# Airbnb_Project

Idaly Ferrales

2023-07-08

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com (http://rmarkdown.rstudio.com).

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
if (!require(tidyverse)) install.packages("tidyverse")
```

```
## Loading required package: tidyverse
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──
```

```
## ✔ ggplot2 3.3.5     ✔ purrr   1.0.1
## ✔ tibble  3.2.1     ✔ dplyr   1.1.2
## ✔ tidyr   1.3.0     ✔ stringr 1.5.0
## ✔ readr   2.1.2     ✔ forcats 0.5.1
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
if (!require(ISLR)) install.packages("ISLR")
```

```
## Loading required package: ISLR
```

```
if (!require(corrplot)) install.packages("corrplot")
```

```
## Loading required package: corrplot
```

```
## corrplot 0.92 loaded
```

```
if (!require(corrgram)) install.packages("corrgram")
```

```
## Loading required package: corrgram
```

```
if (!require(car)) install.packages("car")
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:purrr':
##
##     some
```

```r
library(data.table)
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
## The following object is masked from 'package:purrr':
##
##     transpose
```

```r
library(tidyverse)
library(ISLR)
library(corrplot)
library(corrgram)
library(car)
library(dplyr)
library(ggplot2)
getwd()
```

```
## [1] "/Users/idalyferrales/OMSA folder"
```

```r
setwd("/Users/idalyferrales/OMSA folder")
getwd()
```

```
## [1] "/Users/idalyferrales/OMSA folder"
```

```r
#Import the Airbnb dataset
list.files(path=".", pattern=".csv", all.files=TRUE,
           full.names=TRUE)
```

```
##  [1] "./Airbnb_clean_dataset.csv"
##  [2] "./Airbnb_Data_USA.csv"
##  [3] "./Airbnb_with_factors.csv"
##  [4] "./airbnb_with_fips.csv"
##  [5] "./Berkshire.csv"
##  [6] "./binary_1.csv"
##  [7] "./binary_midterm.csv.numbers"
##  [8] "./binary.csv"
##  [9] "./contrafund_.csv"
## [10] "./contrafund_Exam.csv"
## [11] "./contrafund_final.csv"
## [12] "./contrafund.csv"
## [13] "./Factor_HiTech_Midterm.csv"
## [14] "./factors.csv"
## [15] "./final_data.csv"
## [16] "./github.gatech.edu_raw_MGT-6203-Summer-2023-Canvas_Team-71_main_Brainstorm_walkability_impact_walkabili
ty_index.csv_token=GHSAT0AAAAAAAACWL4HSO7TEOBS7D4WNS5IZFJ3MCQ.txt"
## [17] "./Grades_Data.csv"
## [18] "./sample_airbnb_data_filtered_usa.csv"
## [19] "./UPS_KO_Exam.csv"
```

```r
airbnb_data <- read.csv("./Airbnb_Data_USA.csv")

fips_file <- read.csv("./final_data.csv") #There is a new file with FIPS value
```

# Including Plots

You can also embed plots, for example:

```r
airbnb_data_filtered <- filter(airbnb_data, availability_365 != "0")

airbnb_data_filtered <- airbnb_data_filtered %>% select('price', 'availability_365','id', 'latitude','longitud
e','minimum_nights','number_of_reviews_ltm','room_type','city','calculated_host_listings_count')
colnames(airbnb_data_filtered)
```

```
##  [1] "price"                          "availability_365"              
##  [3] "id"                             "latitude"                      
##  [5] "longitude"                      "minimum_nights"                
##  [7] "number_of_reviews_ltm"          "room_type"                     
##  [9] "city"                           "calculated_host_listings_count"
```

```r
str(airbnb_data_filtered)
```

```
## 'data.frame':    195200 obs. of  10 variables:
##  $ price                         : int  202 235 56 575 110 95 259 100 79 65 ...
##  $ availability_365              : int  128 365 365 365 159 365 32 331 230 365 ...
##  $ id                            : num  958 5858 8142 8339 8739 ...
##  $ latitude                      : num  37.8 37.7 37.8 37.8 37.8 ...
##  $ longitude                     : num  -122 -122 -122 -122 -122 ...
##  $ minimum_nights                : int  2 30 32 9 1 1 5 30 30 32 ...
##  $ number_of_reviews_ltm         : int  59 0 1 0 34 1 13 1 0 1 ...
##  $ room_type                     : chr  "Entire home/apt" "Entire home/apt" "Private room" "Entire home/apt"
...
##  $ city                          : chr  "San Francisco" "San Francisco" "San Francisco" "San Francisco" ...
##  $ calculated_host_listings_count: int  1 1 13 2 2 2 1 1 1 13 ...
```

```r
#Sample using 25%
df1 <- data.frame(airbnb_data_filtered)
#Checking for NaN
df1<- na.omit(df1)

fips <- data.frame(fips_file)
```

```r
#unique(df1$city)
df_merged <- merge(df1, fips, by = 'id')
head(df_merged)
```

```
##     id price.x availability_365 latitude.x longitude.x minimum_nights
## 1 2441      91               24   45.00862   -93.23424             28
## 2 2732     179              362   34.00440  -118.48095              7
## 3 3943      85              289   38.91195   -77.00456              1
## 4 5121      60              365   40.68535   -73.95512             30
## 5 5739     125               82   37.81352  -122.26055              4
## 6 6165      80              365   38.95331   -77.03624             31
##   number_of_reviews_ltm     room_type.x             city
## 1                     5 Entire home/apt Twin Cities MSA
## 2                     2    Private room     Los Angeles
## 3                    17    Private room Washington D.C.
## 4                     0    Private room   New York City
## 5                    23 Entire home/apt         Oakland
## 6                     1    Private room Washington D.C.
##   calculated_host_listings_count     X latitude.y longitude.y         fips
## 1                              2 10008   45.00862   -93.23424 2.705310e+14
## 2                              2 12191   34.00440  -118.48095 6.037702e+13
## 3                              5 26344   38.91195   -77.00456 1.100101e+14
## 4                              2 19658   40.68535   -73.95512 3.604702e+14
## 5                              1 33525   37.81352  -122.26055 6.001404e+13
## 6                              2 15734   38.95331   -77.03624 1.100100e+14
##   state_fips county_fips tract_fips block_fips
## 1         27          53     101200          4
## 2          6          37     702002          2
## 3         11           1       8701          1
## 4         36          47      22900          2
## 5          6           1     403502          1
## 6         11           1       2001          2
##                                                                                          url
## 1  https://geo.fcc.gov/api/census/area?lat=45.00862&lon=-93.23424&censusYear=2020&format=json
## 2  https://geo.fcc.gov/api/census/area?lat=34.0044&lon=-118.48095&censusYear=2020&format=json
## 3  https://geo.fcc.gov/api/census/area?lat=38.91195&lon=-77.00456&censusYear=2020&format=json
## 4  https://geo.fcc.gov/api/census/area?lat=40.68535&lon=-73.95512&censusYear=2020&format=json
## 5 https://geo.fcc.gov/api/census/area?lat=37.81352&lon=-122.26055&censusYear=2020&format=json
## 6  https://geo.fcc.gov/api/census/area?lat=38.95331&lon=-77.03624&censusYear=2020&format=json
##       room_type.y price.y                                         csa_name
## 1 Entire home/apt      91                     Minneapolis-St. Paul, MN-WI
## 2    Private room     179                      Los Angeles-Long Beach, CA
## 3    Private room      85 Washington-Baltimore-Arlington, DC-MD-VA-WV-PA
## 4    Private room      60                    New York-Newark, NY-NJ-CT-PA
## 5 Entire home/apt     125             San Jose-San Francisco-Oakland, CA
## 6    Private room      80 Washington-Baltimore-Arlington, DC-MD-VA-WV-PA
##   walk_index
## 1   14.33333
## 2   19.00000
## 3   15.83333
## 4   13.16667
## 5   13.16667
## 6   13.00000
```

```
unique(df_merged$city)
```

```
##  [1] "Twin Cities MSA"    "Los Angeles"        "Washington D.C."
##  [4] "New York City"      "Oakland"            "Nashville"
##  [7] "San Francisco"      "Cambridge"          "Boston"
## [10] "Seattle"            "Santa Clara County" "Chicago"
## [13] "Denver"             "Jersey City"        "Clark County"
## [16] "Rhode Island"       "San Diego"          "Portland"
## [19] "Santa Cruz County"  "New Orleans"        "Austin"
## [22] "San Mateo County"   "Asheville"          "Broward County"
## [25] "Pacific Grove"      "Columbus"           "Salem"
```

```
df_merged2 <- filter(df_merged, city != c("Cambridge", "Santa Cruz County", "San Mateo County"))
```

