Predicting Airbnb Listing Price

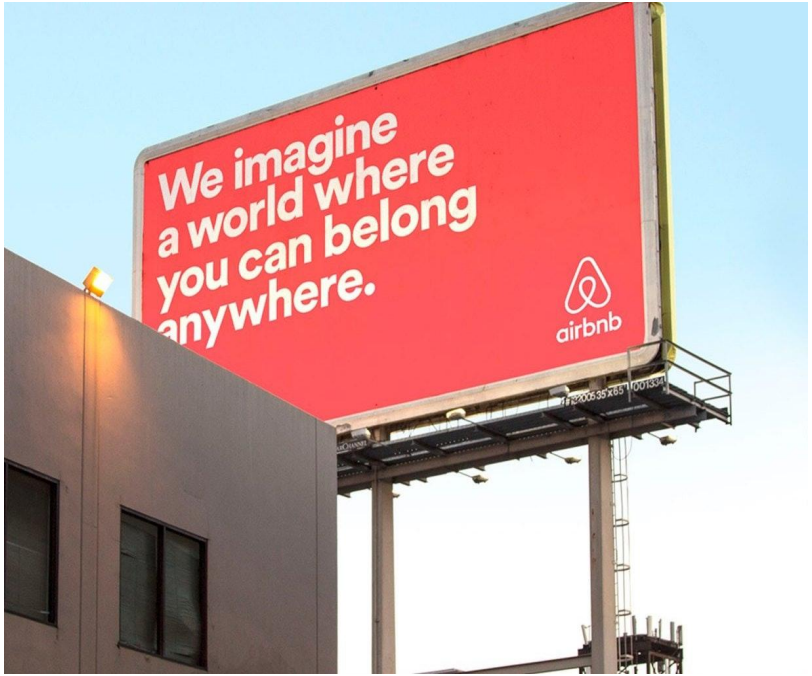
By : Irene Izere

Overview

Airbnb stands for Air Bed and Breakfast. It was started in 2007 by roommates Joe Gebbia and Brian Chesky.

What started as a way to get some extra cash has become a multimillion company with a net worth around $3.5billion at the end of 2020.

Airbnb connects travelers with Airbnb hosts who want to rent out their homes or other property. For guests, Airbnb gives affordable temporary housing options and sometimes fun activities. For hosts, Airbnb is a way to earn extra money.



Problem Statement

The airbnb is a booming business in the hospitality industry right now . Many people are buying house or renting house to make them airbnbs and increase their income.

The aim of this project is to build a regression machine learning model that can help in predicting the listing price of the airbnb and helping the owner understand the important features .

Data Collection

- Dataset from Kaggle.com(https://www.kaggle.com/kritikseth/us-airbnb-open-data)
- Combined listing of house in different city of United State
- Size is 226030 rows and 17 features after cleaning shape is 225448 rows and 15 features

This dataset has one file- AB_US_2020.csv which has columns describing features such as host id, hostname, listing id, listing name, latitude and longitude of listing, the neighbourhood, price, room type, minimum number of nights, number of reviews, last review date, reviews per month, availability, host listings and city.

Exploratory Data Analysis

- Looked at this dataset in two part : the numerical and categorical features and The description for NLP(Nature Language Processing)

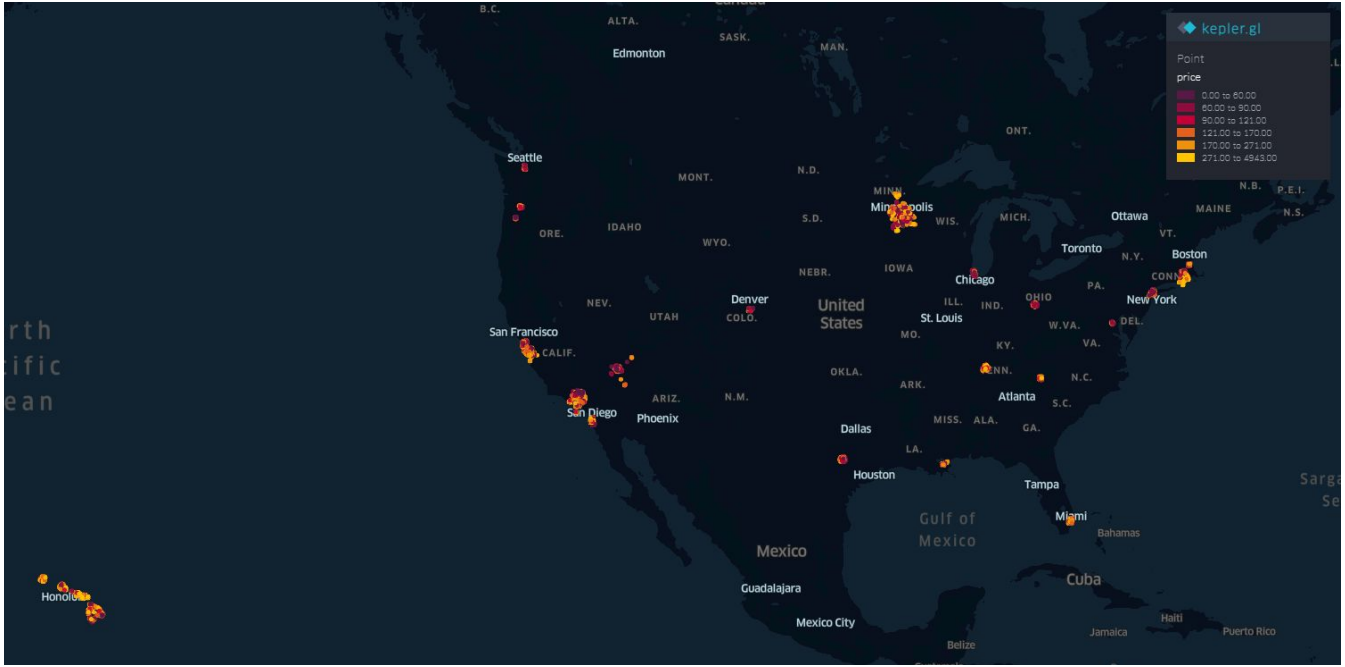
- Analysed the distribution and relationship between features such location.

