# INFO 6105

# Data Science Engineering Method and Tools

# Final Project Report

# Airbnb Price Prediction

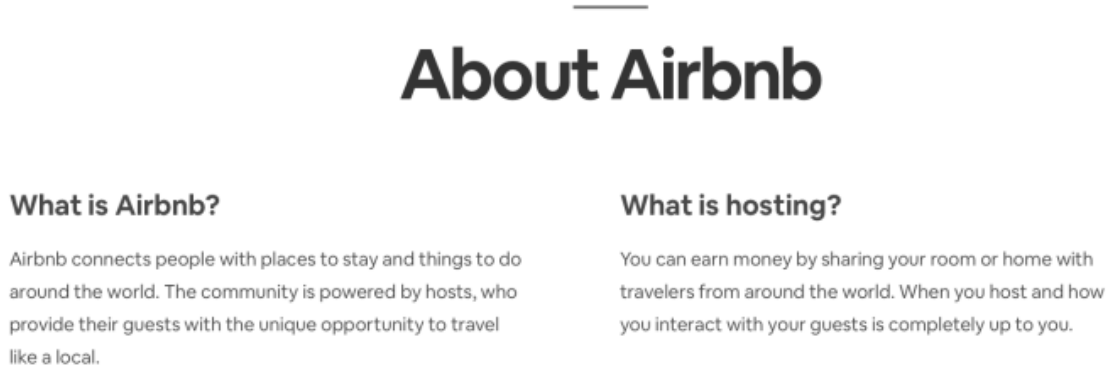

By,
Smit Chandarana
Krishna Modi
Nidhi Sharma
Vaibhavi Khamar
Ziyu Zhou

Under Guidance of:
Prof. Sri Krishnamurty

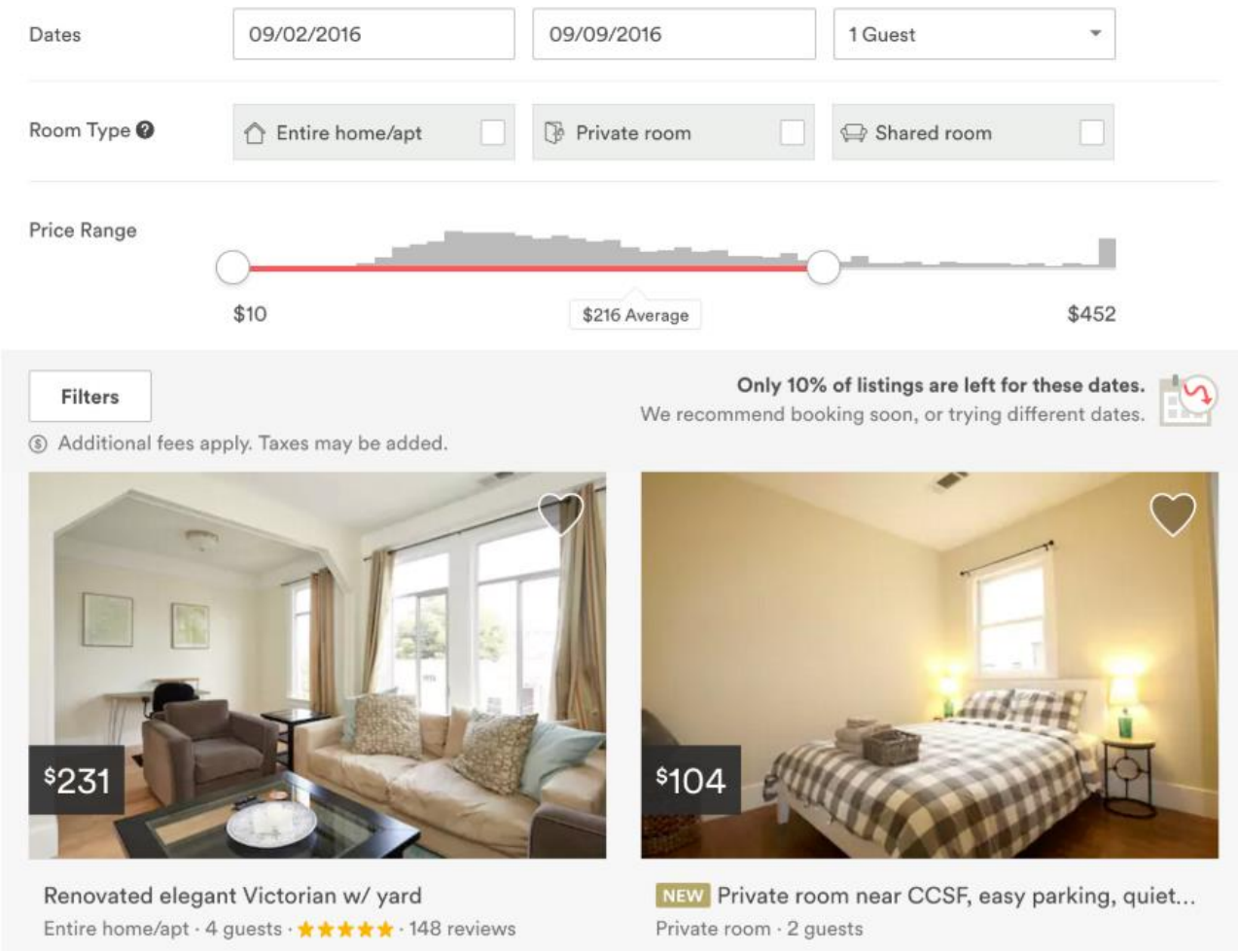
# Contents

## Project Objective:

AirBnB is a marketplace for short term rentals that allows you to list part or all of your living space for others to rent. You can rent everything from a room in an apartment to your entire house on AirBnB. Because most of the listings are on a short-term basis, AirBnB has grown to become a popular alternative to hotels. The company itself has grown from its founding in 2008 to a 30 billion dollar valuation in 2016 and is currently worth more than any hotel chain in the world.




One challenge that hosts looking to rent their living space face is determining the optimal nightly rent price. In many areas, renters are presented with a good selection of listings and can filter on criteria like price, number of bedrooms, room type and more. Since AirBnB is a marketplace, the amount a host can charge on a nightly basis is closely linked to the dynamics of the marketplace.




As a host, if we try to charge above market price for a living space we'd like to rent, then renters will select more affordable alternatives which are similar to ours. If we set our nightly rent price too low, we'll miss out on potential revenue.

Strategy we could use is to:

- find a few listings that are similar to ours,