```
## Warning: There was 1 warning in `filter()`.
## ℹ In argument: `city != c("Cambridge", "Santa Cruz County", "San Mateo
##   County")`.
## Caused by warning in `city != c("Cambridge", "Santa Cruz County", "San Mateo County")`:
## ! longer object length is not a multiple of shorter object length
```

```r
head(df_merged2)
```

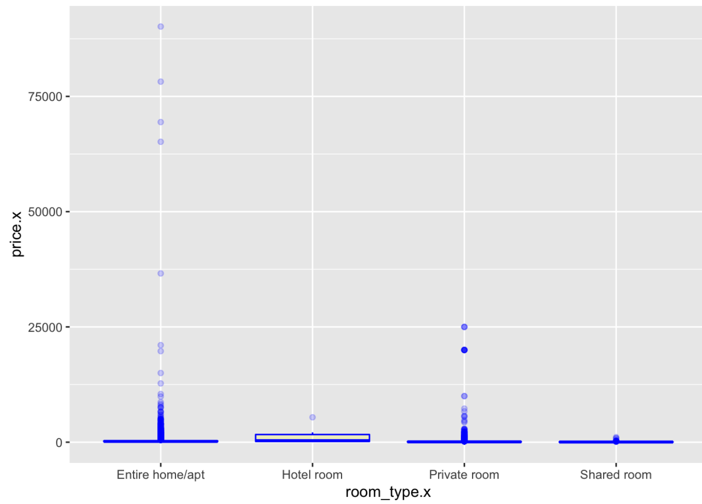
```
##     id price.x availability_365 latitude.x longitude.x minimum_nights
## 1 2441      91               24   45.00862   -93.23424             28
## 2 2732     179              362   34.00440  -118.48095              7
## 3 3943      85              289   38.91195   -77.00456              1
## 4 5121      60              365   40.68535   -73.95512             30
## 5 5739     125               82   37.81352  -122.26055              4
## 6 6165      80              365   38.95331   -77.03624             31
##   number_of_reviews_ltm     room_type.x            city
## 1                     5 Entire home/apt Twin Cities MSA
## 2                     2    Private room     Los Angeles
## 3                    17    Private room Washington D.C.
## 4                     0    Private room   New York City
## 5                    23 Entire home/apt         Oakland
## 6                     1    Private room Washington D.C.
##   calculated_host_listings_count     X latitude.y longitude.y         fips
## 1                              2 10008   45.00862   -93.23424 2.705310e+14
## 2                              2 12191   34.00440  -118.48095 6.037702e+13
## 3                              5 26344   38.91195   -77.00456 1.100101e+14
## 4                              2 19658   40.68535   -73.95512 3.604702e+14
## 5                              1 33525   37.81352  -122.26055 6.001404e+13
## 6                              2 15734   38.95331   -77.03624 1.100100e+14
##   state_fips county_fips tract_fips block_fips
## 1         27          53     101200          4
## 2          6          37     702002          2
## 3         11           1       8701          1
## 4         36          47      22900          2
## 5          6           1     403502          1
## 6         11           1       2001          2
##                                                                                     url
## 1  https://geo.fcc.gov/api/census/area?lat=45.00862&lon=-93.23424&censusYear=2020&format=json
## 2  https://geo.fcc.gov/api/census/area?lat=34.0044&lon=-118.48095&censusYear=2020&format=json
## 3  https://geo.fcc.gov/api/census/area?lat=38.91195&lon=-77.00456&censusYear=2020&format=json
## 4  https://geo.fcc.gov/api/census/area?lat=40.68535&lon=-73.95512&censusYear=2020&format=json
## 5 https://geo.fcc.gov/api/census/area?lat=37.81352&lon=-122.26055&censusYear=2020&format=json
## 6  https://geo.fcc.gov/api/census/area?lat=38.95331&lon=-77.03624&censusYear=2020&format=json
##       room_type.y price.y                                     csa_name
## 1 Entire home/apt      91                     Minneapolis-St. Paul, MN-WI
## 2    Private room     179                      Los Angeles-Long Beach, CA
## 3    Private room      85 Washington-Baltimore-Arlington, DC-MD-VA-WV-PA
## 4    Private room      60                    New York-Newark, NY-NJ-CT-PA
## 5 Entire home/apt     125             San Jose-San Francisco-Oakland, CA
## 6    Private room      80 Washington-Baltimore-Arlington, DC-MD-VA-WV-PA
##   walk_index
## 1   14.33333
## 2   19.00000
## 3   15.83333
## 4   13.16667
## 5   13.16667
## 6   13.00000
```

```r
#Renaming citi values
df_merged2[df_merged2 == "Pacific Grove"] <- "Monterrey"
df_merged2[df_merged2 == "Clark County"] <- "Las Vegas"
df_merged2[df_merged2 == "Broward County"] <- "Fort Lauderdale"
```

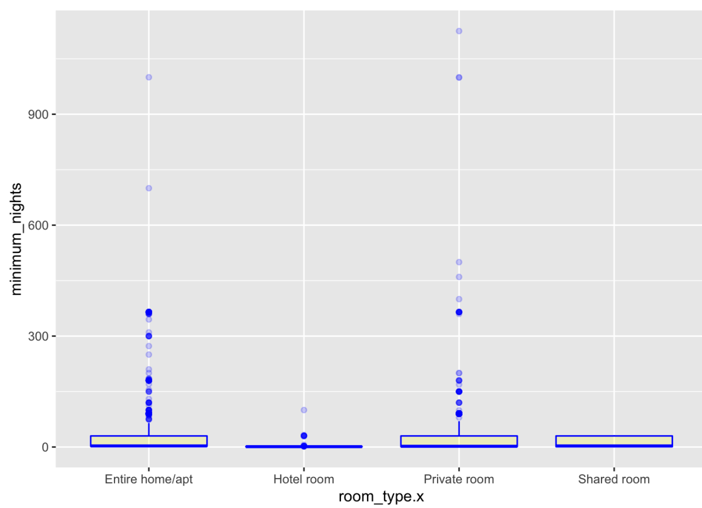
```r
#Outliers found in Price per room type
boxplot_room_type <- ggplot(df_merged2, aes(x=room_type.x, y=price.x)) +
  geom_boxplot(color="blue", fill = "yellow", alpha = 0.2, title= "Price")
```

```
## Warning: Ignoring unknown parameters: title
```

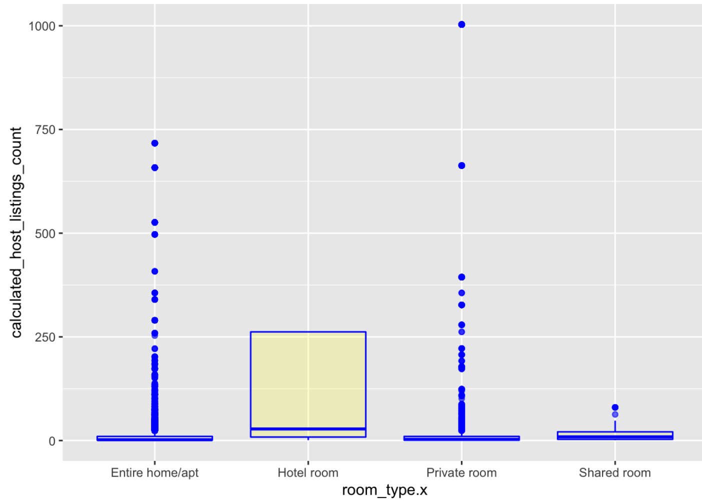
```r
boxplot_room_type
```




```
#Influential points in minimum nights
boxplot_2 <- ggplot(df_merged2, aes(x=room_type.x, y=minimum_nights)) +
  geom_boxplot(color="blue", fill = "yellow", alpha = 0.2)
boxplot_2
```




```
#Influential points in Calculated Host listings
boxplot_3 <- ggplot(df_merged2, aes(x=room_type.x, y=calculated_host_listings_count)) +
  geom_boxplot(color="blue", fill = "yellow", alpha = 0.2)
boxplot_3
```




```
#27 unique cities
unique_cities <- unique(df_merged2$city)
unique_cities
```

```
##  [1] "Twin Cities MSA"    "Los Angeles"        "Washington D.C."
##  [4] "New York City"      "Oakland"            "Nashville"
##  [7] "San Francisco"      "Cambridge"          "Boston"
## [10] "Seattle"            "Santa Clara County" "Chicago"
## [13] "Denver"             "Jersey City"        "Las Vegas"
## [16] "Rhode Island"       "San Diego"          "Portland"
## [19] "Santa Cruz County"  "New Orleans"        "Austin"
## [22] "San Mateo County"   "Asheville"          "Fort Lauderdale"
## [25] "Monterrey"          "Columbus"           "Salem"
```