- Analysed the description of the airbnb to understand how words are used

Exploratory Data Analysis: *Location*

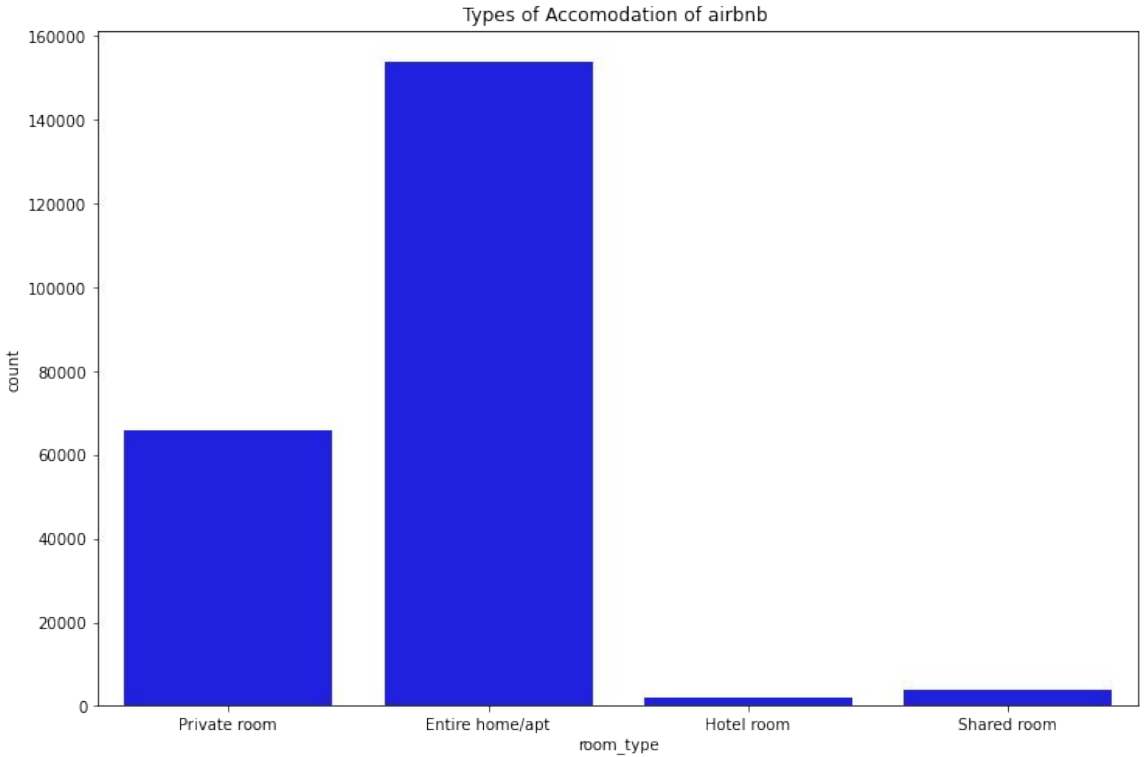
- Airbnbs are now in almost all the major cities
- Their listing price varies.

