# Final Project of Airbnb in Amsterdam

*Xiaofan Xia*

*12/9/2019*

## 1. Overview

### 1.1. Introduction

This project focuses on digging into the Airbnb dataset of Amsterdam. My main objective is to find out the relationship between reviewers' comments and price regarding to different room type and neighborhood. Then, I want to generate a shiny app to show an Airbnb location map of Amsterdam. In this project, I assume that prices of Airbnbs are in the same currency and the same unit.

### 1.2. Motivation

I am planning on take a vacation to Europe next spring break these days. Amsterdam is on top of my visiting lists. I believe Airbnb Amsterdam dataset could give me some help both in finding great neighborhood to stay and also positive reviewed places to stay while my visiting.

### 1.3. Dataset: Airbnb Amsterdam from Kaggle

source: https://www.kaggle.com/erikbruin/airbnb-amsterdam.

### 1.4. Research Questions

a. Is there any relationship between the price and the attitude of reviewers of an Airbnb?

b. Is there any relationship between the price and the neighborhood of the Airbnb located?

c. Is there any difference in prices regarding different room types of Airbnb?

d. Is there any relationship between the price and the availability of an Airbnb?

e. Generate a shiny app to show locations of Amsterdam's Airbnb places.

## 2. Exploratory Data Analysis

### 2.1. Basic Statistics

| | Maximum | Minimum | Mean | Median |
|---|---|---|---|---|
| Availability in nights during a year | 365.00 | 0.00 | 59.913680 | 3.00 |
| Number of reviews per month | 11.85 | 0.01 | 1.062080 | 0.62 |
| Number of reviews in total | 695.00 | 0.00 | 21.560459 | 8.00 |
| Price | 8500.00 | 0.00 | 152.181178 | 125.00 |
| Minimum nights of stay | 1001.00 | 1.00 | 3.328707 | 2.00 |

Here is the table of the numbers of Airbnb places in each neighbourhood.

| neighbourhood | n |
|---|---|
| Bijlmer-Centrum | 111 |
| Bijlmer-Oost | 96 |
| Bos en Lommer | 1145 |
| Buitenveldert - Zuidas | 262 |
| Centrum-Oost | 1730 |
| Centrum-West | 2326 |
| De Aker - Nieuw Sloten | 142 |
| De Baarsjes - Oud-West | 3515 |
| De Pijp - Rivierenbuurt | 2493 |
| Gaasperdam - Driemond | 122 |
| Geuzenveld - Slotermeer | 212 |
| IJburg - Zeeburgereiland | 452 |
| Noord-Oost | 257 |
| Noord-West | 320 |
| Oostelijk Havengebied - Indische Buurt | 959 |
| Osdorp | 159 |
| Oud-Noord | 571 |
| Oud-Oost | 1282 |
| Slotervaart | 393 |
| Watergraafsmeer | 552 |
| Westerpark | 1490 |
| Zuid | 1441 |

## 2.2. Graph Presentations

Here are graph presentations of share rooms. The first plot (Figure 1) illustrates the relationship between the number of reviews of each Airbnb in total and the prices. I can see that Airbnbs with lower prices tend to have more reviews. The second plot (Figure 2) shows that the relationship between the number of reviews per month and the prices of Airbnbs, which is basically consistent with the pattern in Figure 1. The price range of shared rooms is around 0 to 300 units. From Figure 3, most places with prices lower than 100 units are basically sold out through the year.

