

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

- Is there any relationship between the price and the attitude of reviewers of an Airbnb?
- Is there any relationship between the price and the neighborhood of the Airbnb located?
- Is there any difference in prices regarding different room types of Airbnb?
- Is there any relationship between the price and the availability of an Airbnb?
- 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

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

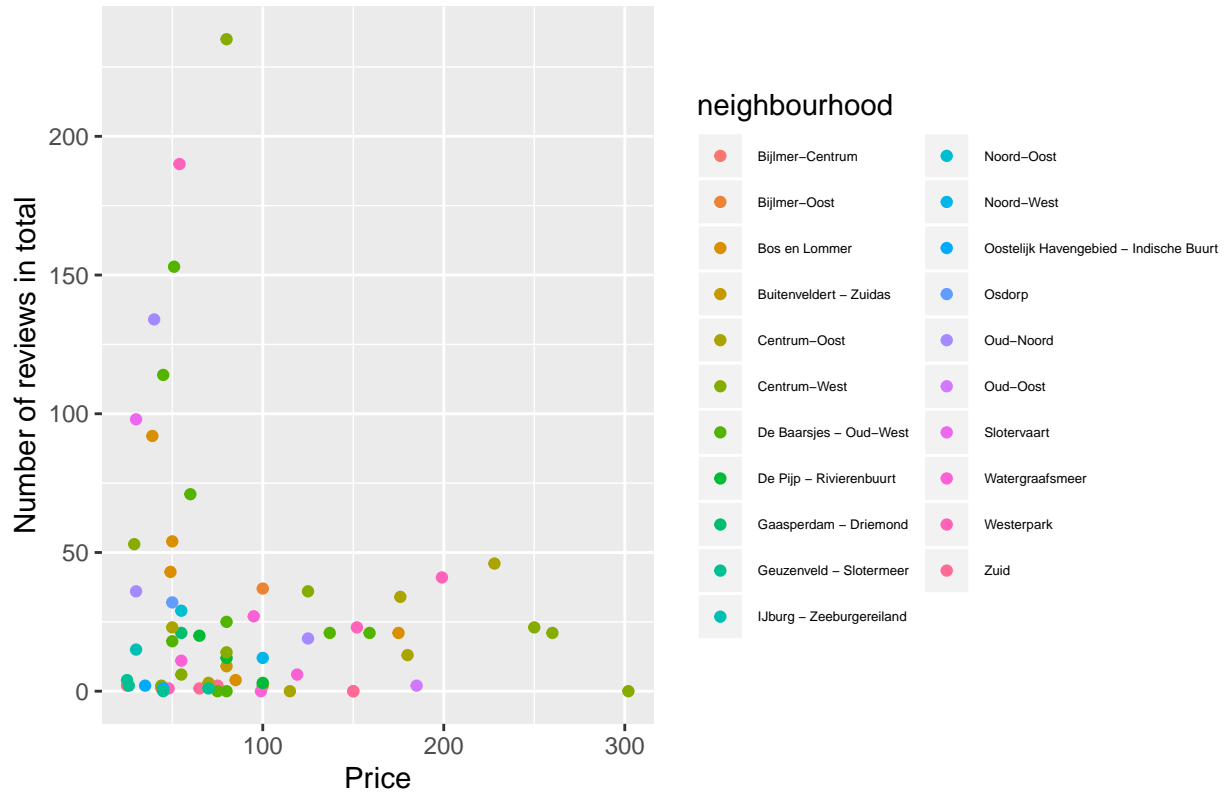
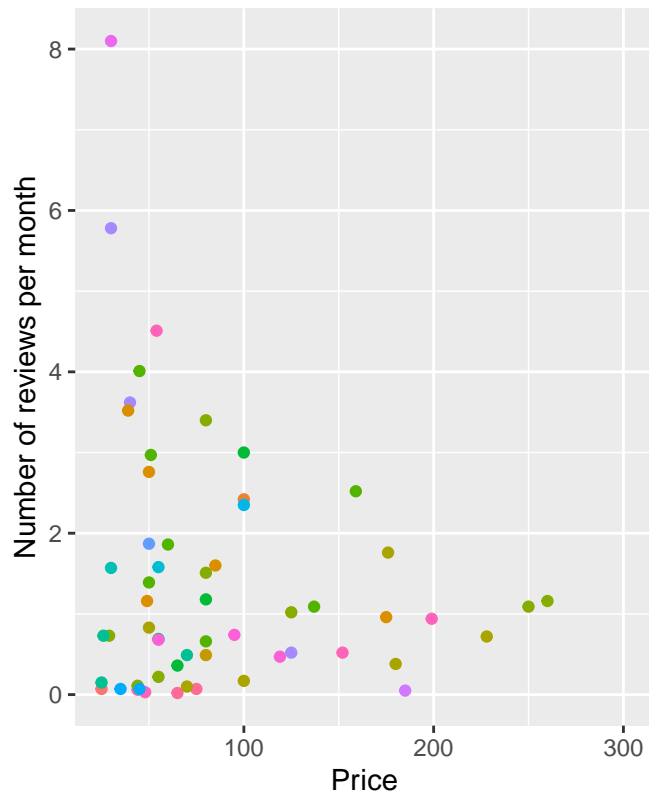