- Build machine learning model to predict the optimal price of the listing,
- set our listing price to this calculated average price.

Airbnb Price Prediction

## Implementation of the Idea:

- Gathering of data
- Exploratory Data Analysis
- Data Cleaning
- Data Pre- Processing
- Pipeline and Docker
- Modeling
    - Machine Learning model used –

      Decision Tree Regressor, Random Forest Regressor and KNN regressor

    - AutoML –

      TPOT and H2o.ai
- Flask Application
- Deploy on Heroku

## Gathering the Data:

While AirBnB doesn't release any data on the listings in their marketplace, a separate group named Inside AirBnB has extracted data on a sample of the listings for many of the major cities on the website. In this project, we'll be working with their dataset from April,2018 to February,2019 on the listings from Massachusetts, a major state of the United States. Here's a direct link to that dataset.

http://insideairbnb.com/get-the-data.html

Each row in the dataset is a specific listing that's available for renting on AirBnB in Massachusetts.

It provides us with following three files:

- Listing – Detailed listing data of Massachusetts. It contains 107 different features.
- Calendar: It contains the different listed price based on day.
- Review: It gives detail reviews of the listing.

All the monthly listings are concatenated into one file to perform the exploratory data analysis, data cleaning and pre-processing and modelling.

## Exploratory Data Analysis:

The Average Price of listing : $187.

