Predicting the price of Airbnb in New York
By Prabal Godbole

1.Introduction

1.1Background

Airbnb provides a platform for hosts to accommodate guests with short-term lodging and tourism-related activities. Guests can search for lodging using filters such as lodging type, dates, location, and price, and can search for specific types of homes, such as bed and breakfasts, unique homes, and vacation homes.Before booking, users must provide personal and payment information.Some hosts also require a scan of government-issued identification before accepting a reservation. Guests can chat with hosts through a secure messaging system. Hosts provide prices and other details for their rental or event listings, such as the allowed number of guests, home type, rules, and amenities. Hosts and guests have the ability to leave reviews about the experience.

1.2 Problem

The purpose of this task is to predict the price of NYC Airbnb rentals based on the data with relevant information.And to show nearby venues to the place.

1.3 Interest

The interested peoples will be the Airbnb Inc. and the owners of the stays and tourists too.

2.Data

2.1 Sources
2.1.1 Kaggle

A dataset from the kaggle consist informations like host names, latitudes, longitudes, price etc.
To get the training data for the model to predict the price of the Airbnb.

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_o
0	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.64749	-73.97237	Private room	149	1	
1	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	
2	3647	THE VILLAGE OF HARLEM....NEW YORK !	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3	
3	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89	1	
4	5022	Entire Apt: Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	40.79851	-73.94399	Entire home/apt	80	10	

2.1.1 Foursquare API

To get the venues around the Airbnb chosen by the user using the latitude and longitude of Airbnb.

	name	categories	address	cross Street	lat	lng	distance
0	Osakana	Fish Market	290 Graham Ave	btwn Powers St & Grand St	40.712218	-73.943838	138
1	Then She Fell	Indie Theater	195 Maujer St	NaN	40.711029	-73.942940	156
2	Ammazza Caffe	Italian Restaurant	702 Grand St	NaN	40.711543	-73.944574	218
3	Tuffet	Bar	286 Graham Ave	btwn Grand St & Powers St	40.712112	-73.943921	147
4	Charter Coffeehouse	Coffee Shop	309 Graham Ave	Devoe & Ainslie	40.713273	-73.944340	209
5	Sage Brooklyn	Thai Restaurant	299-301 Graham Ave	at Ainslie St.	40.713060	-73.944202	187
6	Artist & Craftsman Supply	Arts & Crafts Store	761 Metropolitan Ave	NaN	40.714691	-73.943534	286
7	Easy Lover	Cocktail Bar	790 Metropolitan Ave	NaN	40.714699	-73.942660	267
8	To The World Farm	Market	655 Grand St	btwn Leonard St & Manhattan Ave	40.711445	-73.946143	346
9	Hungry Ghost	Coffee Shop	721 Metropolitan Ave	btwm Manhattan & Graham Ave	40.714567	-73.945016	344

2.2 Data Cleaning

The dataset from the Kaggle was not cleaned. To get more accurate I perform the following operations