Map of the airbnb around the United state

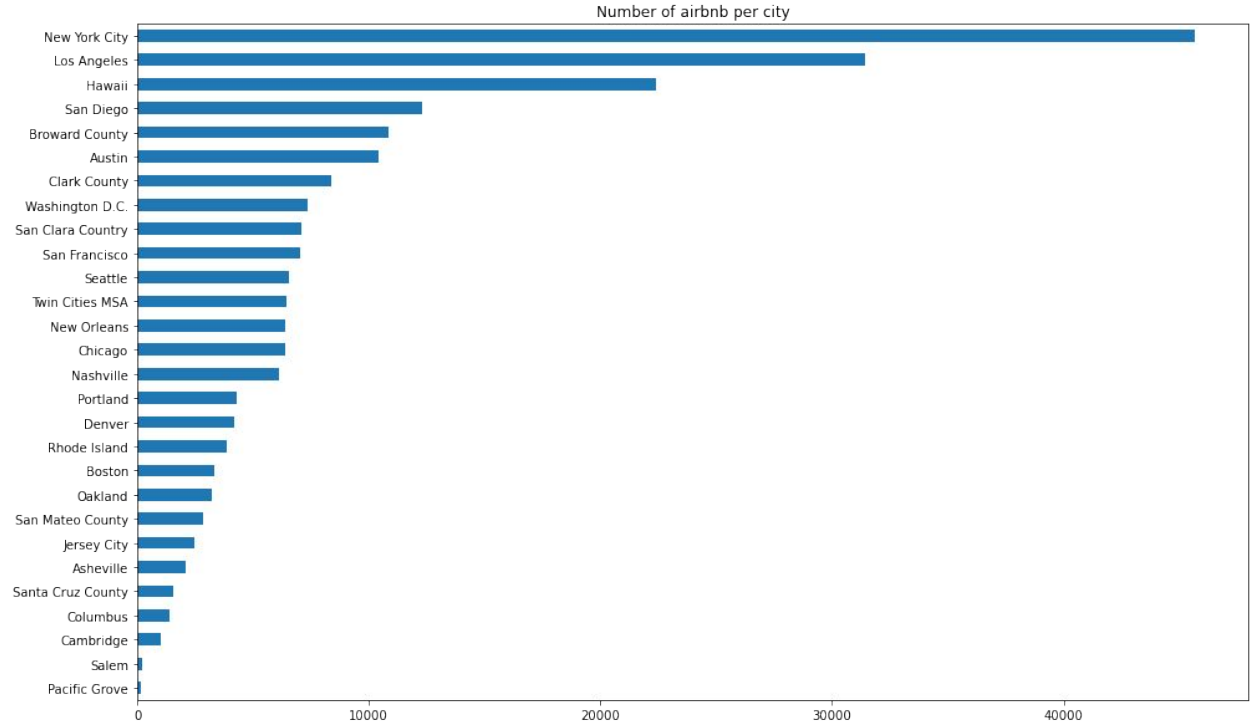


Data analysis

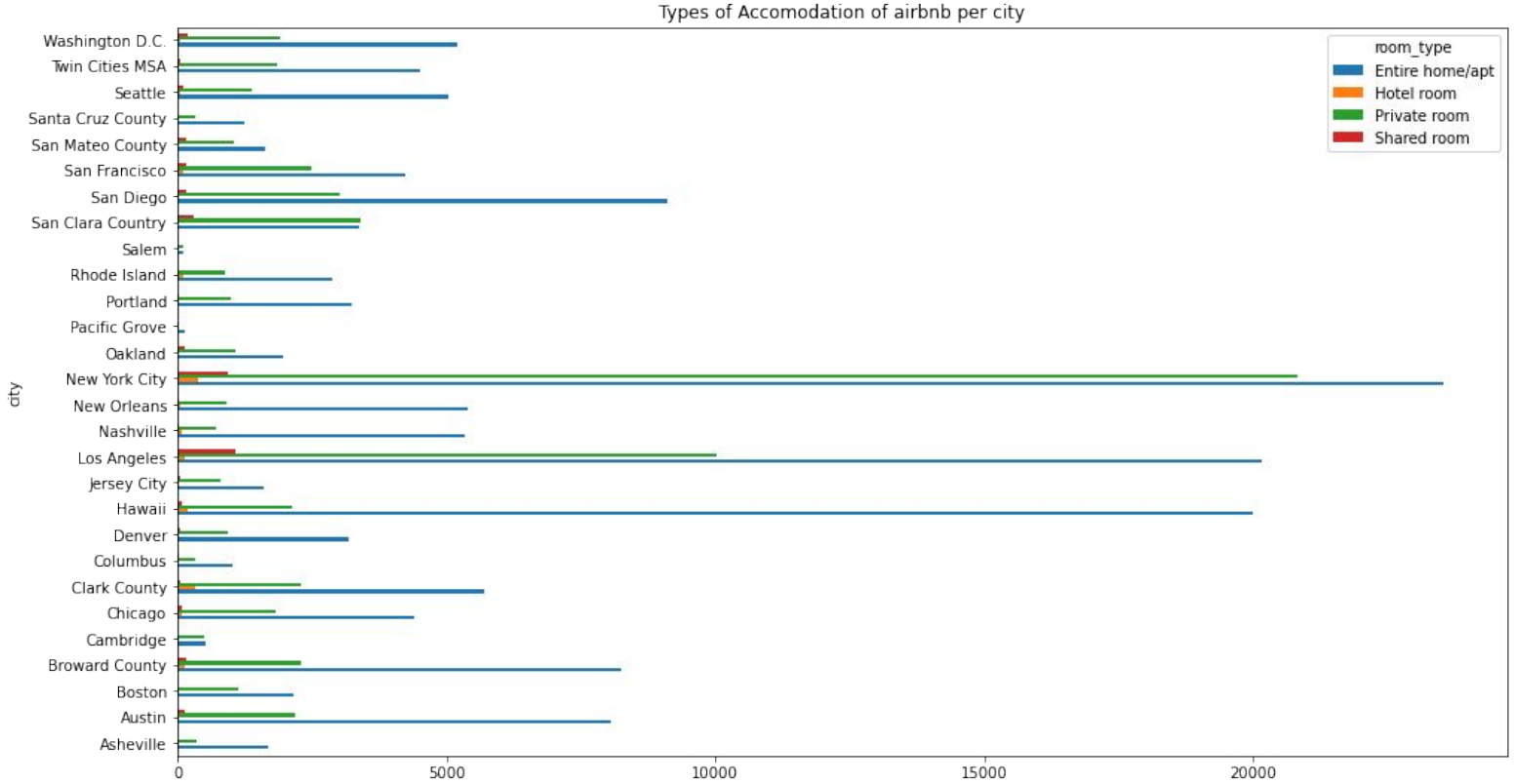
- **Entire place.** If you select this option, you do not have to share your living space, common areas, or bathroom with other guests. Guests who book this type of stay expect privacy, though you may still have a shared outdoor space.
- **Private room.** A private room means that you may live in a house with others but have your own bedroom. An Airbnb host renting out one room of their home should provide information about the house rules, the shared spaces, and the number of other home occupants.
- **Share room.** Travelers on a budget can choose to share a room. This means that you may sleep in one room with two or three beds open to other travelers.



