MA678 Midterm Project

Airbnb Analysis

Dae Hyun Lee

12/2/2019

Contents

Abstract

Airbnb, which started its service in 2008, was an unconventional accommodation rental service that focused on allowing travelers to live like locals with the concept of renting my home or my room.

As one of the travel buffs who love to travel, I use Airbnb a lot even more than Hotel services these days. And, I love that I can experience the traveling area as a resident, not as a tourist, through the Airbnb platform.

The first and most mentioned, when comparing Airbnb to a hotel is the price. One of the reasons why people prefer Airbnb to a hotel is because it is cheaper. And, I've noticed that nicer hotels tend to cost more than average Airbnbs in most cities. However, there were many cases that I found booking a hotel is sometimes even less expensive than renting a house/room through Airbnb.

In many cases, the booking fee set by the host seems unreasonably expensive regarding the condition of the house. Thinking of this made me decide to investigate what are the factors that mostly affect consumers in determining Airbnb.

Also, I compare how the overall Airbnb prices change and differ by time of the year. I initially assume that the overall Airbnb listing prices in Boston would reach the highest in May due to the regional attribute that there are always a lot of people gathering over the graduation period.

1. Introduction

1.1 Overview

Throughout the project, I look for investigating and proving a few research questions and hypotheses, such as "What are the factors that consumers consider the most when deciding where to stay through Airbnb platform." Or "In Massachusetts, the overall Airbnb prices would fluctuate according to the school schedule in Boston and reach their highest point around May when there are many graduation ceremonies."

The data I use for the project is collected from the site 'Inside Airbnb' http://insideairbnb.com/get-the-data.html, which provides more detailed data than most of the other opensource data websites does, such as Kaggle and Tomslee. However, I also use the data from Tomslee to see the difference in years and to better evaluate the models as comparison.

For this project, to better evaluate the effect of factors that I intend to investigate, I restrict the region, especially within the area of Boston.

Below is the data description:

Table 1: List of variables (Tomslee)

room_id	neighborhood	bedrooms	longitude
host_id	reviews	price	last_modified
room_type	overall_satisfaction	minstay	date
borough	accommodates	latitude	Date

Table 2: List of variables (Inside Airbnb)

id	host_is_superhost	beds	require_guest_profile_picture
date	host_listings_count	price	require_guest_phone_verification
Date	host_total_listings_count	cleaning_fee	number_of_reviews
last_scraped	host_verifications	weekly_price	reviews_per_month
calendar_last_scraped	host_has_profile_pic	monthly_price	review_scores_rating
neighbourhood_cleansed	host_identity_verified	security_deposit	review_scores_accuracy
zipcode	calculated_host_listings_count	guests_included	review_scores_cleanliness
latitude	property_type	extra_people	review_scores_checkin
longitude	room_type	minimum_nights	review_scores_communication
host_id	accommodates	maximum_nights	review_scores_location
host_name	bedrooms	instant_bookable	review_scores_value
host_response_time	bathrooms	cancellation_policy	id

1.2 Outline

The outline of this project report is as follows. First, I display the conclusions of exploratory data analyses and state a specific research question of interest. Next, I describe the methods employed to answer the research questions via statistical models. Last, I interpret the models and discuss the results.

Project goals

1. Test and check the factors that mostly affect consumers in determining Airbnb.
2. Investigate how the overall Airbnb prices change and differ by month within the year between Oct 2018 and Sep 2019.
3. Figure out the effect of cleaning fees on other variables, including the price/day.

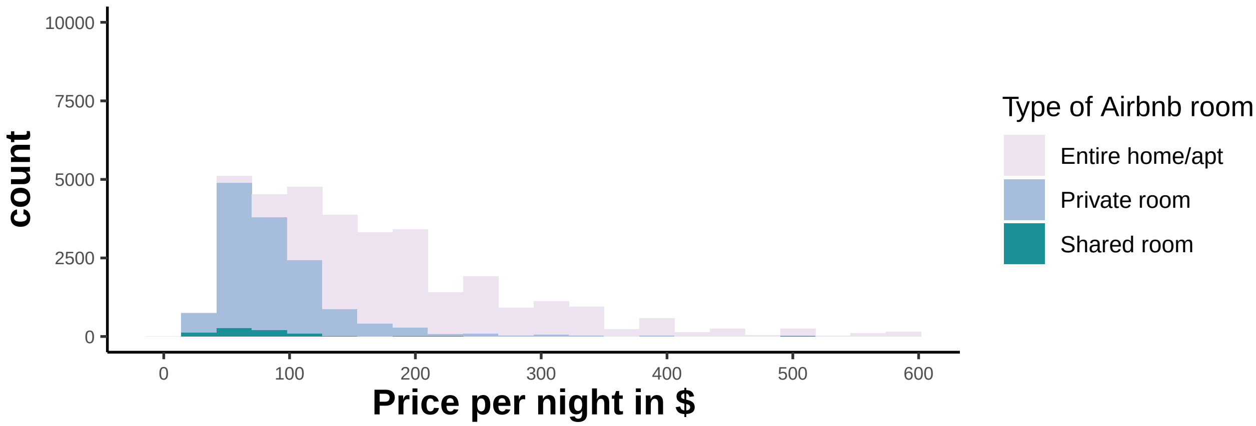
Hypotheses

1. Airbnb prices in Boston would reach the highest in May.
2. Airbnb prices in Boston would reach the lowest in December and January.

2. Exploratory Data Analysis

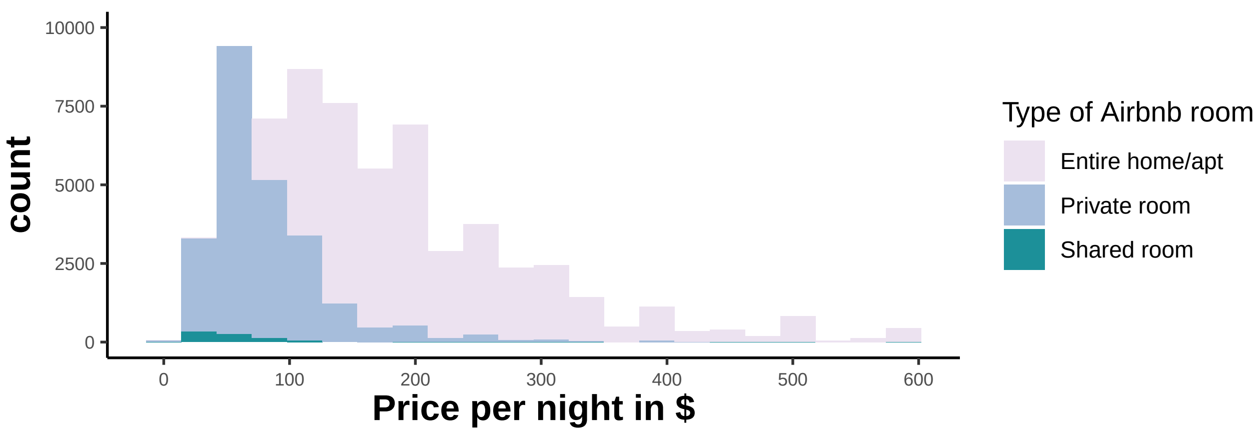
Figure 2.1

Figure 2.1.(tomslee):
Distribution of the airbnb listings over price
within Boston region in 2016




```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    10.0    85.0   145.0   172.4   210.0 10000.0
```

Figure 2.1.(inside):
Distribution of the airbnb listings over price
within Boston region in 2019




```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0    80.0   149.0   195.2   225.0 10000.0
```

This plot gives the general impression of how Airbnb listings are distributed over the price in the Boston area. From this plot, it is evident that there has been a significant increase in the number of Airbnb listings from 2016 to 2019 in Boston. In 2019, the number of Airbnb listings in Boston has almost doubled from 2016.