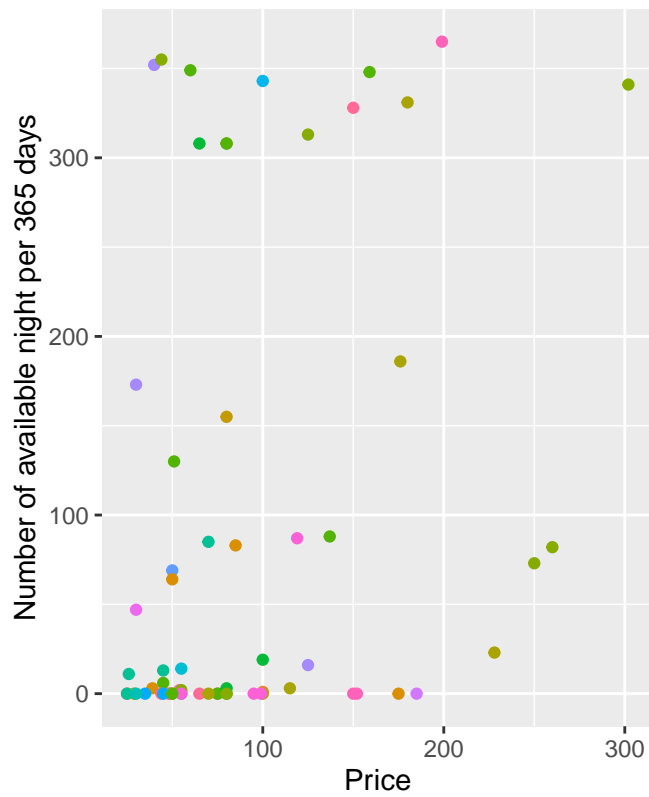
Figure 2



neighbourhood

● Bijlmer-Centrum	● Noord-Oost
● Bijlmer-Oost	● Noord-West
● Bos en Lommer	● Oostelijk Havengebied – Indische Buurt
● Buitenveldert – Zuidas	● Osdorp
● Centrum-Oost	● Oud-Noord
● Centrum-West	● Oud-Oost
● De Baarsjes – Oud-West	● Slotervaart
● De Pijp – Rivierenbuurt	● Watergraafsmeer
● Gaasperdam – Driemond	● Westerpark
● Geuzenveld – Slotermeer	● Zuid
● IJburg – Zeeburgereiland	

Figure 3



neighbourhood

● Bijlmer-Centrum	● Noord-Oost
● Bijlmer-Oost	● Noord-West
● Bos en Lommer	● Oostelijk Havengebied – Indische Buurt
● Buitenveldert – Zuidas	● Osdorp
● Centrum-Oost	● Oud-Noord
● Centrum-West	● Oud-Oost
● De Baarsjes – Oud-West	● Slotervaart
● De Pijp – Rivierenbuurt	● Watergraafsmeer
● Gaasperdam – Driemond	● Westerpark
● Geuzenveld – Slotermeer	● Zuid
● IJburg – Zeeburgereiland	

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 availability and the prices for private rooms.

