Airbnb Data Analysis

Project Report

Divyansh Verma
Data Analyst Intern
iNeuron AI

Table of Contents

1. Introduction

Since its inception in 2008, Airbnb has revolutionized the travel industry by providing a platform for hosts and guests to connect, facilitating unique and personalized travel experiences. This analysis focuses on the Airbnb listings in Amsterdam, for the year 2019. The objective is to explore various aspects of the data, including host earnings, neighborhood popularity, price relationships, and customer reviews, to derive meaningful insights and conclusions.

2. Objectives and Research Questions

2.1. Host Analysis

1. **Top Earners:** Identify the hosts who are the top earners on the platform.
2. **Earnings vs. Prices:** Examine the relationship between the monthly earnings of hosts and the prices they set for their listings.

2.2. Neighborhood Analysis

1. **Popular Locations:** Determine which neighborhoods receive the maximum number of bookings.
2. **Price Relation:** Analyze the relationship between prices and their locations within Amsterdam.

2.3. Reviews Analysis

1. **Quality vs. Price:** Investigate the relationship between the quality of the listings (measured by overall satisfaction) and their prices.

2.4. Price Analysis

1. **Price vs. Amenities:** Examine the relationship between listing prices and the number of amenities offered.
2. **Price vs. Location:** Explore how prices vary with different geographical locations within Amsterdam.

3. Data Overview

The dataset includes various features such as host information, geographical availability, pricing, number of reviews, and overall satisfaction. This data allows for a comprehensive analysis to address the research questions.

Overall Data Properties

```
RangeIndex: 18723 entries, 0 to 18722
Data columns (total 21 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   room_id               18723 non-null  int64
 1   survey_id             18723 non-null  int64
 2   host_id               18723 non-null  int64
 3   room_type             18723 non-null  object
 4   country               0 non-null      float64
 5   city                  18723 non-null  object
 6   borough               0 non-null      float64
 7   neighborhood          18723 non-null  object
 8   reviews               18723 non-null  int64
 9   overall_satisfaction  18723 non-null  float64
 10  accommodates          18723 non-null  int64
 11  bedrooms              18723 non-null  float64
 12  bathrooms             0 non-null      float64
 13  price                 18723 non-null  float64
 14  minstay               0 non-null      float64
 15  name                  18671 non-null  object
 16  last_modified         18723 non-null  object
 17  latitude              18723 non-null  float64
 18  longitude             18723 non-null  float64
 19  location              18723 non-null  object
 20  total_earnings        18723 non-null  float64
dtypes: float64(10), int64(5), object(6)
memory usage: 3.0+ MB
```

Value Counts of various Categorical Variables

```
Value counts for column 'host_id':
host_id
48703385     93
113977564    88
1464510      71
107745142    64
84453740     61
             ..
41902443      1
8305721       1
9901234       1
23343555      1
29724632      1
Name: count, Length: 15943, dtype: int64
```

```
Value counts for column 'reviews':
reviews
0      2984
1      1510
2      1246
3      1103
4       925
       ...
242       1
191       1
334       1
309       1
188       1
Name: count, Length: 284, dtype: int64
```

```
Value counts for column 'neighborhood':
neighborhood
De Baarsjes / Oud West                     3289
De Pijp / Rivierenbuurt                    2378
Centrum West                               2225
Centrum Oost                               1730
Westerpark                                 1430
Noord-West / Noord-Midden                  1418
Oud Oost                                   1169
Bos en Lommer                               988
Oostelijk Havengebied / Indische Buurt      921
Watergraafsmeer                             517
Oud Noord                                   494
Ijburg / Eiland Zeeburg                     378
Slotervaart                                 349
Buitenveldert / Zuidas                      250
Noord West                                  241
Noord Oost                                  221
Geuzenveld / Slotermeer                     195
Osdorp                                      163
De Aker / Nieuw Sloten                      114
Bijlmer Centrum                              99
Bijlmer Oost                                 97
Gaasperdam / Driemond                        42
Westpoort                                    15
Name: count, dtype: int64
```

```
Value counts for column 'accommodates':
accommodates
2     10024
4      5579
3      1585
6       476
5       471
1       367
8       105
7        52
16       20
10       16
12       10
9         8
14        6
11        2
13        1
17        1
Name: count, dtype: int64
```

```
Value counts for column 'overall_satisfaction':
overall_satisfaction
5.0    7708
0.0    5748
4.5    4559
4.0     577
3.5     109
3.0      19
1.5       1
2.5       1
1.0       1
Name: count, dtype: int64
```

```
Value counts for column 'bedrooms':
bedrooms
1.0     11101
2.0      4456
3.0      1444
0.0      1154
4.0       473
5.0        62
6.0        19
10.0        5
7.0         4
8.0         3
9.0         2
Name: count, dtype: int64
```

```
Value counts for column 'room_type':
room_type
Entire home/apt    14978
Private room        3682
Shared room           63
Name: count, dtype: int64
```