To better have an idea of how the listings are distributed over the price, I restricted the range of the Airbnb listing price from zero to 600 dollars per night. In this way, it was easier to read the trends or the patterns from the plot.

Figure 2.2

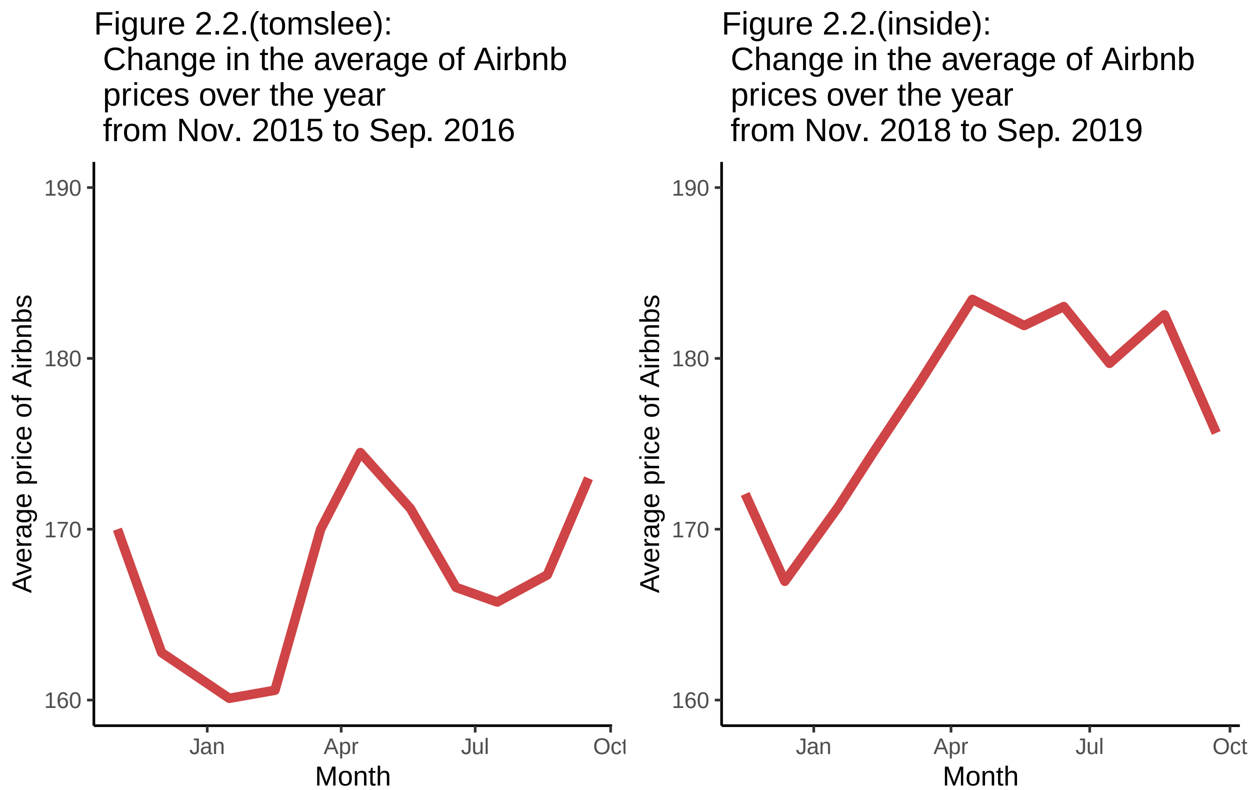


Figure 2.2.(tomslee) and Figure 2.2.(inside) shows the average of the total Airbnb nightly rate over the two different years in Boston.

The figure 2.2.(tomslee) is the average Airbnb price changing over the year between November 2015 and September 2016. Also, the figure 2.2.(inside) is the average price changing from November 2018 to September 2019.

From the plots, the average nightly price for the Airbnb listings in Boston dropped to the lowest point around in January and reached the highest roughly around the month of May and for both years.

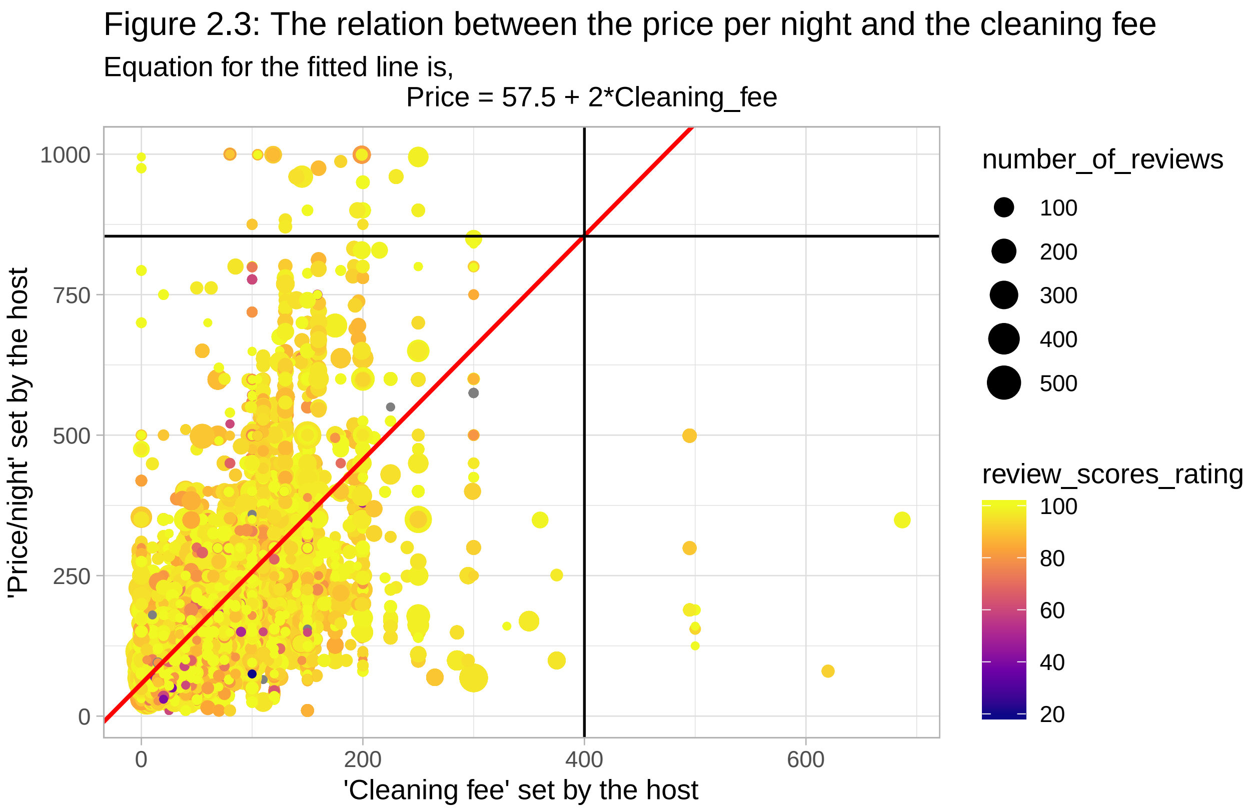
Many visitors or tourists visit Boston every year because of its regional attributes of having numerous schools and colleges around it. Since most schools are holding their graduation ceremonies roughly from April to May, Boston is mostly crowded with many tourists, visitors, and locals trying to attend graduation ceremonies for their friends or families. As a result, we can see that the price of accommodation such as hotels and Airbnb increases as well.

