Proyecto Final. Precios Airbnb

Integrantes de equipo:

Corona Jimenez Andrea

Altamirano Paredes Michel Alejandro

Lira Gonzalez Rosa Linda

Introducción

Como es bien sabido, en los últimos años, los costos de alquiler y arrendamiento de viviendas en la Ciudad de México se han visto afectados de manera considerable, presentando una tendencia al alza que es mayormente visible en plataformas como Airbnb. Gracias a su gran volumen de propiedades en arrendamiento, Airbnb podría llegar a influir en los costos de viviendas circundantes a las registradas en la plataforma, lo que resultaría de gran interés tanto para arrendatarios como arrendadores.

Objetivo

Este trabajo tiene por objetivo ajustar un modelo de regresión lineal a una base de datos de Airbnb´s de la Ciudad de México para mostrarle a la población general como es que el precio de uno de estos arrendamientos podrá variar en un futuro. Sin embargo, ya que muchas personas no están familiarizadas con lenguajes de programación, se decidió crear una calculadora interactiva que le permita a un usuario estimar el precio de un Airbnb en el futuro. Así pues, procedamos con el desarrollo del trabajo.

1. Las variables

Analizando la base de datos original en Excel es posible notar que existen múltiples variables que resultan irrelevantes para nuestro modelo, pues varias de ellas no nos aportan valores medibles que podamos incluir en el mismo. Debido a esto, se descartaron un total de 44 variables, cuyos motivos fueron los siguientes:

```r
#Librerías utilizadas
library(corrplot)
```

```
Warning: package 'corrplot' was built under R version 4.3.3
```

```
corrplot 0.92 loaded
```

```r
library(MASS)
library(dplyr)
```

```

Attaching package: 'dplyr'

The following object is masked from 'package:MASS':

    select

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```r
library(readxl)
library(nortest)
library(ggplot2)
```

```
Warning: package 'ggplot2' was built under R version 4.3.2
```

```
library(car)
```

```
Loading required package: carData


Attaching package: 'car'

The following object is masked from 'package:dplyr':

    recode
```

```
library(DescTools)
```

```
Warning: package 'DescTools' was built under R version 4.3.3


Attaching package: 'DescTools'

The following object is masked from 'package:car':

    Recode
```

```
#Usa el comado file.choose() para abrir la ubicación del xlsx
#Lectura del excel
excel <- "C:\\Users\\andre\\OneDrive\\Documents\\2024-2\\Datos_modelos.xlsx"
datos <- read_excel(excel, sheet = "Datos")
attach(datos);
```

Las variables que en ella aparecen fueren elegidas por su relación con el precio, descartando así un total de 44 variables que resultaban irrelevantes para el modelo. Los motivos por los que se descartaron el resto de variables se muestran a continuación:

- Letras: Debido a que muchas de estas variables presentaban únicamente palabras, links y caracteres que no eran cuantificables ni ajustables al modelo.
- Porcentajes: Al igual que el punto anterior los porcentajes al tener caracteres fueron ignorados para el modelo por contener datos que no eran medibles en R.
- Repetición de datos: Debido a que las columnas del máximo y mínimo de noches aparecían 6 veces con los mismos datos

Aunado a esto y, dado que el porcentaje con respecto al total no era muy elevado, se eliminaron los valores NA que contenía la base de datos.

```
#Primera limpieza
datos <- na.omit(datos)
datos <- datos %>%
  dplyr::select(-id)
```

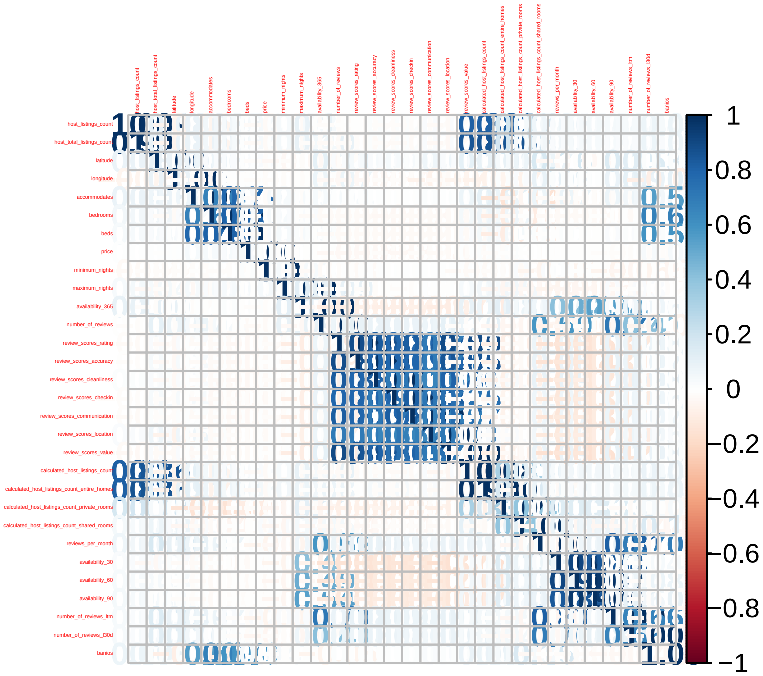
2. Análisis exploratorio de las variables

Para ser más precisos con los datos que estábamos trabajando se discutió sobre que variables podían llegar a estar más correlacionadas entre sí, siendo estas las siguientes:

- 1 y 2.- Ambas variables parametrizan datos muy similares
- 3 y 4.- Ambas tienen datos relativos a la ubicación del Airbnb
- 5, 6 y 7.- Hacen alusión a la cantidad de huéspedes que puede albergar el Airbnb
- 6 y 30.- En una mala tipificación de las habitaciones los baños podrían haber sido tomados como tales, o tal vez porque las habitaciones tengan un baño integrado
- 5, 6, 7 y 30.- Igual que el motivo anterior, una mala tipificación de los datos podría hacer que estas variables estén relacionadas
- 9, 10, 11, 25, 26 y 27.- Mientras mayor o menor sea el número de noches que alguien se hospede en el Airbnb menor o mayor será la cantidad de días que estuvo disponible
- 12 - 19.- Todas estas variables hacen alusión a las reviews del Airbnb
- 3, 4 y 18.- Las variables están relacionadas con la ubicación del Airbnb
- 6, 7, 15 y 30.- La limpieza del Airbnb es notoria en las recámaras, las camas y los baños
- 1, 2, 20, 21, 22, 23.- Son estadísticas del hospedador de cada Airbnb
- 11, 25, 26, 27.- Son las disponibilidades acumuladas en ciertos periodos de tiempo, siendo el más alto de un año, por lo que si descubrimos correlación podríamos quedarnos con el de periodicidad más grande
- 12, 24, 28, 29.- Hacen alusión al número de Reviews que el Airbnb ha tenido a lo largo de diferentes periodos de tiempo.

Para comprobar si nuestras suposiciones eran correctas se realizaron gráficos de correlación entre cada una de las variables anteriormente expuestas y, tras realizar 11 gráficos pudimos ver que nuestras suposiciones eran correctas, por lo que las variables reiterativas fueron eliminadas buscando cumplir con los supuestos del modelo.

```
#Gráfico de correlaciones
corrplot(cor(datos), method = "number", tl.cex = .2)
```



```
#Segunda limpieza después de ver correlaciones
attach(datos)
```

```
The following objects are masked from datos (pos = 3):

    accommodates, availability_30, availability_365, availability_60,
    availability_90, banios, bedrooms, beds,
    calculated_host_listings_count,
    calculated_host_listings_count_entire_homes,
    calculated_host_listings_count_private_rooms,
    calculated_host_listings_count_shared_rooms, host_listings_count,
    host_total_listings_count, latitude, longitude, maximum_nights,
    minimum_nights, number_of_reviews, number_of_reviews_l30d,
    number_of_reviews_ltm, price, review_scores_accuracy,
    review_scores_checkin, review_scores_cleanliness,
    review_scores_communication, review_scores_location,
    review_scores_rating, review_scores_value, reviews_per_month
```

```
datos <- datos %>%
  dplyr::select(-host_listings_count,
                -beds,
                -availability_30,
                -availability_60,
                -availability_90,
                -review_scores_rating,
                -review_scores_accuracy,
                -review_scores_communication,
                -review_scores_value,
                -calculated_host_listings_count,
                -calculated_host_listings_count_entire_homes,
                -number_of_reviews_ltm)
```

Nota: Para ver los 12 gráficos consultar el "Documento expandido"

3. Datos atípicos y el primer modelo

Teniendo ahora una base de datos con menos variables se procedió a analizar las escalas de cada una de ellas.