Data Description (Quantitative Variables)

	reviews	satisfaction	accommodates	bedrooms	price	total_earnings
count	18723	18723	18723	18723	18723	18723
mean	16.74	3.30	2.92	1.43	166.59	2480.49
std	33.52	2.21	1.32	0.87	108.94	4838.25
min	0	0	1	0	12	0
25%	2	0	2	1	108	236
50%	6	4.5	2	1	144	936
75%	17	5	4	2	192	2640
max	532	5	17	10	6000	107280

4. Exploratory Data Analysis (EDA)

4.1. Data Summary

- **Hosts:** The dataset includes multiple host-related variables such as host ID, price per night, and number of reviews.
- **Neighborhoods:** Information on different neighborhoods and their respective booking counts and average prices.
- **Reviews:** Data on the number of reviews and overall satisfaction ratings.
- **Prices:** Detailed price data along with related amenities, number of bedrooms, and geographical coordinates.

4.2. Unique Values and Data Statistics

- **Unique Values:** Various columns in the dataset were examined to identify the number of unique values, helping to understand the diversity within the data.
- **Summary Statistics:** Basic statistical measures such as mean, median, and standard deviation were computed for key numerical columns.

4.3. Data Cleaning and Preparation

- **Derived Metrics:** New columns were created, such as `total_customers` (calculated as `reviews * accommodates`) and `total_earnings` (calculated as `price * reviews`), to aid in deeper analysis.

5. Detailed Analysis

5.1. Host Analysis

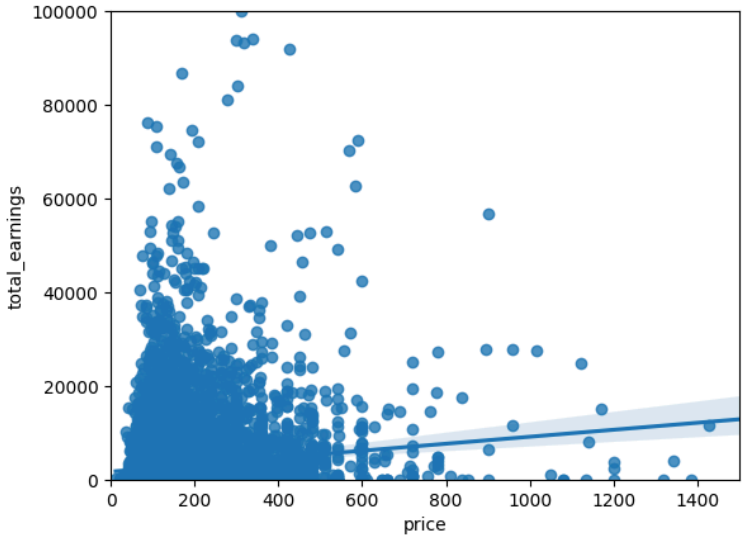
Top Earners

The total earnings for each host were calculated and the top 10 earners were identified. These top earners were visualized to highlight the distribution of earnings among the leading hosts.

	host_id	total_earnings
1292	2674028	162837.0
146	336950	144483.0
646	1464510	121562.0
4651	8558897	118175.0
57	187580	117066.0
1250	2586026	107280.0
415	935723	105712.0
80	225987	101956.0
1047	2234051	100152.0
572	1347048	94064.0

Earnings vs. Prices

A correlation analysis was conducted to examine the relationship between the average price set by hosts and their total earnings. The results showed a positive correlation, indicating that higher prices generally lead to higher total earnings. This relationship was further visualized using a regression plot.