Figure 1

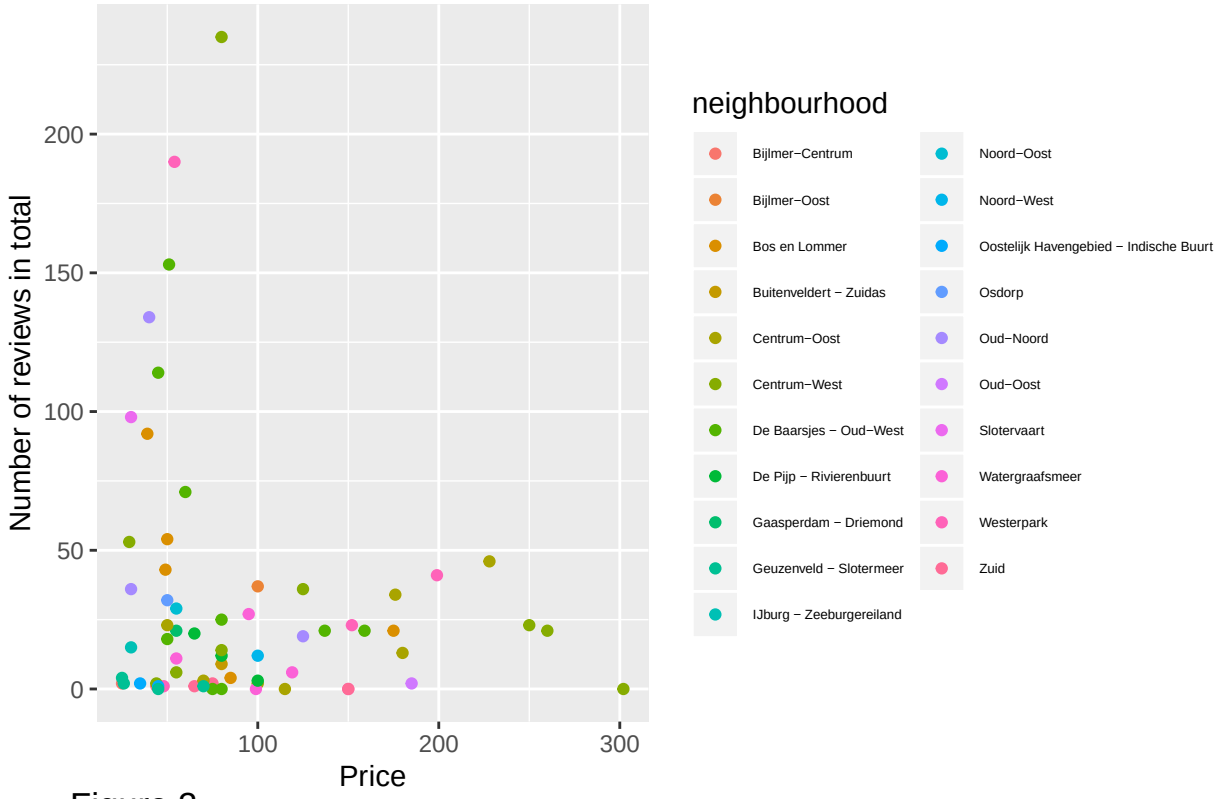



Figure 2

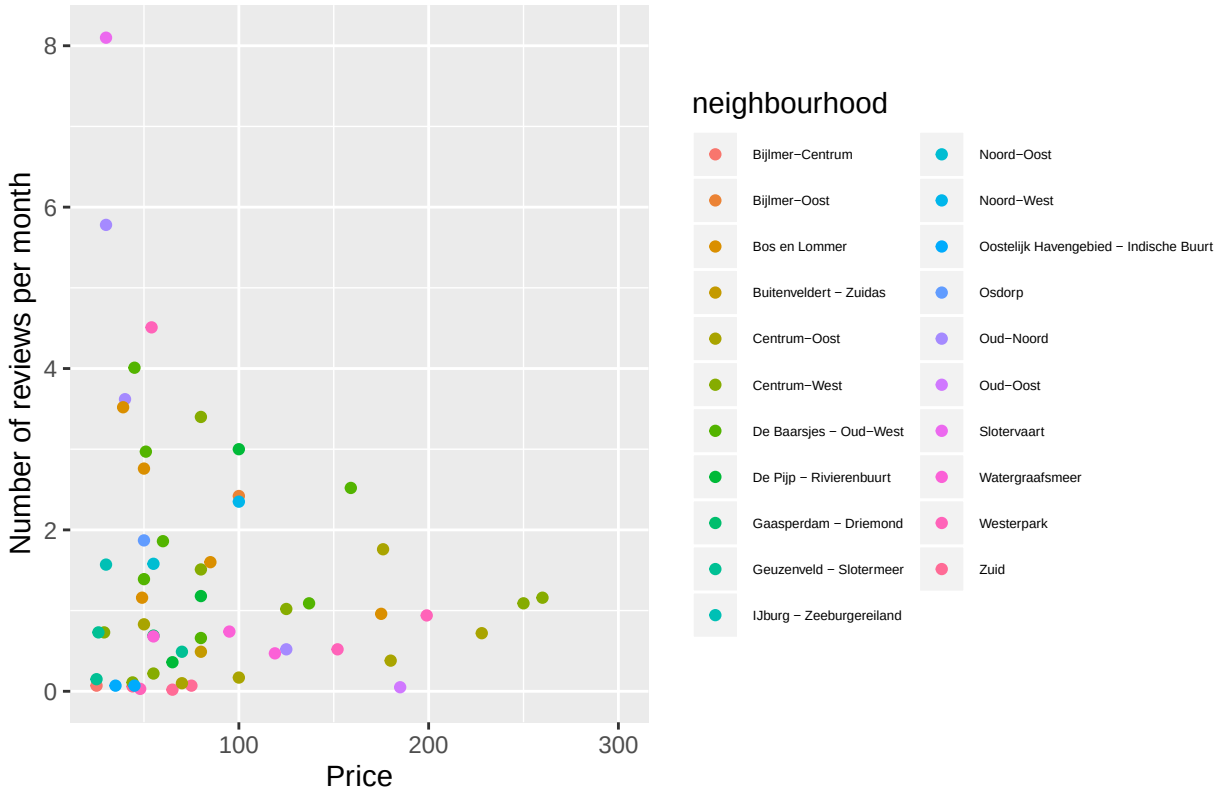
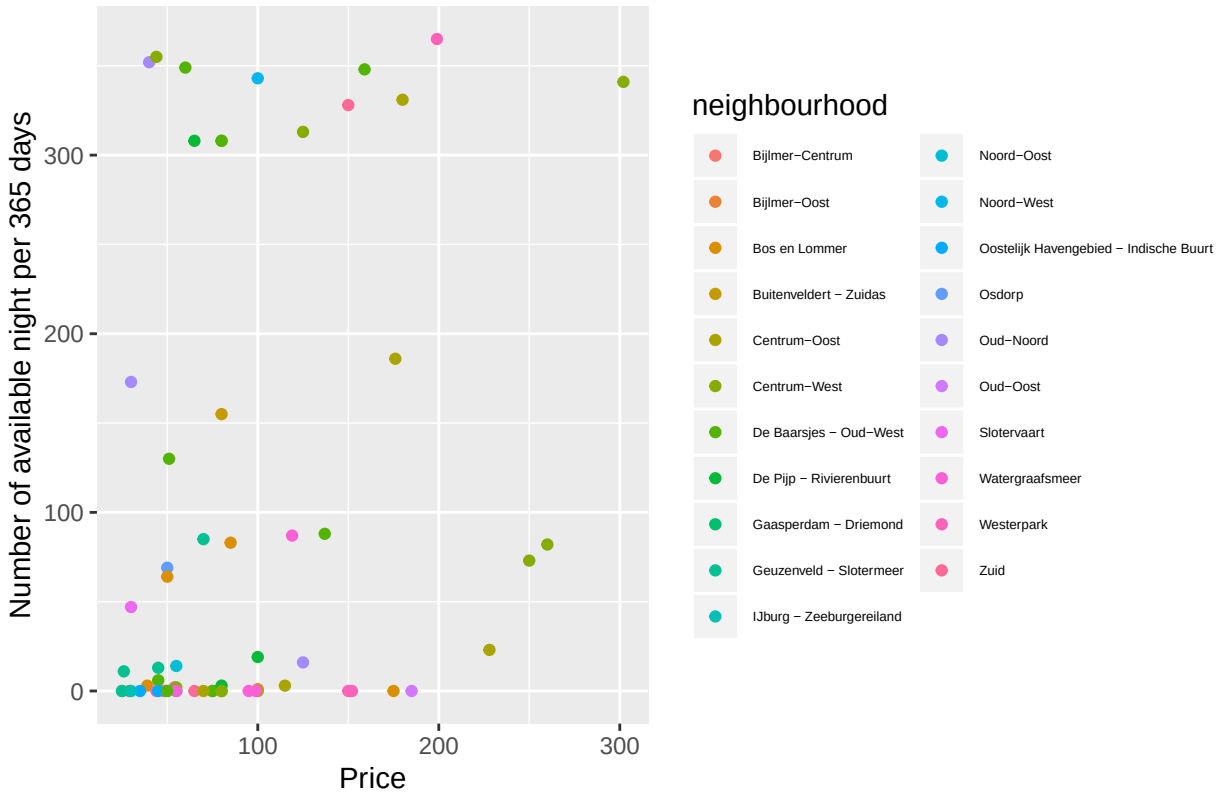


Figure 3


Number of available night per 365 days
0
100
200
300
100
200
300
Price
neighbourhood
Bijlmer−Centrum
Bijlmer−Oost
Bos en Lommer
Buitenveldert − Zuidas
Centrum−Oost
Centrum−West
De Baarsjes − Oud−West
De Pijp − Rivierenbuurt
Gaasperdam − Driemond
Geuzenveld − Slotermeer
IJburg − Zeeburgereiland
Noord−Oost
Noord−West
Oostelijk Havengebied − Indische Buurt
Osdorp
Oud−Noord
Oud−Oost
Slotervaart
Watergraafsmeer
Westerpark
Zuid

Here are the graph presentations of private rooms. The pattern of the relationship between the number of reviews and the prices are similar to the pattern of share rooms. Figure 4 shows the number of reviews in total and Figure 5 shows the number of reviews per month. However, there are a lot more private rooms than shared rooms in Amsterdam, and the price range of private rooms is wider as well. Except one Airbnb's price is 5000 units per night, others' prices are in the range from 0 to 1000 units, and I also neglect the outlier (the room with price at 5000 units) in the plot. From Figure 6, there is no explicit relationship between the availibility and the prices for private rooms.