Habiendo eliminado las variables reiterativas se realizó el sguiente modelo:

```
#Modelo original
modelo1 <- lm(price~host_total_listings_count
              +latitude
              +longitude
              +accommodates
              +bedrooms
              +minimum_nights
              +maximum_nights
              +availability_365
              +number_of_reviews
              +review_scores_cleanliness
              +review_scores_checkin
              +review_scores_location
              +calculated_host_listings_count_private_rooms
              +calculated_host_listings_count_shared_rooms
              +reviews_per_month
              +number_of_reviews_l30d
              +banios)
```

```r
summary(modelo1)
```

```

Call:
lm(formula = price ~ host_total_listings_count + latitude + longitude +
    accommodates + bedrooms + minimum_nights + maximum_nights +
    availability_365 + number_of_reviews + review_scores_cleanliness +
    review_scores_checkin + review_scores_location + calculated_host_listings_count_private_
    calculated_host_listings_count_shared_rooms + reviews_per_month +
    number_of_reviews_l30d + banios)

Residuals:
    Min      1Q  Median      3Q     Max
 -24848   -1079    -520      39 1830428

Coefficients:
                                              Estimate Std. Error t value
(Intercept)                                 -3.948e+05  4.161e+05  -0.949
host_total_listings_count                    1.829e+00  1.583e+00   1.156
latitude                                     4.138e+03  3.291e+03   1.257
longitude                                   -3.238e+03  4.103e+03  -0.789
accommodates                                 1.412e+02  8.496e+01   1.662
bedrooms                                     1.116e+02  1.853e+02   0.602
minimum_nights                               1.359e+01  5.796e+00   2.344
maximum_nights                              -2.866e-01  2.879e-01  -0.995
availability_365                            -5.307e-01  1.101e+00  -0.482
number_of_reviews                            1.780e+00  2.504e+00   0.711
review_scores_cleanliness                   -2.030e+03  4.412e+02  -4.601
review_scores_checkin                        8.127e+02  5.332e+02   1.524
review_scores_location                      -2.161e+01  5.402e+02  -0.040
calculated_host_listings_count_private_rooms -4.979e+01  3.066e+01  -1.624
calculated_host_listings_count_shared_rooms -5.730e+01  6.692e+01  -0.856
reviews_per_month                            2.947e+01  1.112e+02   0.265
number_of_reviews_l30d                      -1.548e+02  9.121e+01  -1.698
banios                                       5.143e+02  2.115e+02   2.431
                                            Pr(>|t|)
(Intercept)                                   0.3427
host_total_listings_count                     0.2479
latitude                                      0.2087
longitude                                     0.4299
accommodates                                  0.0966 .
bedrooms                                      0.5472
```

```
minimum_nights                                0.0191 *
maximum_nights                                0.3195
availability_365                              0.6297
number_of_reviews                             0.4773
review_scores_cleanliness                   4.23e-06 ***
review_scores_checkin                         0.1275
review_scores_location                        0.9681
calculated_host_listings_count_private_rooms   0.1044
calculated_host_listings_count_shared_rooms    0.3919
reviews_per_month                             0.7910
number_of_reviews_l30d                        0.0896 .
banios                                        0.0151 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18610 on 19957 degrees of freedom
Multiple R-squared:  0.004108,  Adjusted R-squared:  0.003259
F-statistic: 4.842 on 17 and 19957 DF,  p-value: 1.57e-10
```

```
seleccion <- stepAIC(modelo1, direction = c("both"));
```

```
Start:  AIC=392780.2
price ~ host_total_listings_count + latitude + longitude + accommodates +
    bedrooms + minimum_nights + maximum_nights + availability_365 +
    number_of_reviews + review_scores_cleanliness + review_scores_checkin +
    review_scores_location + calculated_host_listings_count_private_rooms +
    calculated_host_listings_count_shared_rooms + reviews_per_month +
    number_of_reviews_l30d + banios

                                              Df  Sum of Sq        RSS    AIC
- review_scores_location                       1     554359 6.9102e+12 392778
- reviews_per_month                            1   24312502 6.9103e+12 392778
- availability_365                             1   80513330 6.9103e+12 392778
- bedrooms                                     1  125483056 6.9104e+12 392779
- number_of_reviews                            1  174837817 6.9104e+12 392779
- longitude                                    1  215749397 6.9104e+12 392779
- calculated_host_listings_count_shared_rooms  1  253824793 6.9105e+12 392779
- maximum_nights                               1  343137576 6.9106e+12 392779
- host_total_listings_count                    1  462404486 6.9107e+12 392780
- latitude                                     1  547216556 6.9108e+12 392780
<none>                                                      6.9102e+12 392780
```

```
- review_scores_checkin                         1  804517570 6.9110e+12 392781
- calculated_host_listings_count_private_rooms  1  913064952 6.9111e+12 392781
- accommodates                                  1  956094276 6.9112e+12 392781
- number_of_reviews_l30d                        1  997833706 6.9112e+12 392781
- minimum_nights                                1 1902107935 6.9121e+12 392784
- banios                                        1 2046714743 6.9123e+12 392784
- review_scores_cleanliness                     1 7330167939 6.9176e+12 392799

Step:  AIC=392778.2
price ~ host_total_listings_count + latitude + longitude + accommodates + 
    bedrooms + minimum_nights + maximum_nights + availability_365 + 
    number_of_reviews + review_scores_cleanliness + review_scores_checkin + 
    calculated_host_listings_count_private_rooms + calculated_host_listings_count_shared_room
    reviews_per_month + number_of_reviews_l30d + banios

                                               Df  Sum of Sq        RSS    AIC
- reviews_per_month                             1   24411188 6.9103e+12 392776
- availability_365                              1   80267601 6.9103e+12 392776
- bedrooms                                      1  126341845 6.9104e+12 392777
- number_of_reviews                             1  174605356 6.9104e+12 392777
- longitude                                     1  215314274 6.9104e+12 392777
- calculated_host_listings_count_shared_rooms   1  253629258 6.9105e+12 392777
- maximum_nights                                1  343687211 6.9106e+12 392777
- host_total_listings_count                     1  461898349 6.9107e+12 392778
- latitude                                      1  546904937 6.9108e+12 392778
<none>                                                       6.9102e+12 392778
- calculated_host_listings_count_private_rooms  1  916637488 6.9111e+12 392779
- accommodates                                  1  955563948 6.9112e+12 392779
- number_of_reviews_l30d                        1  997840130 6.9112e+12 392779
- review_scores_checkin                         1 1021629096 6.9113e+12 392779
+ review_scores_location                        1     554359 6.9102e+12 392780
- minimum_nights                                1 1902052074 6.9121e+12 392782
- banios                                        1 2046488717 6.9123e+12 392782
- review_scores_cleanliness                     1 8089570356 6.9183e+12 392800

Step:  AIC=392776.3
price ~ host_total_listings_count + latitude + longitude + accommodates + 
    bedrooms + minimum_nights + maximum_nights + availability_365 + 
    number_of_reviews + review_scores_cleanliness + review_scores_checkin + 
    calculated_host_listings_count_private_rooms + calculated_host_listings_count_shared_room
    number_of_reviews_l30d + banios

                                               Df  Sum of Sq        RSS    AIC
```

```
- availability_365                              1   74636365 6.9103e+12 392774
- bedrooms                                      1  124377034 6.9104e+12 392775
- longitude                                     1  213920223 6.9105e+12 392775
- calculated_host_listings_count_shared_rooms   1  260654110 6.9105e+12 392775
- number_of_reviews                             1  309502558 6.9106e+12 392775
- maximum_nights                                1  366141039 6.9106e+12 392775
- host_total_listings_count                     1  470117738 6.9107e+12 392776
- latitude                                      1  569012696 6.9108e+12 392776
<none>                                                       6.9103e+12 392776
- calculated_host_listings_count_private_rooms  1  901134162 6.9112e+12 392777
- accommodates                                  1  976502697 6.9112e+12 392777
- review_scores_checkin                         1 1014462013 6.9113e+12 392777
- number_of_reviews_l30d                        1 1288579692 6.9115e+12 392778
+ reviews_per_month                             1   24411188 6.9102e+12 392778
+ review_scores_location                        1     653045 6.9103e+12 392778
- minimum_nights                                1 1892821305 6.9121e+12 392780
- banios                                        1 2039276398 6.9123e+12 392780
- review_scores_cleanliness                     1 8067414789 6.9183e+12 392798

Step:  AIC=392774.5
price ~ host_total_listings_count + latitude + longitude + accommodates +
    bedrooms + minimum_nights + maximum_nights + number_of_reviews +
    review_scores_cleanliness + review_scores_checkin + calculated_host_listings_count_privat
    calculated_host_listings_count_shared_rooms + number_of_reviews_l30d +
    banios

                                               Df  Sum of Sq        RSS    AIC
- bedrooms                                      1  126134867 6.9105e+12 392773
- longitude                                     1  212674304 6.9105e+12 392773
- calculated_host_listings_count_shared_rooms   1  264471579 6.9106e+12 392773
- number_of_reviews                             1  331889537 6.9107e+12 392773
- maximum_nights                                1  415978787 6.9107e+12 392774
- host_total_listings_count                     1  446598432 6.9108e+12 392774
- latitude                                      1  581407402 6.9109e+12 392774
<none>                                                       6.9103e+12 392774
- calculated_host_listings_count_private_rooms  1  952872607 6.9113e+12 392775
- accommodates                                  1  969723147 6.9113e+12 392775
- review_scores_checkin                         1 1039385517 6.9114e+12 392775
+ availability_365                              1   74636365 6.9103e+12 392776
- number_of_reviews_l30d                        1 1361460566 6.9117e+12 392776
+ reviews_per_month                             1   18779953 6.9103e+12 392776
+ review_scores_location                        1     381577 6.9103e+12 392776
- minimum_nights                                1 1904831793 6.9122e+12 392778
- banios                                       1 2041877773 6.9124e+12 392778
- review_scores_cleanliness                    1 8042925784 6.9184e+12 392796

Step:  AIC=392772.9
price ~ host_total_listings_count + latitude + longitude + accommodates +
    minimum_nights + maximum_nights + number_of_reviews + review_scores_cleanliness +
    review_scores_checkin + calculated_host_listings_count_private_rooms +
    calculated_host_listings_count_shared_rooms + number_of_reviews_l30d +
    banios

                                              Df  Sum of Sq        RSS    AIC
- longitude                                    1  202514625 6.9107e+12 392771
- calculated_host_listings_count_shared_rooms   1  305926411 6.9108e+12 392772
- number_of_reviews                            1  320768395 6.9108e+12 392772
- maximum_nights                               1  407134715 6.9109e+12 392772
- host_total_listings_count                    1  443176761 6.9109e+12 392772
- latitude                                     1  579976098 6.9110e+12 392773
<none>                                                      6.9105e+12 392773
- calculated_host_listings_count_private_rooms  1  978310409 6.9114e+12 392774
- review_scores_checkin                        1 1022978418 6.9115e+12 392774
+ bedrooms                                     1  126134867 6.9103e+12 392774
+ availability_365                             1   76394198 6.9104e+12 392775