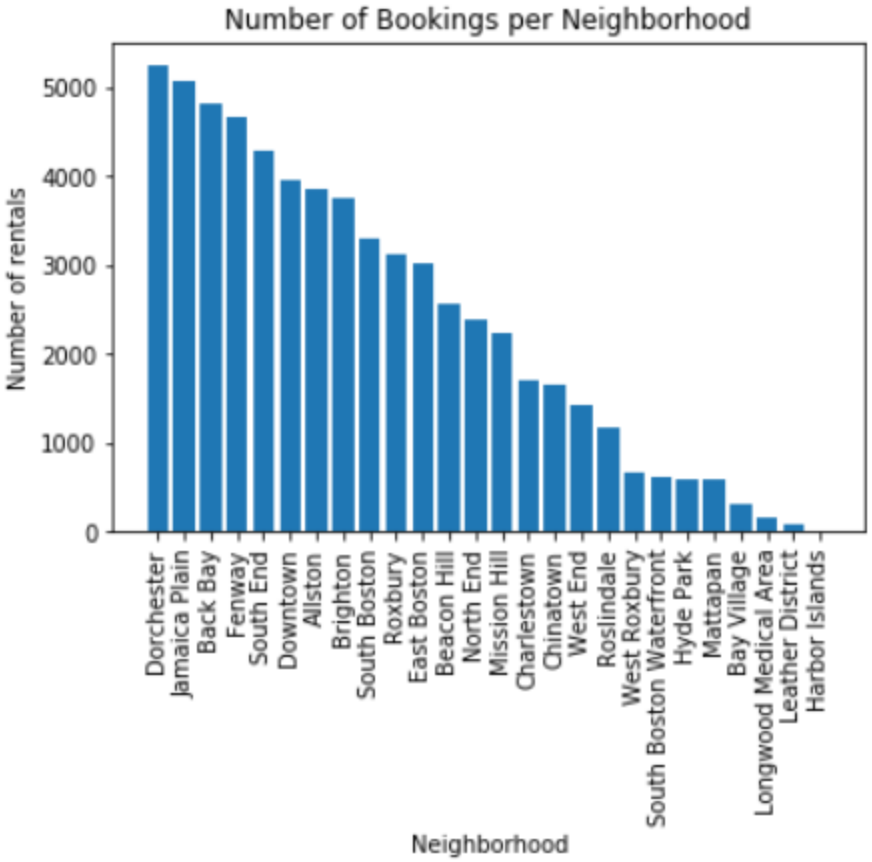
The Maximum Listing Price: $5000

The minimum Listing Price: $0

Average price of Airbnb by neighborhood:

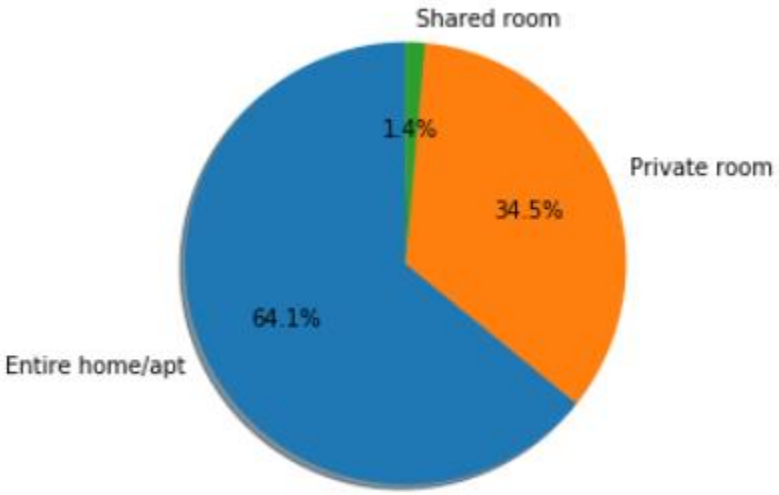

The most expensive listing in Boston are in Leather District, Waterfront and Back Bay.