Also, in the winter, most students, professors, and people who study or work at universities return to their home or travel from vacation time, and most people are expected to leave Boston by December. Therefore, as in the graph above shows, the price of Airbnb reaching the lowest in the month of December within the year seems reasonable.

Lastly, the overall average price has increased from the year between 2015 and 2016 to the year between 2018 and 2019. The overall pattern of rising Airbnb prices in Boston from the previous years can be explained by inflation and its impact on accommodation prices in the United States.

The variable that I used to indicate the timeline of the year for the inside data was stored as 'last_scraped,' which seemed highly reliable to indicate the time-varying prices within the combined datasets that contain redundant Airbnb listings within Boston district. However, for the tomslee datasets, I have manually input the values as the date which were read from the Airbnb website by tomslee.

Figure 2.3



The simple linear regression line fitted to the dataset between the nightly rate and the cleaning fee, which both set by the host, represents that they are positively correlated. According to the regression plot, one unit increase in the cleaning fee would lead to the two-unit increase in the nightly price rate.

The brighter the color of the scatterplots indicates the higher review scores given by the guest. Also, the size of the dots represents the number of reviews that the host received from their guests.

One of the impressive notions that I had while I was looking at this plot was that most of the listings at the very end of the right lower corner received high review scores for the rating. My initial assumption was that there would be many dishonest Airbnb hosts who deliberately set the price lower and additional fees, such as cleaning fees, relatively high to attract consumers. Therefore, hosts who intentionally set the cleaning fee high compared to the nightly price would receive bad reviews. However, it appears that it would not be the case all the time by looking at the plot.

Based on the review scores, I assume those Airbnb listings with a cleaning fee set above 400 dollars would probably be the luxurious Airbnb houses. I suppose that the hosts of the previously mentioned listings may have lowered prices and increased costs as part of their marketing plans. Otherwise, they would not have received that much high review scores if the houses were in bad condition.

At the same time, there are some listings that are colored so dark under the regression line that I fitted in red color. Those were the ones that I wanted to see before I planned for this project. If the price per night can represent the general quality or the condition of the house, including the location, popularity, etc., the cleaning fee should increase or decrease according to the nightly rate. For example, if one looks at the darkest colored listing on the plot under the fitted regression line, that listing has 100 dollars for the cleaning fee and about 100 dollars for the nightly rate. In this case, staying just one night would cost more than the double amount of nightly rate, which seems not a reasonable price to me as a potential guest.

Since the total price of a reservation on Airbnb is based on the nightly rate set by the host plus other fees or costs also determined by the host, I assume many of the hosts would willingly set the price/day lower and cleaning fee higher. Because the higher the cleaning fee is, the more the percentage of tax would be imposed on the guests than to the hosts.

Figure 2.3.1

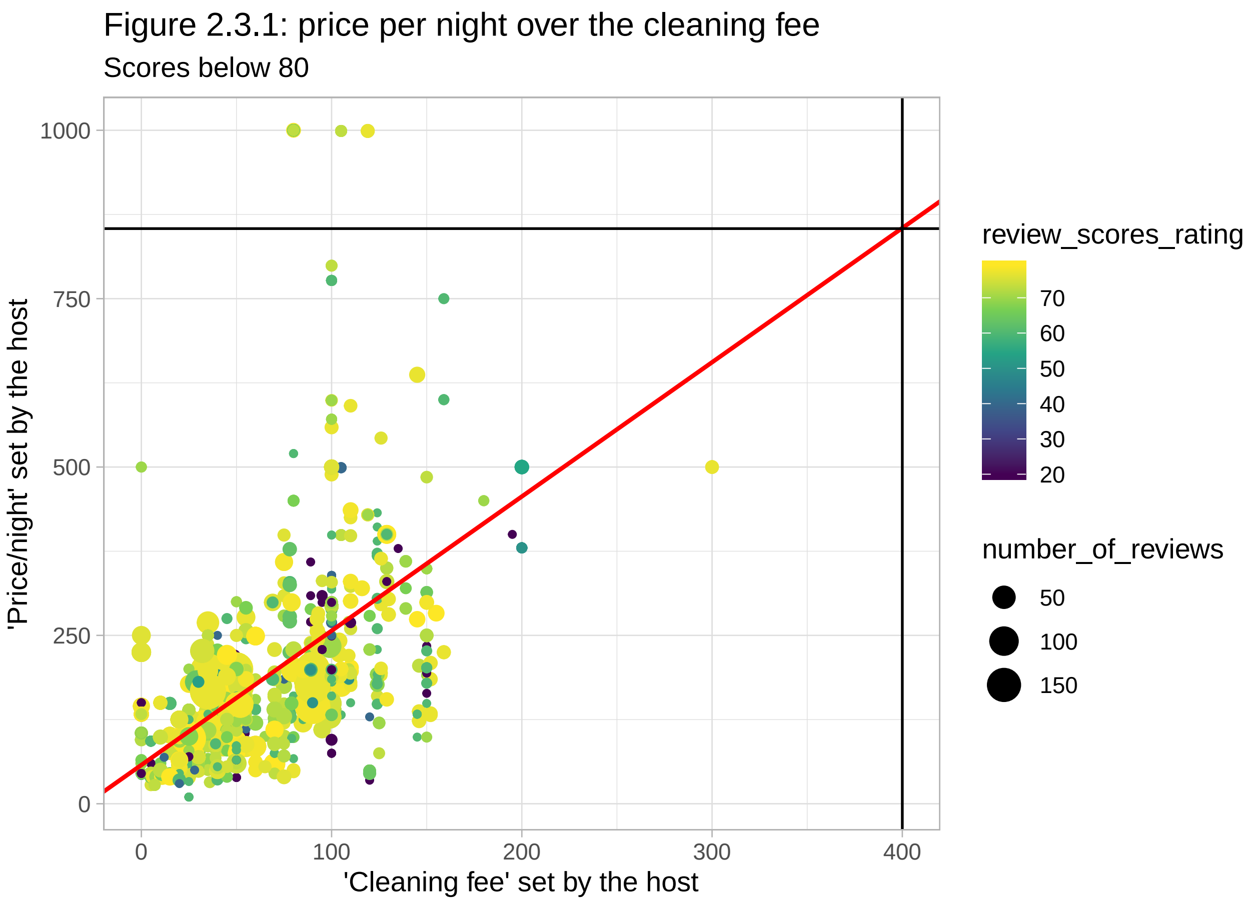


Figure 2.3.1 shows the Airbnb listings, which received review ratings below the score of 85. From this plot, I was able to see that the listings with a comparatively large number of reviews appear to have higher review scores.

Receiving zero for the review score is such an extreme case. Thereby, I intentionally deleted the listings which received zero for the review ratings.