```
#max
max(df_merged2$price.x)
```

```
## [1] 90150
```

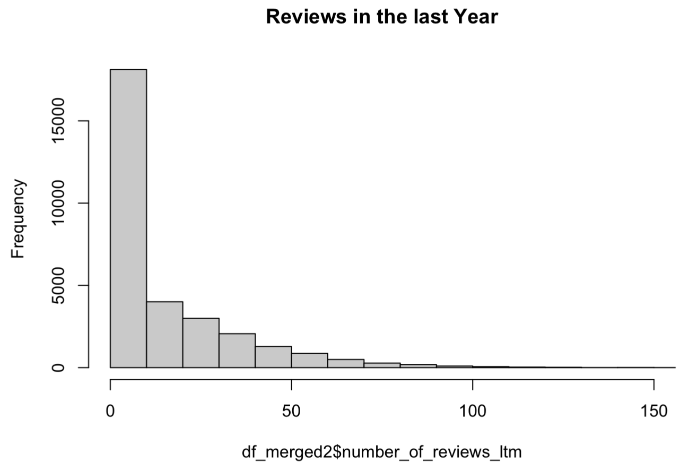
```
#Minimum
min(df_merged2$price.x)
```

```
## [1] 10
```

```
#Visualize quantiles in our data
quantile(df_merged2$price.x, probs = seq(0, 1, 1/20))
```

```
##      0%      5%     10%     15%     20%     25%     30%     35%     40%     45%
##    10.0    46.0    60.0    72.0    82.0    92.0   100.0   112.0   124.0   135.0
##     50%     55%     60%     65%     70%     75%     80%     85%     90%     95%
##   149.0   160.0   178.0   199.0   220.0   250.0   289.0   340.0   434.4   656.0
##    100%
## 90150.0
```

```
#Histogram of Reviews in the last year
hist(df_merged2$number_of_reviews_ltm, breaks = 100, xlim = c(0,150), main = 'Reviews in the last Year')
```



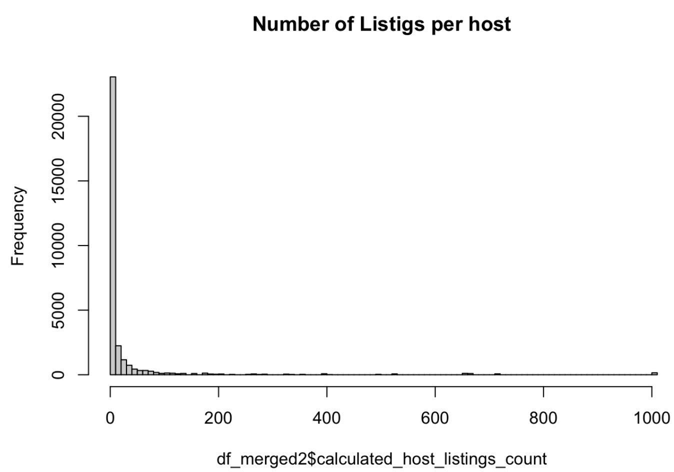

```
#max
max(df_merged2$number_of_reviews_ltm)
```

```
## [1] 1314
```

```
#Minimum
min(df_merged2$number_of_reviews_ltm)
```

```
## [1] 0
```

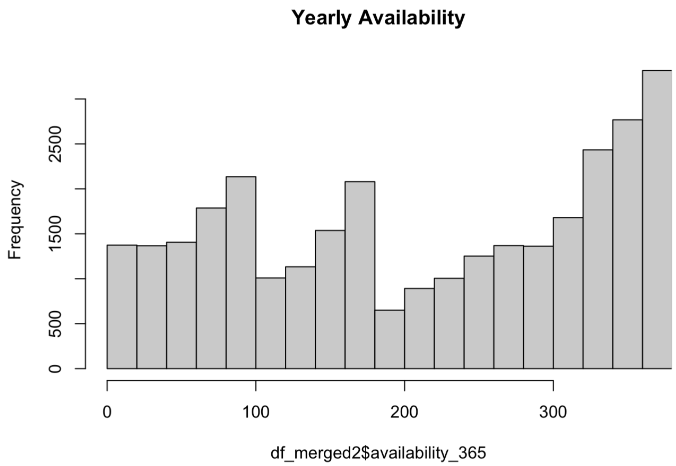
```
#Histogram of Number of Listings per host
hist(df_merged2$calculated_host_listings_count, breaks = 100, xlim = c(0,1003), main = 'Number of Listigs per hos
t')
```




```
#Summary to see min, max of numeric variables
df_summary <- df_merged2[ , c('price.x', 'availability_365', 'minimum_nights', 'number_of_reviews_ltm', 'calculat
ed_host_listings_count')]
summary(df_summary)
```

```
##     price.x        availability_365 minimum_nights    number_of_reviews_ltm
##  Min.   :   10.0   Min.   :  1.0    Min.   :   1.00   Min.   :   0.0
##  1st Qu.:   92.0   1st Qu.: 92.0    1st Qu.:   2.00   1st Qu.:   1.0
##  Median :  149.0   Median :218.0    Median :   3.00   Median :   6.0
##  Mean   :  256.2   Mean   :210.3    Mean   :  12.04   Mean   :  15.1
##  3rd Qu.:  250.0   3rd Qu.:329.0    3rd Qu.:  30.00   3rd Qu.:  23.0
##  Max.   :90150.0   Max.   :365.0    Max.   :1125.00   Max.   :1314.0
##  calculated_host_listings_count
##  Min.   :   1.00
##  1st Qu.:   1.00
##  Median :   2.00
##  Mean   :  28.85
##  3rd Qu.:  10.00
##  Max.   :1003.00
```

```
hist(df_merged2$availability_365, breaks = 20, xlim = c(0,365), main = "Yearly Availability")
```




```
#Histogram of Price
hist(df_merged2$price.x, breaks = 1000, xlim = c(0,600))
```

```r
# Drop columns that will not be part of linear regression
drops <- c('latitude.x', 'latitude.y','longitude.x','longitude.y', 'fips', 'state_fips','county_fips','tract_fip
s', 'url','room_type.y', 'price.y', 'csa_name', 'X', 'block_fips')

df_merged3 <- df_merged2[,!(names(df_merged2)%in% drops)]
```

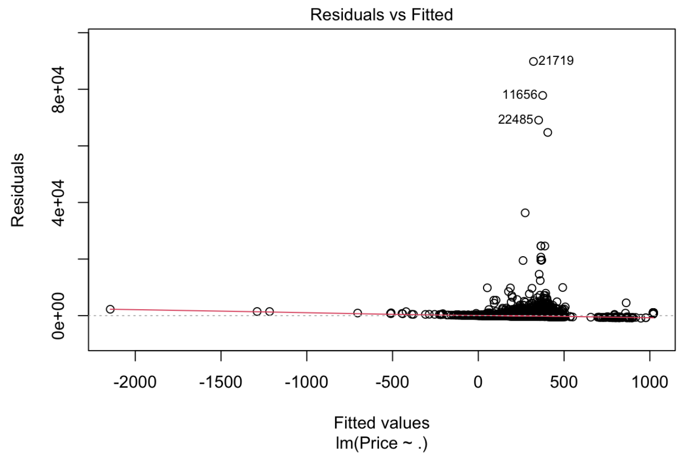
```r
# Renaming some of the columns
#The new dataset name is df_merged3
df_merged3 <- df_merged3 %>%
  rename("Price" = "price.x", "Room_type" = "room_type.x")
```

```r
# To drop the id column
drop_a <- c('id')
#The new name of our dataframe is df_merged_subset with 30,557 rows and 8 variables: Price, availability, number
of nights, number of reviews, room type, city, calculated host listings, and walkability score
df_merged_subset <- df_merged3[,!(names(df_merged3)%in% drop_a)]
```

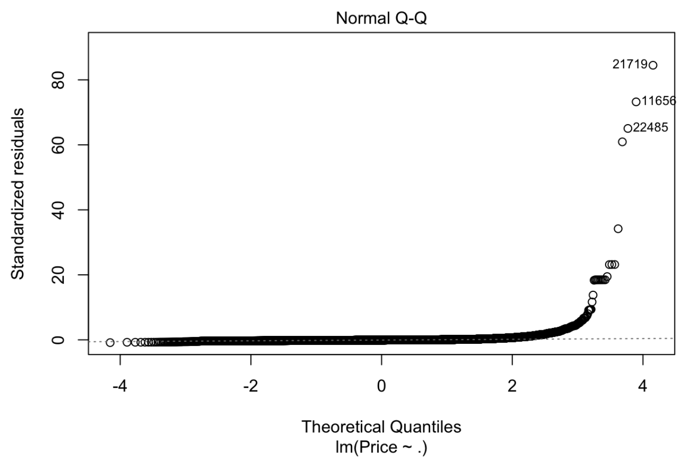
```r
#Simple regression utilizing all variables
simple_regression = lm(Price ~ ., data= df_merged_subset)
summary(simple_regression)
```