1) Replace the missing values with the 0, mean or other data according the need of the column.
2) Drop some rows due missing values which can nott be filled.
3) Finding Outliers using catplot based on the clusters made.
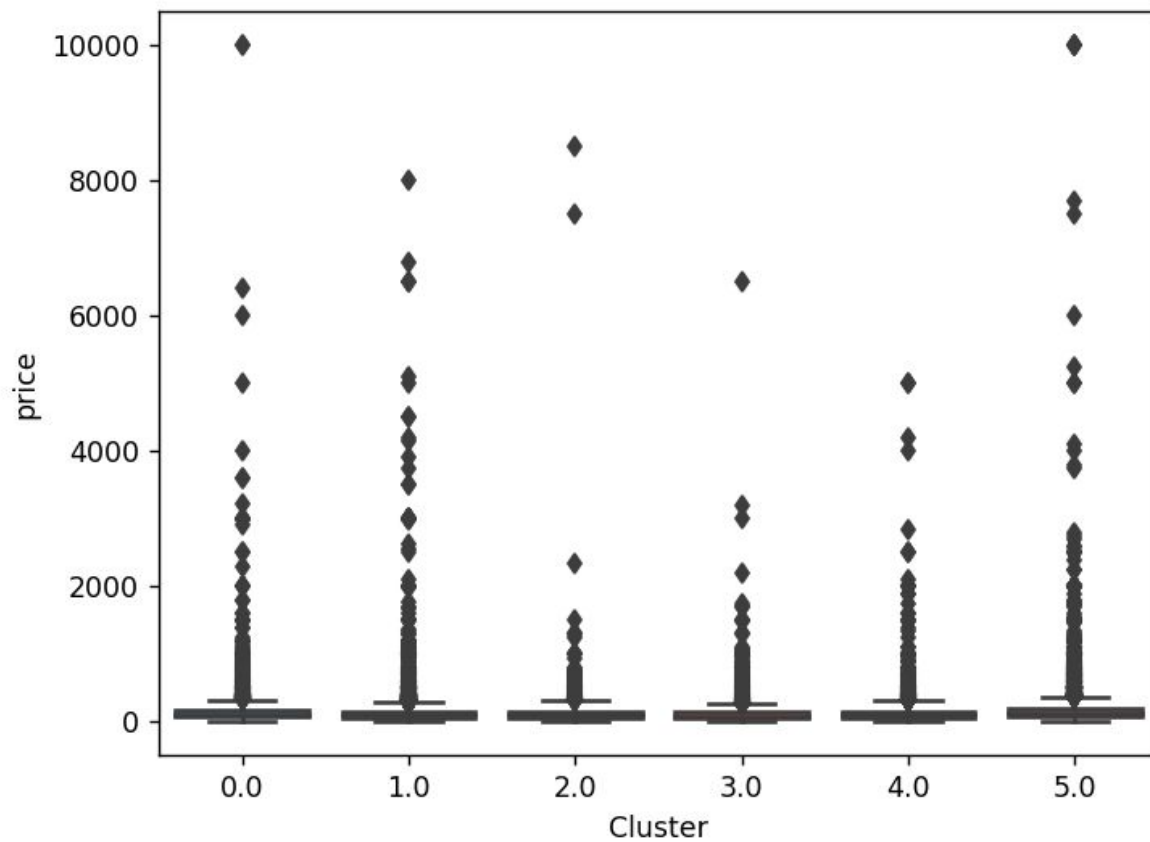
3.Data analysis and manipulation/encoding

3.1 Clustering

Data is clustered into 6 clusters using the latitude and longitude to get more accurate prediction.
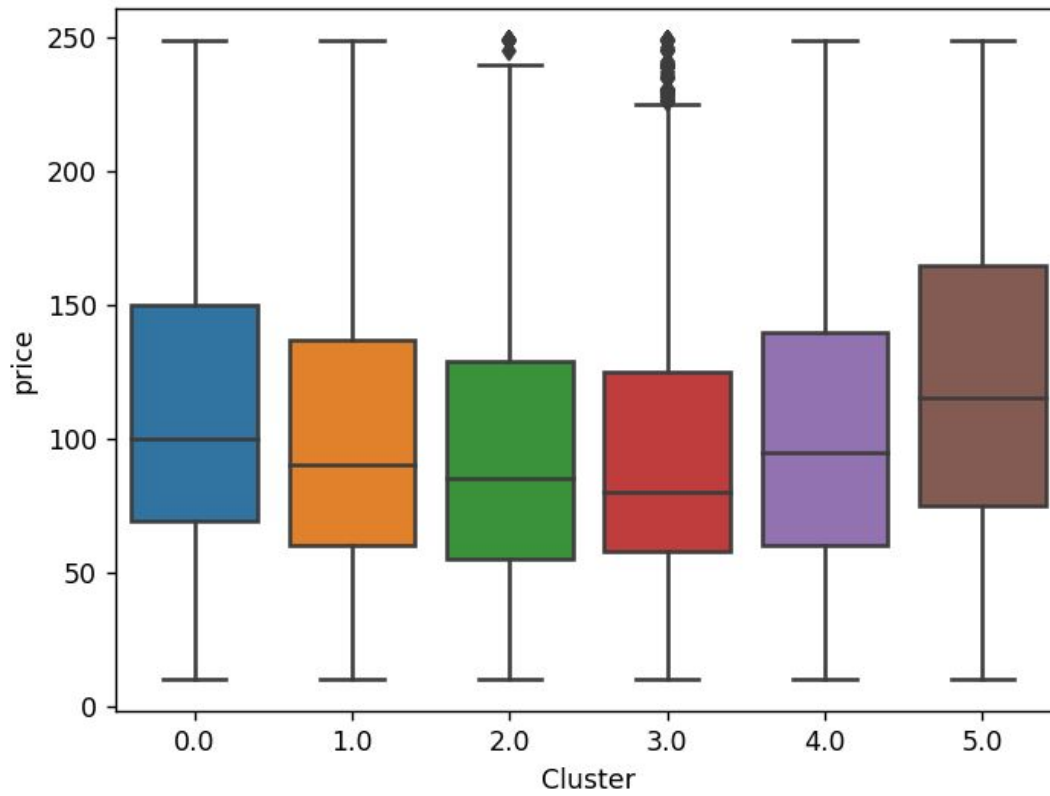
3.2 Removing Outliers