Figure 4

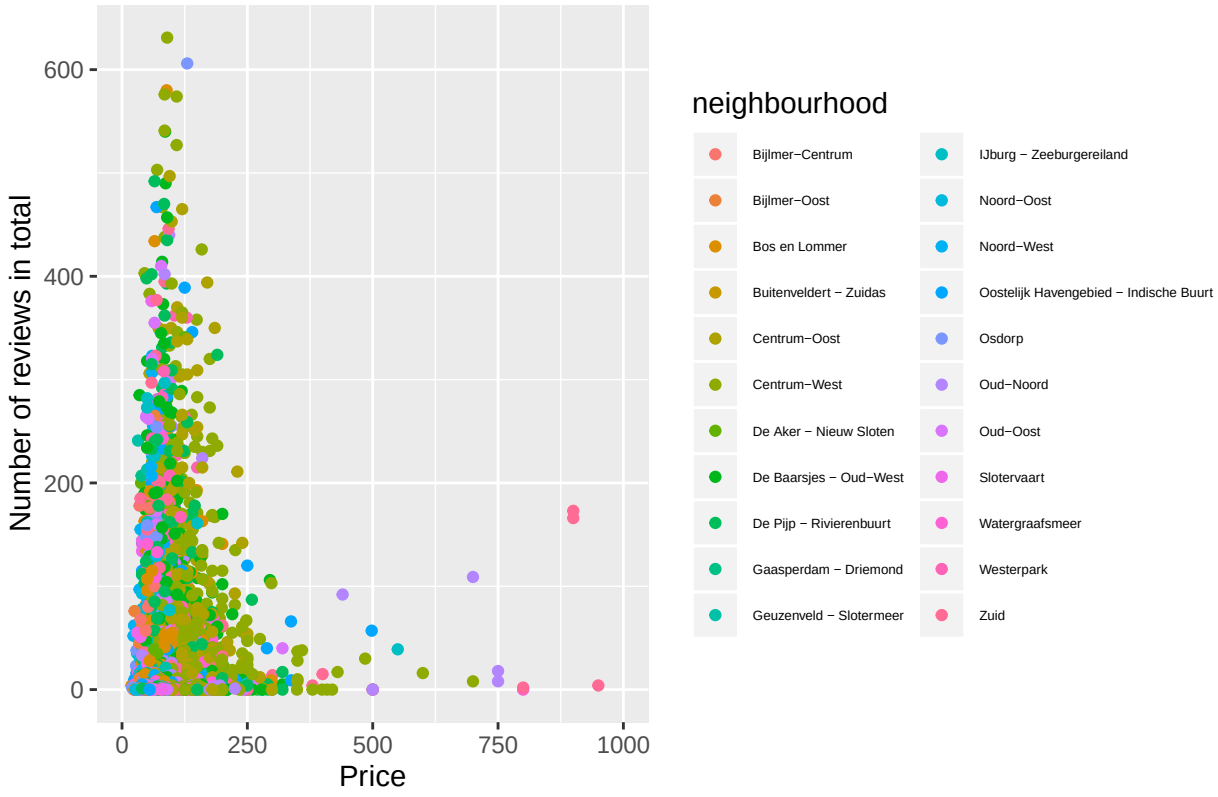


Figure 5

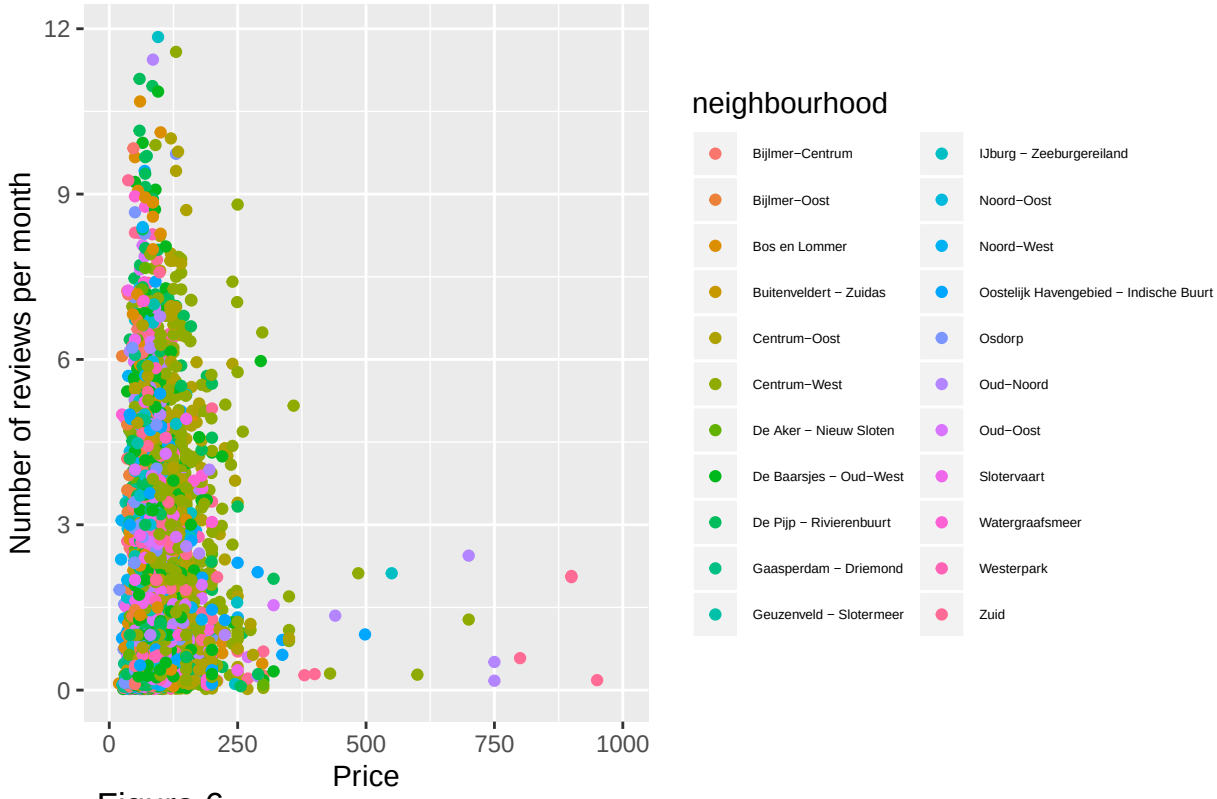
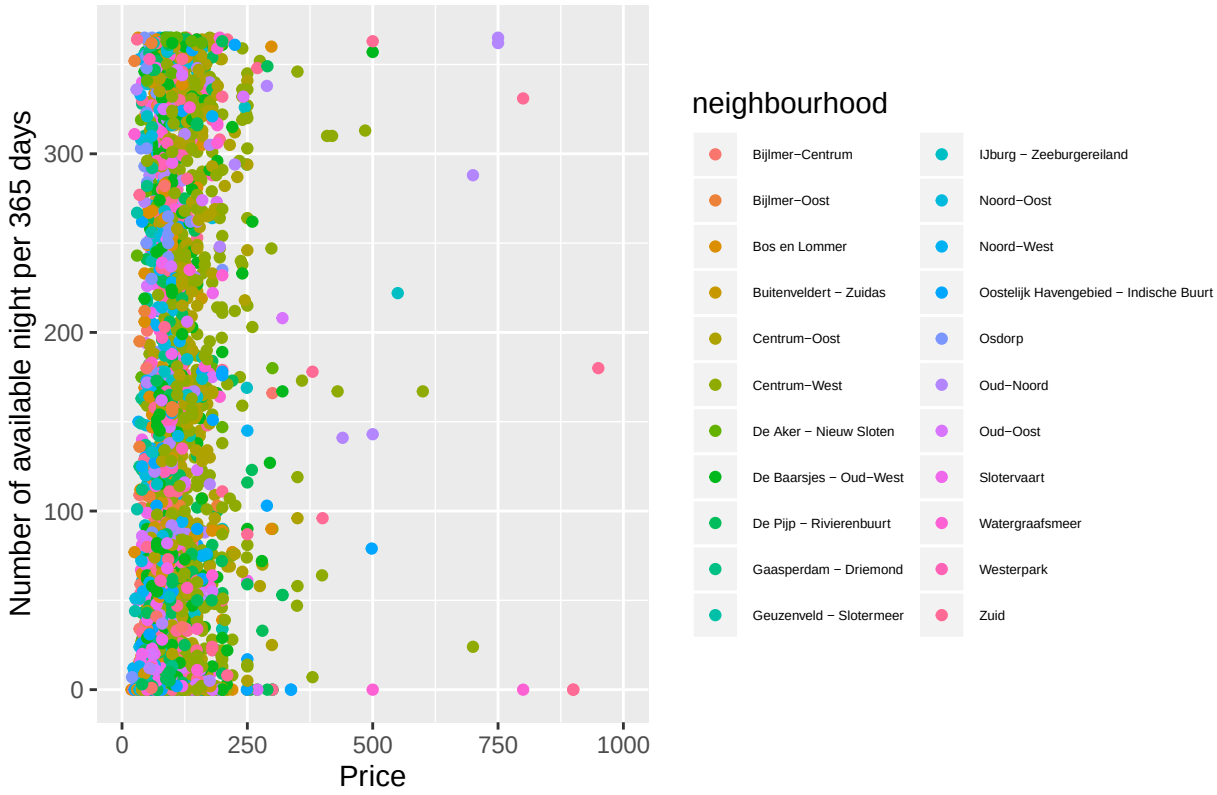



Figure 6



Here are the graph presentations of the entire home or apartments. The price range of entire apartments are much more wider than other room types, but the pattern of relationship between the number of reviews and the prices are very similar in all room types. From Figure 9, there is no explicit relationship between the availability and the prices.