```
##
## Call:
## lm(formula = Price ~ ., data = df_merged_subset)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##   -907   -153    -75     12  89829
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     293.14075   56.46203   5.192 2.10e-07 ***
## availability_365                  0.03848    0.05262   0.731  0.46466
## minimum_nights                   -0.58332    0.25170  -2.318  0.02048 *
## number_of_reviews_ltm            -1.91258    0.26858  -7.121 1.09e-12 ***
## Room_typeHotel room             541.64521   87.39553   6.198 5.81e-10 ***
## Room_typePrivate room          -127.00029   14.58656  -8.707  < 2e-16 ***
## Room_typeShared room           -235.51326   66.51840  -3.541  0.00040 ***
## cityAustin                      226.35092   57.96931   3.905 9.46e-05 ***
## cityBoston                       42.41639   69.82700   0.607  0.54356
## cityCambridge                    61.82447  108.01336   0.572  0.56707
## cityChicago                      18.20587   56.96771   0.320  0.74929
## cityColumbus                    -20.66808   74.68353  -0.277  0.78198
## cityDenver                       15.61530   69.12069   0.226  0.82127
## cityFort Lauderdale             130.37514   53.90820   2.418  0.01559 *
## cityJersey City                  -0.49115   98.21477  -0.005  0.99601
## cityLas Vegas                   230.74942   53.45605   4.317 1.59e-05 ***
## cityLos Angeles                 140.62492   51.26430   2.743  0.00609 **
## cityMonterrey                   132.65967  203.71907   0.651  0.51493
## cityNashville                   162.98045   57.79392   2.820  0.00481 **
## cityNew Orleans                 120.24762   57.65054   2.086  0.03700 *
## cityNew York City                57.19042   51.36242   1.113  0.26552
## cityOakland                     165.08014   70.96428   2.326  0.02001 *
## cityPortland                     -9.01391   64.74304  -0.139  0.88927
## cityRhode Island                137.20365   59.93166   2.289  0.02207 *
## citySalem                       -15.87732  175.18845  -0.091  0.92779
## citySan Diego                   122.35951   56.01354   2.184  0.02894 *
## citySan Francisco               105.15516   61.79325   1.702  0.08882 .
## citySan Mateo County             60.66798   81.13658   0.748  0.45463
## citySanta Clara County           17.27484   60.66605   0.285  0.77584
## citySanta Cruz County            98.41239  111.43850   0.883  0.37718
## citySeattle                      13.76771   64.07604   0.215  0.82987
## cityTwin Cities MSA              31.03695   63.98667   0.485  0.62764
## cityWashington D.C.              42.18284   62.52997   0.675  0.49993
## calculated_host_listings_count    0.15847    0.05930   2.672  0.00754 **
## walk_index                       -6.17778    2.06930  -2.985  0.00283 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1063 on 30522 degrees of freedom
## Multiple R-squared:  0.01157,    Adjusted R-squared:  0.01047
## F-statistic: 10.51 on 34 and 30522 DF,  p-value: < 2.2e-16
```

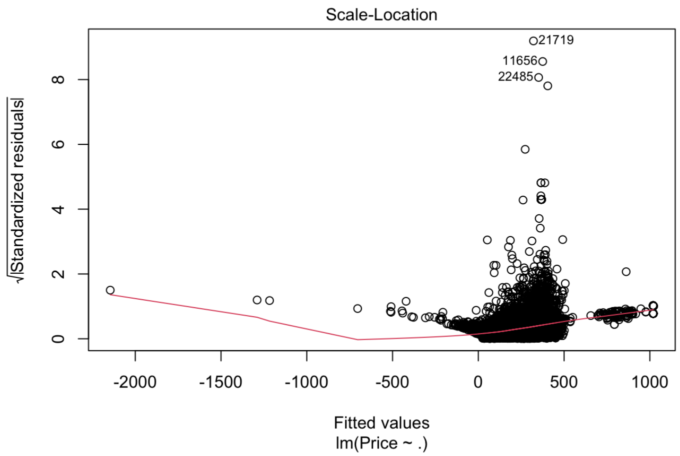
```
plot(simple_regression)
```

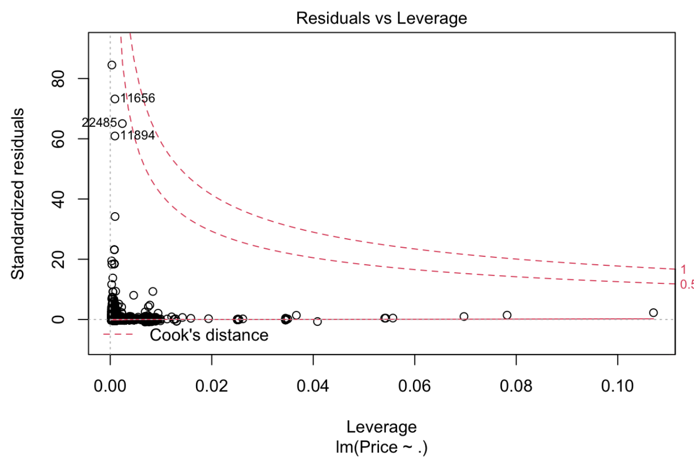


Residuals vs Fitted
21719
11656
22485
Residuals
8e+04
4e+04
0e+00
-2000
-1500
-1000
-500
0
500
1000
Fitted values
lm(Price ~ .)

Normal Q-Q
21719
11656
22485
Standardized residuals
80
60
40
20
0
-4
-2
0
2
4
Theoretical Quantiles
lm(Price ~ .)



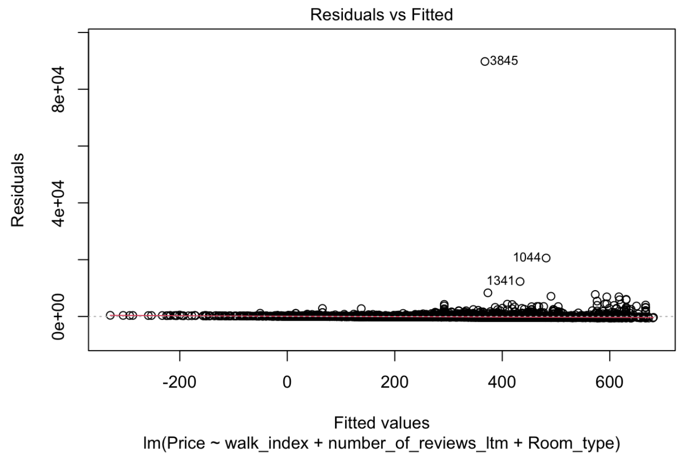



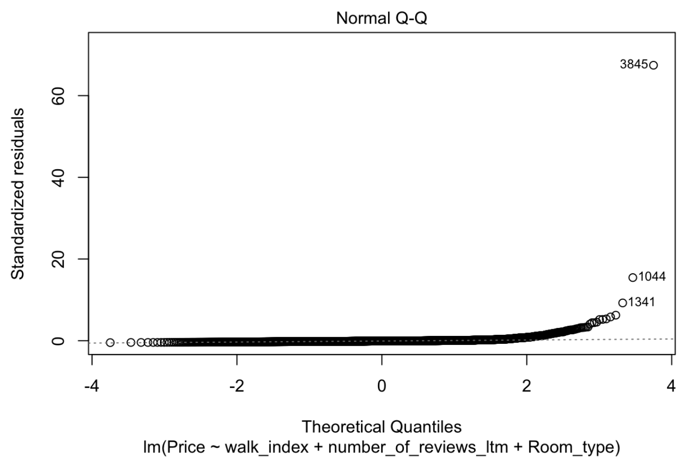

