

# How to Become a Top Earner on Airbnb ?

Anushka Pandey  
PES1UG20CS072

Computer Science & Engineering PES  
University Bangalore, India  
anushkapandey1105@gmail.com

Devika S  
PES1UG20CS123

Computer Science & Engineering PES  
University Bangalore, India  
reachdevikas@gmail.com

**Abstract**— The primary goal of this project is to analyze the Airbnb data statistics of a popular seaport city – Seattle, in the US. It uses exploratory data analysis and data visualization to bring out the patterns within the numerous attributes of the data set. This project also comprises of a model which can predict the rent in a particular neighborhood and also the availability of rooms.

Analysis of such a data set is necessary so as to gain insight of the current demand of the tourists, so that we can implement such a setup in our country as well.

## I. INTRODUCTION

Tourism is one of the most important sources of income for all the nations. It is observed that few nations like Maldives, Bali, Singapore, Nepal etc. rely only on tourism for their economic growth. With tourism, comes the responsibility of making the guests comfortable and giving them high standards of facilities, with the rooms and hotels being the most important aspect of their trip. Airbnb Inc. is an American company which is famous for providing lodgings for vacation stays and other tourist activities.

It provides rental options for different customer segments. Based on their budget, people can rent an entire house or just a room by just a click of a button. Apart from that users can also enter their reviews about their stay at the place and other facilities like cleanliness, location, communication etc. A major question answered in this project is – ‘how is Airbnb affecting the neighborhoods?’ . By the help of our project, we can predict ‘when are the prices high and low’, ‘which neighborhoods are best preferred by the tourists’, ‘relation between average tourist volumes and months of the year’ etc. We have also shown relationships between different attributes of the dataset by the aid of pairplots, histograms, heatmaps, correlation matrix etc.

## II. REVIEW OF LITERATURE

### A. The rise of exponential organization

Analyzing [3], we realize that the digitalization of today’s society has made it possible for organizations to reach beyond the traditional markets. Exponential markets grow 10 times faster than their traditional counterparts and perform better and faster. Ismail et al.(2014)[4] describe the success of these organizations extensively, notable among these are - they have a compelling higher purpose and they dare to experiment, manage algorithms and have empowered autonomous workers. Airbnb, Uber, Gitgap are some of the renowned “exponential organizations”.

### B. The rise of Sharing or Collaborative Economy

Parties like Airbnb and Uber identify themselves as sharing economies wherein the consumers grant each other short term access to underutilized physical assets like rooms and apartments in case of Airbnb and cars in case of Uber. Transfer of ownership is not done in such scenarios. This way it benefits both the temporary customer and the owner. The advent of internet has also made this economy flourish to large extents.

### C. Spike in demand for rental homes

The surge in tourism has led to an increased demand for rental homes and accommodations. Thus, the market for rental homes has widened. Airbnb has captured a huge portion of this market. That is why, this paper aims at finding the methods to become a top earner on the Airbnb platform and maximize revenue and help customers make right decisions when choosing their rooms/apartments. This will also enable the hosts to make smart business decisions.

### D. Using Multiple Linear Regression

From [5], we see that Multiple Linear Regression has been used to predict the price of houses. Significant factors/features that influence the response variable are selected and then the model is built to predict the price. Similarly, we plan on predicting the price of rental rooms after selecting significant attributes like neighborhood, rating, type of room, etc.

## III. DATASET

The chosen dataset, ‘Seattle Airbnb Open Data’ has been procured from [1].

<https://www.kaggle.com/datasets/airbnb/seattle>

This dataset has three csv files – calendar.csv, listings.csv and reviews.csv. We see that there are 92 variables and 3818 observations for listing. Some of the data was collected by scraping Airbnb websites. Airbnb is a \$75 Billion online marketplace for renting out homes/villas/apartments/private rooms. The website charges a commission (3 to 20 percent) for every booking. We can infer quite some details not only about Airbnb’s business and their hosts, but also about Seattle. A lot of insights can be drawn from this dataset that can help improve the host’s business. This will help us answer questions like which neighborhood is popular among tourists and identify neighborhoods to invest in (to maximize the estimated revenue). We also believe we can get a rough idea about the relationship between Airbnb as a listing service and the hosts. We will be able to predict what factors (like cleanliness, amenities provided, surroundings) do customers consider most important while looking for a

