +
maximum_nights + I(maximum_nights^2) + I(maximum_nights^3) +
availability_30 + availability_90 + log(availability_365 +
1) + number_of_reviews + review_scores_rating + I(review_scores_rating^2) +
I(review_scores_rating^3) + review_scores_cleanliness + I(review_scores_checkin^2) +
I(review_scores_checkin^3) + review_scores_communication +
I(review_scores_communication^2) + I(review_scores_communication^3) +
review_scores_location + review_scores_value + instant_bookable +
cancellation_policy + require_guest_profile_picture + require_guest_phone_verification +
reviews_per_month + host_for_date + I(host_for_date^2) +
I(host_for_date^3) + days_since_last_review + I(days_since_last_review^2) +
I(days_since_last_review^3) + log(days_since_first_review +
1) + log(amenities_num + 1) + inter + I(inter^2)

```

Fig.16 - modello Best Subset con relativo AIC

7 Punti influenti

Dall'analisi dei punti influenti nel modello risulta la presenza di 824 osservazioni che superano la soglia limite della *distanza di Cook* (0.0002278034). Dopo l'eliminazione di queste osservazioni risulta esserci ancora un'osservazione con *distanza di Cook* particolarmente elevata e anche questa viene rimossa (Fig.17).

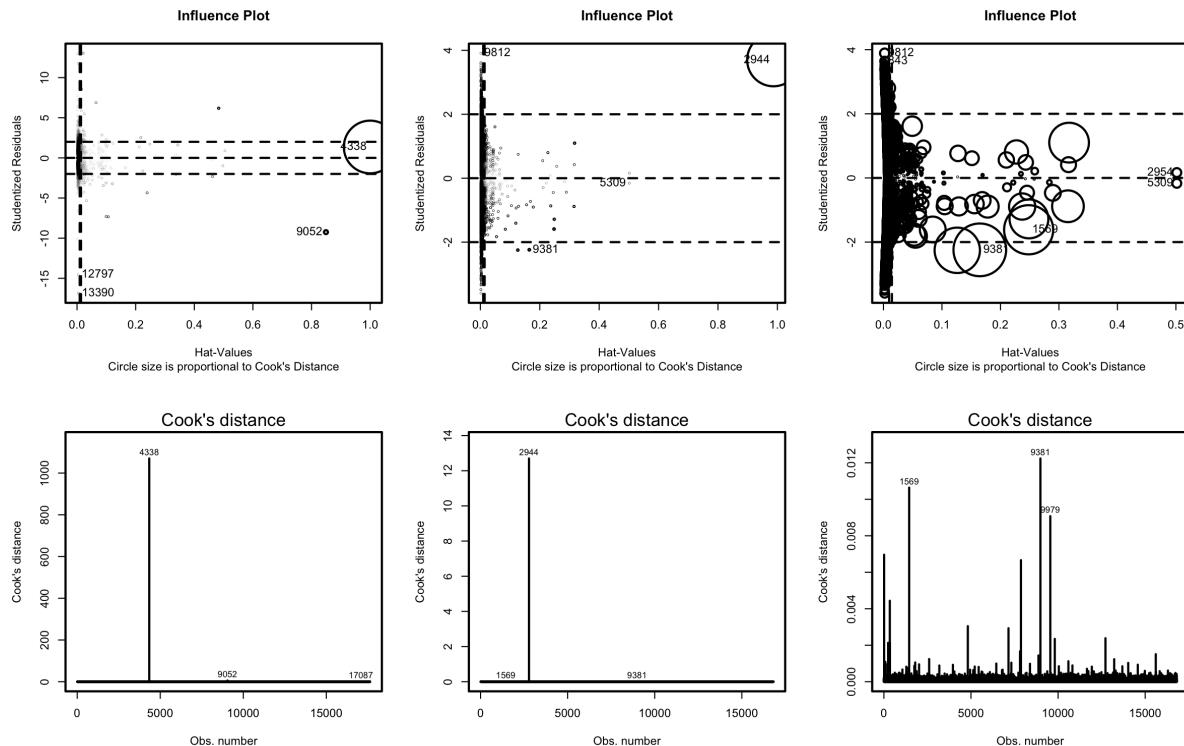


Fig.17 - influence plot e barplot iniziale, dopo la prima e la seconda eliminazione

In seguito alla rimozione delle osservazioni influenti, le seguenti variabili non risultano più significative: *latitude*, *beds*, *beds_type*, *maximum_nights* (tutte le trasformazioni), *review_scores_checkin* (tutte le trasformazioni).