The most number of listing:

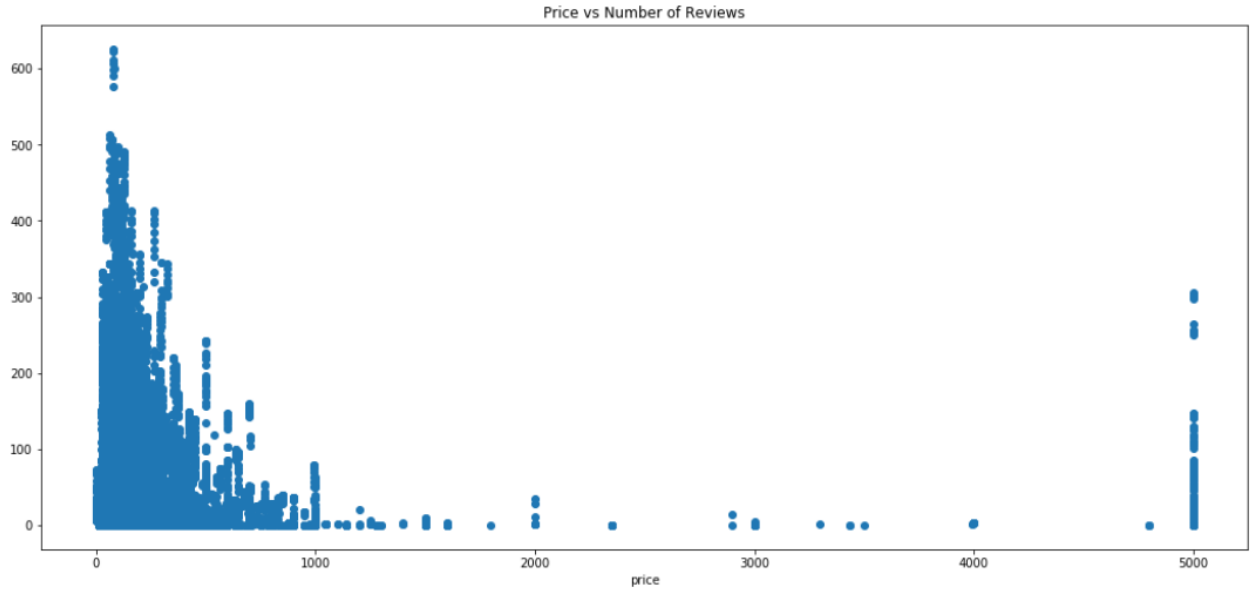



The three most popular areas in Boston are Dorchester, Jamaica Plain, and Back Bay.

Distribution of Rooms:




Price vs No. of reviews




Number of reviews affect the price the most.

The listing in the range of 100 to 300 has the most reviews as they are booked the most.

## Data Cleaning:

We had 107 different columns in the dataset. We discarded non-relatable columns and reduced the number to 25.

```
accommodates

amenities
bathrooms
bed_type
bedrooms
beds
cancellation_policy
cleaning_fee
description
extra_people
guests_included
is_business_travel_ready
minimum_nights
name
neighborhood_overview
neighbourhood_cleansed
number_of_reviews
price
property_type
review_scores_rating
room_type
security_deposit
space
summary
zipcode
```

Then we counted the number of amenities and changed it from string to numeric value.

We implemented on hot encoding to the property type.

## Pipeline and Docker:

Implemented pipeline using Dask where it will accept the raw data as input and return the clean data.
Docker – Command to run the docker on terminal to get the clean data:

docker pull nidhi1993/pipeline
docker run nidhi1993/pipeline

## Modeling:

We implemented Random Forest Regressor, Least Square Regressor, KNN regressor, Decision Tree.

With comparing the MAPE and MSE for all we conclude that Decision Tree is the best model.

MAPE for train is: 1.679198740254469

MAPE for test is: 7.515810609685129

MAE for train is: 2.978320263960393

MAE for test is: 10.887561963137701

MSE for test is: 1265.6177508782146

RMSE for test is: 35.57552179347781

We are finding the best parameters using Grid-Search and get the least MAPE.