Figure 4

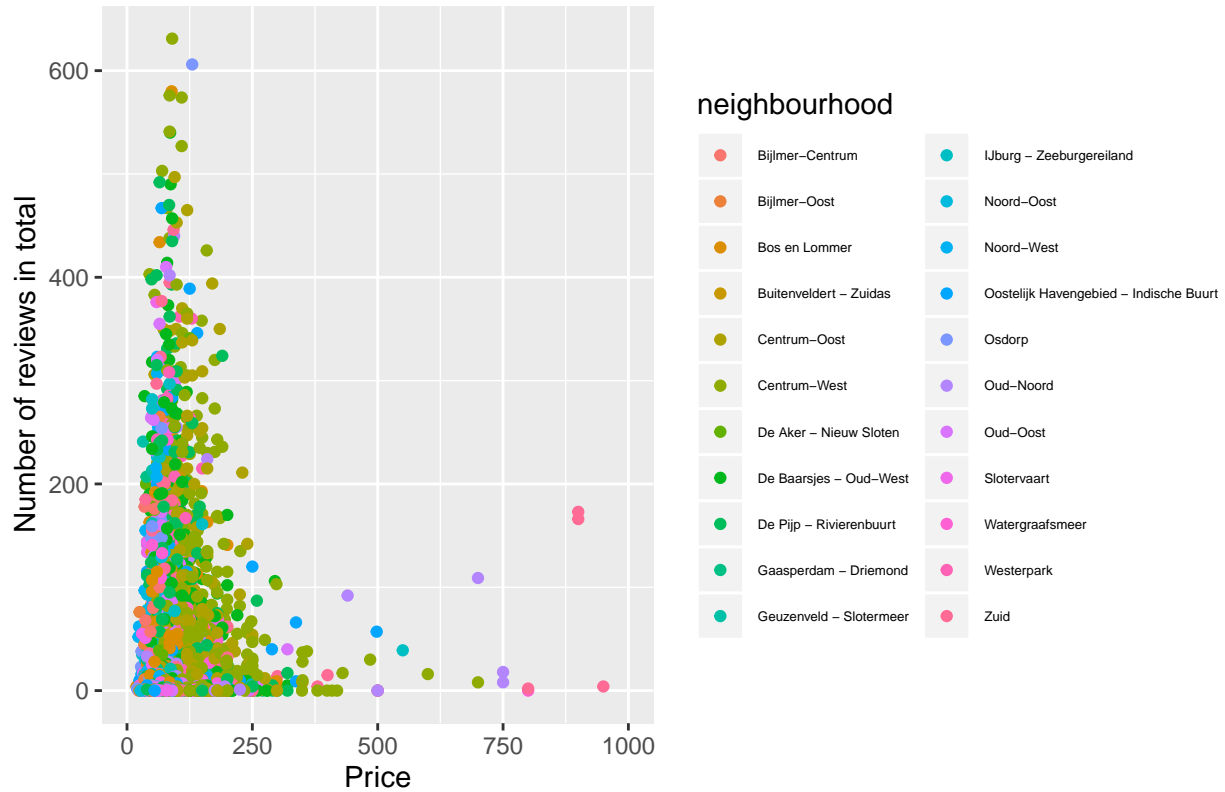
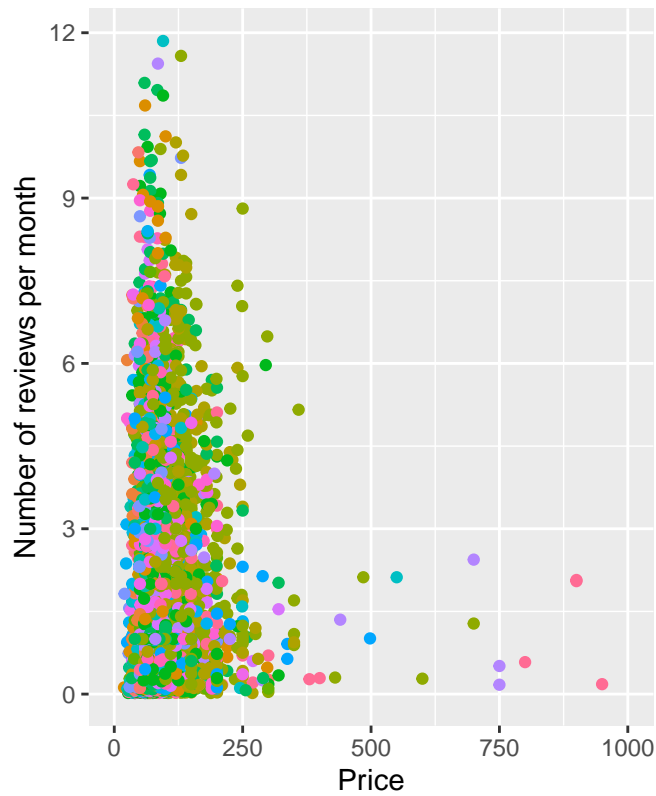