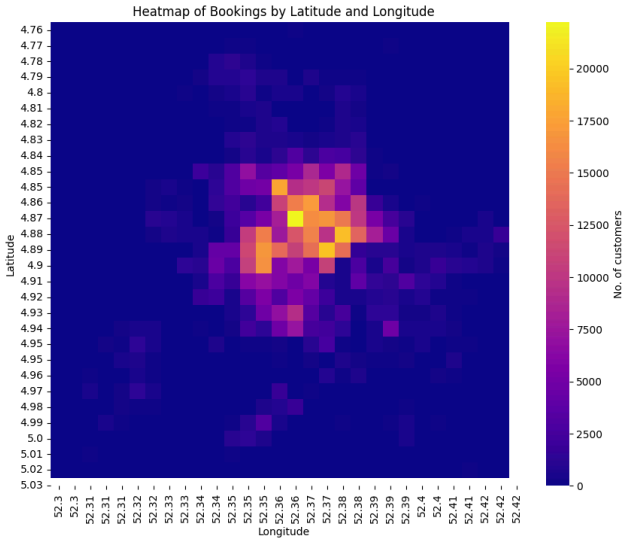


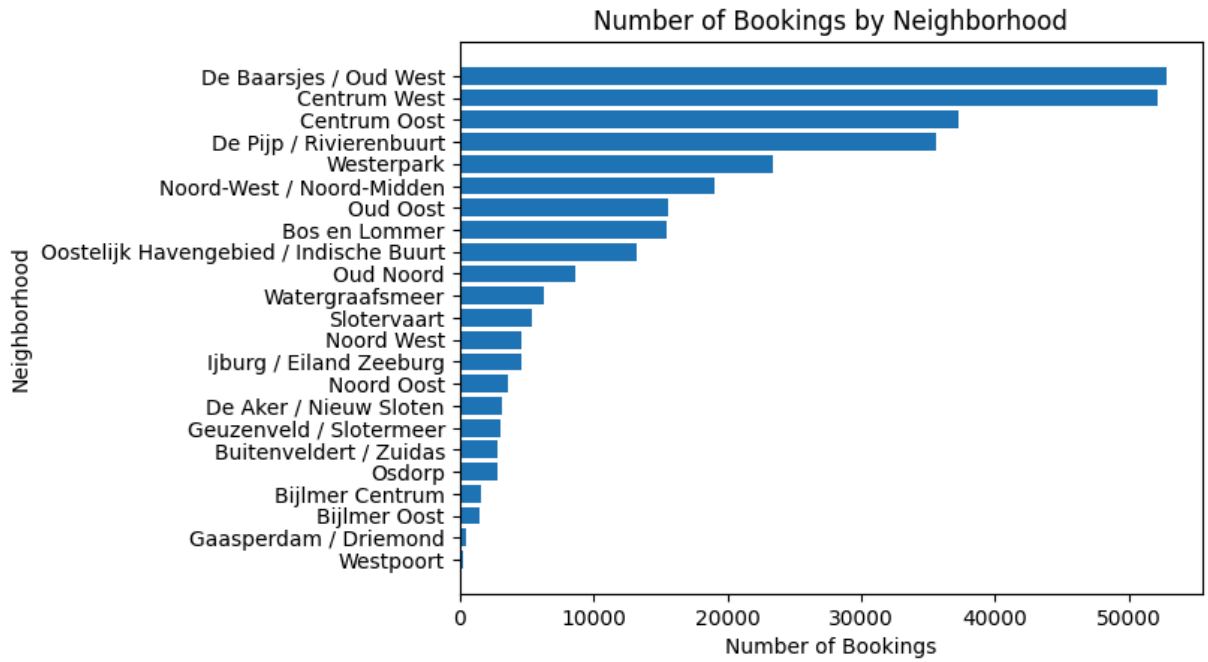
5.2. Neighborhood Analysis

Popular Locations

The number of bookings for each neighborhood was calculated based on the review count. The neighborhoods with the highest number of bookings were identified and visualized.