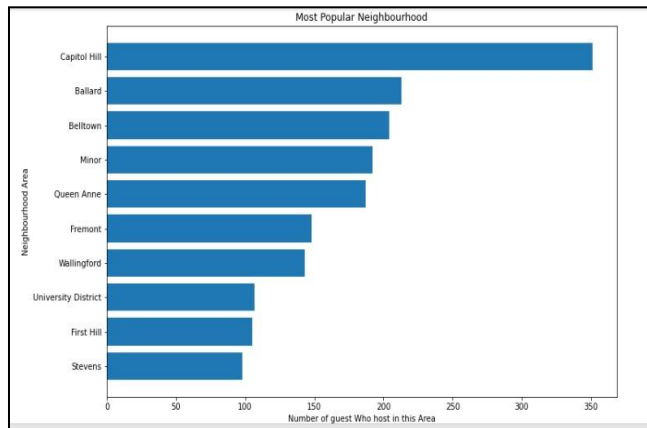
place to rent. This can be used to make informed business decisions.

#### IV. INITIAL INSIGHTS

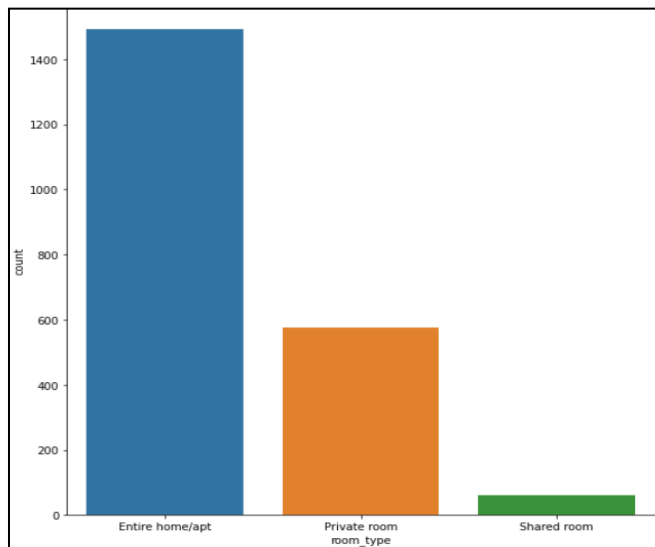
After observing the dataset and having cleaned the dataset, we now perform Exploratory Data Analysis(EDA), with the intention of gaining insights from this process. Null values have been replaced by appropriate metrics.

id	name	description	host_id	host_url	host_name	host_since	host_location	host_about	host_response_time
6	241032	Stylish Queen Anne Apartment one be	956883.0	https://www.airbnb.com/users/show/956883	Maja	2011-05-11	Seattle, Washington, United States	I am an artist, interior designer, and host a b...	within a few hours
1	953095	Charming Queen Anne Apartment	5177328.0	https://www.airbnb.com/users/show/5177328	Andrea	2013-02-21	Seattle, Washington, United States	Living east coast coasters coasters...	within an hour
2	3308979	New Modern Home built in 2013 Amazing view	16708087.0	https://www.airbnb.com/users/show/16708087	Jill	2014-06-12	Seattle, Washington, United States	I love being in Seattle, I grew up in the city	within a few hours
3	7421966	Queen Anne Charm	9801441.0	https://www.airbnb.com/users/show/9801441	Emily	2013-11-06	Seattle, Washington, United States	NaN	NaN

A glimpse of our dataset after the required cleaning and preprocessing



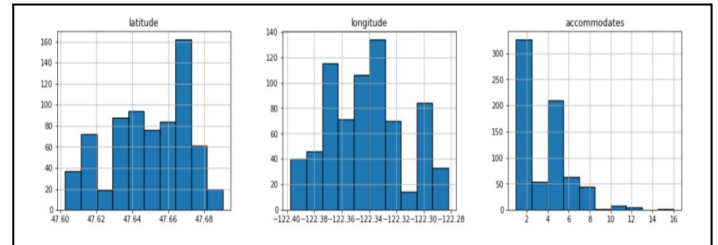
Using the library Matplotlib, we have generated the above plot, which illustrates the popularity of the various neighborhoods in Seattle. Capitol Hill appears to be the most popular neighborhood.