- number_of_reviews_l30d                       1 1364921604 6.9118e+12 392775
+ reviews_per_month                            1   16996576 6.9104e+12 392775
+ review_scores_location                       1    1116768 6.9105e+12 392775
- accommodates                                 1 1673088884 6.9121e+12 392776
- minimum_nights                               1 1907723288 6.9124e+12 392776
- banios                                       1 3411676750 6.9139e+12 392781
- review_scores_cleanliness                    1 8079482545 6.9185e+12 392794

Step:  AIC=392771.4
price ~ host_total_listings_count + latitude + accommodates +
    minimum_nights + maximum_nights + number_of_reviews + review_scores_cleanliness +
    review_scores_checkin + calculated_host_listings_count_private_rooms +
    calculated_host_listings_count_shared_rooms + number_of_reviews_l30d +
    banios

                                              Df  Sum of Sq        RSS    AIC
- number_of_reviews                            1  306400645 6.9110e+12 392770
- calculated_host_listings_count_shared_rooms   1  326101523 6.9110e+12 392770
- maximum_nights                               1  397595990 6.9111e+12 392771
- host_total_listings_count                    1  477963842 6.9111e+12 392771
- latitude                                     1  537241689 6.9112e+12 392771
```

```
<none>                                                        6.9107e+12 392771
- calculated_host_listings_count_private_rooms  1  989880866 6.9116e+12 392772
- review_scores_checkin                         1 1012277787 6.9117e+12 392772
+ longitude                                     1  202514625 6.9105e+12 392773
+ bedrooms                                      1  115975188 6.9105e+12 392773
+ availability_365                              1   75091637 6.9106e+12 392773
+ reviews_per_month                             1   15977289 6.9106e+12 392773
+ review_scores_location                        1        346 6.9107e+12 392773
- number_of_reviews_l30d                        1 1451715034 6.9121e+12 392774
- accommodates                                  1 1626611459 6.9123e+12 392774
- minimum_nights                                1 1920448935 6.9126e+12 392775
- banios                                        1 3591836302 6.9143e+12 392780
- review_scores_cleanliness                     1 7991262869 6.9187e+12 392793

Step:  AIC=392770.3
price ~ host_total_listings_count + latitude + accommodates +
    minimum_nights + maximum_nights + review_scores_cleanliness +
    review_scores_checkin + calculated_host_listings_count_private_rooms +
    calculated_host_listings_count_shared_rooms + number_of_reviews_l30d +
    banios

                                               Df  Sum of Sq        RSS    AIC
- calculated_host_listings_count_shared_rooms   1  328463902 6.9113e+12 392769
- maximum_nights                                1  329194234 6.9113e+12 392769
- host_total_listings_count                     1  458083266 6.9114e+12 392770
- latitude                                      1  596562471 6.9116e+12 392770
<none>                                                        6.9110e+12 392770
- calculated_host_listings_count_private_rooms  1 1013689334 6.9120e+12 392771
- review_scores_checkin                         1 1050404184 6.9120e+12 392771
+ number_of_reviews                             1  306400645 6.9107e+12 392771
- number_of_reviews_l30d                        1 1152251188 6.9121e+12 392772
+ longitude                                     1  188146874 6.9108e+12 392772
+ reviews_per_month                             1  137093255 6.9108e+12 392772
+ bedrooms                                      1  105900428 6.9109e+12 392772
+ availability_365                              1   96575074 6.9109e+12 392772
+ review_scores_location                        1      29334 6.9110e+12 392772
- accommodates                                  1 1678765135 6.9126e+12 392773
- minimum_nights                                1 1900597837 6.9129e+12 392774
- banios                                        1 3493069785 6.9145e+12 392778
- review_scores_cleanliness                     1 7934576527 6.9189e+12 392791

Step:  AIC=392769.3
price ~ host_total_listings_count + latitude + accommodates +
    minimum_nights + maximum_nights + review_scores_cleanliness +
    review_scores_checkin + calculated_host_listings_count_private_rooms +
    number_of_reviews_l30d + banios

                                               Df  Sum of Sq        RSS    AIC
- maximum_nights                                1  318368749 6.9116e+12 392768
- host_total_listings_count                     1  499841711 6.9118e+12 392769
- latitude                                      1  537031988 6.9118e+12 392769
<none>                                                       6.9113e+12 392769
- review_scores_checkin                         1 1042923276 6.9123e+12 392770
+ calculated_host_listings_count_shared_rooms   1  328463902 6.9110e+12 392770
+ number_of_reviews                             1  308763023 6.9110e+12 392770
- number_of_reviews_l30d                        1 1120810345 6.9124e+12 392771
+ longitude                                     1  207632077 6.9111e+12 392771
+ reviews_per_month                             1  152509315 6.9111e+12 392771
+ bedrooms                                      1  145239654 6.9111e+12 392771
+ availability_365                              1  101780683 6.9112e+12 392771
+ review_scores_location                        1      99026 6.9113e+12 392771
- calculated_host_listings_count_private_rooms  1 1740258091 6.9130e+12 392772
- accommodates                                  1 1824876886 6.9131e+12 392773
- minimum_nights                                1 1852913798 6.9131e+12 392773
- banios                                        1 3225026657 6.9145e+12 392777
- review_scores_cleanliness                     1 7898342177 6.9192e+12 392790

Step:  AIC=392768.2
price ~ host_total_listings_count + latitude + accommodates +
    minimum_nights + review_scores_cleanliness + review_scores_checkin +
    calculated_host_listings_count_private_rooms + number_of_reviews_l30d +
    banios

                                               Df  Sum of Sq        RSS    AIC
- host_total_listings_count                     1  459301557 6.9121e+12 392768
- latitude                                      1  569399112 6.9122e+12 392768
<none>                                                       6.9116e+12 392768
- review_scores_checkin                         1 1031012810 6.9126e+12 392769
+ maximum_nights                                1  318368749 6.9113e+12 392769
+ calculated_host_listings_count_shared_rooms   1  317638417 6.9113e+12 392769
- number_of_reviews_l30d                        1 1104304349 6.9127e+12 392769
+ number_of_reviews                             1  241037644 6.9114e+12 392770
+ longitude                                     1  200216272 6.9114e+12 392770
+ reviews_per_month                             1  166603960 6.9114e+12 392770
+ availability_365                              1  147715630 6.9115e+12 392770
+ bedrooms                                      1  137577988 6.9115e+12 392770
+ review_scores_location                        1        186 6.9116e+12 392770
- calculated_host_listings_count_private_rooms  1 1671651621 6.9133e+12 392771
- accommodates                                  1 1771327382 6.9134e+12 392771
- minimum_nights                                1 1807626416 6.9134e+12 392771
- banios                                        1 3288028434 6.9149e+12 392776
- review_scores_cleanliness                     1 7847049869 6.9195e+12 392789

Step:  AIC=392767.5
price ~ latitude + accommodates + minimum_nights + review_scores_cleanliness +
    review_scores_checkin + calculated_host_listings_count_private_rooms +
    number_of_reviews_l30d + banios

                                               Df  Sum of Sq        RSS    AIC
- latitude                                      1  647298458 6.9127e+12 392767
<none>                                                       6.9121e+12 392768
+ host_total_listings_count                     1  459301557 6.9116e+12 392768
- review_scores_checkin                         1  967646743 6.9130e+12 392768
+ calculated_host_listings_count_shared_rooms   1  357696029 6.9117e+12 392768
- number_of_reviews_l30d                        1 1083105824 6.9132e+12 392769
+ maximum_nights                                1  277828595 6.9118e+12 392769
+ longitude                                     1  236386945 6.9118e+12 392769
+ number_of_reviews                             1  227442540 6.9118e+12 392769
+ reviews_per_month                             1  179657202 6.9119e+12 392769
+ bedrooms                                      1  136469344 6.9119e+12 392769
+ availability_365                              1  109601920 6.9120e+12 392769
+ review_scores_location                        1     376682 6.9121e+12 392770
- calculated_host_listings_count_private_rooms  1 1402762384 6.9135e+12 392770
- minimum_nights                                1 1776421786 6.9138e+12 392771
- accommodates                                  1 1922717394 6.9140e+12 392771
- banios                                        1 3331524076 6.9154e+12 392775
- review_scores_cleanliness                     1 7661143327 6.9197e+12 392788

Step:  AIC=392767.4
price ~ accommodates + minimum_nights + review_scores_cleanliness +
    review_scores_checkin + calculated_host_listings_count_private_rooms +
    number_of_reviews_l30d + banios

                                               Df  Sum of Sq        RSS    AIC
<none>                                                       6.9127e+12 392767
+ latitude                                      1  647298458 6.9121e+12 392768
+ host_total_listings_count                     1  537200902 6.9122e+12 392768
- number_of_reviews_l30d                        1  905021911 6.9136e+12 392768
- review_scores_checkin                         1 1013490743 6.9137e+12 392768
```

```
+ maximum_nights                               1  306957867 6.9124e+12 392769
+ calculated_host_listings_count_shared_rooms  1  292418356 6.9124e+12 392769
+ number_of_reviews                            1  276660876 6.9124e+12 392769
+ reviews_per_month                            1  253299212 6.9125e+12 392769
+ longitude                                    1  184651623 6.9125e+12 392769
+ bedrooms                                     1  130309733 6.9126e+12 392769
+ availability_365                             1  126966152 6.9126e+12 392769
- calculated_host_listings_count_private_rooms 1 1311302950 6.9140e+12 392769
+ review_scores_location                       1    3490765 6.9127e+12 392769
- minimum_nights                               1 1763907876 6.9145e+12 392770
- accommodates                                 1 2073656799 6.9148e+12 392771
- banios                                       1 3278290753 6.9160e+12 392775
- review_scores_cleanliness                    1 7682627305 6.9204e+12 392788
```

```r
summary(seleccion)
```

```

Call:
lm(formula = price ~ accommodates + minimum_nights + review_scores_cleanliness +
    review_scores_checkin + calculated_host_listings_count_private_rooms +
    number_of_reviews_l30d + banios)

Residuals:
    Min      1Q  Median      3Q     Max
 -21998   -1061    -552      22 1830686

Coefficients:
                                             Estimate Std. Error t value
(Intercept)                                  6154.460   1669.668   3.686
accommodates                                  182.156     74.429   2.447
minimum_nights                                 13.060      5.786   2.257
review_scores_cleanliness                   -1976.152    419.502  -4.711
review_scores_checkin                         796.235    465.371   1.711
calculated_host_listings_count_private_rooms  -53.706     27.596  -1.946
number_of_reviews_l30d                       -106.279     65.733  -1.617
banios                                        552.808    179.647   3.077
                                            Pr(>|t|)
(Intercept)                                 0.000228 ***
accommodates                                0.014399 *
minimum_nights                              0.024006 *
review_scores_cleanliness                   2.49e-06 ***
```

```
review_scores_checkin                          0.087103 .
calculated_host_listings_count_private_rooms 0.051647 .
number_of_reviews_l30d                         0.105933
banios                                         0.002092 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18610 on 19967 degrees of freedom
Multiple R-squared:  0.003749,  Adjusted R-squared:  0.0034
F-statistic: 10.73 on 7 and 19967 DF,  p-value: 1.428e-13
```

