**Rental Analysis of Airbnb Seattle**

**Manali Nikam**

**7th October,2019**

# **1.Introduction**

## **1.1Background**

Seattle is one of the biggest business and innovation hubs in the country, attracting heavy traffic from tourists and professionals alike. The cities draw people from all walks of life ranging from computer scientists to business owners to startup specialists to tourist groups to college freshmen. Airbnb senses an opportunity to improve their rental programs in these cities and would like to hear your suggestions on how to do so.

## **1.2 Problem:**

The Data that we are working on includes a list of rentals across various locations in Seattle. It gives us information about the number of amenities provided, the number of persons that can be accommodated, room type, property type, the number of reviews that the listing has received.

In this project we will try to find optimal locations for rentals in Seattle neighborhoods listed on Airbnb. We will use Data Science to find which area is most preferred rentals in Seattle neighborhood by the people which can be tourists/professionals.

# **2. Data acquisition and cleaning**

## **2.1 Data sources**

The datasets are sourced from <https://www.kaggle.com/airbnb/seattle> and <https://goo.gl/jcHuwG.>**The data present on the google drive is of year 2016.**

## **2.2 Data cleaning**

Data had NA values and it was inconsistent. There were several columns which were not required or the analysis. Mainly columns which were related to score and reviews were kept.

Columns which gave rough idea about the listing like the location, neighborhood, number of amenities and the price of the rental were kept as it is. The column cancellation policy was converted into dummy as cancellation policy is one of the factors looked upon while finalizing a rental.

The price related fields were all of string type. They were converted to numeric and the $ sign was also removed.NA was filled with 0.

Host which have their host location as Seattle were considered as local hosts.

The count of ‘House’ and ‘Apartment’ types were of high number. The other types such as ‘Tent’,’Logt’,’Treehouse’,’Dorm’;’Yurt’,’Townhouse’,’Camper’,’Bungalow’were significantly less. They were categorized as Other. Bed type column is not a real deciding factor.

Various fields like host\_is\_superhost, host\_has\_profile\_pic , host\_identity\_verified,

instant bookable were converted from categorical to numerical values.

Various calculation like listing duration, price\_per\_person, hosting duration was calculated by taking the difference between relevant columns.

# **3. Data Analysis:**

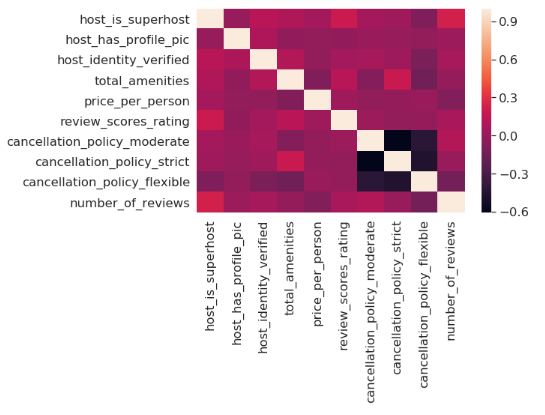
# I used heatmap in order to check the features that are the most relevant ones in deciding the rental. After trial and error, I found out below parameters which were further used to predict the accuracy of the model

host\_is\_superhost','host\_has\_profile\_pic','host\_identity\_verified',

'total\_amenities','price\_per\_person',

'cancellation\_policy\_moderate',

'cancellation\_policy\_strict','cancellation\_policy\_flexible','number\_of\_reviews'



I further used Naïve Bayes to check the accuracy of the model and check if these features are important for getting the number of reviews. The accuracy score was 0.94.