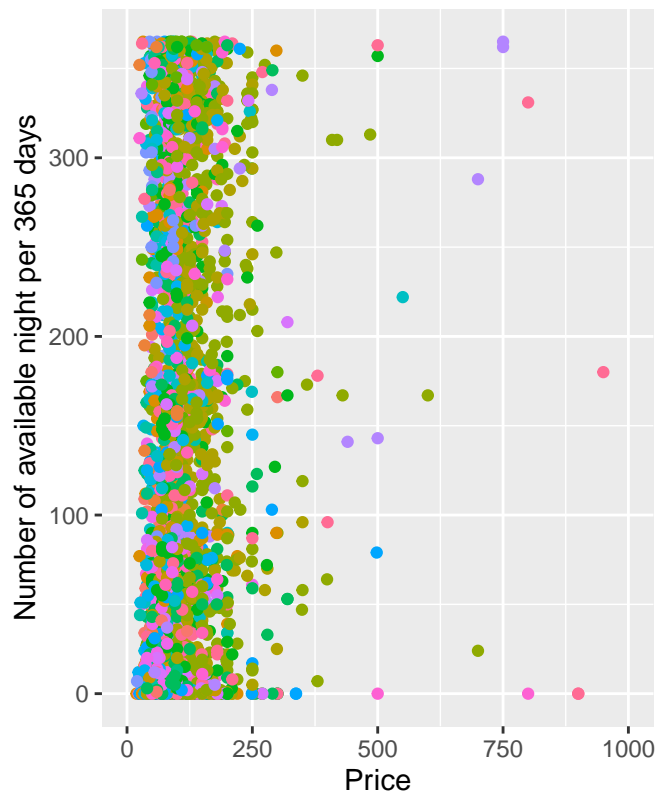
Figure 5



neighbourhood



Figure 6



neighbourhood