Dado que el anterior modelo contempla la base datos, hay que tener en cuenta que dichos datos tienen valores atípicos, por lo que será indispensable quitar para obtner un mejor modelo.

Creamos el siguiente código para eliminar datos atípicos por columna:

```r
#Variables sin datos atípicos

# Crear una función para eliminar valores atípicos basados en el rango intercuartílico en

remove_outliers_iqr <- function(data, multiplier) {
  for (columna in names(data)) {
    q1 <- quantile(data[[columna]], 0.25)
    q3 <- quantile(data[[columna]], 0.75)
    iqr <- q3 - q1
    lower_limit <- q1 - multiplier * iqr
    upper_limit <- q3 + multiplier * iqr
    data <- subset(data, data[[columna]] >= lower_limit & data[[columna]] <= upper_limit)
  }
  return(data)
}
# Llamamos a la función para eliminar valores atípicos en todas las columnas del dataframe

data <- remove_outliers_iqr(datos, 1.5)  # Elimina los valores que están fuera de 1.5 vece
```

Así probamos nuestro modelo 1, pero ahora utilizando una base de datos sin valores atípicos, y seleccionamos un modelo sin considerar variables que no son significantes.

```r
attach(data)
```

```
The following objects are masked from datos (pos = 3):
    accommodates, availability_365, banios, bedrooms,
    calculated_host_listings_count_private_rooms,
    calculated_host_listings_count_shared_rooms,
    host_total_listings_count, latitude, longitude, maximum_nights,
    minimum_nights, number_of_reviews, number_of_reviews_l30d, price,
    review_scores_checkin, review_scores_cleanliness,
    review_scores_location, reviews_per_month

The following objects are masked from datos (pos = 4):

    accommodates, availability_365, banios, bedrooms,
    calculated_host_listings_count_private_rooms,
    calculated_host_listings_count_shared_rooms,
    host_total_listings_count, latitude, longitude, maximum_nights,
    minimum_nights, number_of_reviews, number_of_reviews_l30d, price,
    review_scores_checkin, review_scores_cleanliness,
    review_scores_location, reviews_per_month
```

```r
#Modelo sin atípicos
modelo1 <- lm(price~host_total_listings_count
              +latitude
              +longitude
              +accommodates
              +bedrooms
              +minimum_nights
              +maximum_nights
              +availability_365
              +number_of_reviews
              +review_scores_cleanliness
              +review_scores_checkin
              +review_scores_location
              +calculated_host_listings_count_private_rooms
              +calculated_host_listings_count_shared_rooms
              +reviews_per_month
              +number_of_reviews_l30d
              +banios)
summary(modelo1)
```

```
Call:
lm(formula = price ~ host_total_listings_count + latitude + longitude +
    accommodates + bedrooms + minimum_nights + maximum_nights + 
    availability_365 + number_of_reviews + review_scores_cleanliness + 
    review_scores_checkin + review_scores_location + calculated_host_listings_count_private_
    calculated_host_listings_count_shared_rooms + reviews_per_month + 
    number_of_reviews_l30d + banios)

Residuals:
     Min       1Q   Median       3Q      Max 
-1223.32  -290.61   -84.62   200.74  2326.81 

Coefficients: (1 not defined because of singularities)
                                               Estimate Std. Error t value
(Intercept)                                  -5.459e+05  2.775e+04 -19.674
host_total_listings_count                     1.423e+01  7.847e-01  18.132
latitude                                      1.571e+03  1.622e+02   9.691
longitude                                    -5.160e+03  2.813e+02 -18.346
accommodates                                  1.225e+02  5.786e+00  21.172
bedrooms                                      3.214e+01  1.449e+01   2.218
minimum_nights                               -7.371e+00  4.870e+00  -1.514
maximum_nights                                4.259e-03  1.053e-02   0.404
availability_365                              3.485e-01  4.096e-02   8.508
number_of_reviews                            -6.820e-01  1.787e-01  -3.816
review_scores_cleanliness                     3.189e+02  3.385e+01   9.419
review_scores_checkin                        -1.521e+02  6.294e+01  -2.416
review_scores_location                        6.474e+02  6.008e+01  10.776
calculated_host_listings_count_private_rooms -9.331e+01  4.297e+00 -21.718
calculated_host_listings_count_shared_rooms          NA         NA      NA
reviews_per_month                            -5.054e+00  6.402e+00  -0.789
number_of_reviews_l30d                       -7.070e+00  4.729e+00  -1.495
banios                                        2.115e+02  1.456e+01  14.525
                                             Pr(>|t|)    
(Intercept)                                   < 2e-16 ***
host_total_listings_count                     < 2e-16 ***
latitude                                      < 2e-16 ***
longitude                                     < 2e-16 ***
accommodates                                  < 2e-16 ***
bedrooms                                     0.026605 *  
minimum_nights                               0.130148    
maximum_nights                               0.686019    
availability_365                              < 2e-16 ***
number_of_reviews                            0.000137 ***
review_scores_cleanliness                     < 2e-16 ***
review_scores_checkin                        0.015722 *  
```

```
review_scores_location                          < 2e-16 ***
calculated_host_listings_count_private_rooms    < 2e-16 ***
calculated_host_listings_count_shared_rooms          NA    
reviews_per_month                             0.429878    
number_of_reviews_l30d                        0.134925    
banios                                          < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 433.9 on 7684 degrees of freedom
Multiple R-squared:  0.3946,    Adjusted R-squared:  0.3933 
F-statistic:   313 on 16 and 7684 DF,  p-value: < 2.2e-16
```

```
seleccion <- stepAIC(modelo1, direction = c("both"));
```

```
Start:  AIC=93549.56
price ~ host_total_listings_count + latitude + longitude + accommodates + 
    bedrooms + minimum_nights + maximum_nights + availability_365 + 
    number_of_reviews + review_scores_cleanliness + review_scores_checkin + 
    review_scores_location + calculated_host_listings_count_private_rooms + 
    calculated_host_listings_count_shared_rooms + reviews_per_month + 
    number_of_reviews_l30d + banios


Step:  AIC=93549.56
price ~ host_total_listings_count + latitude + longitude + accommodates + 
    bedrooms + minimum_nights + maximum_nights + availability_365 + 
    number_of_reviews + review_scores_cleanliness + review_scores_checkin + 
    review_scores_location + calculated_host_listings_count_private_rooms + 
    reviews_per_month + number_of_reviews_l30d + banios

                                               Df Sum of Sq        RSS   AIC
- maximum_nights                                1     30767 1446520887 93548
- reviews_per_month                             1    117319 1446607439 93548
<none>                                                      1446490120 93550
- number_of_reviews_l30d                        1    420802 1446910921 93550
- minimum_nights                                1    431318 1446921437 93550
- bedrooms                                      1    925826 1447415946 93552
- review_scores_checkin                         1   1098658 1447588778 93553
- number_of_reviews                             1   2741239 1449231359 93562
- availability_365                              1  13627115 1460117234 93620
- review_scores_cleanliness                     1  16702585 1463192705 93636
- latitude                                      1  17678074 1464168194 93641
- review_scores_location                        1  21858850 1468348969 93663
- banios                                        1  39716258 1486206377 93756
- host_total_listings_count                     1  61889682 1508379802 93870
- longitude                                     1  63362412 1509852532 93878
- accommodates                                  1  84383882 1530874001 93984
- calculated_host_listings_count_private_rooms  1  88787633 1535277753 94006

Step:  AIC=93547.72
price ~ host_total_listings_count + latitude + longitude + accommodates + 
    bedrooms + minimum_nights + availability_365 + number_of_reviews + 
    review_scores_cleanliness + review_scores_checkin + review_scores_location + 
    calculated_host_listings_count_private_rooms + reviews_per_month + 
    number_of_reviews_l30d + banios

                                               Df Sum of Sq        RSS   AIC
- reviews_per_month                             1    132193 1446653079 93546
<none>                                                      1446520887 93548
- number_of_reviews_l30d                        1    420887 1446941773 93548
- minimum_nights                                1    434881 1446955768 93548
+ maximum_nights                                1     30767 1446490120 93550
- bedrooms                                      1    926548 1447447435 93551
- review_scores_checkin                         1   1103685 1447624572 93552
- number_of_reviews                             1   2716146 1449237033 93560
- availability_365                              1  14081628 1460602515 93620
- review_scores_cleanliness                     1  16673092 1463193978 93634
- latitude                                      1  17666620 1464187507 93639
- review_scores_location                        1  21830396 1468351283 93661
- banios                                        1  39702573 1486223460 93754
- host_total_listings_count                     1  62003073 1508523960 93869
- longitude                                     1  63431078 1509951964 93876
- accommodates                                  1  84577833 1531098720 93983
- calculated_host_listings_count_private_rooms  1  88857001 1535377888 94005

Step:  AIC=93546.42
price ~ host_total_listings_count + latitude + longitude + accommodates + 
    bedrooms + minimum_nights + availability_365 + number_of_reviews + 
    review_scores_cleanliness + review_scores_checkin + review_scores_location + 
    calculated_host_listings_count_private_rooms + number_of_reviews_l30d + 
    banios

                                               Df Sum of Sq        RSS   AIC
```

```
- minimum_nights                                1    370566 1447023646 93546
<none>                                                      1446653079 93546
+ reviews_per_month                             1    132193 1446520887 93548
+ maximum_nights                                1     45641 1446607439 93548
- bedrooms                                      1    962735 1447615814 93550
- number_of_reviews_l30d                        1   1021929 1447675008 93550
- review_scores_checkin                         1   1060000 1447713079 93550
- number_of_reviews                             1   3198838 1449851917 93561
- availability_365                              1  13975687 1460628766 93618
- review_scores_cleanliness                     1  16636163 1463289242 93632
- latitude                                      1  17548216 1464201295 93637
- review_scores_location                        1  21913560 1468566639 93660
- banios                                        1  39636627 1486289706 93753
- host_total_listings_count                     1  61889420 1508542500 93867
- longitude                                     1  63515062 1510168141 93875
- accommodates                                  1  84774450 1531427529 93983
- calculated_host_listings_count_private_rooms  1  89256077 1535909156 94005

Step:  AIC=93546.4
price ~ host_total_listings_count + latitude + longitude + accommodates + 
    bedrooms + availability_365 + number_of_reviews + review_scores_cleanliness + 
    review_scores_checkin + review_scores_location + calculated_host_listings_count_private_
    number_of_reviews_l30d + banios

                                               Df Sum of Sq        RSS   AIC
<none>                                                      1447023646 93546
+ minimum_nights                                1    370566 1446653079 93546
+ reviews_per_month                             1     67878 1446955768 93548
+ maximum_nights                                1     44982 1446978664 93548
- bedrooms                                      1    907465 1447931111 93549
- number_of_reviews_l30d                        1    941796 1447965442 93549
- review_scores_checkin                         1   1117206 1448140852 93550
- number_of_reviews                             1   3283850 1450307496 93562
- availability_365                              1  14729023 1461752669 93622
- review_scores_cleanliness                     1  16752559 1463776205 93633
- latitude                                      1  17596789 1464620434 93637
- review_scores_location                        1  21978521 1469002167 93660
- banios                                        1  39379111 1486402757 93751
- longitude                                     1  63398815 1510422461 93875
- host_total_listings_count                     1  63672663 1510696308 93876
- accommodates                                  1  85745632 1532769278 93988
- calculated_host_listings_count_private_rooms  1  89001653 1536025299 94004
```

```r
summary(seleccion)
```

```
Call:
lm(formula = price ~ host_total_listings_count + latitude + longitude + 
    accommodates + bedrooms + availability_365 + number_of_reviews + 
    review_scores_cleanliness + review_scores_checkin + review_scores_location + 
    calculated_host_listings_count_private_rooms + number_of_reviews_l30d + 
    banios)

Residuals:
     Min       1Q   Median       3Q      Max 
-1215.86  -291.61   -84.42   200.10  2329.41 

Coefficients:
                                               Estimate Std. Error t value
(Intercept)                                  -5.457e+05  2.774e+04 -19.674
host_total_listings_count                     1.431e+01  7.782e-01  18.391
latitude                                      1.565e+03  1.619e+02   9.668
longitude                                    -5.160e+03  2.812e+02 -18.352
accommodates                                  1.226e+02  5.743e+00  21.343
bedrooms                                      3.174e+01  1.446e+01   2.196
availability_365                              3.557e-01  4.021e-02   8.846
number_of_reviews                            -7.156e-01  1.713e-01  -4.177
review_scores_cleanliness                     3.191e+02  3.382e+01   9.434
review_scores_checkin                        -1.529e+02  6.275e+01  -2.436
review_scores_location                        6.488e+02  6.004e+01  10.805
calculated_host_listings_count_private_rooms -9.232e+01  4.246e+00 -21.744
number_of_reviews_l30d                       -8.842e+00  3.953e+00  -2.237
banios                                        2.104e+02  1.455e+01  14.464
                                             Pr(>|t|)    
(Intercept)                                   < 2e-16 ***
host_total_listings_count                     < 2e-16 ***
latitude                                      < 2e-16 ***
longitude                                     < 2e-16 ***
accommodates                                  < 2e-16 ***
bedrooms                                       0.0281 *  
availability_365                              < 2e-16 ***
number_of_reviews                            2.99e-05 ***
review_scores_cleanliness                     < 2e-16 ***
review_scores_checkin                          0.0149 *  
review_scores_location                        < 2e-16 ***
```

```
calculated_host_listings_count_private_rooms  < 2e-16 ***
number_of_reviews_l30d                          0.0253 *  
banios                                        < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 433.9 on 7687 degrees of freedom
Multiple R-squared:  0.3944,	Adjusted R-squared:  0.3934 
F-statistic: 385.1 on 13 and 7687 DF,  p-value: < 2.2e-16
```