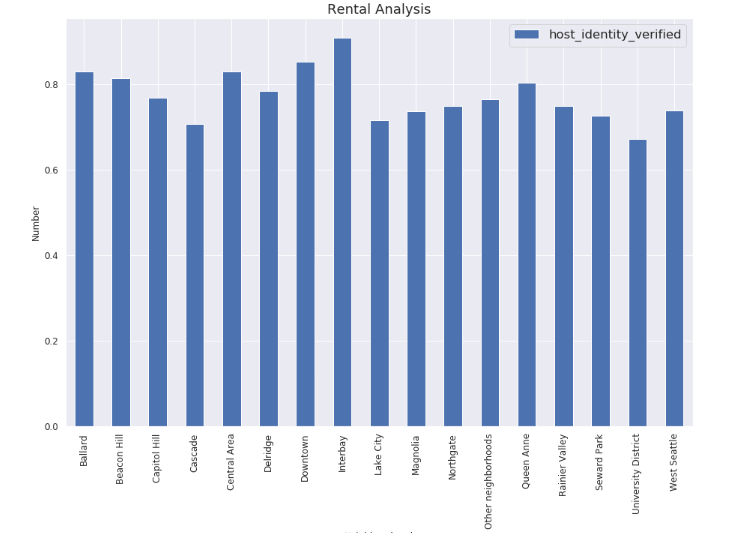
The Classification report is as below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F-1 Score | Support |
| Micro avg | 0.95 | 0.95 | 0.95 | 958 |
| Macro avg | 0.61 | 0.64 | 0.62 | 958 |
| Weighted avg | 0.95 | 0.95 | 0.94 | 958 |

## **3.1 Exploratory analysis:**

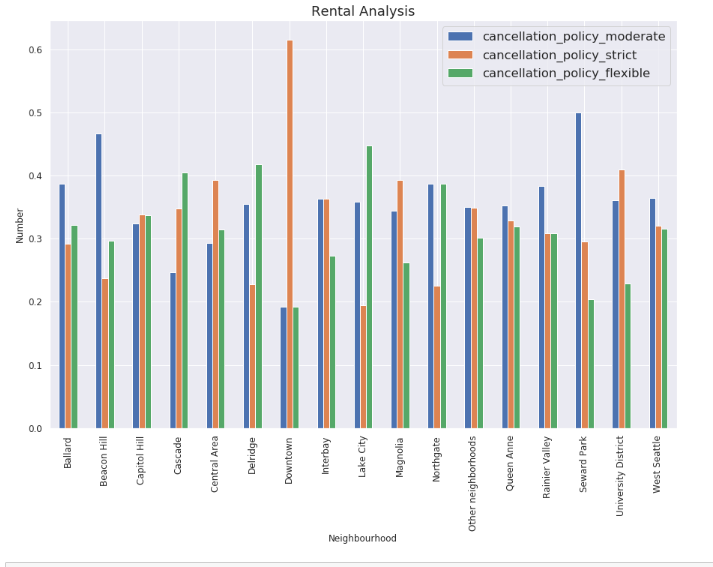
I checked the relationship between the various factors (independent variables) and number of reviews (dependent Variable)

**1.Host Identity Verified:**



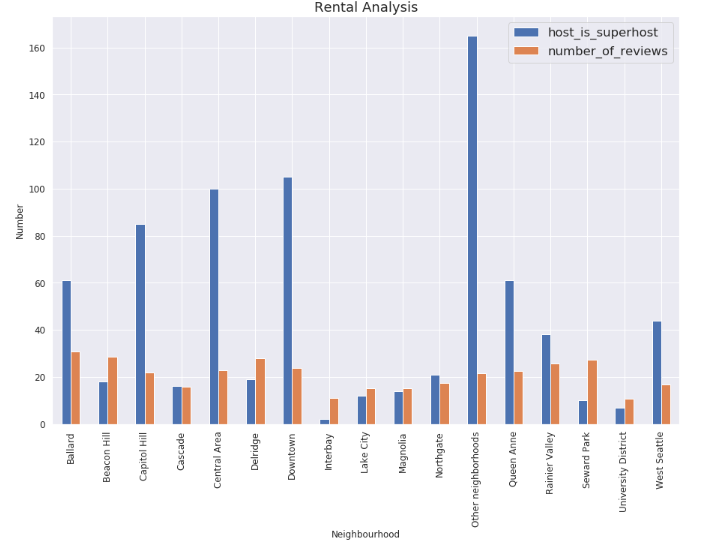
It is observed here that most of the rentals across the neighborhoods have verified hosts meaning that the listings are authentic.