Interpretability: Interpretability is the degree to which a human can understand the cause of a decision or the degree to which a human can consistently predict the model's result

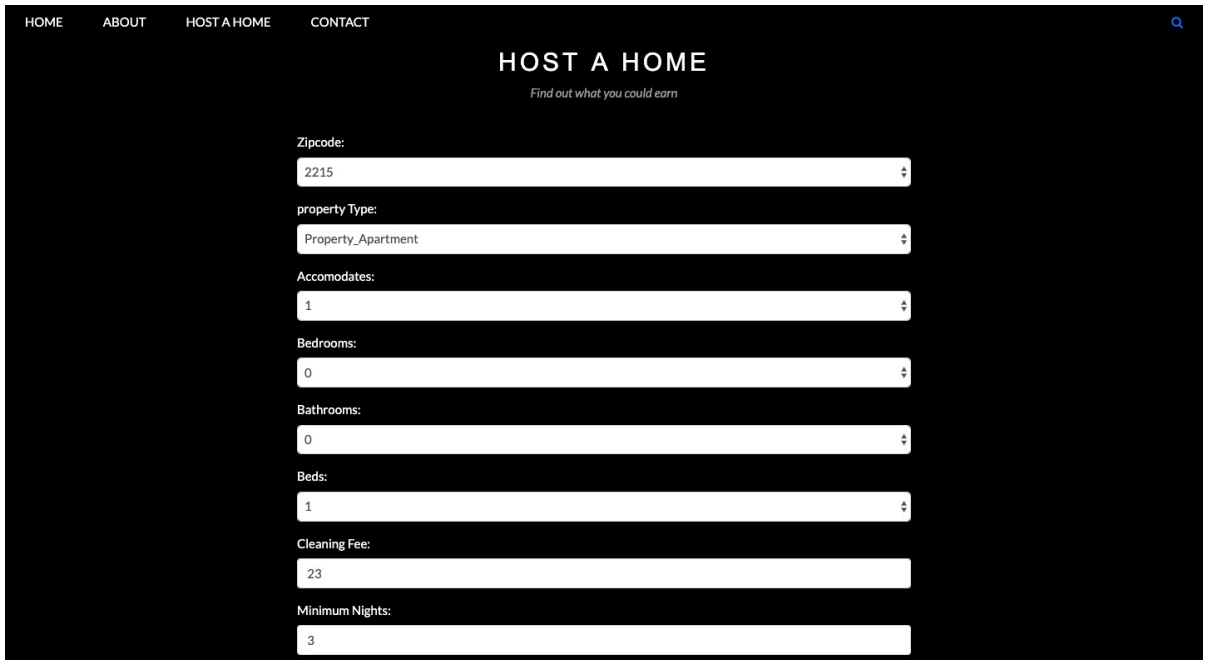
Reproducibility: It is the extent to which consistent results are obtained when an experiment is repeated.

| Model | Interpretable | Reproducible |
|---|---|---|
| Random Forest Regressor | No | Yes |
| Decision Tree Regressor | Yes | Yes |
| KNN Regressor | No | Yes |

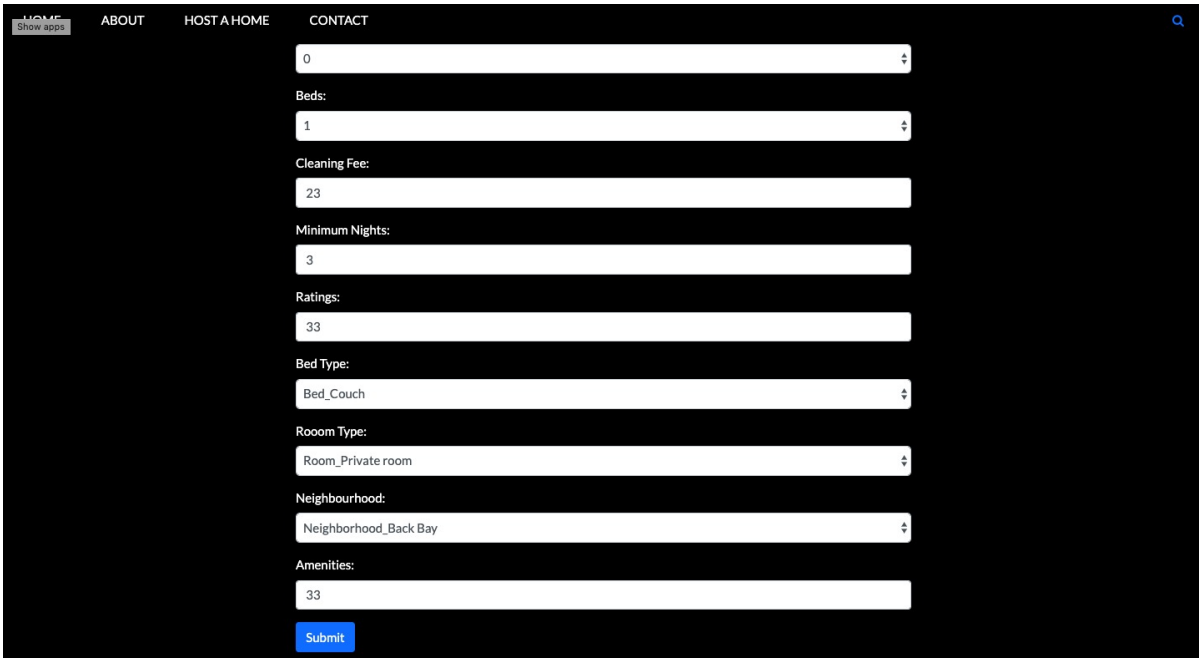
## Web Application and Hosting on Heroku:

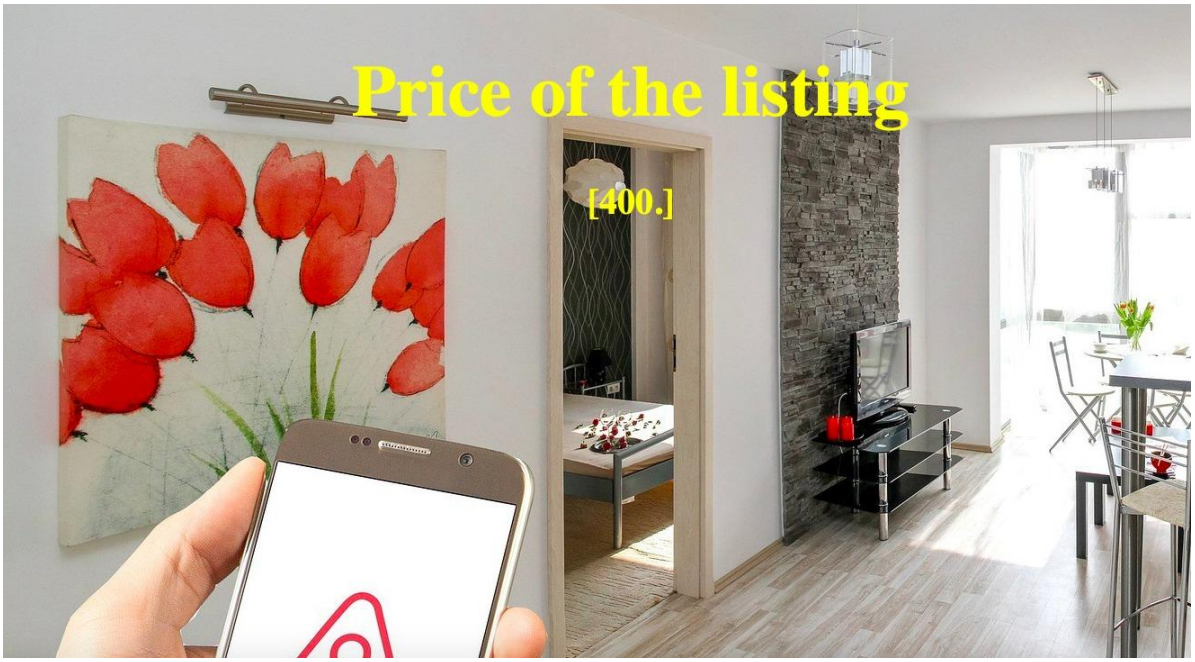
We have created a web application  using flask and hosted it on Heroku.

https://bostonairbnbpriceprediction.herokuapp.com/



Airbnb Price Prediction




It can be accessed from the above link.

The web application takes you the web page where you enter the asked details and it will take you to the final page where it gives you the predicted price.

REFERENCES:

Insideairbnb.com/boston

https://www.numpy.org/

https://scikit-learn.org/stable/

www.stackoverflow.com

https://medium.com/airbnb-engineering/aerosolve-machine-learning-for-humans-55efcf602665