Teniendo ahora la base de datos definitiva, se ajustó un primer modelo para la misma, sin embargo, tras esto nos pudimos percatar de que la base de datos contenía valores atípicos, por lo que resultó indispensable quitarlos para obtener un mejor modelo y, tras eliminarlos, pudimos ver un aumento en la precisión del modelo tanto en su R^2 ajustada como en los p-values de cada una de sus variables.

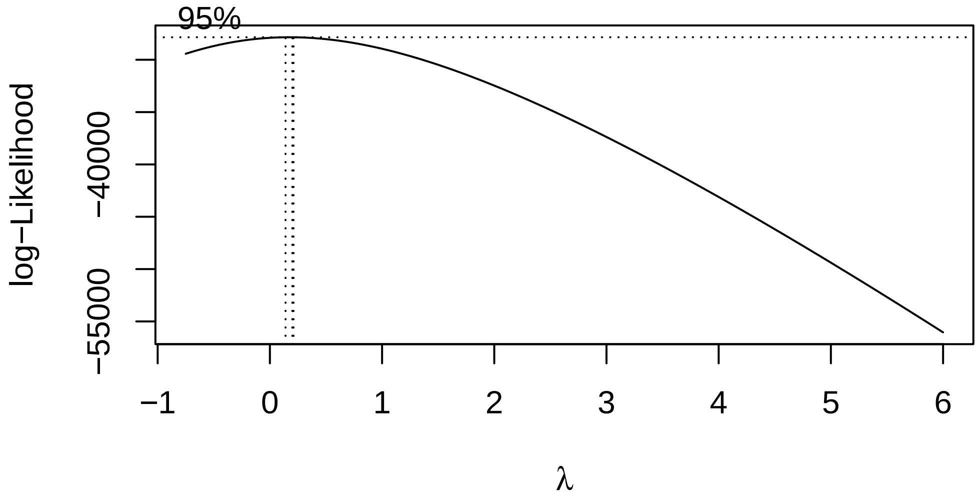
Nota: Para más información de cada transformación consultar el archivo "Documento expandido"

4. Los datos atípicos y los supuestos

Finalmente procedimos a a acotar nuestro modelo, haciéndolo efectivo únicamente para Airbnb's que cumplen ciertas características.

Para tratar de subir un poco nuestra R^2 ajustada, utilizamos la función de Box-Cox, para obtener una potencia (que llamaremos lambda) que nos ayudara a reescalar nuestros datos de la variable precios para así poder trabajar con ellos de manera más fácil. Dicha lambda es la siguiente:

```r
BC <-boxcox(seleccion, lambda = seq(-0.75,6, by = 0.05), plotit = TRUE)
```



```
lamda <- BC$x[which.max(BC$y)]
```

Una vez encontrada la lambda adecuada, transformamos nuestra variable de precios. Así, obtenemos el siguiente modelo, el cual es el mejor según el análisis previo.

Para lograrlo, aplicaremos la siguiente transformación:

$$y = \frac{price^{\lambda} - 1}{\lambda}$$

```
modelo2 <- lm((((price^lamda)-1)/lamda) ~ host_total_listings_count + latitude + longitude
                accommodates + bedrooms + availability_365 + number_of_reviews +
                review_scores_cleanliness + review_scores_checkin + review_scores_location
                calculated_host_listings_count_private_rooms + number_of_reviews_l30d +
                banios)
summary(modelo2)
```

```
Call:
lm(formula = (((price^lamda) - 1)/lamda) ~ host_total_listings_count +
    latitude + longitude + accommodates + bedrooms + availability_365 +
    number_of_reviews + review_scores_cleanliness + review_scores_checkin + 
    review_scores_location + calculated_host_listings_count_private_rooms + 
    number_of_reviews_l30d + banios)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.7521 -1.1399 -0.1166  1.0424  7.6814 

Coefficients:
                                               Estimate Std. Error t value
(Intercept)                                  -2.259e+03  1.054e+02 -21.437
host_total_listings_count                     5.677e-02  2.957e-03  19.199
latitude                                      5.802e+00  6.151e-01   9.433
longitude                                    -2.161e+01  1.068e+00 -20.228
accommodates                                  5.696e-01  2.182e-02  26.101
bedrooms                                      1.263e-02  5.493e-02   0.230
availability_365                              1.507e-03  1.528e-04   9.862
number_of_reviews                            -2.145e-03  6.510e-04  -3.295
review_scores_cleanliness                     1.277e+00  1.285e-01   9.937
review_scores_checkin                        -6.342e-01  2.384e-01  -2.660
review_scores_location                        2.537e+00  2.281e-01  11.121
calculated_host_listings_count_private_rooms -4.686e-01  1.613e-02 -29.046
number_of_reviews_l30d                       -2.827e-02  1.502e-02  -1.882
banios                                        6.706e-01  5.527e-02  12.133
                                             Pr(>|t|)    
(Intercept)                                   < 2e-16 ***
host_total_listings_count                     < 2e-16 ***
latitude                                      < 2e-16 ***
longitude                                     < 2e-16 ***
accommodates                                  < 2e-16 ***
bedrooms                                     0.818192    
availability_365                              < 2e-16 ***
number_of_reviews                            0.000988 ***
review_scores_cleanliness                     < 2e-16 ***
review_scores_checkin                        0.007833 ** 
review_scores_location                        < 2e-16 ***
calculated_host_listings_count_private_rooms  < 2e-16 ***
number_of_reviews_l30d                       0.059881 .  
banios                                        < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.649 on 7687 degrees of freedom
```

```
Multiple R-squared:  0.4471,    Adjusted R-squared:  0.4462 
F-statistic: 478.2 on 13 and 7687 DF,  p-value: < 2.2e-16
```

```
seleccion <- stepAIC(modelo2, direction = c("both"));
```

```
Start:  AIC=7713.96
(((price^lamda) - 1)/lamda) ~ host_total_listings_count + latitude + 
    longitude + accommodates + bedrooms + availability_365 + 
    number_of_reviews + review_scores_cleanliness + review_scores_checkin + 
    review_scores_location + calculated_host_listings_count_private_rooms + 
    number_of_reviews_l30d + banios

                                              Df Sum of Sq   RSS    AIC
- bedrooms                                     1      0.14 20893 7712.0
<none>                                                     20893 7714.0
- number_of_reviews_l30d                       1      9.63 20902 7715.5
- review_scores_checkin                        1     19.23 20912 7719.0
- number_of_reviews                            1     29.51 20922 7722.8
- latitude                                     1    241.85 21135 7800.6
- availability_365                             1    264.32 21157 7808.8
- review_scores_cleanliness                    1    268.36 21161 7810.2
- review_scores_location                       1    336.13 21229 7834.9
- banios                                       1    400.10 21293 7858.0
- host_total_listings_count                    1   1001.87 21895 8072.7
- longitude                                    1   1112.13 22005 8111.4
- accommodates                                 1   1851.59 22744 8365.9
- calculated_host_listings_count_private_rooms  1   2293.00 23186 8513.9

Step:  AIC=7712.01
(((price^lamda) - 1)/lamda) ~ host_total_listings_count + latitude + 
    longitude + accommodates + availability_365 + number_of_reviews + 
    review_scores_cleanliness + review_scores_checkin + review_scores_location + 
    calculated_host_listings_count_private_rooms + number_of_reviews_l30d + 
    banios

                                              Df Sum of Sq   RSS    AIC
<none>                                                     20893 7712.0
- number_of_reviews_l30d                       1       9.7 20903 7713.6
+ bedrooms                                     1       0.1 20893 7714.0
- review_scores_checkin                        1      19.2 20912 7717.1
- number_of_reviews                            1      29.7 20923 7721.0
- latitude                                       1     241.7 21135 7798.6
- availability_365                               1     264.7 21158 7807.0
- review_scores_cleanliness                      1     268.2 21161 7808.2
- review_scores_location                         1     336.1 21229 7832.9
- banios                                         1     451.2 21344 7874.6
- host_total_listings_count                      1    1008.8 21902 8073.2
- longitude                                      1    1112.4 22005 8109.5
- calculated_host_listings_count_private_rooms   1    2301.0 23194 8514.6
- accommodates                                   1    3367.4 24260 8860.8
```

```
summary(seleccion)
```

```
Call:
lm(formula = (((price^lamda) - 1)/lamda) ~ host_total_listings_count +
    latitude + longitude + accommodates + availability_365 +
    number_of_reviews + review_scores_cleanliness + review_scores_checkin +
    review_scores_location + calculated_host_listings_count_private_rooms +
    number_of_reviews_l30d + banios)

Residuals:
    Min      1Q  Median      3Q     Max
-5.7439 -1.1389 -0.1156  1.0403  7.6815

Coefficients:
                                               Estimate Std. Error t value
(Intercept)                                  -2.260e+03  1.054e+02 -21.440
host_total_listings_count                     5.671e-02  2.943e-03  19.267
latitude                                      5.799e+00  6.149e-01   9.431
longitude                                    -2.161e+01  1.068e+00 -20.232
accommodates                                  5.729e-01  1.628e-02  35.201
availability_365                              1.508e-03  1.528e-04   9.869
number_of_reviews                            -2.151e-03  6.504e-04  -3.308
review_scores_cleanliness                     1.276e+00  1.285e-01   9.935
review_scores_checkin                        -6.341e-01  2.384e-01  -2.660
review_scores_location                        2.537e+00  2.281e-01  11.121
calculated_host_listings_count_private_rooms -4.688e-01  1.611e-02 -29.098
number_of_reviews_l30d                       -2.832e-02  1.502e-02  -1.886
banios                                        6.746e-01  5.236e-02  12.885
                                             Pr(>|t|)
(Intercept)                                   < 2e-16 ***
```

```
host_total_listings_count                          < 2e-16 ***
latitude                                           < 2e-16 ***
longitude                                          < 2e-16 ***
accommodates                                       < 2e-16 ***
availability_365                                   < 2e-16 ***
number_of_reviews                                 0.000945 ***
review_scores_cleanliness                          < 2e-16 ***
review_scores_checkin                             0.007840 **
review_scores_location                             < 2e-16 ***
calculated_host_listings_count_private_rooms       < 2e-16 ***
number_of_reviews_l30d                            0.059334 .
banios                                             < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.649 on 7688 degrees of freedom
Multiple R-squared:  0.4471,    Adjusted R-squared:  0.4463
F-statistic: 518.1 on 12 and 7688 DF,  p-value: < 2.2e-16
```