Figure 2.4

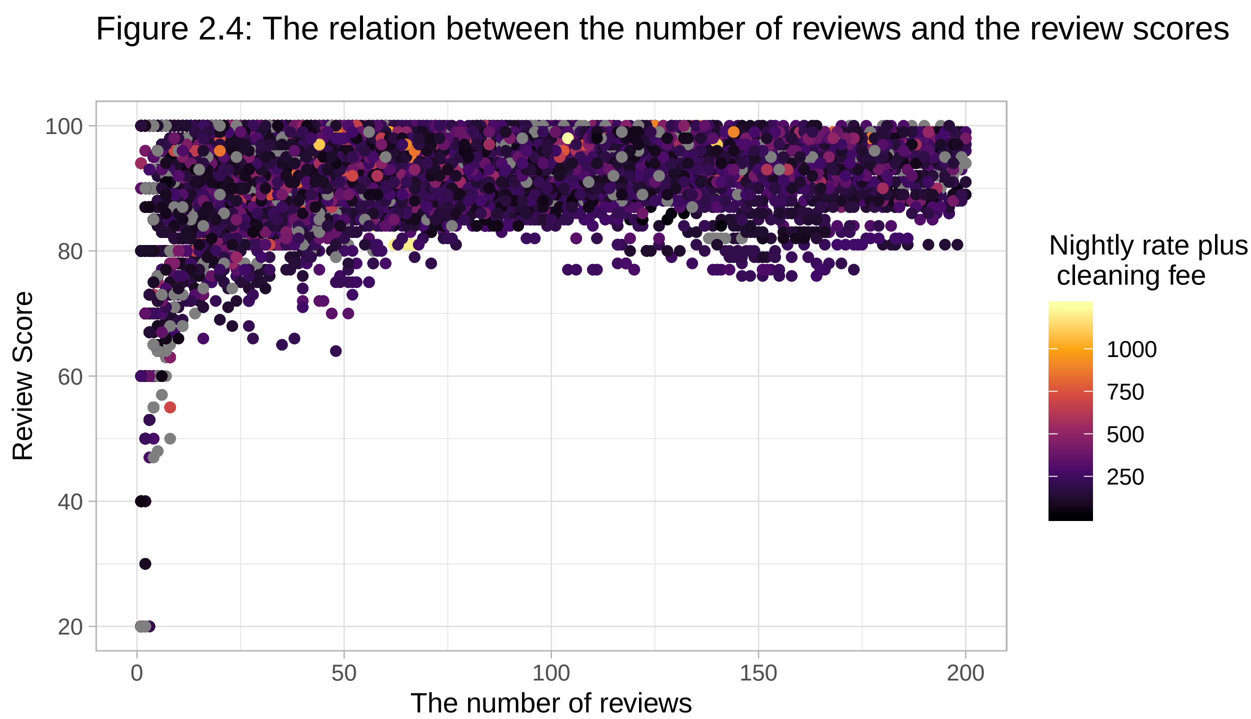
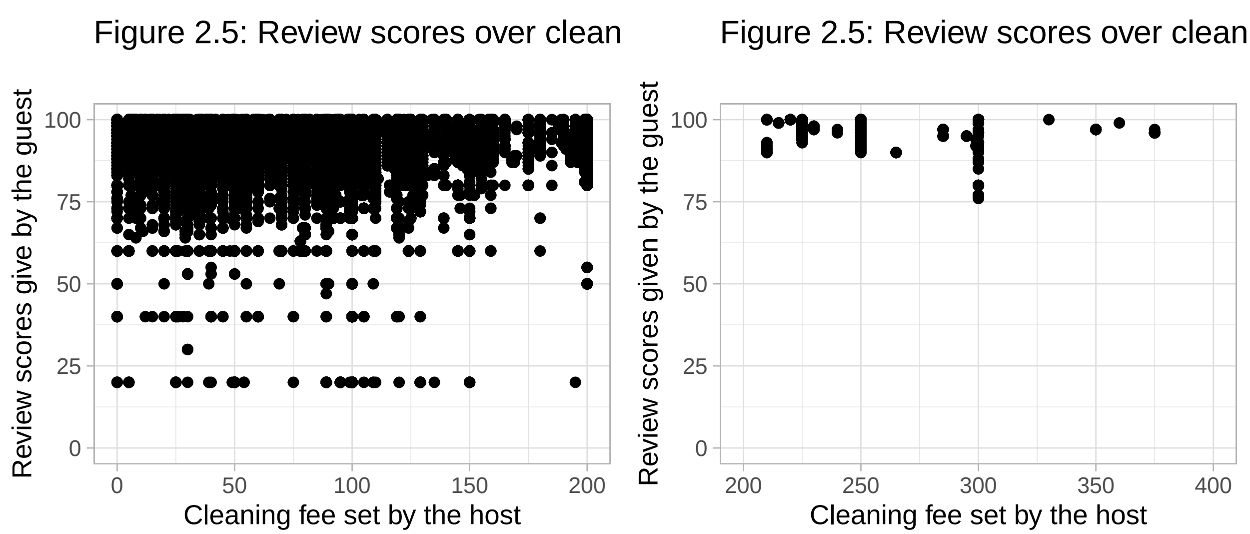


Figure 2.4 shows the relationship between the number of reviews and the review score ratings. Similar to the result from the figure 2.4, individual listings appear to have a positive relationship with the number of reviews.

Figure 2.5



Lastly, figure 2.5 shows that there is not any clear patterns or the relationship between the cleaning fee and the review scores within the range of cleaning fee between zero to 200 dollars.

3. Methods

I had initially decided to use Boston Airbnb open data from Kaggle, multiple regional Airbnb data from Tom Slee.

However, one of the problems that I found from those datasets is that the price for the individual Airbnb house/room is not explicitly indicated. It was not clear if the Airbnb price in the datasets is whether the total booking price set by the host, including the fee for the cleaning and service all together. I believed that it might not be reliable to make careful analyses or predictions based on the price listed in the previous datasets.

The data I found from the website called 'Inside Airbnb' has a variable that holds the price of the cleaning fee set by the host for each Airbnb listings. And, the data contains much more observations than the ones from Tom Slee.

For both data from Inside Airbnb and Tom Slee, I created a variable holding the value as a date of indicating when the data was scraped from the Airbnb Website.

Since most of the information in Inside Airbnb datasets were stored as a factor level, I had to make significant adjustments to the datasets. For example, the value of the price is even stored as a factor level with '$' sign as well as ',' sign in between the numbers as '$100,231.00'.

The cleaning and wrangling process that I had conducted for both Tom Slee and Inside Airbnb datasets is separated from this project Rmd file and saved in other files. Due to the amount of the work that I had made for just cleaning and wrangling, I decided to separate the part of the cleaning process from this analysis file.

The method of data cleaning and wrangling for Tom Slee datasets is in the file named 'tomslee_data_cleaning' as well as that of data cleaning for Inside Airbnb datasets is in the file 'insideairbnb_data_cleaning.R.'

4. Models

Model 1

What are the factors that mostly effect consumers in determining Airbnb.

Review scores the for Airbnb listings in Boston will be higher if the cleaning fee is high and the nightly rate is low.

```
##
## Call:
## lm(formula = review_scores_rating ~ number_of_reviews + cleaning_fee +
##     price + instant_bookable + review_scores_accuracy + review_scores_cleanliness +
##     review_scores_checkin + review_scores_communication + review_scores_location +
##     review_scores_value, data = dat_inside)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -44.487  -1.562   0.380   1.805  33.519
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -4.6921747  0.3147582 -14.907  < 2e-16 ***
## number_of_reviews           -0.0021121  0.0003004  -7.031 2.07e-12 ***
## cleaning_fee                 0.0046076  0.0004627   9.959  < 2e-16 ***
## price                        0.0004480  0.0001544   2.901  0.00372 **
## instant_bookable            -0.4946233  0.0365697 -13.525  < 2e-16 ***
## review_scores_accuracy       2.8742372  0.0340946  84.302  < 2e-16 ***
## review_scores_cleanliness    1.9281299  0.0279012  69.106  < 2e-16 ***
## review_scores_checkin        0.5450584  0.0328463  16.594  < 2e-16 ***
## review_scores_communication  2.1226662  0.0339710  62.485  < 2e-16 ***
## review_scores_location       0.2021874  0.0268543   7.529 5.20e-14 ***
## review_scores_value          2.6147503  0.0300560  86.996  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.034 on 49351 degrees of freedom
##   (96 observations deleted due to missingness)
## Multiple R-squared:  0.7486, Adjusted R-squared:  0.7486
## F-statistic: 1.47e+04 on 10 and 49351 DF,  p-value: < 2.2e-16
```

- number of reviews: On average review score for the Airbnb listing decrease by 0.002 with every one more review on the listing.
- cleaning fee: one unit increase in the cleaning fee will lead to 0.005 increase in the review score.
- price: an increase in the price will have a significantly small amount of positive effect on the review score.
- instant bookable: instant bookable has a negative effect on the review score.
- review scores for six categories: increase in all the review scores for six categories will have a positive effective on the overall review scores.