Outliers are found by the plotting the catplot between the price and clusters
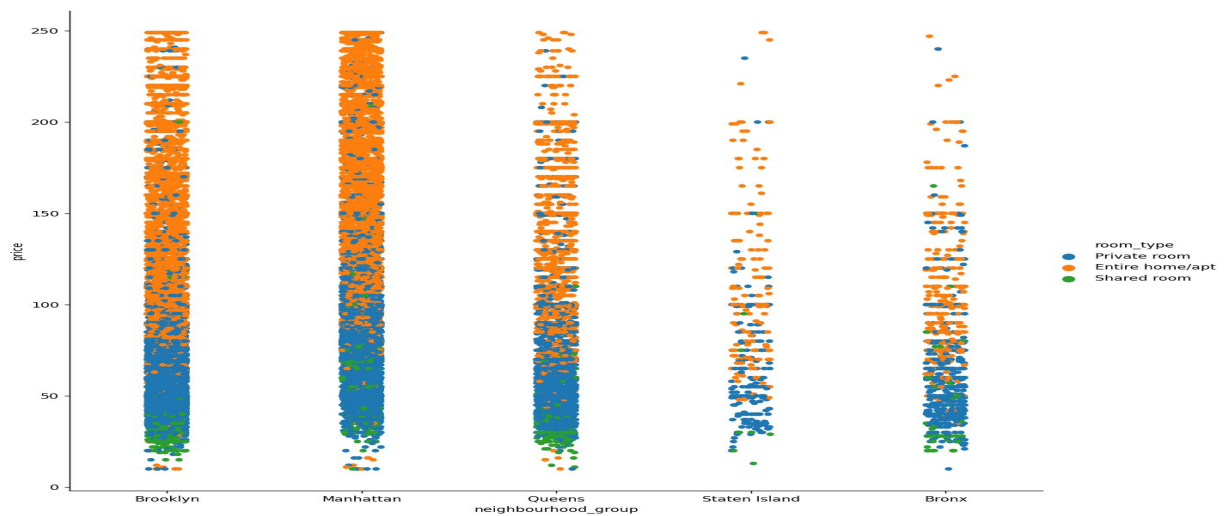
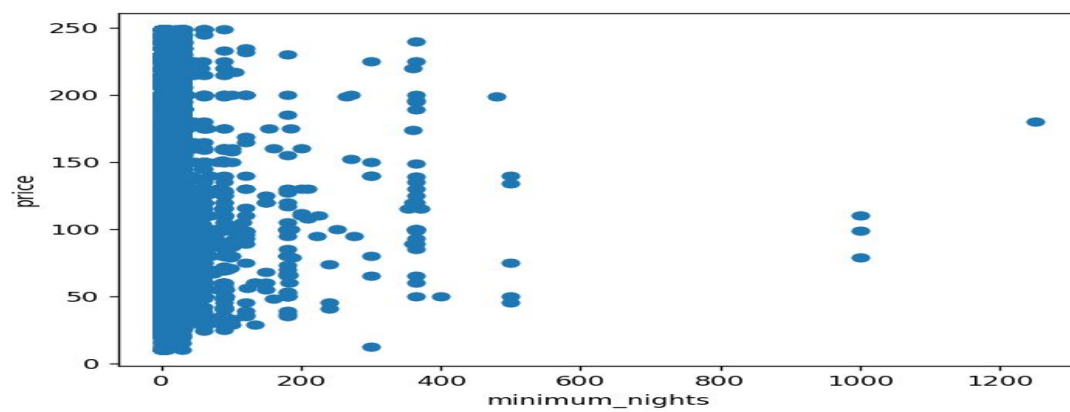Data before removing outliers