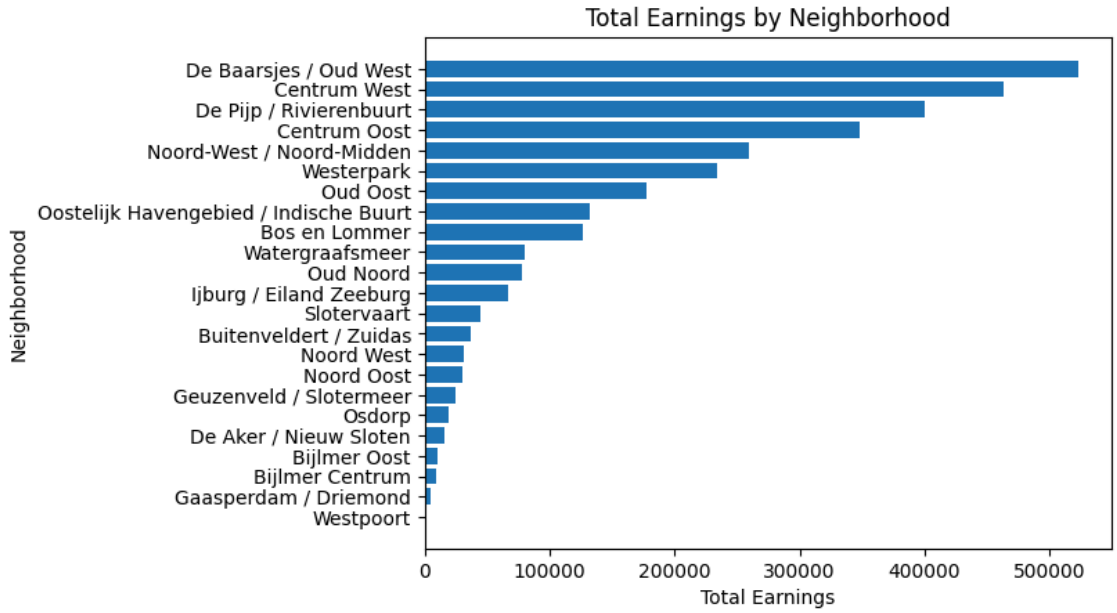
neighborhood	no_of_bookings
De Baarsjes / Oud West	3289
De Pijp / Rivierenbuurt	2378
Centrum West	2225
Centrum Oost	1730
Westerpark	1430
Noord-West / Noord-Midden	1418
Oud Oost	1169
Bos en Lommer	988
Oostelijk Havengebied / Indische Buurt	921
Watergraafsmeer	517
Oud Noord	494
Ijburg / Eiland Zeeburg	378
Slotervaart	349
Buitenveldert / Zuidas	250
Noord West	241
Noord Oost	221
Geuzenveld / Slotermeer	195
Osdorp	163
De Aker / Nieuw Sloten	114
Bijlmer Centrum	99
Bijlmer Oost	97
Gaasperdam / Driemond	42
Westpoort	15





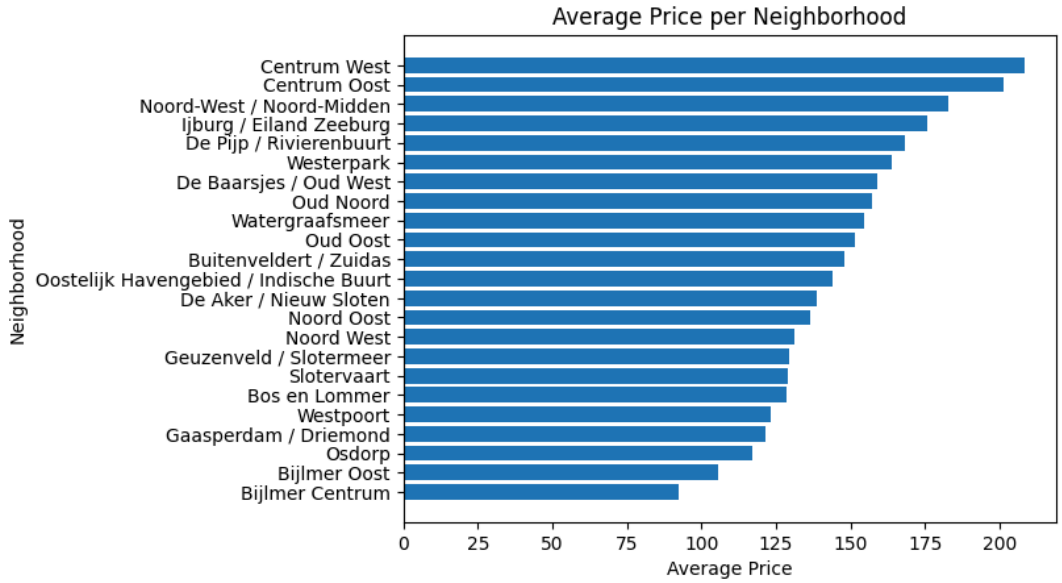
Top Earning Neighborhoods

The top neighborhoods by total earnings were determined and plotted to show which areas are the most lucrative for hosts.