As a result, both cleaning fee and price does not have much effect on the review scores unlike what I expected from the EDA. Also, the higher number of reviews does not necessarily increase the review scores. Moreover, the relationship between the cleaning fee set by the host is more likely to effect review scores in a positive way.

Model 2

What are the factors that can predict the nightly rate of the Airbnb listings in Boston.

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## price ~ review_scores_rating + accommodates + bedrooms + number_of_reviews +
##     reviews_per_month + (1 + Date) + (1 | room_type)
##    Data: dat_inside
##
## REML criterion at convergence: 607100.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.2592 -0.4705 -0.1167  0.2808  7.1075
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  room_type (Intercept)  4416     66.45
##  Residual              12574    112.13
## Number of obs: 49447, groups:  room_type, 3
##
## Fixed effects:
##                       Estimate Std. Error t value
## (Intercept)          -6.864e+02  9.432e+01  -7.278
## review_scores_rating -3.334e-01  6.320e-02  -5.275
## accommodates          1.361e+01  4.185e-01  32.516
## bedrooms              2.946e+01  9.165e-01  32.145
## number_of_reviews    -4.310e-02  1.045e-02  -4.124
## reviews_per_month    -8.942e+00  3.053e-01 -29.284
## Date                  4.367e-02  4.772e-03   9.152
##
## Correlation of Fixed Effects:
##             (Intr) rvw_s_ accmmd bedrms nmbr__ rvws__
## rvw_scrs_rt -0.082
## accommodats  0.012  0.036
## bedrooms    -0.015 -0.047 -0.787
## nmbr_f_rvws  0.067 -0.036  0.047 -0.015
## rvws_pr_mnt -0.051 -0.064 -0.169  0.124 -0.613
## Date        -0.911  0.022 -0.017  0.016 -0.071  0.056
```

```
##                            2.5 %        97.5 %
## .sig01                28.7083726  160.35949315
## .sigma               111.4315414  112.82927487
## (Intercept)         -867.3798564 -505.49584489
## review_scores_rating   -0.4572768   -0.20953263
## accommodates           12.7897917   14.43031682
## bedrooms               27.6625482   31.25512182
## number_of_reviews      -0.0635636   -0.02259962
## reviews_per_month      -9.5406569   -8.34377868
## Date                    0.0343214    0.05302645
```

Random effect is significant since none of the confidence interval crosses zero.

Model 3

The likelihood of booking Airbnb would reach the highest in the month of May.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   0.000   0.410   1.250   2.058   3.100  53.080   14264 
```

```
## 
## Call:
## glm(formula = popularity ~ factor(Date), data = dat_m2)
## 
## Deviance Residuals: 
##     Min       1Q   Median       3Q      Max  
## -0.2782  -0.2556  -0.2360  -0.2261   0.7739  
## 
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)    
## (Intercept)             0.273322   0.006861  39.837  < 2e-16 ***
## factor(Date)2018-11-17  0.004834   0.009618   0.503 0.615280    
## factor(Date)2018-12-13 -0.010902   0.009650  -1.130 0.258595    
## factor(Date)2019-01-17 -0.030202   0.009606  -3.144 0.001667 ** 
## factor(Date)2019-02-09 -0.037350   0.009641  -3.874 0.000107 ***
## factor(Date)2019-03-12 -0.047241   0.009620  -4.911 9.09e-07 ***
## factor(Date)2019-04-15 -0.046530   0.009597  -4.849 1.25e-06 ***
## factor(Date)2019-05-19 -0.041582   0.009572  -4.344 1.40e-05 ***
## factor(Date)2019-06-14 -0.025325   0.009550  -2.652 0.008008 ** 
## factor(Date)2019-07-14 -0.017756   0.009561  -1.857 0.063309 .  
## factor(Date)2019-08-19 -0.008678   0.009545  -0.909 0.363282    
## factor(Date)2019-09-22 -0.019119   0.009772  -1.957 0.050399 .  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for gaussian family taken to be 0.1872106)
## 
##     Null deviance: 9268.9  on 49446  degrees of freedom
## Residual deviance: 9254.8  on 49435  degrees of freedom
## AIC: 57489
## 
## Number of Fisher Scoring iterations: 2
```

The probability of making reviews greater than the third qualtile value from its distribution during the specific month for Airbnb listings in Boston is,

$Pr(popularity = 1) = logit^{-1}(0.273322 + 0.004834 * Nov. + 0.030202 * Dec. - 0.037350 * Jan. - 0.037350 * Feb. - 0.047241 * Mar. - 0.046530 * Apr. - 0.041582 * May. - 0.025325 * Jun. - 0.017756 * Jul. - 0.008678 * Aug. - 0.019119 * Sep.)$.

The regression coefficient for making reviews in April is -0.0465, which indicates that the month of April has a multiplicative effect of $e^{-0.046530} = 0.9545$ on the odds of booking 'Airbnb' compared to other months over the year in Boston.

The regression coefficient for making reviews in May is -0.0416, which indicates that the month of May has a multiplicative effect of $e^{-0.041582} = 0.9593$ on the odds of booking 'Airbnb' compared to other months over the year in Boston.

To quickly interpret the coefficient for April on probability scale, I divide the coefficient estimate for April by four: $\frac{0.046530}{4} = 0.0116325$.

To quickly interpret the coefficient for May on probability scale, I divide the coefficient estimate for May by four: $\frac{0.041582}{4} = 0.0103955$.

Thus, in April, people are 1.2% more likely to book Airbnbs in Boston and make reviews. Also, in May, people are 1.03% more likely to book Airbnbs in Boston and make reviews.

To test the popularity of Airbnb listings in Boston, I dichotomized the values in the variable 'review per month' either as 1 or 0 depending on the number of reviews the listing has received within the month.

Created another variable which contains 1 and if the listing received greater number of comments or reviews than the third quartile value of its distribution. For example, if the number of reviews per month exceeds the value of 3.1 than I gave 1 for that Airbnb listing.

Since there were not any variables holding the information of any number of times that each listings were booked within the month, I thought it could be reasonable to rely on the 'review per month' variable to see the popularity of Airbnb booking in Boston Area. Because the customers must have stayed from the listed houses/rooms in order to make any reviews or to see the reviews made by the host to the guest who stayed.

Based on the regression results, I have a moderate evidence that the number of booking Airbnb, as well as leaving reviews, would reach the highest in April or in May, especially in Boston.

Therefore, I would like to conclude that the model also supports the hypothesis that Airbnb prices in Boston would reach the highest around in May and drop to the lowest around December or January.

5. Conclusions

This project entailed testing a set of hypotheses pertaining to the effect that a particular set of factors had on the Airbnb prices and reviews.

I took the approach of first cleaning and manipulating the data so as to facilitate its use with visualization techniques that would, then, go on to serve as a prelude to model fitting. Performing analysis on each of the hypotheses resulted in some hypothesis being moderately supported by the data.

In summary, I can only infer that the regional attributes of Boston within the scope of the provided hypotheses do affect the price of Airbnb listings.