Data after removing outliers

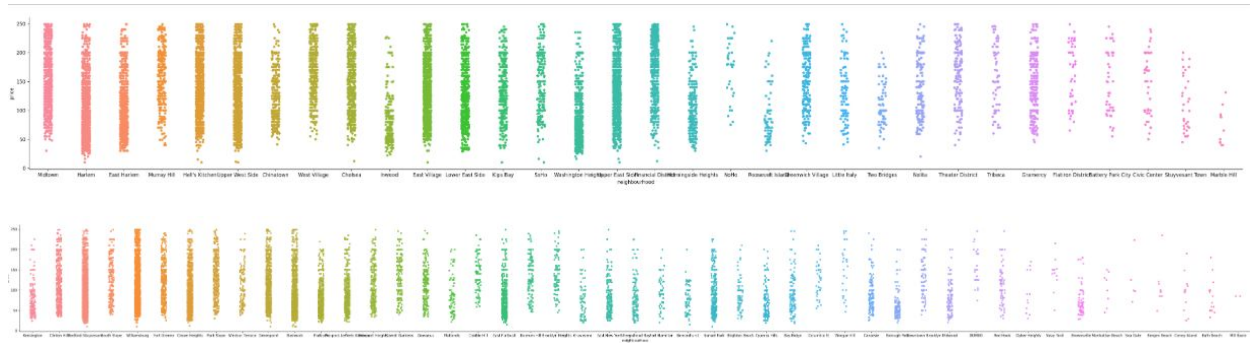## 3.3 Exploration of data
3.3.1 Prices in neighbourhood groups according to different types of Airbnbs

It can be seen that the price of shared rooms are the lowest while the private room is mid to high and apt is mostly high.