```
#Interacting with subset of cities for possible linear models
# Creating a subset for the city of Los Angeles
df_LA <- filter(df_merged_subset, city == 'Los Angeles')
```

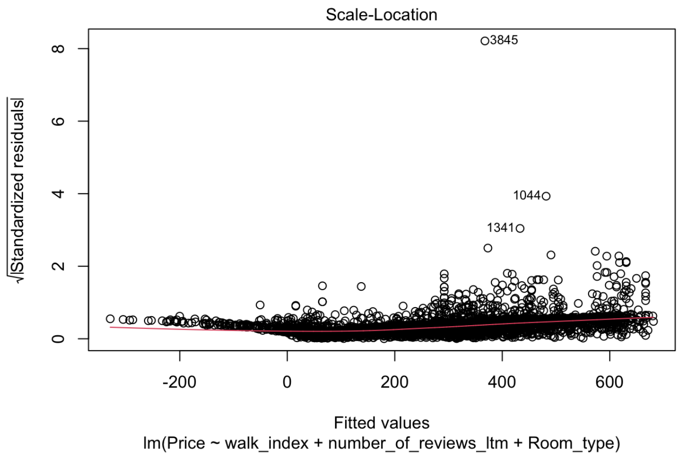
```
regression_la = lm(Price ~ walk_index + number_of_reviews_ltm + Room_type, data= df_LA)
summary(regression_la)
```

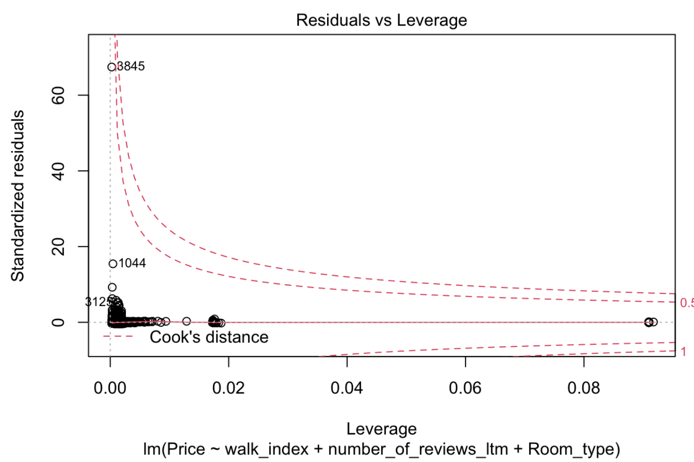
```
##
## Call:
## lm(formula = Price ~ walk_index + number_of_reviews_ltm + Room_type,
##     data = df_LA)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##   -618   -209    -93     15  89782
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            802.0673    84.0556   9.542  < 2e-16 ***
## walk_index             -27.0825     5.5238  -4.903 9.71e-07 ***
## number_of_reviews_ltm   -3.3729     0.9448  -3.570  0.00036 ***
## Room_typeHotel room   -117.4407   402.4490  -0.292  0.77044
## Room_typePrivate room -280.7717    40.0949  -7.003 2.81e-12 ***
## Room_typeShared room  -304.2824   176.2674  -1.726  0.08436 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1331 on 5620 degrees of freedom
## Multiple R-squared:  0.01383,    Adjusted R-squared:  0.01295
## F-statistic: 15.76 on 5 and 5620 DF,  p-value: 1.931e-15
```

```
plot(regression_la)
```


Residuals vs Fitted
Residuals
8e+04
4e+04
0e+00
3845
1044
1341
-200
0
200
400
600
Fitted values
lm(Price ~ walk_index + number_of_reviews_ltm + Room_type)


Normal Q-Q
Standardized residuals
60
40
20
0
3845
1044
1341
-4
-2
0
2
4
Theoretical Quantiles
lm(Price ~ walk_index + number_of_reviews_ltm + Room_type)




```
#Trying with a new dataset df_cleaned that focuses on pricing under 25,000USD
df_cleaned <- subset(df_merged3, Price < 25000)
```

```
#This is a for loop that creates linear regression by adding one variable at a time. Used to see the impact of ad
ding variables to the model,
mod_summaries <- list()
for(i in 3:ncol(df_cleaned)) {                 # Head of for-loop

  predictors_i <- colnames(df_cleaned)[3:i]    # Create vector of predictor names
  mod_summaries[[i - 1]] <- summary(     # Store regression model summary in list
    lm(Price ~ ., df_cleaned[ , c("Price", predictors_i)]))

}
```

```
mod_summaries
```

```
## [[1]]
## NULL
##
## [[2]]
##
## Call:
## lm(formula = Price ~ ., data = df_cleaned[, c("Price", predictors_i)])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -254.4  -149.6   -93.8     8.3 20787.7
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      211.84896    6.17258  34.321  < 2e-16 ***
## availability_365   0.14687    0.02556   5.747 9.18e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 530.1 on 30547 degrees of freedom
## Multiple R-squared:  0.00108,    Adjusted R-squared:  0.001047
## F-statistic: 33.03 on 1 and 30547 DF,  p-value: 9.176e-09
##
##
## [[3]]
##
## Call:
## lm(formula = Price ~ ., data = df_cleaned[, c("Price", predictors_i)])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -253.7  -149.1   -93.6     6.7 20795.7
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      216.07275    6.24283  34.611  < 2e-16 ***
## availability_365   0.15741    0.02566   6.135 8.63e-10 ***
## minimum_nights    -0.53476    0.11982  -4.463 8.12e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 530 on 30546 degrees of freedom
## Multiple R-squared:  0.001731,   Adjusted R-squared:  0.001666
## F-statistic: 26.48 on 2 and 30546 DF,  p-value: 3.228e-12
##
##
## [[4]]
##
## Call:
## lm(formula = Price ~ ., data = df_cleaned[, c("Price", predictors_i)])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -276.0  -150.5   -86.9    11.6 20780.3
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           252.25744    6.81180  37.032  < 2e-16 ***
## availability_365        0.12773    0.02569   4.973 6.64e-07 ***
## minimum_nights         -0.86699    0.12216  -7.097 1.30e-12 ***
## number_of_reviews_ltm  -1.71803    0.13127 -13.088  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 528.5 on 30545 degrees of freedom
## Multiple R-squared:  0.007298,   Adjusted R-squared:  0.0072
## F-statistic: 74.85 on 3 and 30545 DF,  p-value: < 2.2e-16
##
##
## [[5]]
##
## Call:
## lm(formula = Price ~ ., data = df_cleaned[, c("Price", predictors_i)])
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -900.8  -134.9   -81.5     6.1 20741.8
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            280.88683    6.99546  40.153  < 2e-16 ***
## availability_365         0.14728    0.02546   5.784 7.37e-09 ***
## minimum_nights          -0.77806    0.12107  -6.427 1.32e-10 ***
## number_of_reviews_ltm   -1.93053    0.13048 -14.795  < 2e-16 ***
## Room_typeHotel room    596.69951   42.76806  13.952  < 2e-16 ***
## Room_typePrivate room -127.75578    6.98727 -18.284  < 2e-16 ***
## Room_typeShared room  -229.08078   32.68522  -7.009 2.46e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 523.5 on 30542 degrees of freedom
## Multiple R-squared:  0.02601,    Adjusted R-squared:  0.02582
## F-statistic: 135.9 on 6 and 30542 DF,  p-value: < 2.2e-16
##
##
## [[6]]
##
## Call:
## lm(formula = Price ~ ., data = df_cleaned[, c("Price", predictors_i)])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -897.1  -134.4   -65.7    17.0 20708.9
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             192.94043   24.77648   7.787 7.06e-15 ***
## availability_365          0.14473    0.02557   5.661 1.52e-08 ***
## minimum_nights           -0.67102    0.12274  -5.467 4.62e-08 ***
## number_of_reviews_ltm    -1.75062    0.13098 -13.366  < 2e-16 ***
## Room_typeHotel room     565.95513   42.66061  13.266  < 2e-16 ***
## Room_typePrivate room  -121.00440    7.10800 -17.024  < 2e-16 ***
## Room_typeShared room   -224.32352   32.54057  -6.894 5.54e-12 ***
## cityAustin              163.13334   28.29388   5.766 8.21e-09 ***
## cityBoston               28.73887   34.00204   0.845  0.39800
## cityCambridge            50.05707   52.72975   0.949  0.34247
## cityChicago               4.52736   27.62845   0.164  0.86984
## cityColumbus            -34.99839   36.45011  -0.960  0.33698
## cityDenver                2.60936   33.69629   0.077  0.93828
## cityFort Lauderdale     121.67358   26.33354   4.620 3.84e-06 ***
## cityJersey City         -15.15456   47.97480  -0.316  0.75209
## cityLas Vegas           200.22386   26.12418   7.664 1.85e-14 ***
## cityLos Angeles         118.61403   24.95609   4.753 2.01e-06 ***
## cityMonterrey           104.17450   99.53012   1.047  0.29526
## cityNashville            89.75658   28.23601   3.179  0.00148 **
## cityNew Orleans          74.81748   28.02948   2.669  0.00761 **
## cityNew York City        50.02857   25.05375   1.997  0.04585 *
## cityOakland              -8.51985   34.64966  -0.246  0.80577
## cityPortland            -35.05234   31.36643  -1.118  0.26378
## cityRhode Island        138.76691   29.31476   4.734 2.21e-06 ***
## citySalem               -36.02903   85.65768  -0.421  0.67404
## citySan Diego           113.40160   27.33698   4.148 3.36e-05 ***
## citySan Francisco        84.01518   29.95173   2.805  0.00503 **
## citySan Mateo County     77.22609   39.63163   1.949  0.05135 .
## citySanta Clara County   10.95326   29.60187   0.370  0.71137
## citySanta Cruz County   116.84549   54.45360   2.146  0.03190 *
## citySeattle              -4.77845   31.11615  -0.154  0.87795
## cityTwin Cities MSA      25.52820   31.28819   0.816  0.41456
## cityWashington D.C.      27.82213   30.45117   0.914  0.36090
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 520.2 on 30516 degrees of freedom
## Multiple R-squared:  0.03906,    Adjusted R-squared:  0.03805
## F-statistic: 38.76 on 32 and 30516 DF,  p-value: < 2.2e-16
##
##
## [[7]]
##
```