Appendix

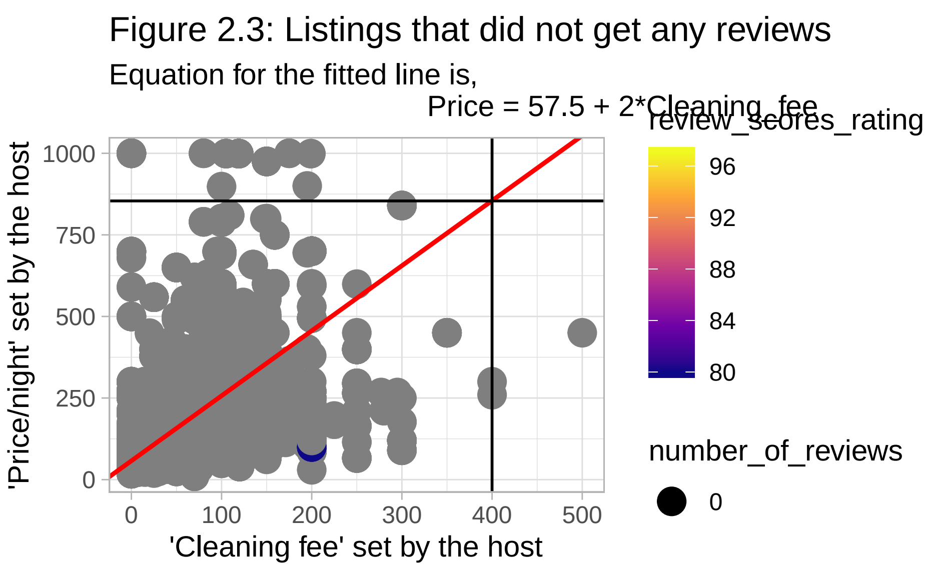
Additional figures for EDA

Summary of Reg_2.3

```
##
## Call:
## lm(formula = price ~ cleaning_fee, data = data_inside)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1213.4   -75.3   -30.2    25.1  9942.5
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  57.54806    1.92401   29.91   <2e-16 ***
## cleaning_fee  1.99328    0.02197   90.71   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 271.5 on 62339 degrees of freedom
##   (11624 observations deleted due to missingness)
## Multiple R-squared:  0.1166, Adjusted R-squared:  0.1166
## F-statistic:  8229 on 1 and 62339 DF,  p-value: < 2.2e-16
```

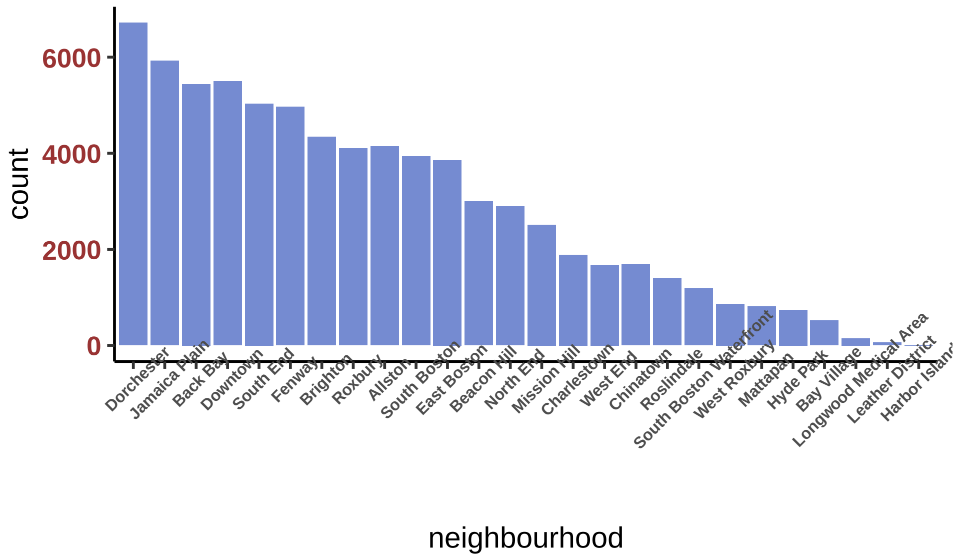
Summary of the Linear Regression fitted to test the relationship between the price and the cleaning fee.

Figure 2.3.2



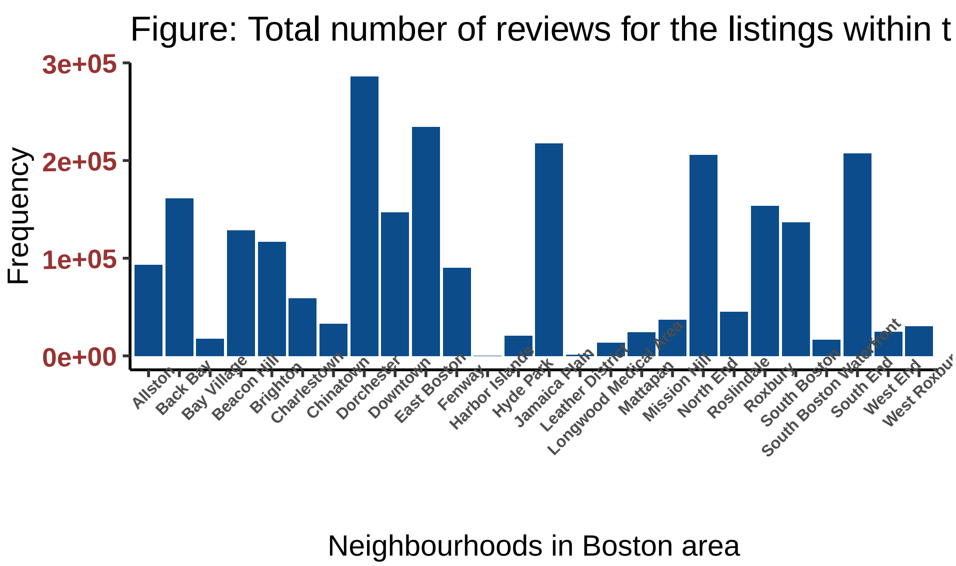
There are many listings that does not have any reviews.

Figure.A



It is clear to see the overall number of listings around Boston Area.