Price Analysis per Neighborhood

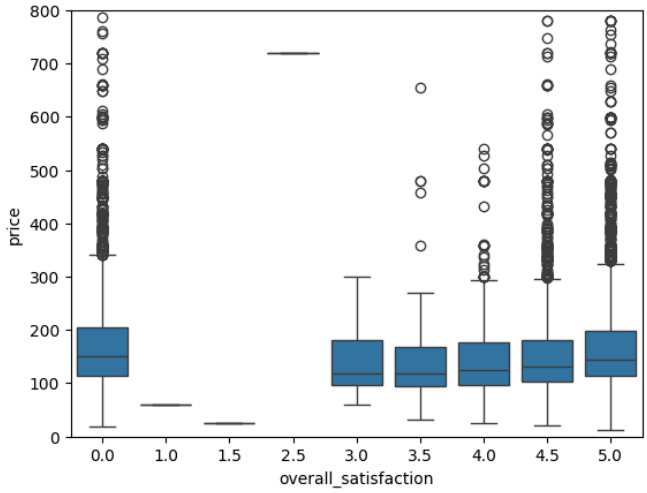
The average price per night for each neighborhood was calculated and visualized. This helped in understanding the price distribution across different neighborhoods.



5.3. Reviews Analysis

Quality vs. Price

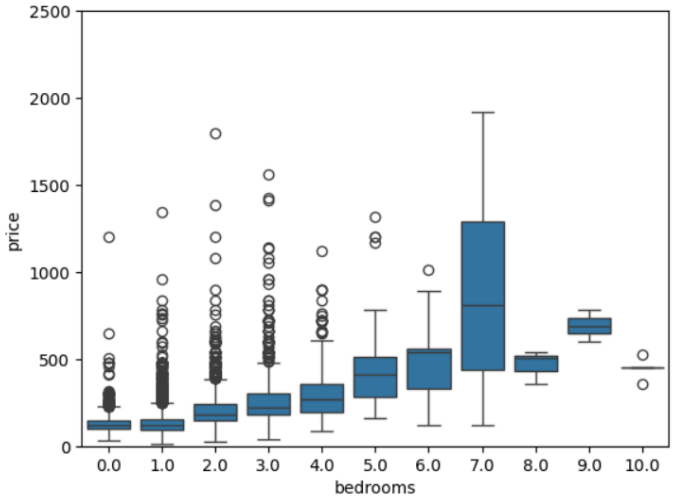
The relationship between overall satisfaction (a proxy for quality) and price was examined. The correlation analysis showed a positive but weak relationship, indicating that higher prices do not necessarily guarantee higher satisfaction. This was visualized using a box plot.



5.4. Price Analysis

Price vs. Bedrooms

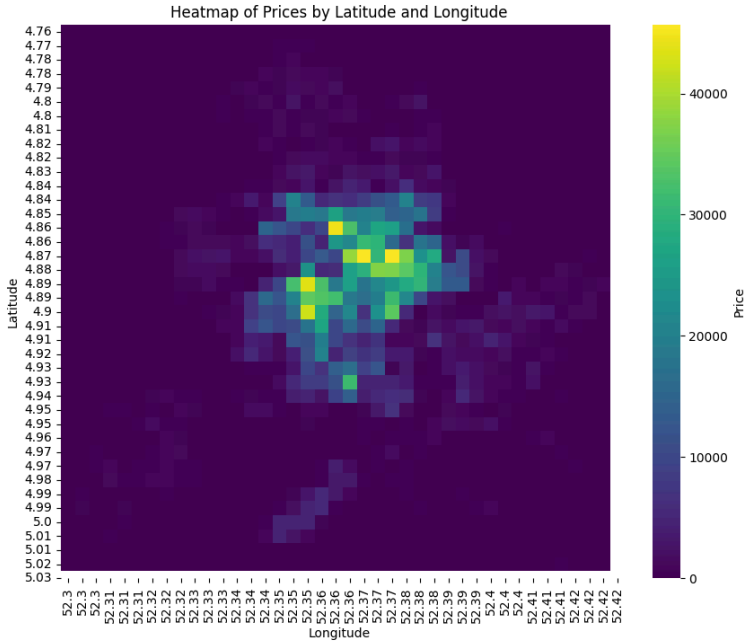
The distribution of prices with respect to the number of bedrooms was analyzed. A correlation analysis showed a positive relationship, indicating that listings with more bedrooms tend to have higher prices. This was further supported by a box plot.