Tras este reescalamiento, se eliminaron nuevamente los valores atípicos mediante dos métodos siendoel primero el uso de la distancia de Cook para su eliminación y el segundo acotando el costo de arrendamiento de un Airbnb, acotando a su vez a nuestro modelo para Airbnb´s cuyo costo anteriormente mencionado no supere los $2385 MXN (en escala transformada).

```
# Calcular la distancia de Cook del mejor modelo
distancia_cook <- cooks.distance(seleccion)

# Definir el umbral para la distancia de Cook
umbral <- 4/length(distancia_cook)  # Utilizamos el umbral 4/n, donde n es el número de ob

# Identificar las observaciones influyentes
observaciones_influyentes <- which(distancia_cook > umbral)

# Paso 4: Eliminar las observaciones influyentes del conjunto de datos
datos_filtrados <- data[-observaciones_influyentes, ]
```

```
summary(datos_filtrados$price)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
    180     600     932    1015    1314    2750 
```

Finalmente con el resúmen anterior, observemos que con la siguiente fórmula

$$1314 + (1.5(1314 - 600)) = 2385$$

Así eliminamos precios que esten por encima de dicho valor, con el siguiente código:

```
#Quitamos los últimos valore atípicos en los precios
data_clean <- subset(datos_filtrados, datos_filtrados$price <= 2385)
#Volvemos a eliminar datos atípicos
data_2 <- remove_outliers_iqr(data_clean, 1.5)  # Elimina los valores que están fuera de 1
```

Finalmente nuestro mejor modelo queda de la siguiente manera:

```
attach(data_2)
```

```
The following objects are masked from data:

    accommodates, availability_365, banios, bedrooms,
    calculated_host_listings_count_private_rooms,
    calculated_host_listings_count_shared_rooms,
    host_total_listings_count, latitude, longitude, maximum_nights,
    minimum_nights, number_of_reviews, number_of_reviews_l30d, price,
    review_scores_checkin, review_scores_cleanliness,
    review_scores_location, reviews_per_month


The following objects are masked from datos (pos = 4):

    accommodates, availability_365, banios, bedrooms,
    calculated_host_listings_count_private_rooms,
    calculated_host_listings_count_shared_rooms,
    host_total_listings_count, latitude, longitude, maximum_nights,
    minimum_nights, number_of_reviews, number_of_reviews_l30d, price,
    review_scores_checkin, review_scores_cleanliness,
    review_scores_location, reviews_per_month


The following objects are masked from datos (pos = 5):

    accommodates, availability_365, banios, bedrooms,
    calculated_host_listings_count_private_rooms,
    calculated_host_listings_count_shared_rooms,
    host_total_listings_count, latitude, longitude, maximum_nights,
    minimum_nights, number_of_reviews, number_of_reviews_l30d, price,
    review_scores_checkin, review_scores_cleanliness,
    review_scores_location, reviews_per_month
```

```
mejor_modelo_ajustado <- lm((((price^lamda) - 1)/lamda) ~ host_total_listings_count +
                        latitude + longitude + accommodates + availability_365 +
                        number_of_reviews + review_scores_cleanliness + review_scores_chec
                        review_scores_location + calculated_host_listings_count_private_ro
                        number_of_reviews_l30d + banios)
summary(mejor_modelo_ajustado)
```

```
Call:
lm(formula = (((price^lamda) - 1)/lamda) ~ host_total_listings_count +
    latitude + longitude + accommodates + availability_365 +
    number_of_reviews + review_scores_cleanliness + review_scores_checkin +
    review_scores_location + calculated_host_listings_count_private_rooms +
    number_of_reviews_l30d + banios)

Residuals:
    Min      1Q  Median      3Q     Max
-4.1785 -0.9435 -0.0778  0.9031  5.3215

Coefficients:
                                              Estimate Std. Error t value
(Intercept)                                 -2.250e+03  1.237e+02 -18.183
host_total_listings_count                    9.689e-02  7.029e-03  13.783
latitude                                     8.512e+00  6.684e-01  12.735
longitude                                   -2.095e+01  1.254e+00 -16.712
accommodates                                 4.496e-01  1.804e-02  24.923
availability_365                             9.828e-04  1.609e-04   6.110
number_of_reviews                           -1.507e-03  8.487e-04  -1.776
review_scores_cleanliness                    1.431e+00  1.612e-01   8.880
review_scores_checkin                       -8.546e-02  3.241e-01  -0.264
review_scores_location                       2.615e+00  2.999e-01   8.719
calculated_host_listings_count_private_rooms -1.118e+00  3.427e-02 -32.634
number_of_reviews_l30d                      -1.353e-02  1.660e-02  -0.815
banios                                       6.992e-01  5.463e-02  12.797
                                            Pr(>|t|)
(Intercept)                                  < 2e-16 ***
host_total_listings_count                    < 2e-16 ***
latitude                                     < 2e-16 ***
```

```
longitude                                     < 2e-16 ***
accommodates                                  < 2e-16 ***
availability_365                             1.08e-09 ***
number_of_reviews                              0.0758 .
review_scores_cleanliness                     < 2e-16 ***
review_scores_checkin                          0.7920
review_scores_location                        < 2e-16 ***
calculated_host_listings_count_private_rooms  < 2e-16 ***
number_of_reviews_l30d                         0.4153
banios                                        < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.372 on 4733 degrees of freedom
Multiple R-squared:  0.539, Adjusted R-squared:  0.5378
F-statistic: 461.2 on 12 and 4733 DF,  p-value: < 2.2e-16
```

Nuevamente tomamos el mejor modelo con la variable seleccion:

```
seleccion <- stepAIC(mejor_modelo_ajustado, direction = c("both"));
```

```
Start:  AIC=3012.33
(((price^lamda) - 1)/lamda) ~ host_total_listings_count + latitude +
    longitude + accommodates + availability_365 + number_of_reviews +
    review_scores_cleanliness + review_scores_checkin + review_scores_location +
    calculated_host_listings_count_private_rooms + number_of_reviews_l30d +
    banios

                                               Df Sum of Sq     RSS    AIC
- review_scores_checkin                         1      0.13  8904.4 3010.4
- number_of_reviews_l30d                        1      1.25  8905.5 3011.0
<none>                                                       8904.3 3012.3
- number_of_reviews                             1      5.93  8910.2 3013.5
- availability_365                              1     70.23  8974.5 3047.6
- review_scores_location                        1    143.02  9047.3 3086.0
- review_scores_cleanliness                     1    148.36  9052.6 3088.8
- latitude                                      1    305.11  9209.4 3170.2
- banios                                        1    308.11  9212.4 3171.8
- host_total_listings_count                     1    357.42  9261.7 3197.1
- longitude                                     1    525.44  9429.7 3282.4
- accommodates                                  1   1168.58 10072.9 3595.6
- calculated_host_listings_count_private_rooms  1   2003.59 10907.9 3973.5
```

```
Step:  AIC=3010.4
(((price^lamda) - 1)/lamda) ~ host_total_listings_count + latitude +
    longitude + accommodates + availability_365 + number_of_reviews +
    review_scores_cleanliness + review_scores_location + calculated_host_listings_count_priva
    number_of_reviews_l30d + banios

                                               Df Sum of Sq     RSS    AIC
- number_of_reviews_l30d                        1      1.24  8905.7 3009.1
<none>                                                       8904.4 3010.4
- number_of_reviews                             1      5.81  8910.2 3011.5
+ review_scores_checkin                         1      0.13  8904.3 3012.3
- availability_365                              1     70.76  8975.2 3046.0
- review_scores_location                        1    149.68  9054.1 3087.5
- review_scores_cleanliness                     1    158.41  9062.8 3092.1
- latitude                                      1    306.52  9210.9 3169.0
- banios                                        1    308.67  9213.1 3170.1
- host_total_listings_count                     1    358.88  9263.3 3195.9
- longitude                                     1    526.32  9430.7 3280.9
- accommodates                                  1   1170.44 10074.9 3594.5
- calculated_host_listings_count_private_rooms  1   2003.60 10908.0 3971.6

Step:  AIC=3009.06
(((price^lamda) - 1)/lamda) ~ host_total_listings_count + latitude +
    longitude + accommodates + availability_365 + number_of_reviews +
    review_scores_cleanliness + review_scores_location + calculated_host_listings_count_priva
    banios

                                               Df Sum of Sq     RSS    AIC
<none>                                                       8905.7 3009.1
+ number_of_reviews_l30d                        1      1.24  8904.4 3010.4
+ review_scores_checkin                         1      0.12  8905.5 3011.0
- number_of_reviews                             1      7.58  8913.2 3011.1
- availability_365                              1     69.54  8975.2 3044.0
- review_scores_location                        1    149.91  9055.6 3086.3
- review_scores_cleanliness                     1    157.88  9063.5 3090.5
- latitude                                      1    305.35  9211.0 3167.1
- banios                                        1    309.40  9215.1 3169.1
- host_total_listings_count                     1    358.13  9263.8 3194.2
- longitude                                     1    530.65  9436.3 3281.7
- accommodates                                  1   1169.84 10075.5 3592.8
- calculated_host_listings_count_private_rooms  1   2006.32 10912.0 3971.3
```

```r
summary(seleccion)
```

```
Call:
lm(formula = (((price^lamda) - 1)/lamda) ~ host_total_listings_count +
    latitude + longitude + accommodates + availability_365 +
    number_of_reviews + review_scores_cleanliness + review_scores_location +
    calculated_host_listings_count_private_rooms + banios)

Residuals:
    Min      1Q  Median      3Q     Max
-4.1836 -0.9444 -0.0743  0.9033  5.3290

Coefficients:
                                              Estimate Std. Error t value
(Intercept)                                 -2.256e+03  1.235e+02 -18.275
host_total_listings_count                    9.685e-02  7.019e-03  13.799
latitude                                     8.494e+00  6.666e-01  12.742
longitude                                   -2.101e+01  1.251e+00 -16.797
accommodates                                 4.489e-01  1.800e-02  24.940
availability_365                             9.697e-04  1.595e-04   6.080
number_of_reviews                           -1.630e-03  8.120e-04  -2.007
review_scores_cleanliness                    1.417e+00  1.546e-01   9.162
review_scores_location                       2.598e+00  2.910e-01   8.928
calculated_host_listings_count_private_rooms -1.116e+00  3.418e-02 -32.661
banios                                       7.003e-01  5.460e-02  12.826
                                            Pr(>|t|)
(Intercept)                                  < 2e-16 ***
host_total_listings_count                    < 2e-16 ***
latitude                                     < 2e-16 ***
longitude                                    < 2e-16 ***
accommodates                                 < 2e-16 ***
availability_365                            1.29e-09 ***
number_of_reviews                             0.0448 *
review_scores_cleanliness                    < 2e-16 ***
review_scores_location                       < 2e-16 ***
calculated_host_listings_count_private_rooms  < 2e-16 ***
banios                                       < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.371 on 4735 degrees of freedom
```

```
Multiple R-squared:  0.5389,    Adjusted R-squared:  0.538
F-statistic: 553.5 on 10 and 4735 DF,  p-value: < 2.2e-16
```

```r
modelo_final<- lm((((price^lamda) - 1)/lamda) ~ host_total_listings_count +
    latitude + longitude + accommodates + availability_365 +
    + review_scores_cleanliness + review_scores_location +
    calculated_host_listings_count_private_rooms + banios)
summary(modelo_final)
```