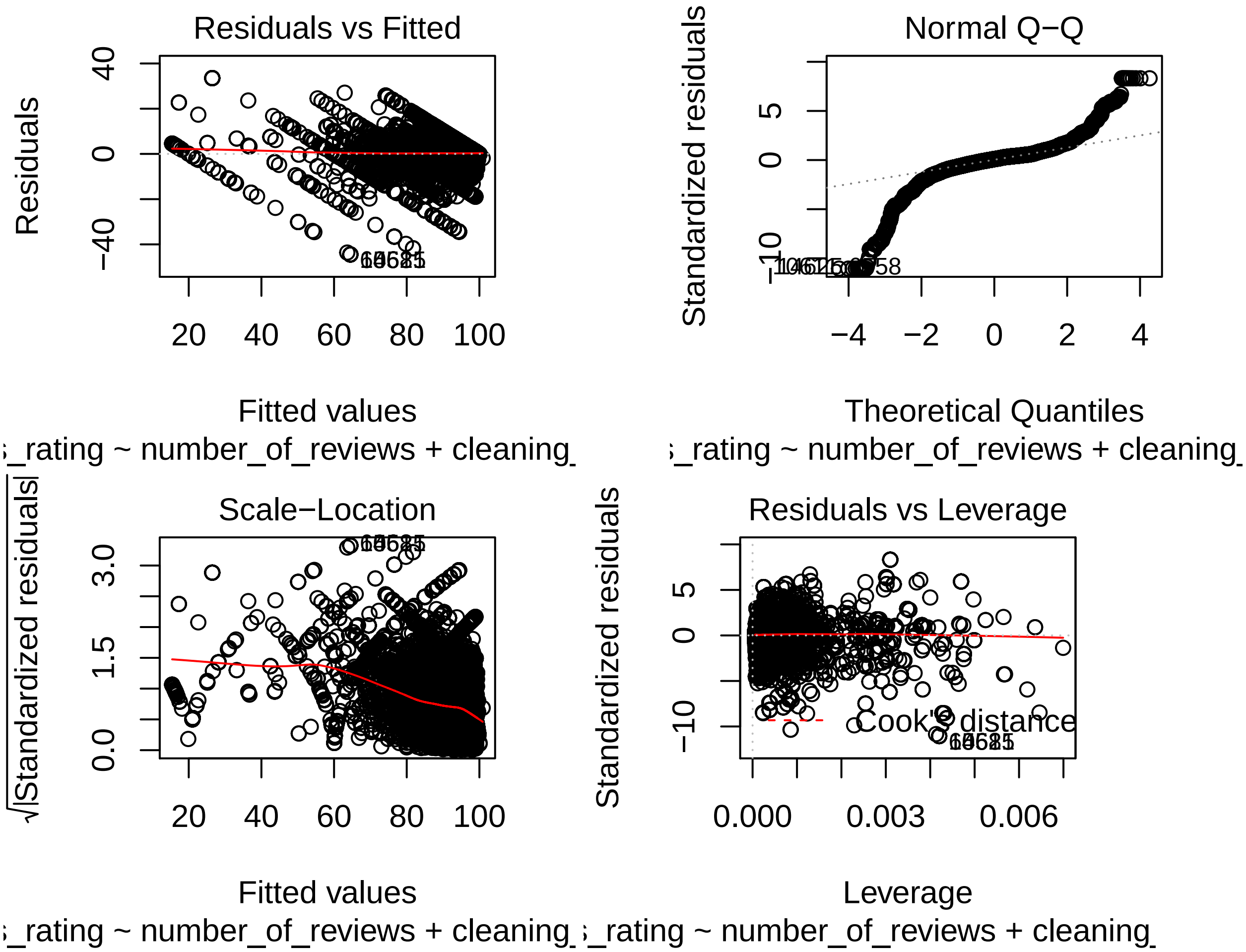
Figure.B



It shows the total number of reviews over the neighbourhoods in Boston. It was useful to see which districts of the Boston area are typically famous to stay among the Airbnb users.

Model Checking

Model 1




```
## Start:  AIC=137702.6
## review_scores_rating ~ number_of_reviews + cleaning_fee + price +
##     instant_bookable + review_scores_accuracy + review_scores_cleanliness +
##     review_scores_checkin + review_scores_communication + review_scores_location +
##     review_scores_value
##
##                               Df Sum of Sq    RSS    AIC
## <none>                                     803022 137703
## - price                        1       137 803159 137709
## - number_of_reviews            1       804 803827 137750
## - review_scores_location       1       922 803945 137757
## - cleaning_fee                 1      1614 804636 137800
## - instant_bookable             1      2977 805999 137883
## - review_scores_checkin        1      4481 807503 137975
## - review_scores_communication  1     63530 866552 141459
## - review_scores_cleanliness    1     77707 880729 142260
## - review_scores_accuracy       1    115640 918662 144342
## - review_scores_value          1    123149 926171 144743
```

```
##
## Call:
## lm(formula = review_scores_rating ~ number_of_reviews + cleaning_fee +
##     price + instant_bookable + review_scores_accuracy + review_scores_cleanliness +
##     review_scores_checkin + review_scores_communication + review_scores_location +
```

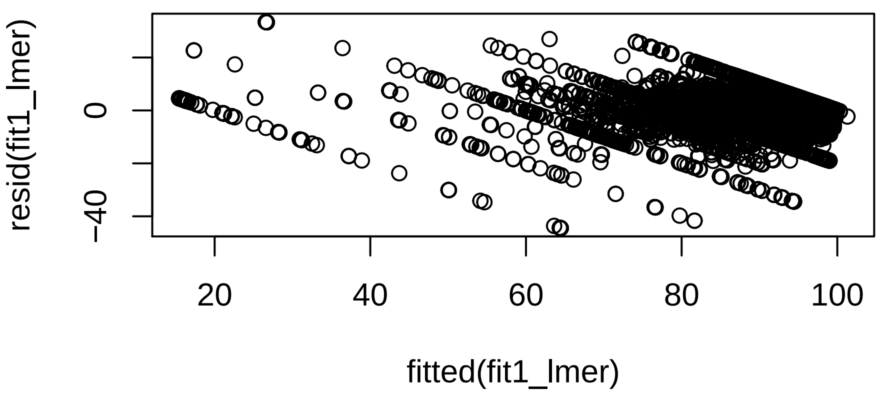
```
##     review_scores_value, data = dat_inside)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -44.487  -1.562   0.380   1.805  33.519
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -4.6921747  0.3147582 -14.907  < 2e-16 ***
## number_of_reviews          -0.0021121  0.0003004  -7.031 2.07e-12 ***
## cleaning_fee                0.0046076  0.0004627   9.959  < 2e-16 ***
## price                       0.0004480  0.0001544   2.901  0.00372 **
## instant_bookable           -0.4946233  0.0365697 -13.525  < 2e-16 ***
## review_scores_accuracy      2.8742372  0.0340946  84.302  < 2e-16 ***
## review_scores_cleanliness   1.9281299  0.0279012  69.106  < 2e-16 ***
## review_scores_checkin       0.5450584  0.0328463  16.594  < 2e-16 ***
## review_scores_communication  2.1226662  0.0339710  62.485  < 2e-16 ***
## review_scores_location      0.2021874  0.0268543   7.529 5.20e-14 ***
## review_scores_value         2.6147503  0.0300560  86.996  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.034 on 49351 degrees of freedom
##   (96 observations deleted due to missingness)
## Multiple R-squared:  0.7486, Adjusted R-squared:  0.7486
## F-statistic: 1.47e+04 on 10 and 49351 DF,  p-value: < 2.2e-16
```