8 Controllo Eteroschedasticità

Viene svolto il test di *Breusch-Pagan* per valutare l'ipotesi di omoschedasticità dei residui del modello, che viene però rifiutata con un p-value di molto inferiore a 0.1 (*Fig.18*).

```
studentized Breusch-Pagan test

data: lm10
BP = 436.18, df = 69, p-value < 2.2e-16
```

Fig.18 - Esito del test di *Breusch-Pagan*

Si procede quindi alla correzione degli standard error tramite l'approccio di *White* (*Fig.19*).

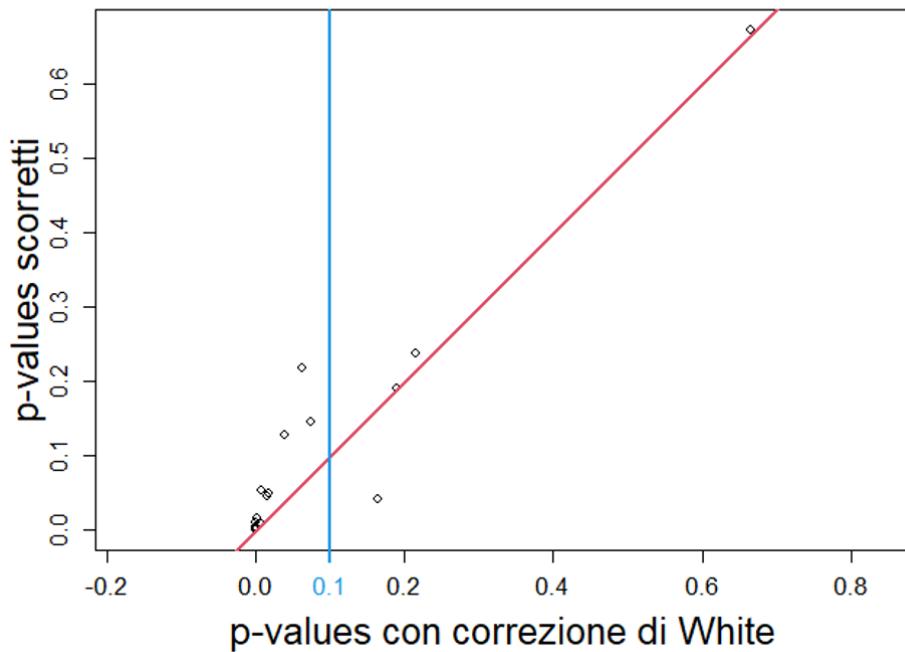


Fig.19 - p-values di ogni parametro del modello prima e dopo la correzione di *White*

In seguito alla correzione di *White* risulta che la variabile *guest_included*² non è più significativa in quanto il suo p-value passa da 0.042079 a 0.1632009.

9 Controllo inferenza con Bootstrap

Viene effettuata una stima bootstrap dei coefficienti del modello con l'estrazione di 1500 campioni per valutare la robustezza del modello stimato con *OLS*. Dal confronto tra le stime bootstrap e le stime *OLS* i valori e le significatività di tutti i coefficienti vengono confermati (*Fig.20*).