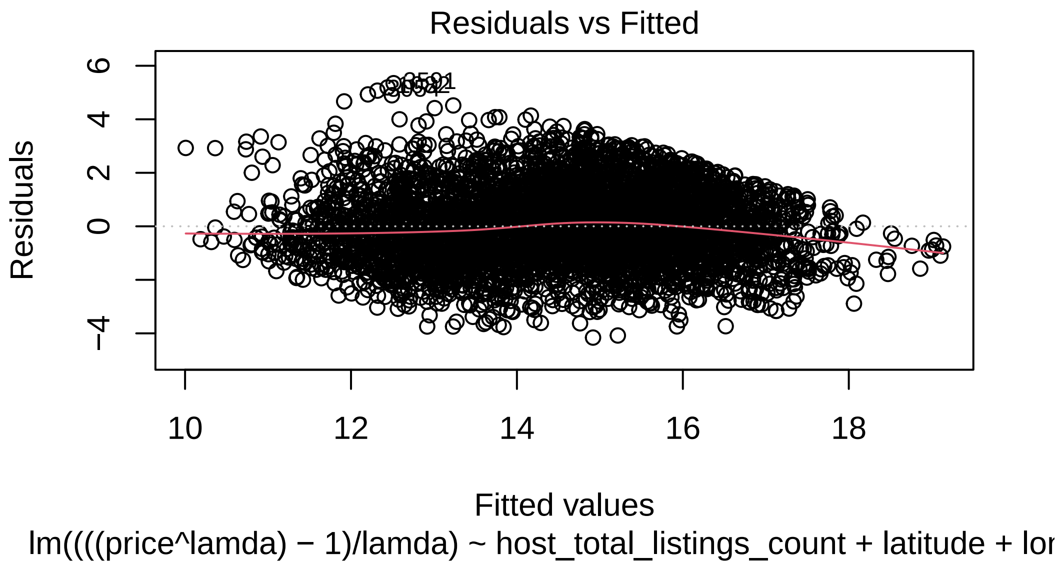
```
Call:
lm(formula = (((price^lamda) - 1)/lamda) ~ host_total_listings_count +
    latitude + longitude + accommodates + availability_365 +
    +review_scores_cleanliness + review_scores_location + calculated_host_listings_count_priv
    banios)

Residuals:
    Min      1Q  Median      3Q     Max
-4.1586 -0.9596 -0.0736  0.9105  5.3492

Coefficients:
                                               Estimate Std. Error t value
(Intercept)                                  -2.258e+03  1.235e+02 -18.287
host_total_listings_count                     9.588e-02  7.004e-03  13.689
latitude                                      8.513e+00  6.668e-01  12.767
longitude                                    -2.102e+01  1.252e+00 -16.799
accommodates                                  4.465e-01  1.796e-02  24.853
availability_365                              9.943e-04  1.591e-04   6.251
review_scores_cleanliness                     1.467e+00  1.526e-01   9.611
review_scores_location                        2.709e+00  2.858e-01   9.479
calculated_host_listings_count_private_rooms -1.114e+00  3.417e-02 -32.600
banios                                        7.028e-01  5.460e-02  12.872
                                             Pr(>|t|)
(Intercept)                                   < 2e-16 ***
host_total_listings_count                     < 2e-16 ***
latitude                                      < 2e-16 ***
longitude                                     < 2e-16 ***
accommodates                                  < 2e-16 ***
availability_365                             4.43e-10 ***
review_scores_cleanliness                     < 2e-16 ***
review_scores_location                        < 2e-16 ***
```

```
calculated_host_listings_count_private_rooms  < 2e-16 ***
banios                                        < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.372 on 4736 degrees of freedom
Multiple R-squared:  0.5385,    Adjusted R-squared:  0.5377
F-statistic: 614.1 on 9 and 4736 DF,  p-value: < 2.2e-16
```

Finalmente, procederemos a la validación de nuestro modelo y, a continuación, realizaremos una interpretación detallada del mismo:

Linealidad:

```
#Hacemos la prueba
plot(modelo_final,1)
```



```
cor.test((((price^lamda) - 1)/lamda), host_total_listings_count)
```

```

	Pearson's product-moment correlation

data:  (((price^lamda) - 1)/lamda) and host_total_listings_count
t = 13.486, df = 4744, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1646058 0.2194092
sample estimates:
      cor 
0.1921573 
```

```r
cor.test((((price^lamda) - 1)/lamda), latitude)
```

```

	Pearson's product-moment correlation

data:  (((price^lamda) - 1)/lamda) and latitude
t = 14.793, df = 4744, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1826212 0.2370170
sample estimates:
      cor 
0.2099816 
```

```r
cor.test((((price^lamda) - 1)/lamda), longitude)
```

```

	Pearson's product-moment correlation

data:  (((price^lamda) - 1)/lamda) and longitude
t = -16.794, df = 4744, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.2635561 -0.2098437
sample estimates:
       cor 
-0.2368809 
```

```r
cor.test((((price^lamda) - 1)/lamda), accommodates)
```

```

    Pearson's product-moment correlation

data:  (((price^lamda) - 1)/lamda) and accommodates
t = 45.585, df = 4744, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5318038 0.5713839
sample estimates:
      cor 
0.5519046 
```

```r
cor.test((((price^lamda) - 1)/lamda), availability_365)
```

```

    Pearson's product-moment correlation

data:  (((price^lamda) - 1)/lamda) and availability_365
t = 5.4775, df = 4744, p-value = 4.538e-08
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.05093873 0.10748430
sample estimates:
       cor 
0.07927529 
```

```r
cor.test((((price^lamda) - 1)/lamda), review_scores_cleanliness)
```

```

    Pearson's product-moment correlation

data:  (((price^lamda) - 1)/lamda) and review_scores_cleanliness
t = 5.7414, df = 4744, p-value = 9.972e-09
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.05474806 0.11125860
sample estimates:
       cor 
0.08307011 
```

```r
cor.test((((price^lamda) - 1)/lamda), review_scores_location)
```

```

    Pearson's product-moment correlation

data:  (((price^lamda) - 1)/lamda) and review_scores_location
t = 4.9214, df = 4744, p-value = 8.883e-07
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.04290655 0.09952063
sample estimates:
       cor 
0.07127099 
```

```r
cor.test((((price^lamda) - 1)/lamda), calculated_host_listings_count_private_rooms)
```

```

    Pearson's product-moment correlation

data:  (((price^lamda) - 1)/lamda) and calculated_host_listings_count_private_rooms
t = -45.505, df = 4744, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.5707371 -0.5311151
sample estimates:
       cor 
-0.5512368 
```

```r
cor.test((((price^lamda) - 1)/lamda), banios)
```

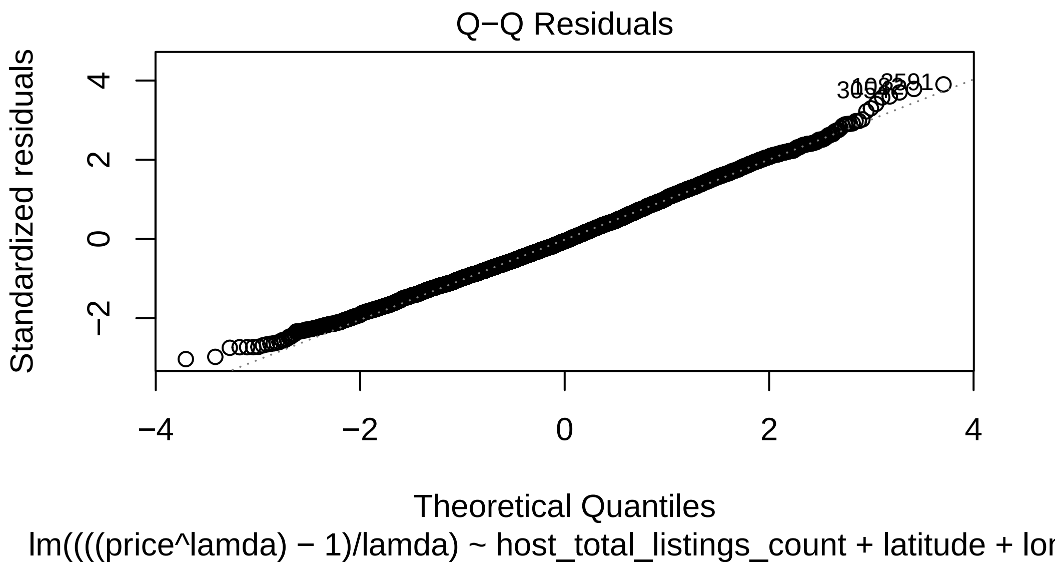
```

    Pearson's product-moment correlation

data:  (((price^lamda) - 1)/lamda) and banios
t = 27.648, df = 4744, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
```

```
95 percent confidence interval:
 0.3477585 0.3967702
sample estimates:
      cor
0.3725241
```

Al analizar la correlación entre los precios y variables independientes, observamos que los p-value son menores a la significancia de 0.05, por lo que concluimos que hay linealidad en nuestro modelo.