```
## Call:
## lm(formula = Price ~ ., data = df_cleaned[, c("Price", predictors_i)])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -886.9  -135.7   -65.3    17.4 20711.8
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    1.941e+02  2.477e+01   7.836 4.82e-15 ***
## availability_365               1.334e-01  2.575e-02   5.181 2.23e-07 ***
## minimum_nights                -7.069e-01  1.231e-01  -5.742 9.44e-09 ***
## number_of_reviews_ltm         -1.715e+00  1.313e-01 -13.059  < 2e-16 ***
## Room_typeHotel room            5.563e+02  4.273e+01  13.019  < 2e-16 ***
## Room_typePrivate room         -1.232e+02  7.132e+00 -17.279  < 2e-16 ***
## Room_typeShared room          -2.227e+02  3.254e+01  -6.845 7.77e-12 ***
## cityAustin                     1.626e+02  2.829e+01   5.749 9.09e-09 ***
## cityBoston                     2.676e+01  3.400e+01   0.787 0.431269
## cityCambridge                  4.836e+01  5.272e+01   0.917 0.358958
## cityChicago                   -3.037e-03  2.765e+01   0.000 0.999912
## cityColumbus                  -3.520e+01  3.644e+01  -0.966 0.334104
## cityDenver                     2.836e+00  3.369e+01   0.084 0.932921
## cityFort Lauderdale            1.214e+02  2.633e+01   4.609 4.05e-06 ***
## cityJersey City               -1.342e+01  4.797e+01  -0.280 0.779684
## cityLas Vegas                  1.984e+02  2.612e+01   7.594 3.19e-14 ***
## cityLos Angeles                1.147e+02  2.497e+01   4.595 4.35e-06 ***
## cityMonterrey                  1.043e+02  9.951e+01   1.048 0.294541
## cityNashville                  8.791e+01  2.823e+01   3.114 0.001850 **
## cityNew Orleans                7.456e+01  2.802e+01   2.661 0.007803 **
## cityNew York City              4.933e+01  2.505e+01   1.969 0.048930 *
## cityOakland                   -7.528e+00  3.464e+01  -0.217 0.827968
## cityPortland                  -3.466e+01  3.136e+01  -1.105 0.269075
## cityRhode Island               1.388e+02  2.931e+01   4.735 2.20e-06 ***
## citySalem                     -3.497e+01  8.564e+01  -0.408 0.683041
## citySan Diego                  1.131e+02  2.733e+01   4.137 3.53e-05 ***
## citySan Francisco              8.512e+01  2.995e+01   2.842 0.004482 **
## citySan Mateo County           7.679e+01  3.962e+01   1.938 0.052629 .
## citySanta Clara County         8.679e+00  2.960e+01   0.293 0.769382
## citySanta Cruz County          1.170e+02  5.444e+01   2.149 0.031613 *
## citySeattle                   -5.569e+00  3.111e+01  -0.179 0.857927
## cityTwin Cities MSA            2.633e+01  3.128e+01   0.842 0.399912
## cityWashington D.C.            2.727e+01  3.045e+01   0.896 0.370465
## calculated_host_listings_count  1.069e-01  2.901e-02   3.686 0.000228 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 520.1 on 30515 degrees of freedom
## Multiple R-squared:  0.03949,    Adjusted R-squared:  0.03845
## F-statistic: 38.02 on 33 and 30515 DF,  p-value: < 2.2e-16
##
##
## [[8]]
##
## Call:
## lm(formula = Price ~ ., data = df_cleaned[, c("Price", predictors_i)])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -889.9  -136.3   -65.2    17.8 20698.4
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    257.77278   27.60797   9.337  < 2e-16 ***
## availability_365                 0.13238    0.02573   5.144 2.71e-07 ***
## minimum_nights                  -0.69315    0.12308  -5.632 1.80e-08 ***
## number_of_reviews_ltm           -1.69222    0.13133 -12.886  < 2e-16 ***
## Room_typeHotel room            562.96026   42.73275  13.174  < 2e-16 ***
## Room_typePrivate room         -124.44713    7.13304 -17.447  < 2e-16 ***
## Room_typeShared room          -220.99360   32.52448  -6.795 1.11e-11 ***
## cityAustin                     173.09257   28.34751   6.106 1.03e-09 ***
## cityBoston                      43.84127   34.14220   1.284 0.199124
## cityCambridge                   66.55002   52.81352   1.260 0.207645
## cityChicago                     18.06843   27.85461   0.649 0.516557
## cityColumbus                   -21.85483   36.51677  -0.598 0.549519
## cityDenver                      17.78909   33.79680   0.526 0.598646
```

```
## cityFort Lauderdale            129.08454   26.35861   4.897 9.77e-07 ***
## cityJersey City                  0.64773   48.02245   0.013 0.989238
## cityLas Vegas                  204.59741   26.13955   7.827 5.15e-15 ***
## cityLos Angeles                126.62819   25.06613   5.052 4.40e-07 ***
## cityMonterrey                  132.05669   99.60912   1.326 0.184933
## cityNashville                   95.69513   28.26198   3.386 0.000710 ***
## cityNew Orleans                 91.14923   28.19178   3.233 0.001225 **
## cityNew York City               59.46460   25.11388   2.368 0.017900 *
## cityOakland                      5.55324   34.71943   0.160 0.872925
## cityPortland                   -11.62514   31.65636  -0.367 0.713450
## cityRhode Island               142.26613   29.30379   4.855 1.21e-06 ***
## citySalem                      -18.93306   85.65897  -0.221 0.825072
## citySan Diego                  123.15102   27.38802   4.497 6.93e-06 ***
## citySan Francisco              106.51421   30.21406   3.525 0.000424 ***
## citySan Mateo County            64.91911   39.67201   1.636 0.101767
## citySanta Clara County          19.55687   29.66293   0.659 0.509706
## citySanta Cruz County          102.70164   54.48823   1.885 0.059461 .
## citySeattle                     14.32399   31.33022   0.457 0.647535
## cityTwin Cities MSA             31.77660   31.28650   1.016 0.309797
## cityWashington D.C.             42.61936   30.57426   1.394 0.163339
## calculated_host_listings_count    0.11083    0.02901   3.821 0.000133 ***
## walk_index                      -5.27681    1.01186  -5.215 1.85e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 519.9 on 30514 degrees of freedom
## Multiple R-squared:  0.04035,    Adjusted R-squared:  0.03928
## F-statistic: 37.73 on 34 and 30514 DF,  p-value: < 2.2e-16
```

```
#Cooks distance function to identify outliers in our dataset
cooksD <- cooks.distance(simple_regression)
influential <- cooksD[(cooksD > (3 * mean(cooksD, na.rm = TRUE)))]
```

```
#New York  #.0276 R squared

df_NWY <- filter(df_cleaned, city == 'New York City')

Regression_NWY = lm(Price ~ walk_index + number_of_reviews_ltm + Room_type, data= df_NWY)
summary(Regression_NWY)
```

```
##
## Call:
## lm(formula = Price ~ walk_index + number_of_reviews_ltm + Room_type,
##     data = df_NWY)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -236.5   -92.8   -51.9     3.3 19485.2
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            200.5468    30.8519   6.500 8.76e-11 ***
## walk_index               3.9772     2.1364   1.862 0.062714 .
## number_of_reviews_ltm   -0.6420     0.2332  -2.753 0.005919 **
## Room_typeHotel room     80.9048    70.5480   1.147 0.251514
## Room_typePrivate room -132.7424    11.4669 -11.576  < 2e-16 ***
## Room_typeShared room  -187.0109    50.5092  -3.703 0.000216 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 403 on 5242 degrees of freedom
## Multiple R-squared:  0.02853,    Adjusted R-squared:  0.0276
## F-statistic: 30.78 on 5 and 5242 DF,  p-value: < 2.2e-16
```

```
#Creating a list of dataframes per city to later create linear regression per city
LOF <- split(df_cleaned, df_cleaned$city)
```