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

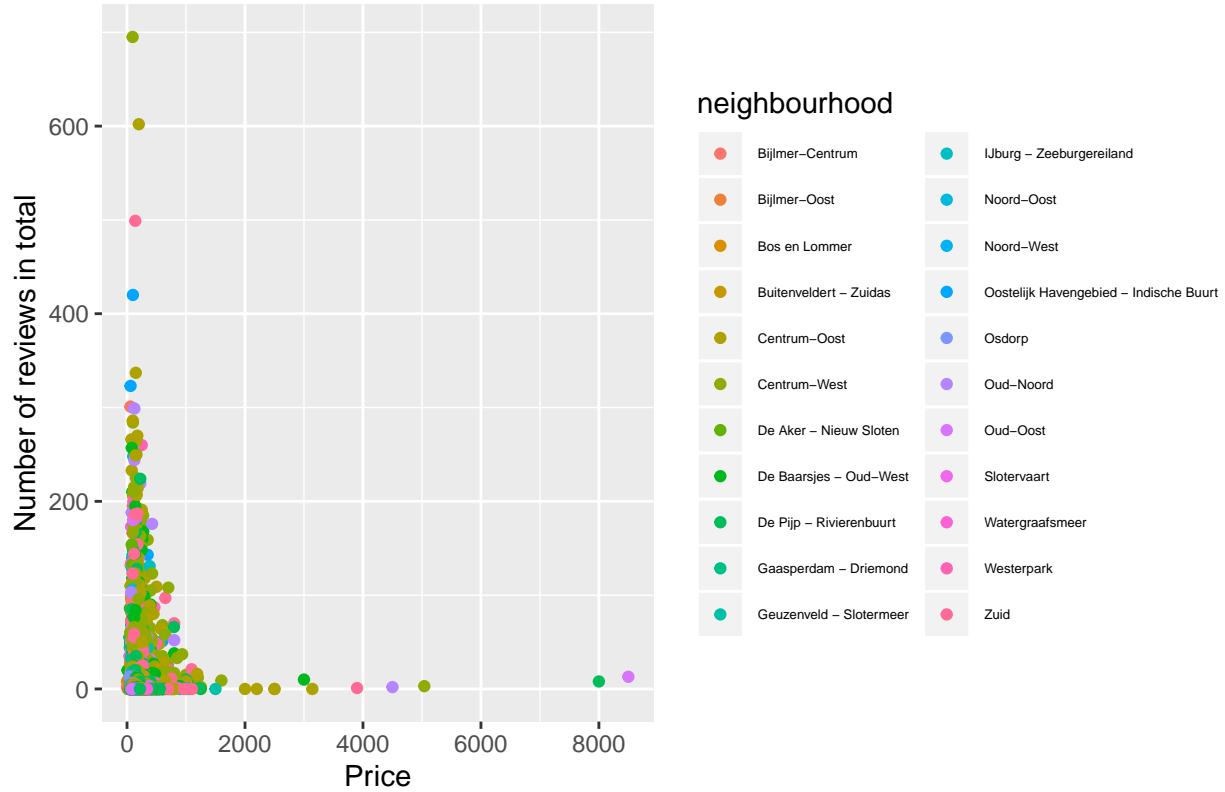
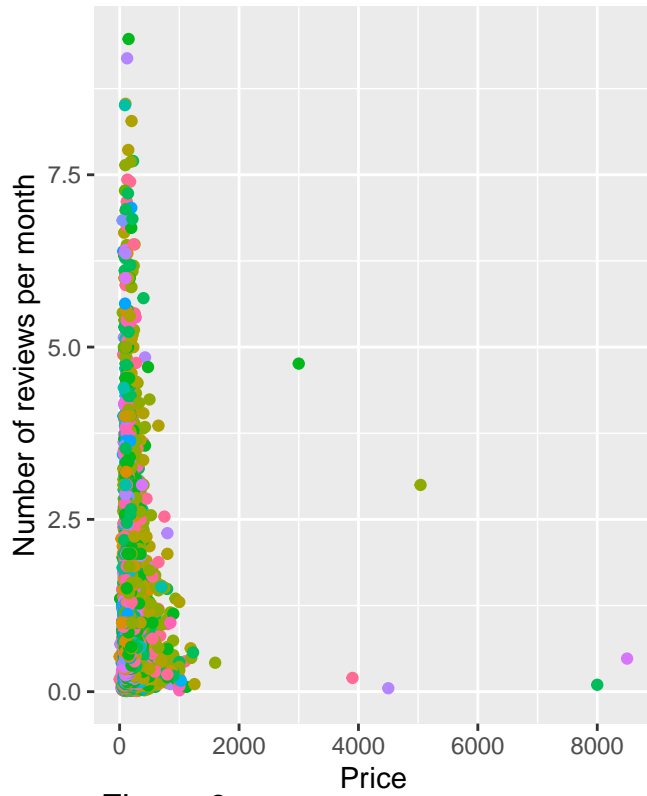