Figure 7

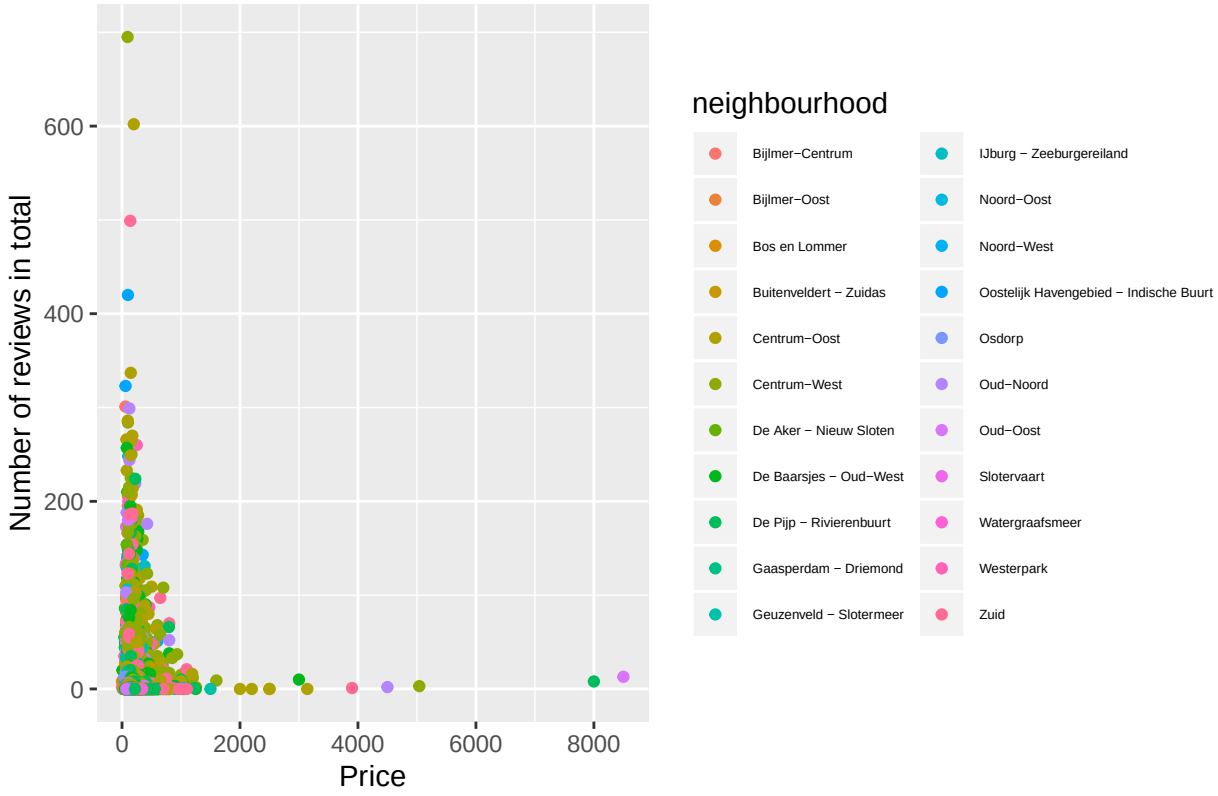


Figure 8

