Airbnb Price Prediction

# OVERVIEW

The problem this capstone project aims at solving is predicting an accurate price to be set for an Airbnb property by the host.

Airbnb is a home-sharing platform that allows home-owners and renters (‘hosts’) to put their properties (‘listings’) online, so that guests can pay to stay in them. Hosts are expected to set their own prices for their listings.

# THE CLIENT

# The client is the Airbnb host (owner/renter) and the project aims at helping the host to accurately predict the optimum price for their property.

# Airbnb pricing is important to get right, particularly in big cities where there is lots of competition and even small differences in prices can make a big difference.

It is also a difficult thing to do correctly — price too high and no one will book. Price too low and you’ll be missing out on a lot of potential income.

The project aims at solving this problem by using supervised machine learning algorithm predominantly linear regression.

# THE DATA

The data is already available to us in the form of a Kaggle Dataset at

<https://www.kaggle.com/navaneesh/airbnb#train.csv>

<https://www.kaggle.com/rudymizrahi/airbnb-listings-in-major-us-cities-deloitte-ml#train.csv>

There isn’t a need for additional data mining or web scraping.

The data provided is in the form of CSV files and are listed below:

1. **train.csv**​: The training set of users
2. **test.csv**​: The test set of users.

# APPROACH AND MILESTONES

The approach to solving this problem is subject to change as I progress and learn new concepts and approaches. The tentative broad stages involved are listed below

## Data Wrangling

* The first step would be to load the given datasets and clean them.
* This would be followed by various wrangling methods to arrive at data on which analysis and prediction can be performed.

## Exploratory Data Analysis

* This step would involve creation of visualisations for the cleaned data and try to come up with a few preliminary analysis as well as significant trends and features of the dataset

## Statistical Analysis

* In continuation of step 2 to perform furthur exploratory analysis using statistical methods to identify relationships between different variables.

## Machine Learning

* The next step would be to build a predictive model using supervised learning methods namely regression to solve the requirement and predict Airbnb prices

# DELIVERABLES

The following should be considered as deliverables as part of the project:

1. **Jupyter Notebook**

* Contains all the code involved as part of wrangling, analysis and building predictive models.

1. **Project Report**

* ​A Document highlighting the entire process of the project.

1. **Presentation**

* A Slide Deck to presented to the clients as the final product of the analysis performed and the model built.