```r
df_list <- list(LOF$Asheville, LOF$Austin, LOF$Boston, LOF$Cambridge, LOF$`Clark County`, LOF$Chicago, LOF$Columb
us, LOF$Denver, LOF$`Fort Lauderdale`, LOF$`Jersey City`, LOF$`Los Angeles`, LOF$Monterrey, LOF$Nashville, LOF$`N
ew Orleans`,LOF$`New York City`, LOF$Oakland, LOF$Portland, LOF$`Rhode Island`, LOF$Salem, LOF$`San Diego`, LOF$`
San Francisco`, LOF$`San Mateo County`, LOF$`Santa Clara County`, LOF$`Santa Cruz County`, LOF$Seattle, LOF$`Twin
Cities MSA`, LOF$`Washington D.C.`)
```

```r
#R-square .04
regression_Asheville = lm(Price ~ walk_index + number_of_reviews_ltm + Room_type, data= LOF$Asheville)
summary(regression_Asheville)
```

```
##
## Call:
## lm(formula = Price ~ walk_index + number_of_reviews_ltm + Room_type,
##     data = LOF$Asheville)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -171.29  -73.43  -33.31   34.04 2049.07
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           217.0505    19.7905  10.967  < 2e-16 ***
## walk_index             -1.6096     1.4486  -1.111 0.267076
## number_of_reviews_ltm  -1.0330     0.2672  -3.866 0.000126 ***
## Room_typeHotel room   111.9034    74.4953   1.502 0.133729
## Room_typePrivate room -59.0670    20.6760  -2.857 0.004469 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 148.2 on 470 degrees of freedom
## Multiple R-squared:  0.05193,    Adjusted R-squared:  0.04386
## F-statistic: 6.435 on 4 and 470 DF,  p-value: 4.771e-05
```

```r
#R-square .05
regression_austin = lm(Price ~ walk_index + number_of_reviews_ltm + Room_type, data= LOF$Austin)
summary(regression_austin)
```

```
##
## Call:
## lm(formula = Price ~ walk_index + number_of_reviews_ltm + Room_type,
##     data = LOF$Austin)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -463.6 -216.2  -98.6   41.8 9931.4
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            614.6939    69.2691   8.874  < 2e-16 ***
## walk_index             -12.5293     4.7710  -2.626  0.00875 **
## number_of_reviews_ltm   -3.8720     0.7229  -5.356 1.02e-07 ***
## Room_typePrivate room -297.6656    49.7977  -5.977 3.00e-09 ***
## Room_typeShared room  -401.6624   108.0340  -3.718  0.00021 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 520.6 on 1176 degrees of freedom
## Multiple R-squared:  0.0591, Adjusted R-squared:  0.0559
## F-statistic: 18.47 on 4 and 1176 DF,  p-value: 9.911e-15
```

```r
#R-square .17
regression_boston = lm(Price ~ walk_index + number_of_reviews_ltm + Room_type, data= LOF$Boston)
summary(regression_boston)
```

```
##
## Call:
## lm(formula = Price ~ walk_index + number_of_reviews_ltm + Room_type,
##     data = LOF$Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -169.90  -61.13  -21.86   25.86 2456.63
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            103.4788    56.1921   1.842   0.0662 .
## walk_index               7.8537     3.5856   2.190   0.0290 *
## number_of_reviews_ltm   -0.1269     0.3485  -0.364   0.7160
## Room_typeHotel room     23.1164   157.2831   0.147   0.8832
## Room_typePrivate room -142.4806    14.9832  -9.509   <2e-16 ***
## Room_typeShared room  -198.3731   157.3548  -1.261   0.2081
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 156.9 on 463 degrees of freedom
## Multiple R-squared:  0.1795, Adjusted R-squared:  0.1707
## F-statistic: 20.26 on 5 and 463 DF,  p-value: < 2.2e-16
```

```
#R-square .27
regression_cam = lm(Price ~ walk_index + number_of_reviews_ltm + Room_type, data= LOF$Cambridge)
summary(regression_cam)
```

```
##
## Call:
## lm(formula = Price ~ walk_index + number_of_reviews_ltm + Room_type,
##     data = LOF$Cambridge)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -142.65  -49.45  -18.82   30.01  409.32
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            260.98508   67.51163   3.866 0.000181 ***
## walk_index              -2.45057    4.30552  -0.569 0.570314
## number_of_reviews_ltm   -0.00826    0.33729  -0.024 0.980504
## Room_typePrivate room -115.60357   16.60042  -6.964 1.95e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 88.75 on 119 degrees of freedom
## Multiple R-squared:  0.2953, Adjusted R-squared:  0.2775
## F-statistic: 16.62 on 3 and 119 DF,  p-value: 4.389e-09
```

```
#R-square .05
regression_chicago = lm(Price ~ walk_index + number_of_reviews_ltm + Room_type, data= LOF$Chicago)
summary(regression_chicago)
```

```
##
## Call:
## lm(formula = Price ~ walk_index + number_of_reviews_ltm + Room_type,
##     data = LOF$Chicago)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -183.4  -81.6  -42.5    8.3 4766.2
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)             19.6826    40.5151   0.486   0.6272
## walk_index              11.4700     2.5543   4.490 7.69e-06 ***
## number_of_reviews_ltm   -0.4252     0.1726  -2.463   0.0139 *
## Room_typeHotel room      0.2105    76.1859   0.003   0.9978
## Room_typePrivate room -101.5074    14.4945  -7.003 3.87e-12 ***
## Room_typeShared room  -138.1820    57.9324  -2.385   0.0172 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 214.7 on 1402 degrees of freedom
## Multiple R-squared:  0.05624,    Adjusted R-squared:  0.05288
## F-statistic: 16.71 on 5 and 1402 DF,  p-value: 4.609e-16
```

```
#R-square .12
regression_colu = lm(Price ~ walk_index + number_of_reviews_ltm + Room_type, data= LOF$Columbus)
summary(regression_colu)
```

```
##
## Call:
## lm(formula = Price ~ walk_index + number_of_reviews_ltm + Room_type,
##     data = LOF$Columbus)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -103.51  -47.25  -17.18   13.72  664.46
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           161.8202    23.0168   7.031 1.07e-11 ***
## walk_index             -1.1788     1.4981  -0.787    0.432
## number_of_reviews_ltm  -0.2746     0.1989  -1.380    0.168
## Room_typePrivate room -79.4152    10.8388  -7.327 1.61e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 81.43 on 353 degrees of freedom
## Multiple R-squared:  0.1363, Adjusted R-squared:  0.129
## F-statistic: 18.57 on 3 and 353 DF,  p-value: 3.289e-11
```

```
#R-square .002
regression_denver = lm(Price ~ walk_index + number_of_reviews_ltm + Room_type, data= LOF$Denver)
summary(regression_denver)
```

```
##
## Call:
## lm(formula = Price ~ walk_index + number_of_reviews_ltm + Room_type,
##     data = LOF$Denver)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -152.5  -79.7  -36.7   12.0 4371.8
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)             85.0381    87.8932   0.968   0.3338
## walk_index               6.3020     5.7844   1.089   0.2765
## number_of_reviews_ltm   -0.7407     0.4259  -1.739   0.0826 .
## Room_typeHotel room    -26.6182   235.2536  -0.113   0.9100
## Room_typePrivate room  -37.7340    29.1288  -1.295   0.1958
## Room_typeShared room  -154.5613   166.6397  -0.928   0.3541
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 234.8 on 474 degrees of freedom
## Multiple R-squared:  0.01263,    Adjusted R-squared:  0.002215
## F-statistic: 1.213 on 5 and 474 DF,  p-value: 0.302
```

```
#R-square .023
regression_fl = lm(Price ~ walk_index + number_of_reviews_ltm + Room_type, data= LOF$`Fort Lauderdale`)
summary(regression_fl)
```

```
##
## Call:
## lm(formula = Price ~ walk_index + number_of_reviews_ltm + Room_type,
##     data = LOF$`Fort Lauderdale`)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -356.8  -176.9   -92.4    26.2 14616.7
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            439.7381    50.3045   8.742  < 2e-16 ***
## walk_index              -4.2317     3.5983  -1.176   0.2397
## number_of_reviews_ltm   -4.1788     0.6889  -6.066 1.54e-09 ***
## Room_typeHotel room    -86.9110   215.7379  -0.403   0.6871
## Room_typePrivate room -131.5260    32.0060  -4.109 4.11e-05 ***
## Room_typeShared room  -311.7701   124.9226  -2.496   0.0126 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 526.4 on 2229 degrees of freedom
## Multiple R-squared:  0.02581,    Adjusted R-squared:  0.02363
## F-statistic: 11.81 on 5 and 2229 DF,  p-value: 2.688e-11
```