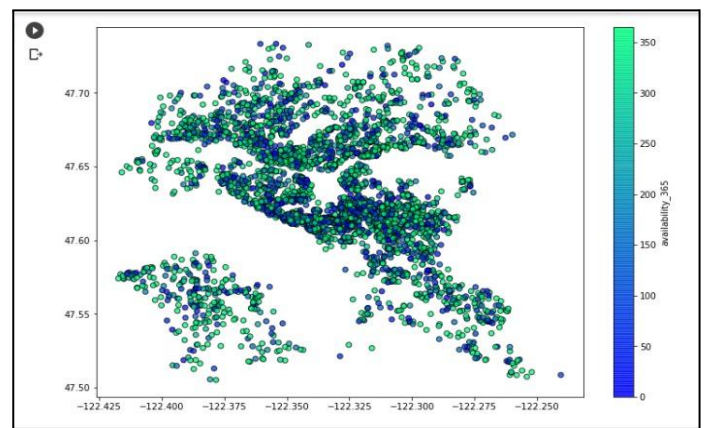
The above categorical plot was created using the Seaborn function catplot. We infer from this plot that most of the listings are of entire homes/apartments, followed by private rooms.

amenities	count
3 Wireless Internet	3667.0
6 Heating	3627.0
5 Kitchen	3423.0
12 Smoke Detector	3281.0
17 Essentials	3237.0
9 Dryer	2997.0
8 Washer	2992.0
2 Internet	2811.0

The above table shows the number of amenities provided by the hosts all over Seattle.



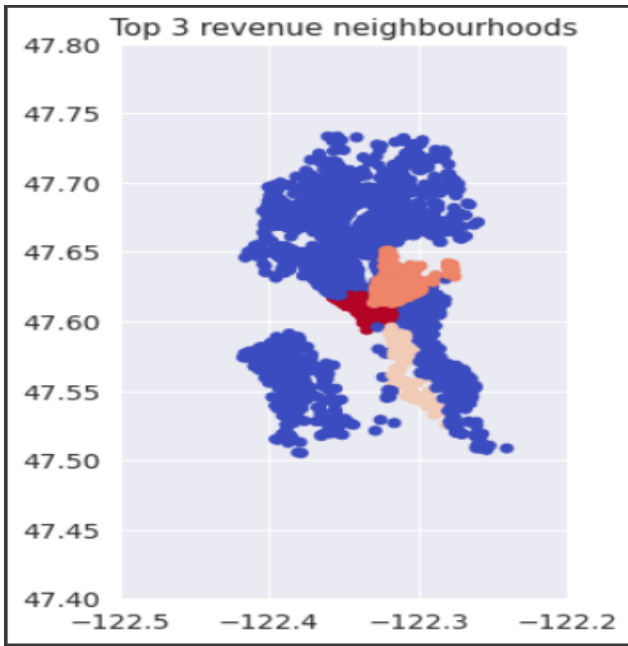
Histogram of a few attributes like latitude and longitude



The above scatter plot is plotted between latitude and longitude, taking availability as the hue. We can see distinct clusters. It shows the availability of rooms around the year in different neighborhoods, based on their locations.

neighbourhood_group_cleansed	estimated_revenue
Downtown	7247.666038
Capitol Hill	7064.079365
Beacon Hill	6606.983051
Ballard	6078.726087
Queen Anne	6038.752542
Central Area	4636.371274
Other neighborhoods	4409.843829
Cascade	4075.134831
Seward Park	4063.500000
Rainier Valley	3827.345912
Delridge	3641.189873
Magnolia	3587.819672
West Seattle	3370.783251
Northgate	2962.362500
Lake City	2476.432836
Interbay	2105.727273
University District	1558.557377

The above table shows the overall estimated revenue of listings in each neighborhood.