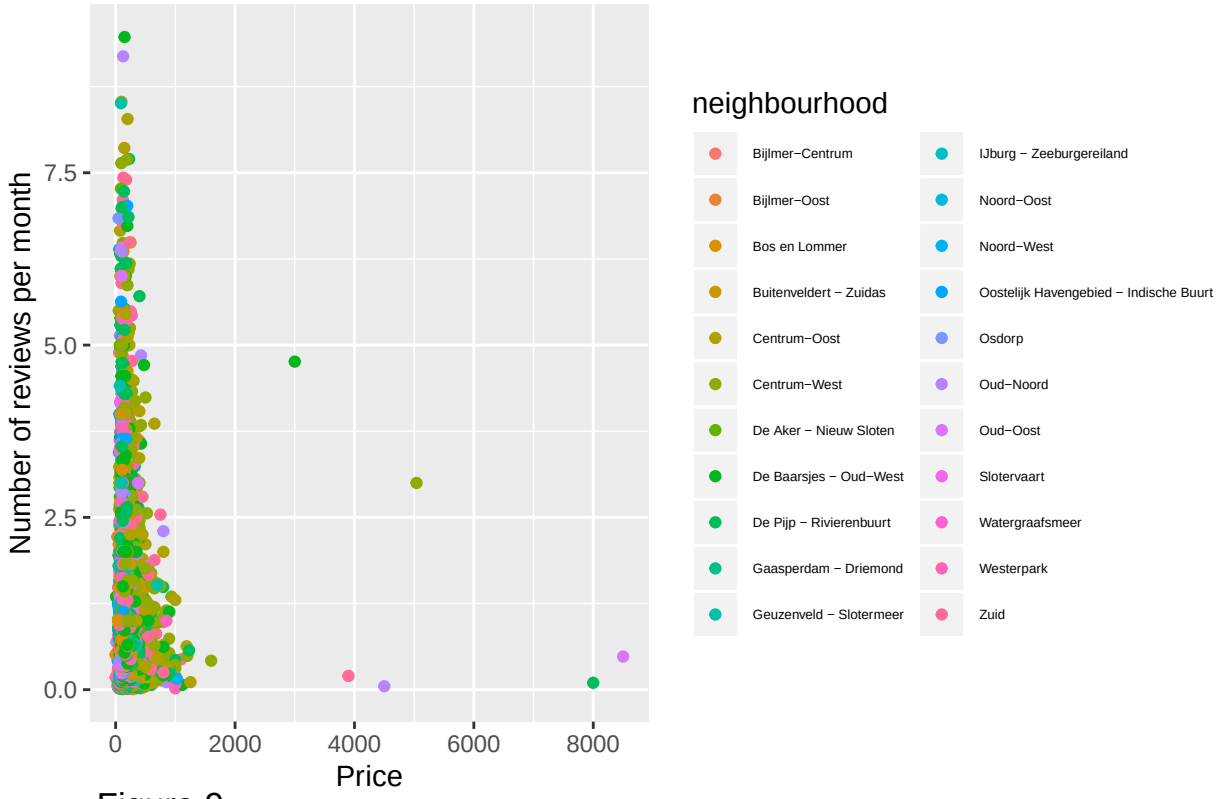
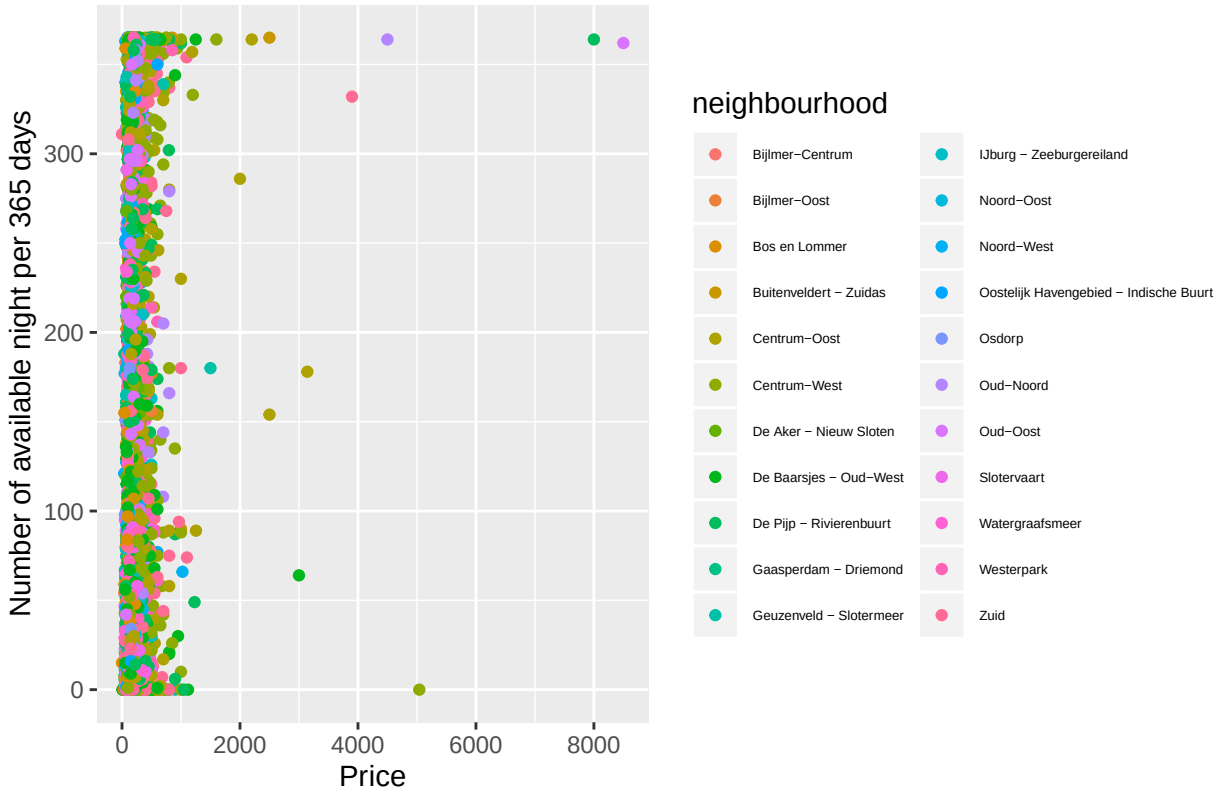