3.3.2 Prices in different neighbourhoods
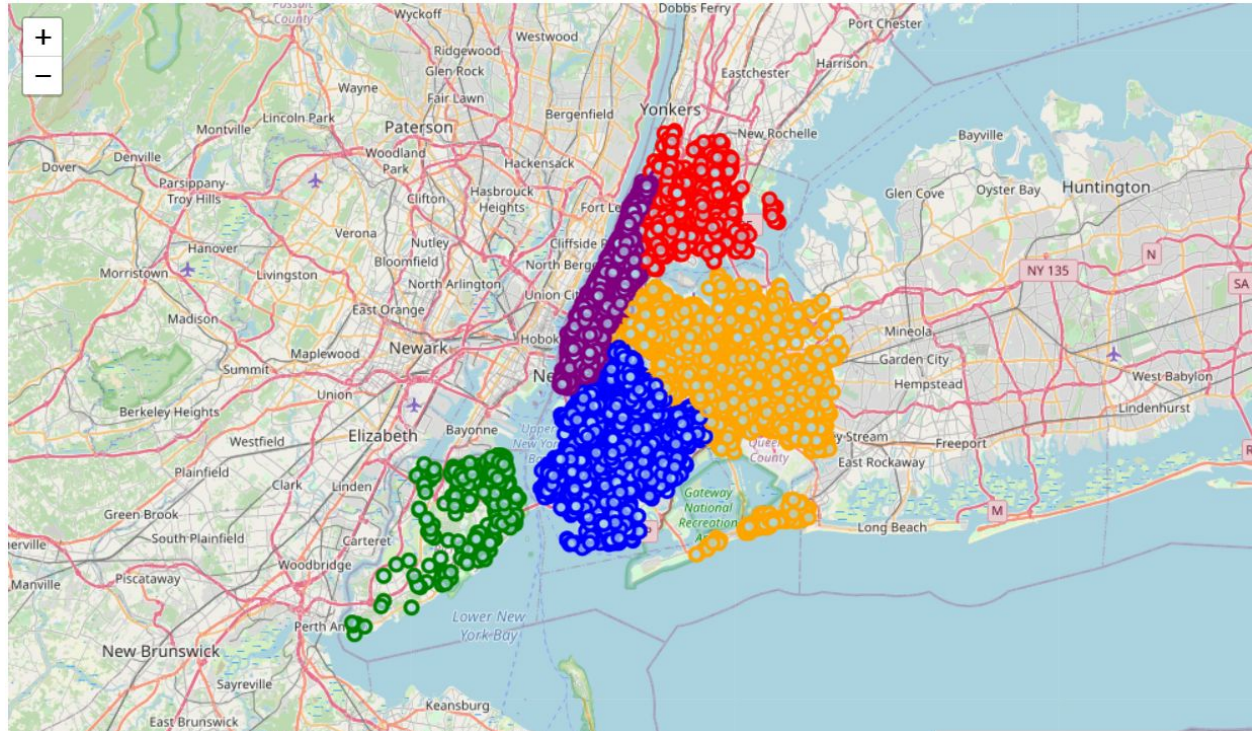

3.3.3 Price vs different features
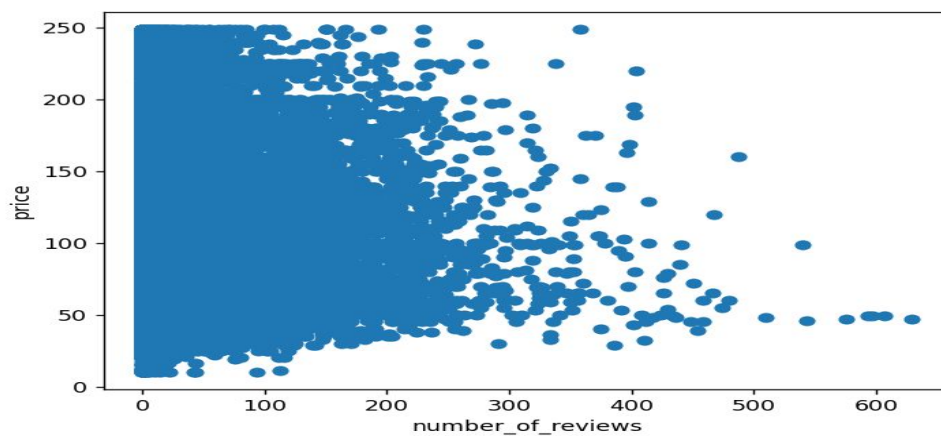
3.3.4 Airbnbs in different neighbourhood groups

```
Manhattan          21660
Brooklyn           20095
Queens              5666
Bronx               1090
Staten Island        373
```

Map of Airbnb to according to different Airbnbs

3.4 Encoding data

Converting the data types of the columns in desirable types

```
id                                int64
name                             object
host_id                           int64
host_name                        object
neighbourhood_group              object
neighbourhood                    object
latitude                        float64
longitude                       float64
room_type                        object
price                             int64
minimum_nights                    int64
number_of_reviews                 int64
last_review                      object
reviews_per_month               float64
calculated_host_listings_count    int64
availability_365                  int64
Cluster                         float64
dtype: object
```

To

```
id                                int64
name                             object
host_id                           int64
host_name                        object
neighbourhood_group               int64
neighbourhood                     int64
latitude                        float64
longitude                       float64
room_type                         int64
price                             int64
minimum_nights                    int64
number_of_reviews                 int64
last_review                      object
reviews_per_month               float64
calculated_host_listings_count    int64
availability_365                  int64
Cluster                         float64
dtype: object
```

3.4 Splitting data and making predictive model

Data splitted into test and train parts.