**2.Cancellation Policy:**



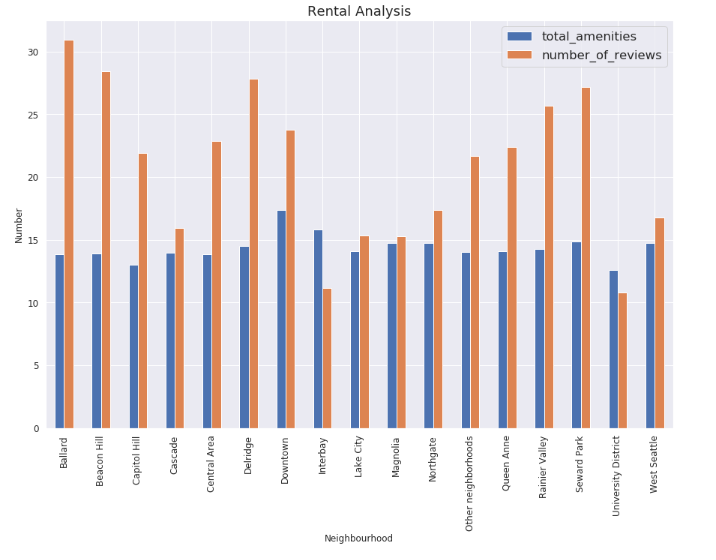
The cancellation policy that most tourists seek will be from moderate to flexible. As per our analysis, the neighborhoods that have high number of reviews have moderate cancellation policy.

**3.Host\_is\_superhost:**



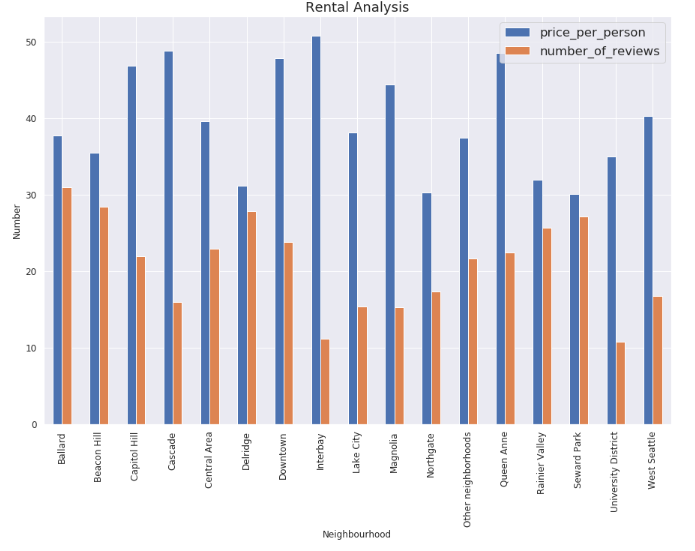
Host is superhost is one of the factor which can give a higher score to a particular rental hence it cannot be ignored.

**4.Total number of amenities:**



The neigborhoods having high number of reviews have upto 15 amenities.These are common amenities which a visitor looks for in a particular rental

**5.Price per person:**



As seen in this graph, the price is one of the factor. An average price between 30 to 38$ is how the listings are rated in top 5 neighborhoods.

# **4. Conclusion:**

Based on above analysis it was found that Ballard, Beacon Hill, Delridge, Rainier Valley and Seward Park is the most preferred neighborhood for tourists in Airbnb Seattle. They all offer rentals with below facilities:

1.Cancellation policy for most of the rentals in these areas is moderate.

2.The total number of amenities up to 15

3.Price is between 30$ to 38$ per person

4.If the host is super host, the listing has more reviews

5. The host identity is verified for most of the rentals especially in Ballard and Beacon

Hill thus making the listings

as legal ones hence getting more preference by the travelers.

6. Also most of the rentals in these areas have a profile picture which further adds to the authenticity of the listing.

Based on above points, efforts can be made to increase listings in these neighborhoods so that Airbnb can improve their business as well. To improve business on existing rentals in other neighborhoods, amenities and price needs to be worked upon.