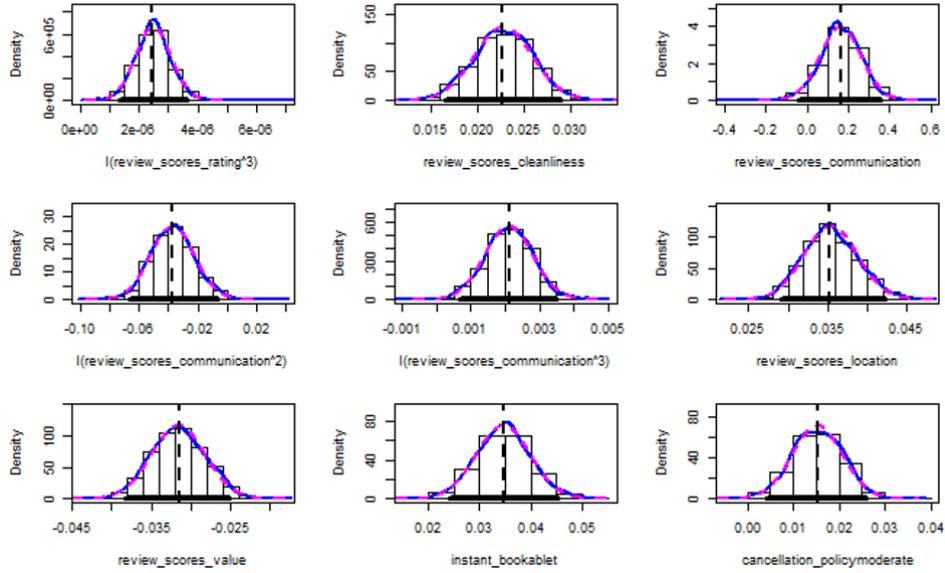


Fig.20 - distribuzioni delle stime bootstrap relative ad alcuni coefficienti

10 Conclusioni

Confrontando il modello iniziale con quello finale risulta un ottimo miglioramento dei grafici relativi alle diagnostiche sui residui (*Fig.22*). Il modello finale presenta inoltre un adattamento del 65% rispetto al 21% del modello iniziale (R^2_{adj}).

Inoltre, in *Fig.21* e in *Tab.1* vengono mostrati rispettivamente gli scatterplot delle osservazioni rispetto al target osservato/previsto e le stime dei coefficienti di alcune variabili più significative.

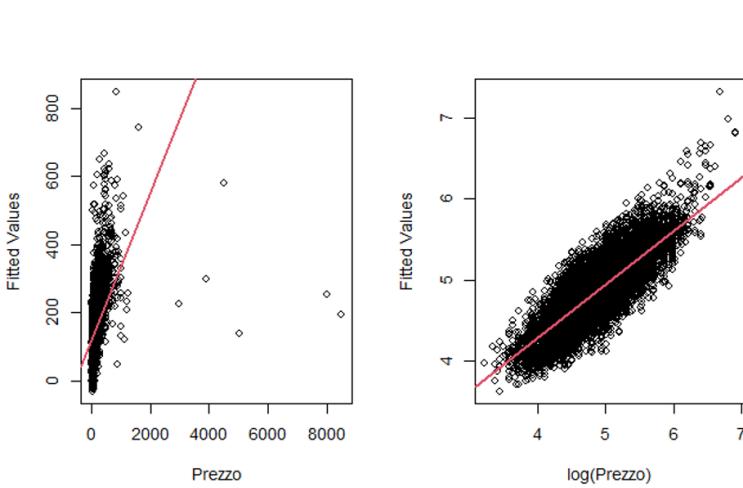


Fig.21 - scatterplot delle osservazioni rispetto al target osservato e previsto rispettivamente del modello iniziale e finale

Variabile	Estimate	Std. Error	t value	Pr (> t)
<i>bedrooms</i>	0.107	0.008	13.653	< 2 × 10 ⁻¹⁶
<i>room_type</i> (Private Room)	-0.335 (-28%)	0.006	-51.615	< 2 × 10 ⁻¹⁶
<i>room_type</i> (Shared Room)	-0.530 (-41%)	0.052	-10.213	< 2 × 10 ⁻¹⁶
log (<i>amenities_num</i> + 1)	0.066	0.005	12.804	< 2 × 10 ⁻¹⁶