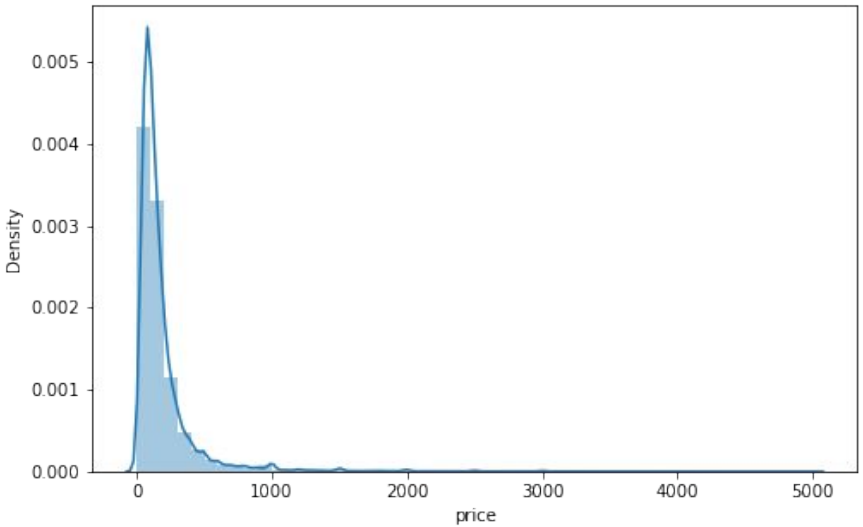
Data analysis

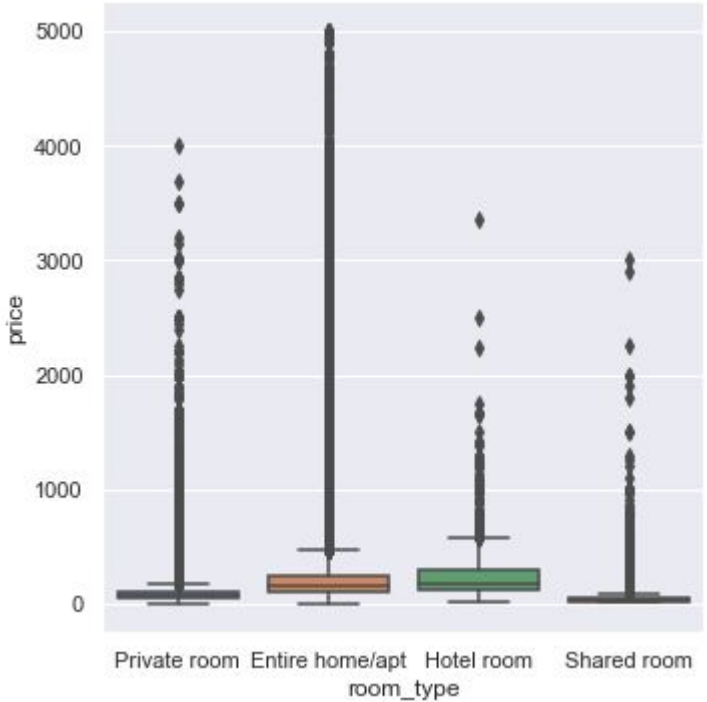


Most of the city that have a highest number of Airbnbs are touristic , entertainment or vacation destinations.


Types of Accomodation of airbnb per city
room_type
Entire home/apt
Hotel room
Private room
Shared room
city
Washington D.C.
Twin Cities MSA
Seattle
Santa Cruz County
San Mateo County
San Francisco
San Diego
San Clara Country
Salem
Rhode Island
Portland
Pacific Grove
Oakland
New York City
New Orleans
Nashville
Los Angeles
Jersey City
Hawaii
Denver
Columbus
Clark County
Chicago
Cambridge
Broward County
Boston
Austin
Asheville
0
5000
10000
15000
20000