Then I used **RandomForest Regression** to get the price with accuracy of **0.7898807316942643** or **78%.**

And the taking the latitude and longitude of desirable Airbnb to get the all venues around it.

	name	categories	address	crossStreet	lat	lng	distance
0	Osakana	Fish Market	290 Graham Ave	btwn Powers St & Grand St	40.712218	-73.943838	138
1	Then She Fell	Indie Theater	195 Maujer St	NaN	40.711029	-73.942940	156
2	Ammazza Caffe	Italian Restaurant	702 Grand St	NaN	40.711543	-73.944574	218
3	Tuffet	Bar	286 Graham Ave	btwn Grand St & Powers St	40.712112	-73.943921	147
4	Charter Coffeehouse	Coffee Shop	309 Graham Ave	Devoe & Ainslie	40.713273	-73.944340	209

And getting them on the map.
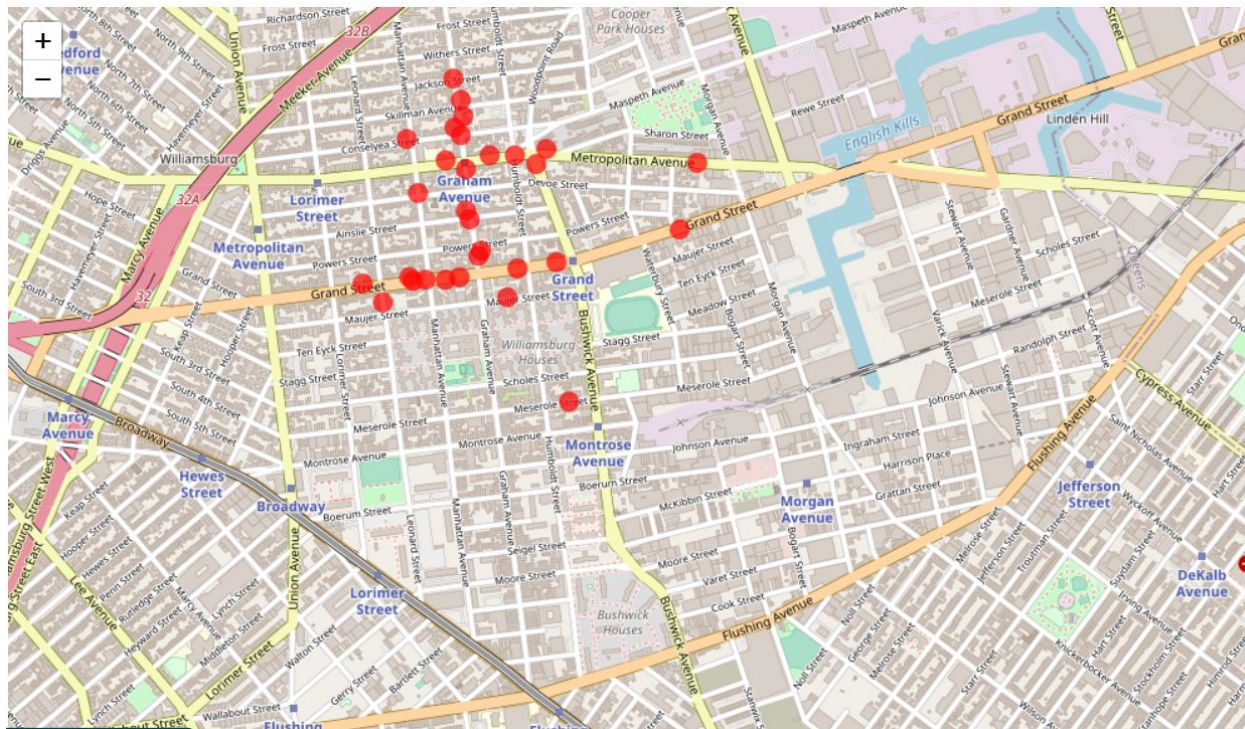
4.Results

1. Model to predict price of Airbnb in Newyork with 78% accuracy
2. Location , Information about the nearby venues of Airbnb.

5.Discussion

The only thing I like to discuss about machine learning algorithms.
I actually used Lasso first but the accuracy was too low so I used Random Forest I could've increased accuracy more but then it would've consumed more resources and time.

6.Observations and Conclusion

1. Prices in Brooklyn and Manhattan are higher than others.
2. Private rooms and shared rooms are cheaper compared to whole Apt but that's not always the case.
3. Most of the Airbnb in New York cost less than $250.