Tab.1 - stime dei coefficienti di alcune variabili più significative

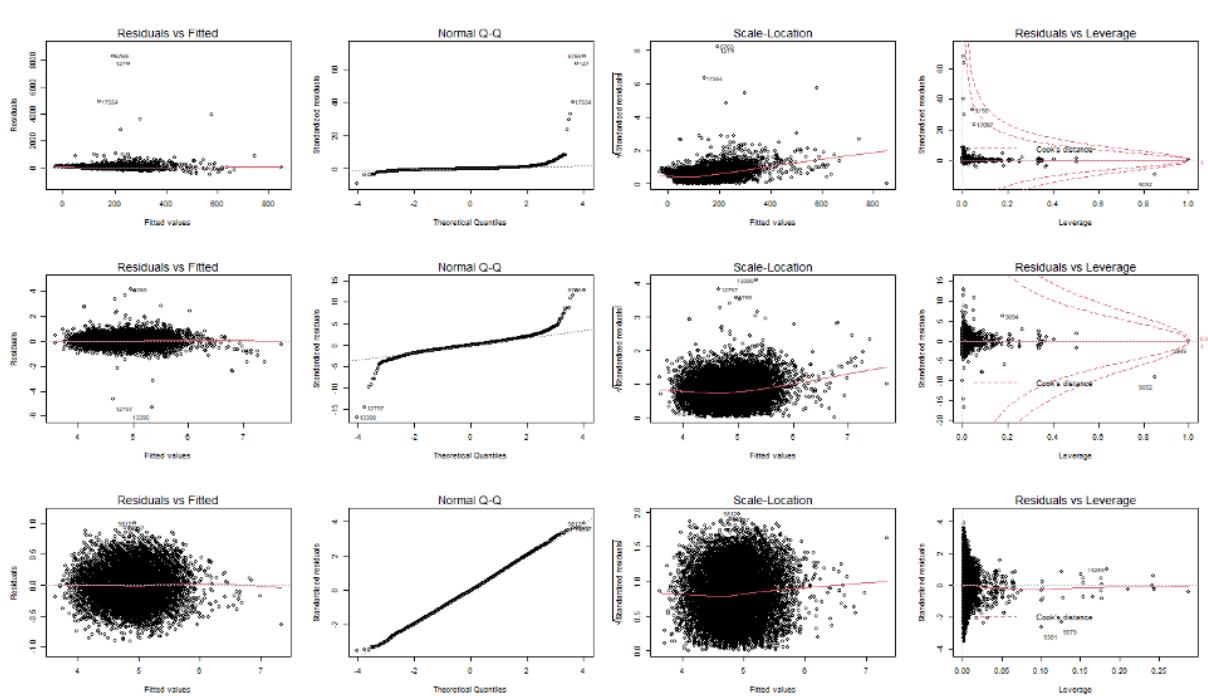


Fig.22 - grafici delle diagnostiche sui residui del modello iniziale (prima riga), del modello dopo la trasformazione del target (seconda riga) e del modello finale (terza riga)

11 Modello logistico

La variabile target si ottiene binarizzando *review_scores_rating* nel seguente modo: le osservazioni con voto medio complessivo superiore al 95 appartengono alla classe 1 e quelle con voto inferiore alla soglia appartengono invece alla classe 0. Le osservazioni di classe 1 sono il 59%, quelle di classe 0 sono il 41%. Le covariate inserite nel modello logistico sono le più rilevanti tra quelle usate nel modello precedente. Tutte le covariate sono statisticamente significative sia per il *test di Wald* (Fig.23) che per il *Likelihood Ratio Test (LRT)* (Fig.24). *R² di McFadden* è pari al 2.2%.

Infine, in (Fig.25) sono mostrati gli intervalli di confidenza degli *Odds Ratio* delle variabili del modello logistico.