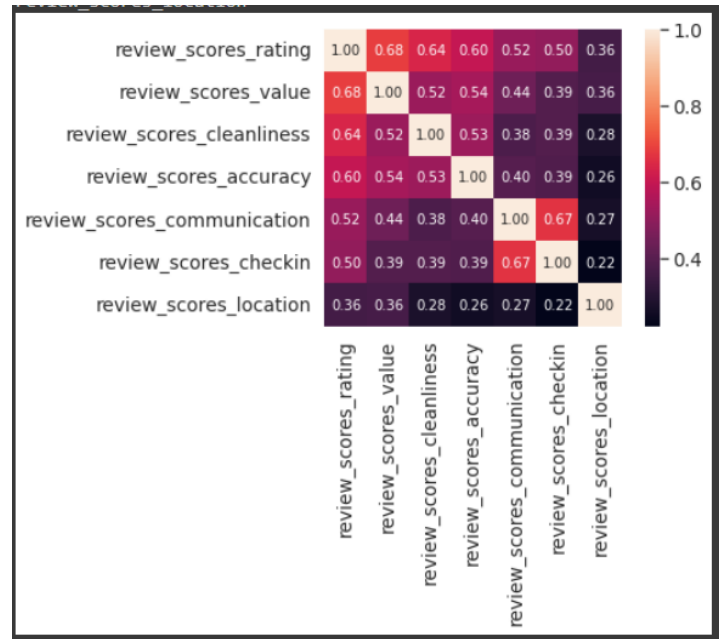


So the inference that can be made from the above table and scatter plot is that the neighborhoods of Downtown, Capitol Hill and Beacon Hill have the highest revenue.

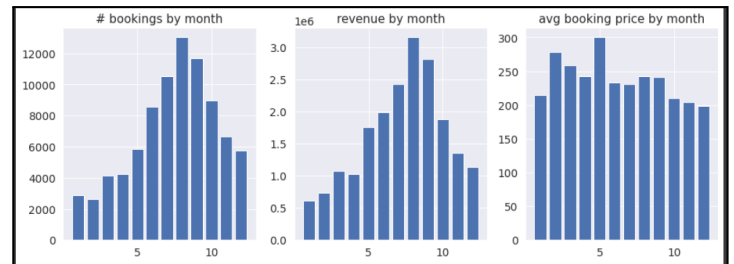


From the above correlation matrix, we can infer that -

1. The overall rating (described by the attribute `review_scores_rating`) has a very small positive correlation with the price. This means that overall rating does not change on changing the price set by the host. This implies that an inexpensive room can have a high overall rating and an expensive room can have a low rating. An expensive room can also have a high rating and an inexpensive room can have a low rating.
2. The overall rating (described by the attribute `review_scores_rating`) has a very small positive correlation of 0.02 with the estimated revenue. This implies that having a high overall rating does not guarantee a high estimated revenue. The overall rating column had several missing values. This could possibly be explained by the fact that most guests prefer not to rate the rooms and services.



1. The above correlation matrix has been used to analyze the correlation between overall ratings and other rating parameters like cleanliness, location, communication and others. We can see that overall ratings have a strong positive correlation with cleanliness and accuracy.



The above plots describe the number of bookings by month, revenue per month and the average booking price per month. The number of bookings per month and revenue per month are slightly skewed to the left. From these plots, we can infer that July, August and September are the best months to maximize revenue as most of the rooms are booked during these months.

## V. PROBLEM STATEMENT AND APPROACH

The problem statement is “How to become a top earner on the platform Airbnb?” We intend to compare top performer's acts ( the ones that have more bookings and great listing's rating ) to the low performers and figure out the methods the latter can opt to improve their performance. This can also be used by hosts to increase services provided by them to create “premium” features. We also seek answers to questions like which the most popular neighborhood is, whether the ratings of a listing are affected by the price and see if we can increase the estimated revenue.

#### ACKNOWLEDGMENT

We would like to express our profound gratitude to Dr. Gowri Srinivasa and the entire Data Analytics team, for encouraging and providing us with this opportunity to get hands-on experience in the field, and guiding us along the way. We would also like to thank the Computer Science and Engineering department at PES University, for always inspiring us to conduct frequent research and inculcating a problem-solving discipline in us.

#### REFERENCES

- [1] <https://www.kaggle.com/datasets/airbnb/seattle>
- [2] <https://towardsdatascience.com>
- [3] [https://www.researchgate.net/publication/298305479\\_Airbnb\\_The\\_future\\_of\\_networked\\_hospitality\\_businesses](https://www.researchgate.net/publication/298305479_Airbnb_The_future_of_networked_hospitality_businesses)
- [4] Ismail, S., Malone, M., van Geest, Y. and Diamandis, P. (2014), Exponential Organizations: Why New Organizations are Ten Times Better, Faster and Cheaper than Yours (and What to Do About It), Diversion Books, New York, NY
- [5] <https://doi.org/10.1155/2021/7678931>