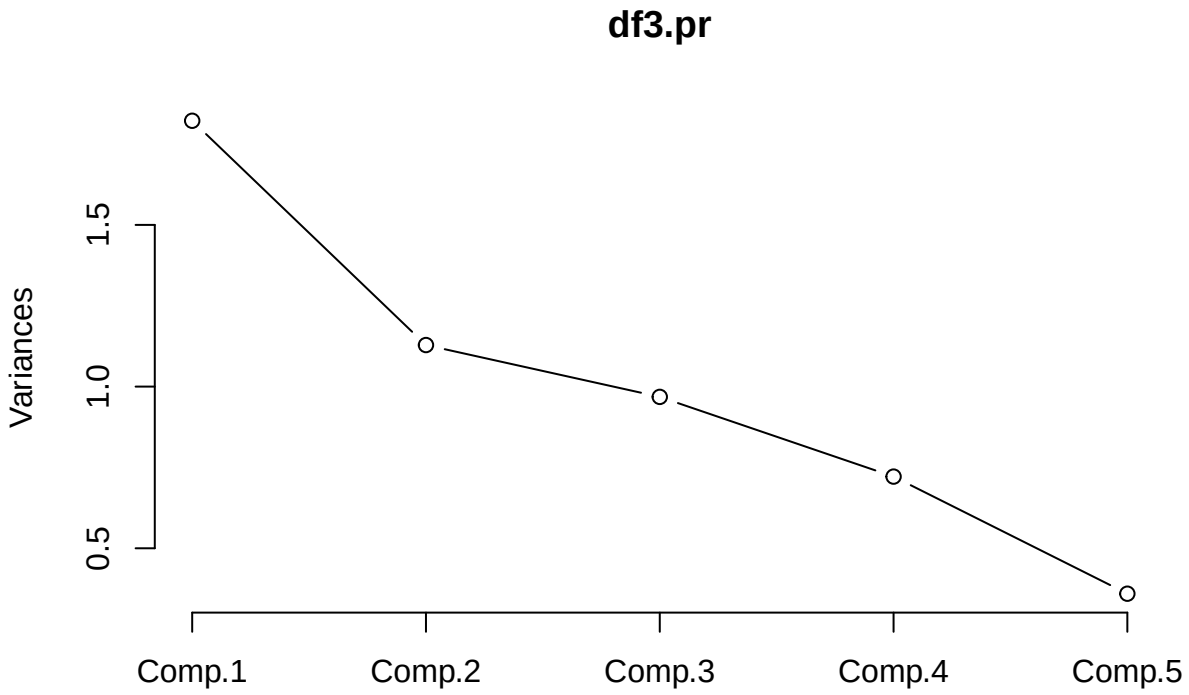

Figure 9



## 3. Principal Component Analysis

I conduct a principal component analysis (PCA) for the numeric columns in the dataset. From the result of the analysis, I can tell that the cumulative proportion becomes 92.8% after the 4th component, and from the scree plot, the changes are not very stable. Therefore, it is not very necessary to reduce the dimension in this project.

| | Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 |
|---|---|---|---|---|---|
| Standard deviation | 1.3498195 | 1.0623095 | 0.9839606 | 0.8495234 | 0.5996810 |
| Proportion of Variance | 0.3644025 | 0.2257003 | 0.1936357 | 0.1443380 | 0.0719235 |
| Cumulative Proportion | 0.3644025 | 0.5901028 | 0.7837385 | 0.9280765 | 1.0000000 |




## 4. Text Analysis

I conduct a text analysis on three Airbnb places from each room type (shared room, entire home/apt, and private room) to obtain that what elements relate the most to each Airbnb and also what people care most about those Airbnb places.

### 4.1. Shared room

From this word cloud, I can easily tell that the host has a boat in this place, and probably the host name is Simon. Most of the comments reflect a positive attitude towards this Airbnb.

## 4.2. Private room

From this word cloud, I believe that Daniel might be the host of this private room and his place is clean and nice, which is also recommended by the reviewers.

### 4.3. Entire homes/ apartments

From the word cloud, I think Alex might be the host of this apartment. His place is clean, nice and comfortable and I believe the apartment's location is most noticable element for people who stay here.


## 5. Conclusion and Discussion

After the analysis, I find out that places with lower prices tend to have more reviews and reviewers write about the elements they care about most in their comments of the Airbnb places. Some certain neighbourhoods have more Airbnb places than others, which also leads to more reviews and comments for those neighbourhoods.