```

              Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.0556135  0.0701457 -0.793  0.42788
room_typePrivate room -0.4198824  0.0415653 -10.102 < 2e-16 ***
room_typeShared room -0.5038174  0.2715502 -1.855  0.06355 .
bathrooms       0.1380235  0.0530713  2.601  0.00930 **
bedrooms        -0.0707453  0.0225910 -3.132  0.00174 **
price            0.0010536  0.0002245  4.692 2.70e-06 ***
cleaning_fee    -0.0045388  0.0008608 -5.273 1.34e-07 ***
cancellation_policy_moderate -0.1565587  0.0438822 -3.568  0.00036 ***
cancellation_policy_strict_14_with_grace_period -0.2931359  0.0439407 -6.671 2.54e-11 ***
cancellation_policy_super_strict_60      -2.7573084  0.5580366 -4.941 7.77e-07 ***
amenities_num     0.0320533  0.0018444 17.379 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 23583  on 17390  degrees of freedom
Residual deviance: 23055  on 17380  degrees of freedom
AIC: 23077

```

Fig.23 - summary del modello logistico con test di Wald

```

Model:
target ~ room_type + bathrooms + bedrooms + price + cleaning_fee +
cancellation_policy + amenities_num
          Df Deviance   AIC   LRT Pr(>Chi)
<none>           23055 23077
room_type         2    23159 23177 103.92 < 2.2e-16 ***
bathrooms        1    23063 23083  7.63  0.005728 **
bedrooms         1    23065 23085  9.83  0.001717 **
price            1    23083 23103 27.64 1.458e-07 ***
cleaning_fee     1    23083 23103 27.92 1.265e-07 ***
cancellation_policy 3    23131 23147 76.08 < 2.2e-16 ***
amenities_num    1    23374 23394 318.68 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Fig.24 - Likelihood Ratio Test (LRT) eseguito sulle singole covariate del modello

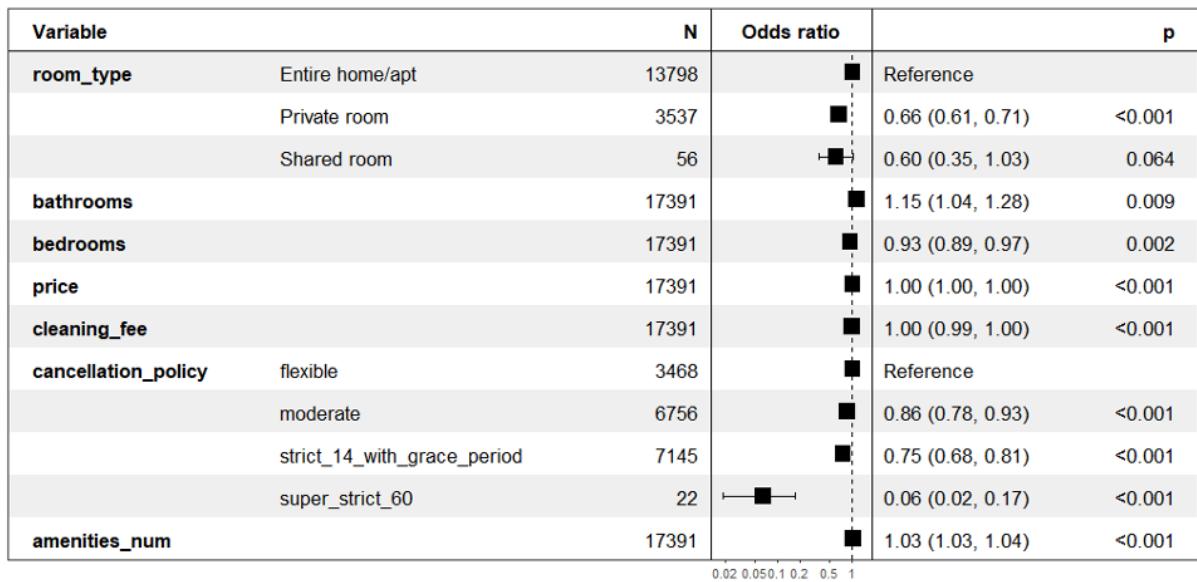


Fig.25 - intervalli di confidenza degli Odds Ratio delle variabili del modello logistico