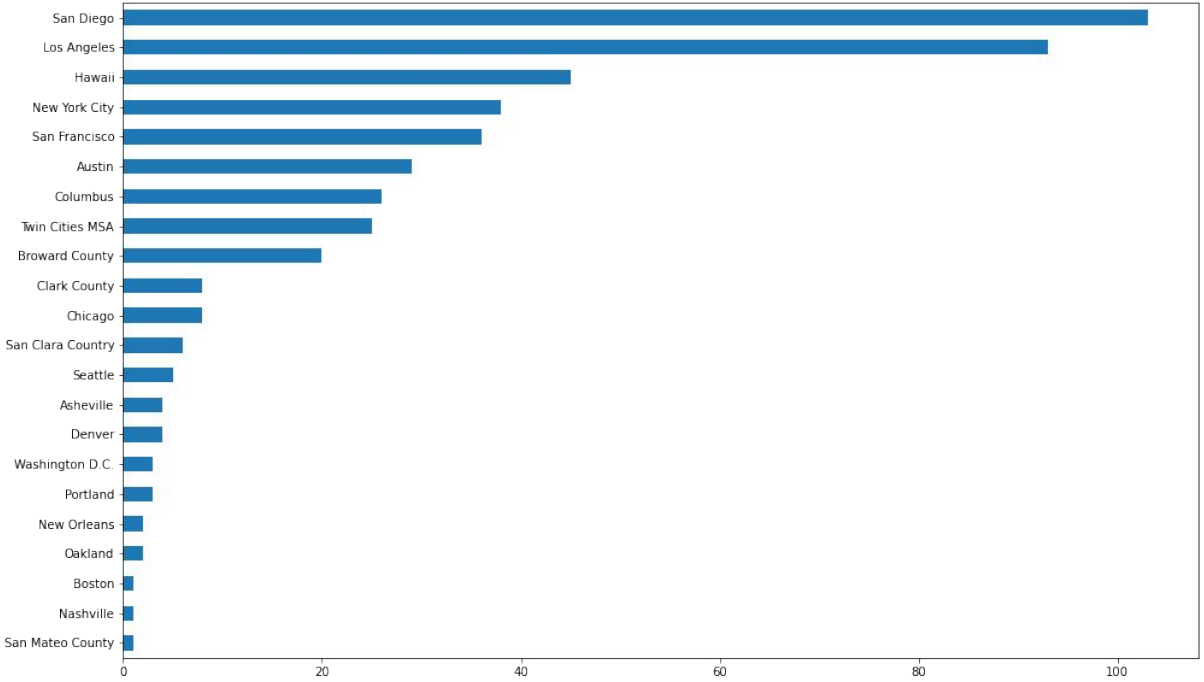
Distribution of the price





The distribution of the price is very skewed , and have many outliers which is an indication on how variable the price can be.