Figure 8



neighbourhood

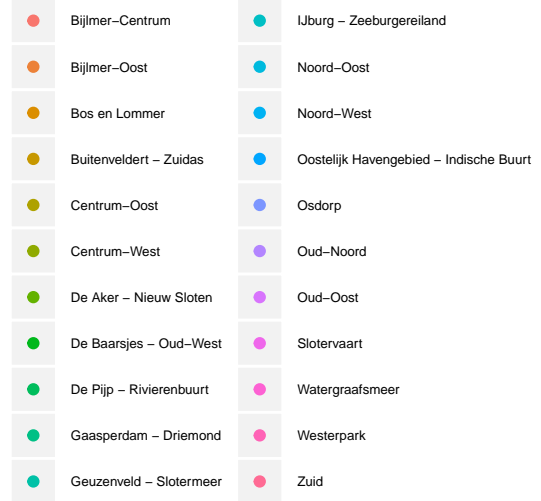
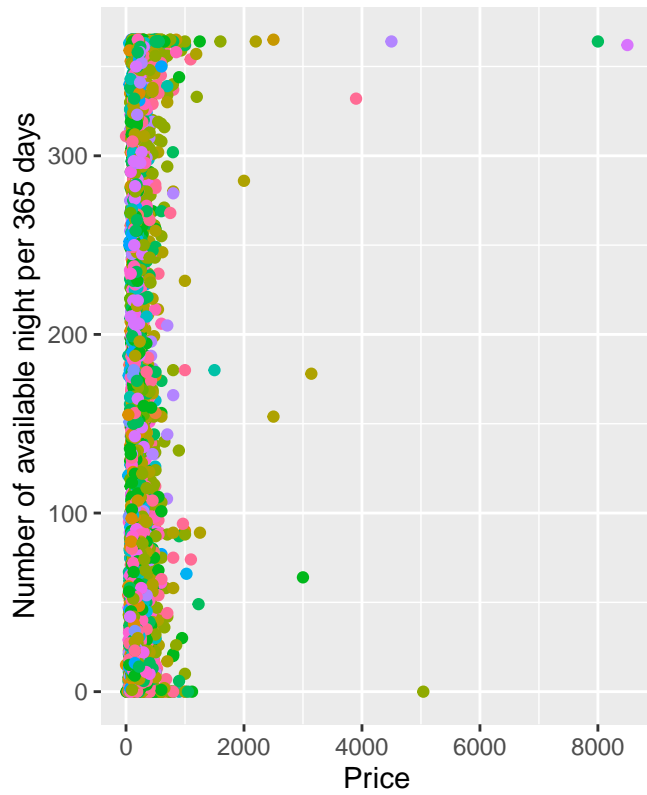


Figure 9



neighbourhood



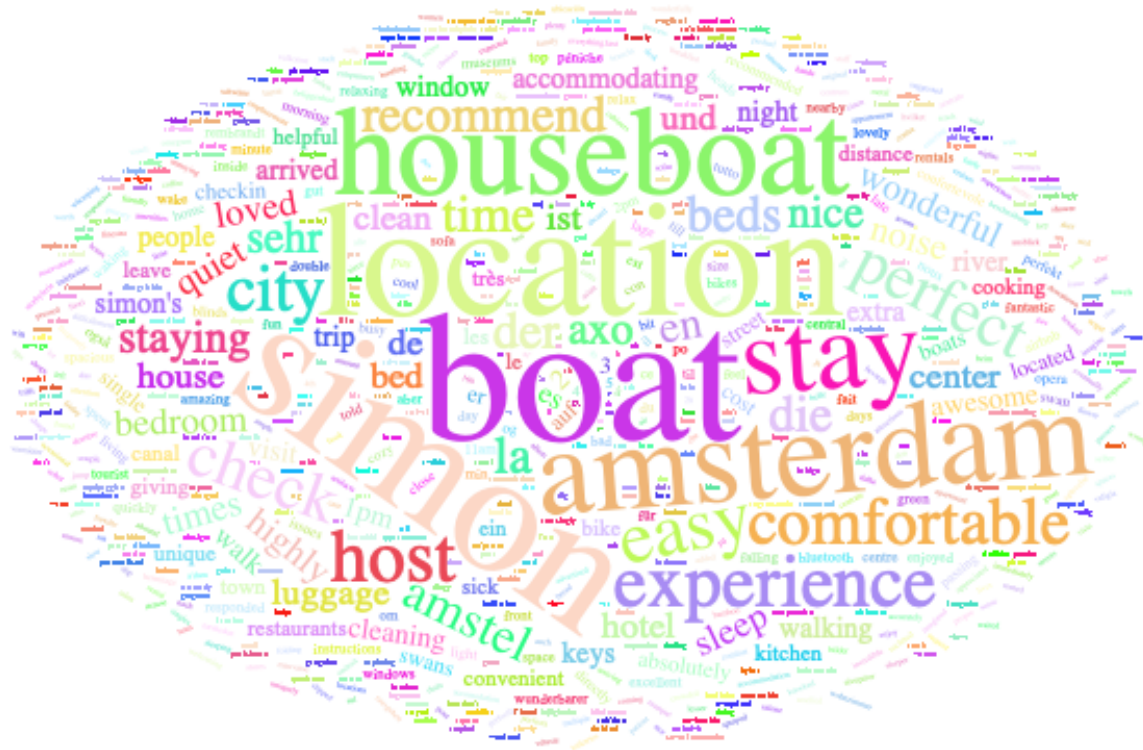
3. Principal Component Analysis

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 noticeable 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.