Model 1 - Multilevel

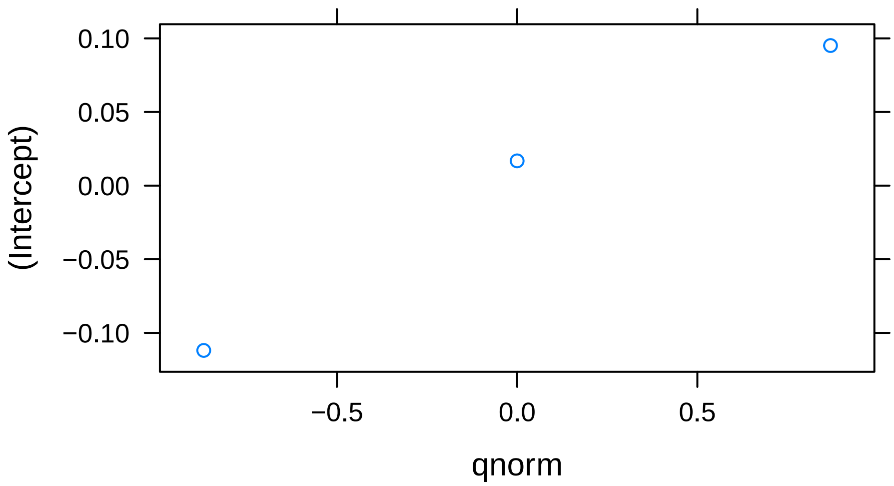
Multilevel linear model for the model.1

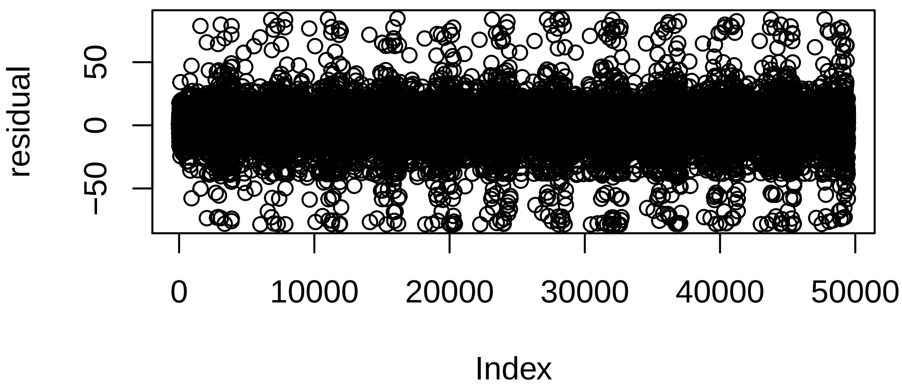
```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## review_scores_rating ~ number_of_reviews + cleaning_fee + price +
##     Date + instant_bookable + review_scores_accuracy + review_scores_cleanliness +
##     review_scores_checkin + review_scores_communication + review_scores_location +
##     review_scores_value + (1 | room_type)
##    Data: dat_inside
##
## REML criterion at convergence: 277845.1
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -11.0291  -0.3875   0.0923   0.4496   8.2911
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  room_type (Intercept)  0.01617 0.1272
##  Residual              16.26393 4.0329
## Number of obs: 49362, groups:  room_type, 3
##
## Fixed effects:
##                              Estimate Std. Error t value
## (Intercept)                 3.1859604  3.1050596   1.026
## number_of_reviews          -0.0019555  0.0003021  -6.474
```

```
## cleaning_fee                 0.0054190  0.0005007  10.823
## price                        0.0006197  0.0001590   3.898
## Date                        -0.0004452  0.0001718  -2.592
## instant_bookable            -0.4968311  0.0365685 -13.586
## review_scores_accuracy       2.8731789  0.0340884  84.286
## review_scores_cleanliness    1.9334765  0.0279323  69.220
## review_scores_checkin        0.5417150  0.0328475  16.492
## review_scores_communication  2.1196413  0.0339703  62.397
## review_scores_location       0.2234431  0.0272293   8.206
## review_scores_value          2.6037642  0.0301412  86.386
##
## Correlation of Fixed Effects:
##             (Intr) nmbr__ clnng_ price  Date   instn_ rvw_scrs_cc
## nmbr_f_rvws  0.059
## cleaning_fe -0.002  0.094
## price        0.029  0.067 -0.378
## Date        -0.994 -0.050 -0.006 -0.034
## instnt_bkbl -0.001 -0.047 -0.038 -0.020 -0.007
## rvw_scrs_cc -0.010 -0.036  0.008 -0.004  0.004  0.023
## rvw_scrs_cl -0.026 -0.020 -0.071 -0.030  0.008 -0.008 -0.283
## rvw_scrs_ch -0.034 -0.052 -0.010  0.048  0.003 -0.006 -0.135
## rvw_scrs_cm -0.017 -0.030  0.002  0.038  0.002  0.023 -0.250
## rvw_scrs_lc -0.008 -0.045  0.017 -0.084 -0.037 -0.042 -0.099
## rvw_scrs_vl -0.011  0.034  0.063  0.020  0.016  0.028 -0.299
##             rvw_scrs_cl rvw_scrs_ch rvw_scrs_cm rvw_scrs_l
## nmbr_f_rvws
## cleaning_fe
## price
## Date
## instnt_bkbl
## rvw_scrs_cc
## rvw_scrs_cl
## rvw_scrs_ch -0.102
## rvw_scrs_cm -0.035      -0.396
## rvw_scrs_lc  0.015      -0.050      -0.028
## rvw_scrs_vl -0.259      -0.029      -0.210      -0.219
```




```
## $room_type
```

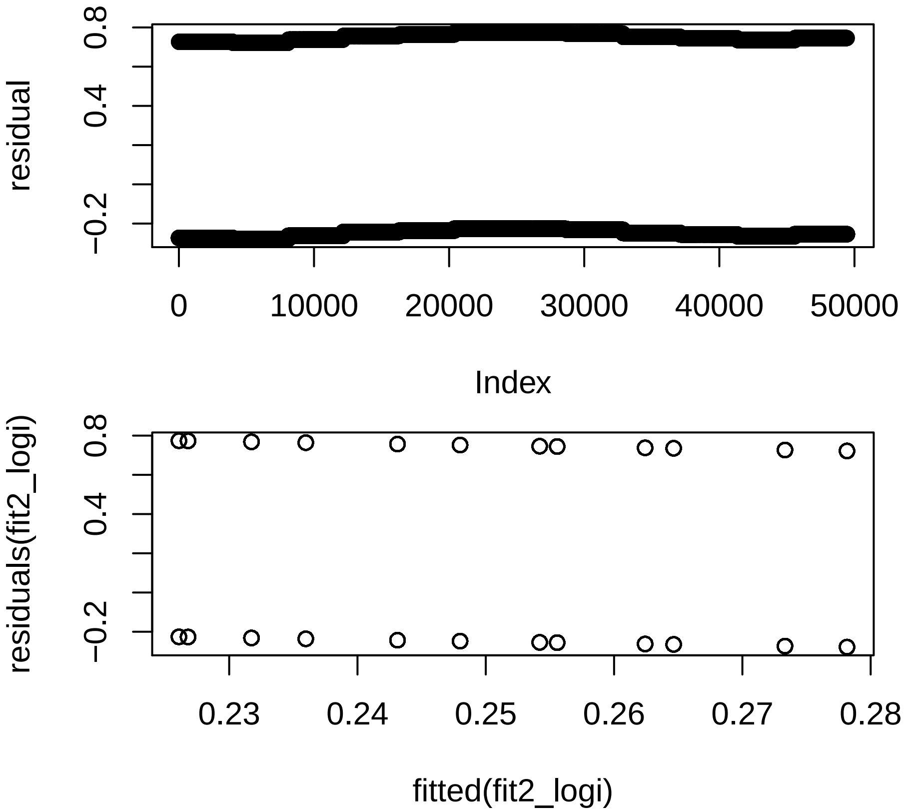




Model 2

Model 3

	2.5 %	97.5 %
(Intercept)	0.2598743	0.2867689
factor(Date)2018-11-17	-0.0140179	0.0236854
factor(Date)2018-12-13	-0.0298146	0.0080115
factor(Date)2019-01-17	-0.0490294	-0.0113753
factor(Date)2019-02-09	-0.0562458	-0.0184542
factor(Date)2019-03-12	-0.0660949	-0.0283871
factor(Date)2019-04-15	-0.0653391	-0.0277206
factor(Date)2019-05-19	-0.0603424	-0.0228224
factor(Date)2019-06-14	-0.0440434	-0.0066074
factor(Date)2019-07-14	-0.0364950	0.0009839
factor(Date)2019-08-19	-0.0273864	0.0100304
factor(Date)2019-09-22	-0.0382715	0.0000328



residual
0.8
0.4
−0.2
0
10000
20000
30000
40000
50000
Index
residuals(fit2_logi)
0.8
0.4
−0.2
0.23
0.24
0.25
0.26
0.27
0.28
fitted(fit2_logi)