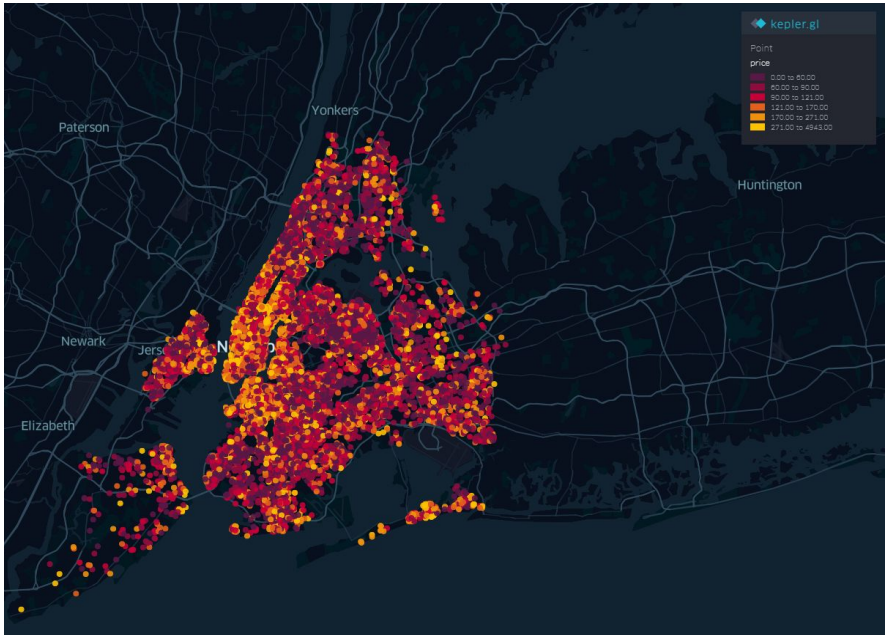
Data analysis

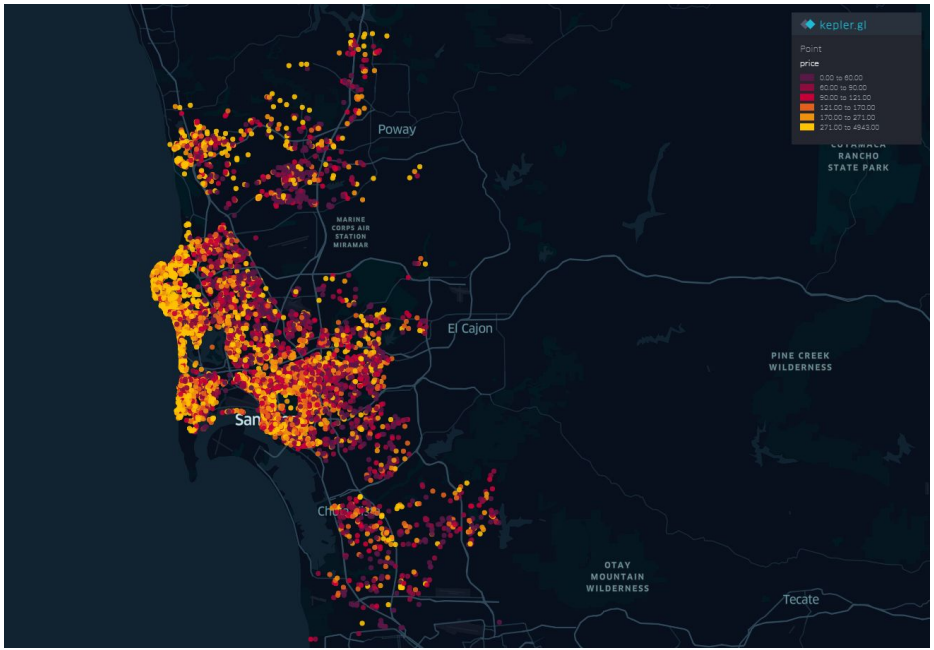


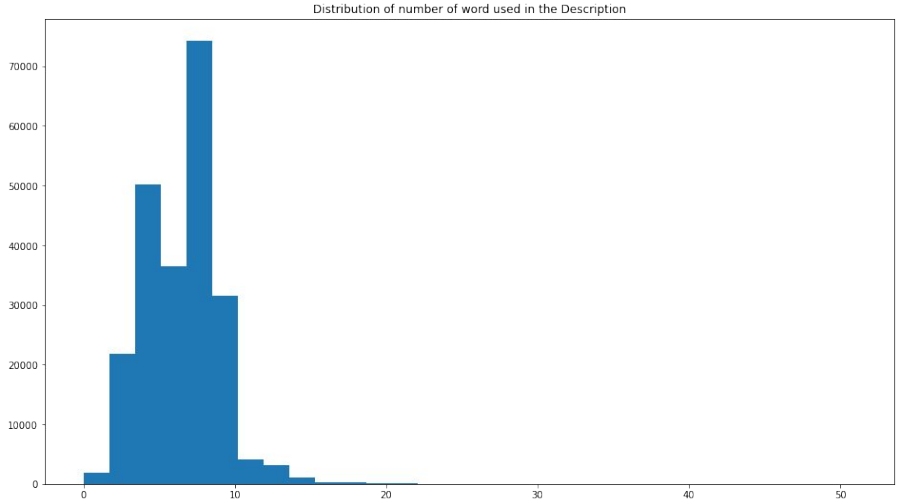
The cities with a highest number of airbnbs that have the price over $5000 per night .

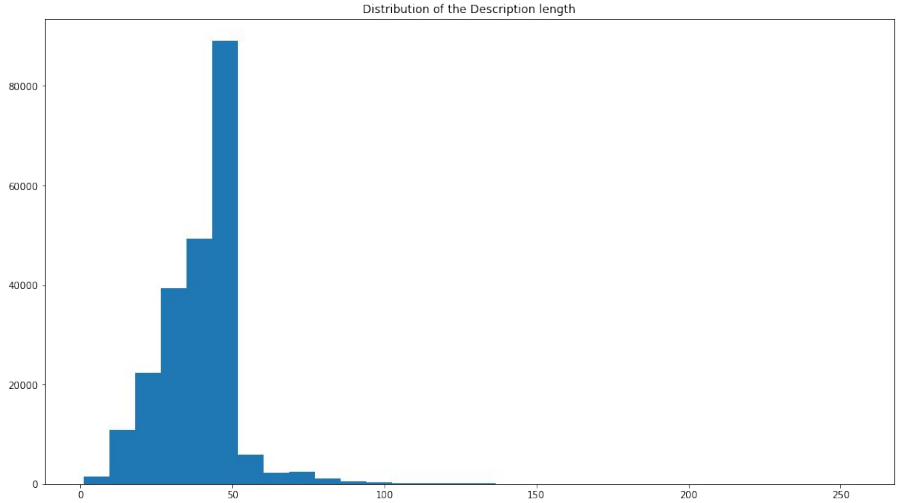
Closer view

The Airbnb in san Diego and New York are located with their price . As it become brighter the price is increasing.