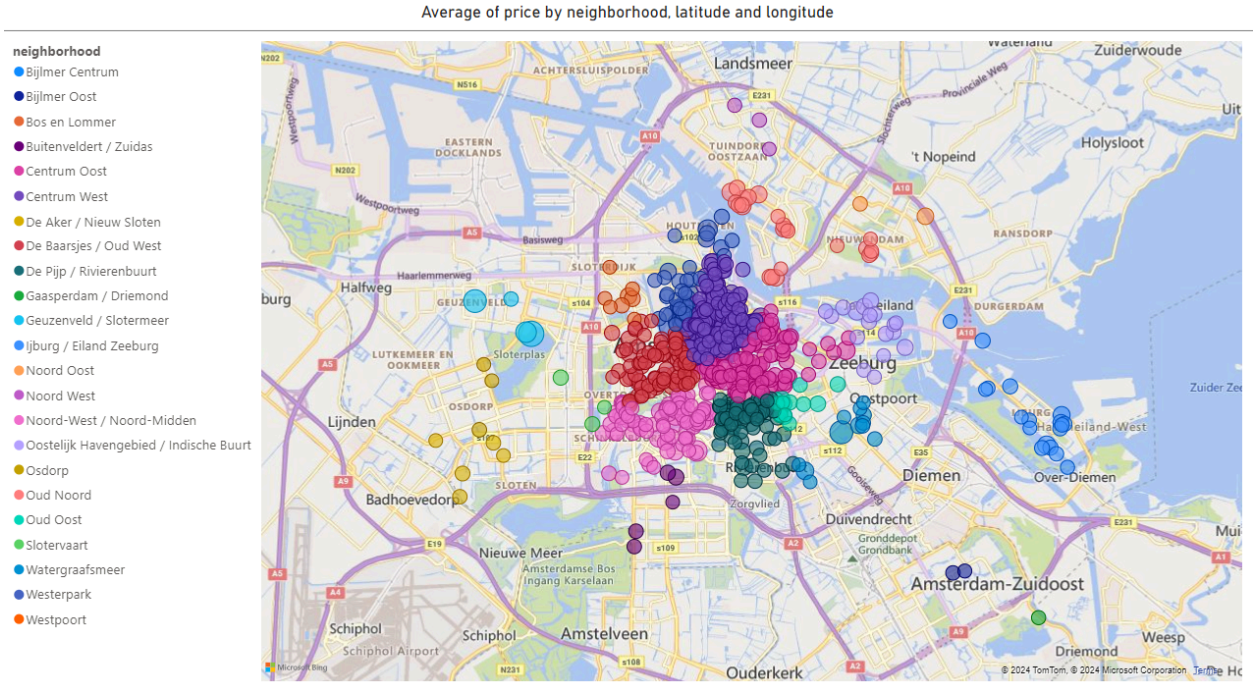



```
The correlation between number of bedrooms and price is 0.4461436487654141
```

Price vs. Location

A heatmap was created to visualize the distribution of prices across different geographical locations in Amsterdam. This highlighted areas with higher and lower prices.





Insights from the analysis

- **Host Analysis**
    - Top Earners: The host with the ID 2674028 is the highest earner, with total earnings amounting to $162,837.
    - Earnings vs. Prices: There is a weak positive correlation (0.124) between earnings and prices. This indicates that prices have a minimal impact on the total earnings of Airbnb hosts.
- **Neighborhood Analysis**
    - Popular Locations: The most popular area for booking Airbnbs in Amsterdam is De Baarsjes/Oud West. This popularity is reflected in its earnings, making it the highest-earning neighborhood in the city.
    - Price Relation: The correlation between the popularity of a location and its prices is moderate. This can be observed by comparing the number of bookings per neighborhood with the average prices in those neighborhoods.
- **Reviews Analysis**
    - Quality vs. Price: When considering missing data as zeroes, it is evident from the box plot that the average price of rooms with a 5-star rating is higher than those with lower ratings.
- **Price Analysis**
    - Price vs. Amenities: Up to seven bedrooms, there is a trend showing that more bedrooms generally lead to higher prices.

6. Conclusion

In summary, the analysis reveals that the top-earning Airbnb host, with an ID of 2674028, has earned $162,837. Despite a faint correlation (0.124) between earnings and prices, indicating minimal price impact on total earnings, neighborhood popularity significantly influences earnings, with De Baarsjes/Oud West being the highest-earning area in Amsterdam.

A moderate correlation exists between neighborhood popularity and prices. Reviews show that rooms with 5-star ratings tend to be priced higher on average. Additionally, prices generally increase with the number of bedrooms up to seven, suggesting that more amenities lead to higher rates.