```
#R-square .06. Regression on subset of LA with high influential variables
regression_LA = lm(Price ~ walk_index + number_of_reviews_ltm + Room_type + availability_365 + minimum_nights + c
alculated_host_listings_count,  data= LOF$`Los Angeles`)
summary(regression_LA)
```

```
##
## Call:
## lm(formula = Price ~ walk_index + number_of_reviews_ltm + Room_type +
##     availability_365 + minimum_nights + calculated_host_listings_count,
##     data = LOF$`Los Angeles`)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -653.0  -195.7   -78.8    31.0 20551.2
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    697.21891   40.36685  17.272  < 2e-16 ***
## walk_index                     -26.88713    2.41203 -11.147  < 2e-16 ***
## number_of_reviews_ltm           -3.15980    0.43249  -7.306 3.14e-13 ***
## Room_typeHotel room           -104.72579  175.30146  -0.597 0.550262
## Room_typePrivate room         -250.36990   18.13393 -13.807  < 2e-16 ***
## Room_typeShared room          -294.16711   76.78409  -3.831 0.000129 ***
## availability_365                 0.38256    0.06754   5.664 1.55e-08 ***
## minimum_nights                  -0.35793    0.25187  -1.421 0.155342
## calculated_host_listings_count  -0.12345    0.04260  -2.898 0.003768 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 579.7 on 5616 degrees of freedom
## Multiple R-squared:  0.06883,    Adjusted R-squared:  0.0675
## F-statistic: 51.89 on 8 and 5616 DF,  p-value: < 2.2e-16
```

```
#Regression on Monterrey with certain variables, we see based on the R-squared this is not a good model, the mode
l is
#predicting worse than the mean of our model
regression_monterrey = lm(Price ~ walk_index + number_of_reviews_ltm + Room_type, data= LOF$Monterrey)
summary(regression_monterrey)
```

```
##
## Call:
## lm(formula = Price ~ walk_index + number_of_reviews_ltm + Room_type,
##     data = LOF$Monterrey)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -184.03  -64.80  -14.78   20.37  473.79
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)             50.799    260.902   0.195    0.847
## walk_index              12.211     14.990   0.815    0.423
## number_of_reviews_ltm    1.058      1.237   0.856    0.401
## Room_typeHotel room   -136.357    158.217  -0.862    0.397
## Room_typePrivate room   65.688     73.847   0.890    0.383
##
## Residual standard error: 151.2 on 24 degrees of freedom
## Multiple R-squared:  0.1304, Adjusted R-squared:  -0.01448
## F-statistic: 0.9001 on 4 and 24 DF,  p-value: 0.4794
```

```
#R-square .04
regression_nashville = lm(Price ~ walk_index + number_of_reviews_ltm + Room_type, data= LOF$Nashville)
summary(regression_nashville)
```

```
##
## Call:
## lm(formula = Price ~ walk_index + number_of_reviews_ltm + Room_type,
##     data = LOF$Nashville)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -255.7 -123.3  -51.8   45.1 4698.1
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            177.2185    29.6572   5.976 3.03e-09 ***
## walk_index               8.3835     2.0642   4.061 5.20e-05 ***
## number_of_reviews_ltm   -0.6363     0.1764  -3.607 0.000323 ***
## Room_typeHotel room    -81.4789   124.3929  -0.655 0.512587
## Room_typePrivate room -152.9519    26.6984  -5.729 1.28e-08 ***
## Room_typeShared room  -151.4721   195.8622  -0.773 0.439463
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 276.5 on 1185 degrees of freedom
## Multiple R-squared:  0.05236,    Adjusted R-squared:  0.04836
## F-statistic:  13.1 on 5 and 1185 DF,  p-value: 1.977e-12
```

```
#R-square .04
regression_nol = lm(Price ~ walk_index + number_of_reviews_ltm + Room_type, data= LOF$`New Orleans`)
summary(regression_nol)
```

```
##
## Call:
## lm(formula = Price ~ walk_index + number_of_reviews_ltm + Room_type,
##     data = LOF$`New Orleans`)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -911.6 -120.9  -62.0   36.9 5706.9
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)             64.4292    60.3004   1.068  0.28551
## walk_index              13.5667     3.9218   3.459  0.00056 ***
## number_of_reviews_ltm   -1.0760     0.3842  -2.801  0.00518 **
## Room_typeHotel room    664.4559   125.2691   5.304 1.34e-07 ***
## Room_typePrivate room  -81.8884    26.1062  -3.137  0.00175 **
## Room_typeShared room  -172.9859   116.0114  -1.491  0.13618
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 305.3 on 1251 degrees of freedom
## Multiple R-squared:  0.04689,    Adjusted R-squared:  0.04308
## F-statistic: 12.31 on 5 and 1251 DF,  p-value: 1.133e-11
```

```
regression_nyc = lm(Price ~ walk_index + number_of_reviews_ltm + Room_type, data= LOF$`New York City`)
summary(regression_nyc)
```

```
##
## Call:
## lm(formula = Price ~ walk_index + number_of_reviews_ltm + Room_type,
##     data = LOF$`New York City`)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -236.5   -92.8   -51.9     3.3 19485.2
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            200.5468    30.8519   6.500 8.76e-11 ***
## walk_index               3.9772     2.1364   1.862 0.062714 .
## number_of_reviews_ltm   -0.6420     0.2332  -2.753 0.005919 **
## Room_typeHotel room     80.9048    70.5480   1.147 0.251514
## Room_typePrivate room -132.7424    11.4669 -11.576  < 2e-16 ***
## Room_typeShared room  -187.0109    50.5092  -3.703 0.000216 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 403 on 5242 degrees of freedom
## Multiple R-squared:  0.02853,    Adjusted R-squared:  0.0276
## F-statistic: 30.78 on 5 and 5242 DF,  p-value: < 2.2e-16
```

```
regression_oak = lm(Price ~ walk_index + number_of_reviews_ltm + Room_type, data= LOF$Oakland)
summary(regression_oak)
```

```
##
## Call:
## lm(formula = Price ~ walk_index + number_of_reviews_ltm + Room_type,
##     data = LOF$Oakland)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -186.21  -66.36  -23.64   17.30 2180.23
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            265.0381    40.7260   6.508 2.15e-10 ***
## walk_index              -4.7651     2.7374  -1.741  0.08245 .
## number_of_reviews_ltm   -1.1452     0.4161  -2.752  0.00617 **
## Room_typePrivate room -111.9217    17.9358  -6.240 1.06e-09 ***
## Room_typeShared room  -154.2319    51.2295  -3.011  0.00276 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 150.6 on 426 degrees of freedom
## Multiple R-squared:  0.107,  Adjusted R-squared:  0.09856
## F-statistic: 12.75 on 4 and 426 DF,  p-value: 8.18e-10
```

```
regression_portland = lm(Price ~ walk_index + number_of_reviews_ltm + Room_type, data= LOF$Portland)
summary(regression_portland)
```

```
##
## Call:
## lm(formula = Price ~ walk_index + number_of_reviews_ltm + Room_type,
##     data = LOF$Portland)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -100.01  -43.75  -19.65   14.12 1360.54
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            110.9507    28.9823   3.828 0.000142 ***
## walk_index               1.7515     1.7287   1.013 0.311341
## number_of_reviews_ltm   -0.3889     0.1328  -2.929 0.003521 **
## Room_typeHotel room    -13.0032    64.2199  -0.202 0.839606
## Room_typePrivate room  -58.1078     9.9133  -5.862  7.3e-09 ***
## Room_typeShared room  -118.0016    90.7316  -1.301 0.193872
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 90.58 on 648 degrees of freedom
## Multiple R-squared:  0.06132,    Adjusted R-squared:  0.05408
## F-statistic: 8.467 on 5 and 648 DF,  p-value: 8.947e-08
```

```
regression_wa = lm(Price ~ walk_index + number_of_reviews_ltm + Room_type, data= LOF$`San Mateo County`)
summary(regression_wa)
```

```
##
## Call:
## lm(formula = Price ~ walk_index + number_of_reviews_ltm + Room_type,
##     data = LOF$`San Mateo County`)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -235.48  -92.22  -37.23   34.37 1119.47
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            359.723     38.510   9.341  < 2e-16 ***
## walk_index              -6.797      3.413  -1.992  0.04743 *
## number_of_reviews_ltm   -0.972      0.520  -1.869  0.06272 .
## Room_typePrivate room -169.993     25.152  -6.759 8.78e-11 ***
## Room_typeShared room  -236.190     84.890  -2.782  0.00578 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 184.2 on 266 degrees of freedom
## Multiple R-squared:  0.1732, Adjusted R-squared:  0.1608
## F-statistic: 13.93 on 4 and 266 DF,  p-value: 2.48e-10
```

```
write.csv(df_cleaned, "Airbnb_clean_dataset.csv")
```