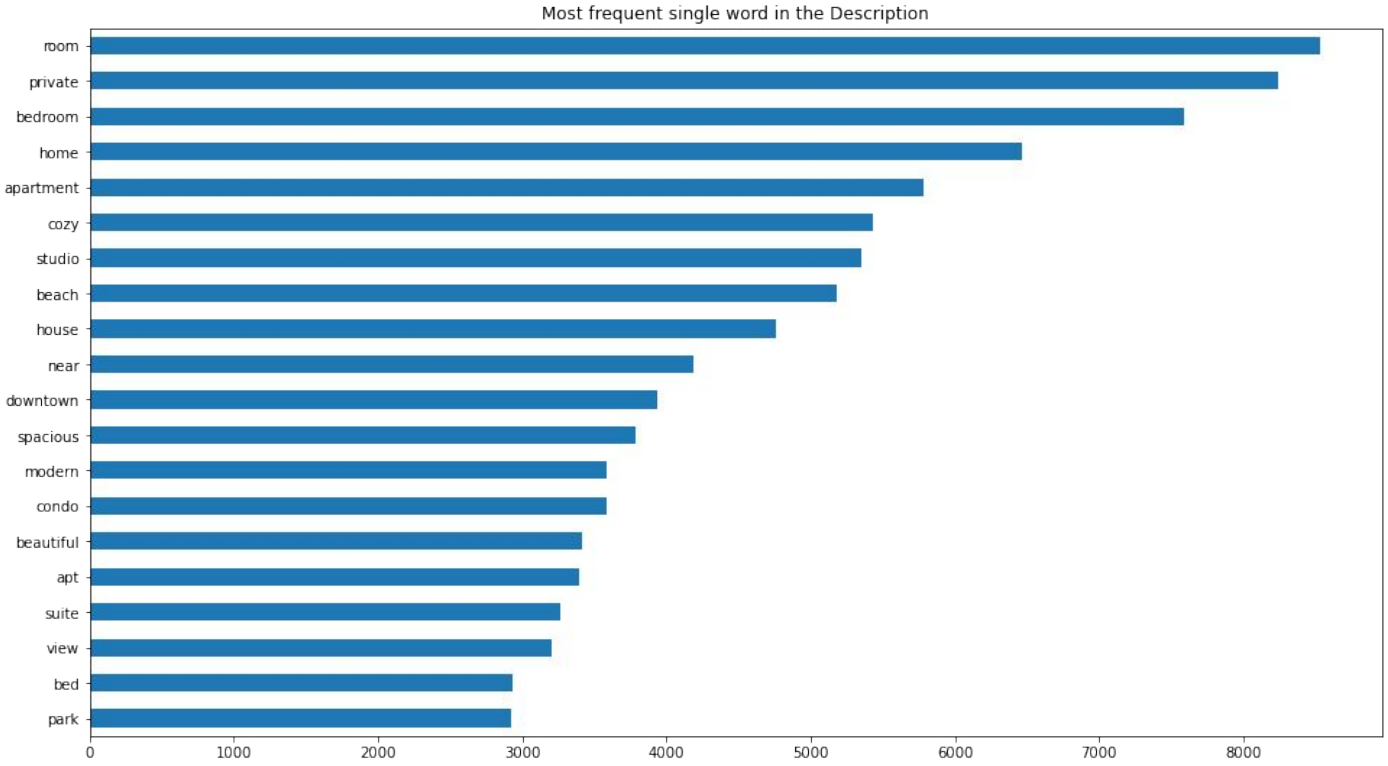




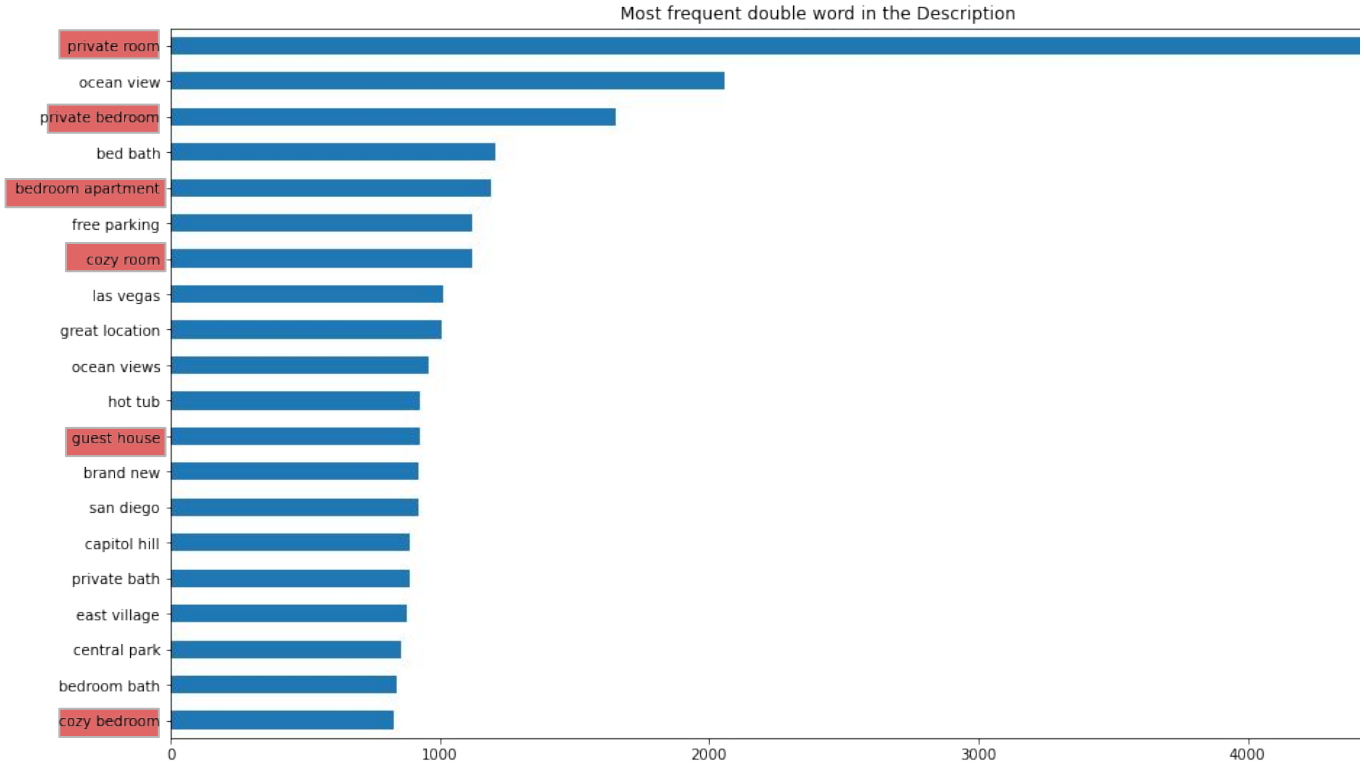




Most of the Description are between 1-10 words , so it is a short description .