Normalidad

```r
plot(modelo_final, 2)
```



```r
shapiro.test(modelo_final$residuals)
```

```

	Shapiro-Wilk normality test

data:  modelo_final$residuals
```

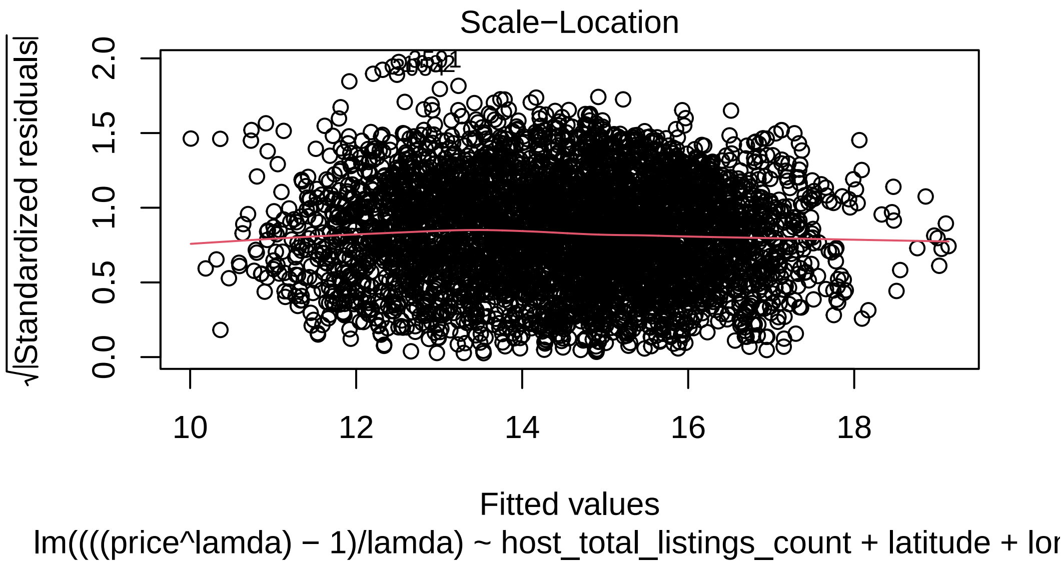
```
W = 0.9971, p-value = 6.541e-08
```

Para validar este supuesto, se realizó una prueba Shapiro-Wilk, notando que nuestro p-value < 0.05, por lo que se rechaza la hipótesis nula: ***“H_0 Los residuos siguen una distribución normal”,*** sin embargo, tras realizar un gráfico QQ podemos decir con seguridad que nuestro modelo si cumple con el supuesto de normalidad ya que nuestros datos se ajustan perfectamente a la gráfica.

Nota: Esta discrepancia suele darse cuando los datos se enfrentan a un reescalamiento severo, lo que hace que las pruebas de bondad de ajuste den falsos negativos o positivos.

Homocedasticidad

```
plot(modelo_final, 3)
```



Para validar este supuesto se realizó un gráfico de dispersión, observando que los puntos no siguen algún tipo de patrón, además de que la línea roja va de manera horizontal, por lo que nuestro modelo cumple con la Homocedasticidad.

Multicolinealidad

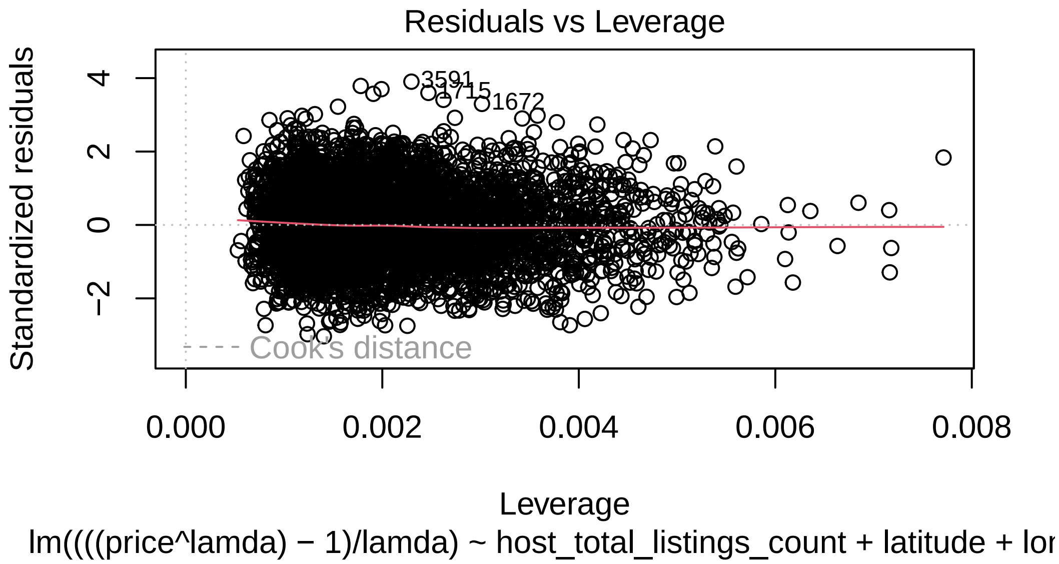
```r
vif(modelo_final)
```

```
                   host_total_listings_count
                                    1.026718
                                    latitude
                                    1.022833
                                   longitude
                                    1.041577
                                accommodates
                                    1.573110
                            availability_365
                                    1.014967
                   review_scores_cleanliness
                                    1.074754
                      review_scores_location
                                    1.098347
calculated_host_listings_count_private_rooms
                                    1.291497
                                      banios
                                    1.296549
```

Este supuesto fue validado prácticamente al inicio cuando las variables más fuertemente correlacionadas fueron eliminadas, corroborado por la función vif que nos ayuda a determinar si un modelo de regresión lineal presenta colinealidad o no, y, dado que los datos obtenidos para nuestras variables son menores a 5, podemos concluir que nuestro modelo cumple también con este supuesto.

Valores influyentes

```r
plot(modelo_final, 5)
```



En un caso similar al anterior, al haber eliminado los valores atípicos del modelo y al haberlo acotado solamente a ciertos Airbnb's, nos podemos percatar de que este supuesto también se cumple.

Interpretación

```
summary(modelo_final)
```

```
Call:
lm(formula = (((price^lamda) - 1)/lamda) ~ host_total_listings_count +
    latitude + longitude + accommodates + availability_365 +
    +review_scores_cleanliness + review_scores_location + calculated_host_listings_count_pri
    banios)

Residuals:
    Min      1Q  Median      3Q     Max
-4.1586 -0.9596 -0.0736  0.9105  5.3492

Coefficients:
```

```
                                               Estimate Std. Error t value
(Intercept)                                  -2.258e+03  1.235e+02 -18.287
host_total_listings_count                     9.588e-02  7.004e-03  13.689
latitude                                      8.513e+00  6.668e-01  12.767
longitude                                    -2.102e+01  1.252e+00 -16.799
accommodates                                  4.465e-01  1.796e-02  24.853
availability_365                              9.943e-04  1.591e-04   6.251
review_scores_cleanliness                     1.467e+00  1.526e-01   9.611
review_scores_location                        2.709e+00  2.858e-01   9.479
calculated_host_listings_count_private_rooms -1.114e+00  3.417e-02 -32.600
banios                                        7.028e-01  5.460e-02  12.872
                                             Pr(>|t|)    
(Intercept)                                   < 2e-16 ***
host_total_listings_count                     < 2e-16 ***
latitude                                      < 2e-16 ***
longitude                                     < 2e-16 ***
accommodates                                  < 2e-16 ***
availability_365                             4.43e-10 ***
review_scores_cleanliness                     < 2e-16 ***
review_scores_location                        < 2e-16 ***
calculated_host_listings_count_private_rooms  < 2e-16 ***
banios                                        < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.372 on 4736 degrees of freedom
Multiple R-squared:  0.5385,	Adjusted R-squared:  0.5377 
F-statistic: 614.1 on 9 and 4736 DF,  p-value: < 2.2e-16
```

Tras haber realizado la validación de supuestos de nuestro modelo obtuvimos la siguiente ecuación:

$$y = \frac{price^{\lambda} - 1}{\lambda} = -225.8 + 0.09588 * (host\ total\ listings\ count)$$

$$+8.513*(latitude) - 21.02*(longitude) + 0.4465*(accommodates) + 0.0009943*(availability\ 365)$$

$$+1.467 * (review\ scores\ cleanliness) + 2.709 * (review\ scores\ location)$$

Viendo que nuestra R^2 ajustada es del 0.5377, implica que nuestro modelo nos explica en un 53.77% nuestro problema de predecir el precio de un Airbnb, y además es muy significante, dado que nuestro p-value es menor que un nivel de significancia 0.05, así como también lo es en cada una de nuestras variables explicativas.

Finalmente, podemos concluir lo siguiente de nuestro modelo:

- Decimos que con cada 0.09588 de Airbnb's que posea el anfitrión el precio aumentará.
- Decimos por cada 8.513 que la latitud se mueva en la CDMX, el precio aumentará.
- Dado que la longitud en la CDMX es negativa, decimos que precio aumenta con 21.02 por cada vez la longitud se mueva.
- Decimos que aproximadamente por cada media persona que se hospede en el AIRBNB, el precio igual aumentará, por lo que una persona completa casi duplica el precio.
- Vemos que mientras casi no haya disponibilidad el precio aumentará en 0.0009943
- Vemos que por cada punto que se le da a la limpieza de un Airbnb, nuestro precio aumentará con una razón de 1.467
- Análogamente al punto anterior, mientras mayor sea la calificación que tenga por ubicación el Airbnb, mayor será el precio con una razón de 2.709