- *Type of Accommodation*
- *Location*
- *Qualitative adjective*



- *Type of Accommodation*
- *Location*
- *amenities*

Modeling

Based on our problem statement to predict price , We need to build a regression model.

We built and compare 4 models using a linear regression as a baseline .

- Linear regression,Decision tree , Random Forest ,XGboost

Evaluating

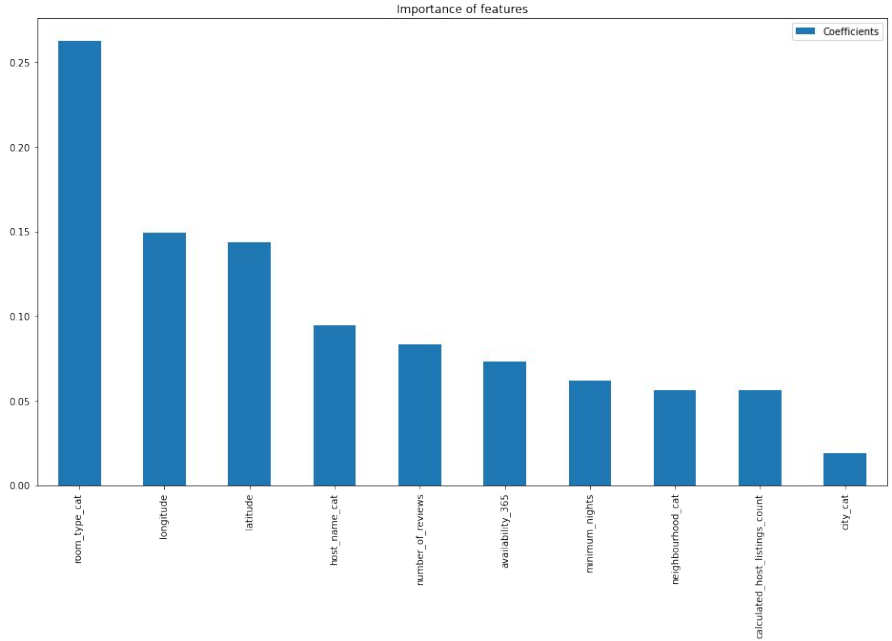
Model	Training RMSE	Testing RMSE
Linear Regression	287.6	290.5
Decision Tree	5.1	320.6
Random Forest	129.6	234.2
XGBoost	211.13	236.07

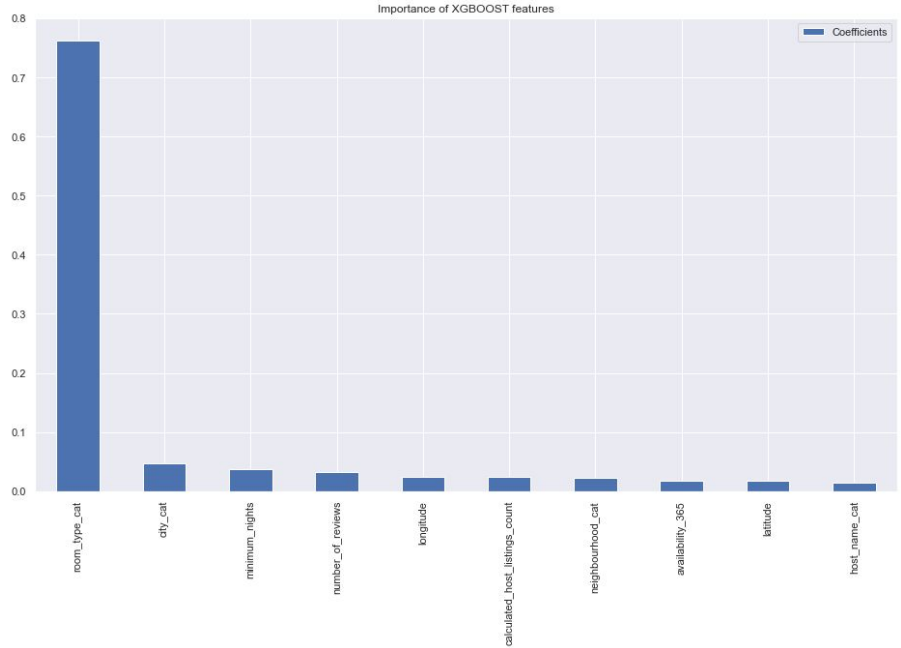
Evaluation metrics : RMSE
The smaller the better

Evaluation metrics : R^2 score
The higher the best

Model	Training score	Testing score
Linear Regression	0.292	0.291
Decision Tree	0.99	0.17
Random Forest	0.94	0.57
XGBoost	0.69	0.58

Features Importances





Conclusions and Recommendations

After our analysis and model training , the best model to use to predict the listing price of airbnb is Xgboost model with the accuracy of 58% .

The model also help determine the features importances .

To maximize the listing price the type of accommodations,and the location.

Q & A?