**Airbnb Suggested Pricing**

Renting accommodations has become a profitable source of revenue for many house owners whose homes are often vacant. Monetizing property that is otherwise empty has proved to be quite successful on both the supply and demand of the accommodations market. With the growing dynamics of users who want to rent a house for a couple of days and owners who want to rent out their property, it is a challenge to understand how to price the property to fit market needs. We also want to take into account the supply and demand changes with seasonality. Holidays attract more customers and drive higher prices, but it is unclear how much of a premium we should pay per holiday. With better price suggestion estimates, Airbnb home providers can reach an equilibrium price that optimizes profit and affordability. The objective of this project is the predict the optimal price of a property on a specified date.

Milestones:

**1. Project Selection**: Form teams of 2 or 3 and select a project from the provided list.

**2. Literature Study**: Go through the following resources for background and write a half to 1 page summary for each one

Tang, Sangani. "Neighborhood and Price Prediction for San Francisco Airbnb Listings"

- http://cs229.stanford.edu/proj2015/236_report.pdf

A previous project on this topic:

- https://www.mapr.com/blog/predicting-airbnb-listing-prices-scikit-learn-and-apache-spark

**3. Data Exploration and Cleaning:**

1. Data Sources
   - ➔ The primary source of data for this project is Airbnb data from January 2015.
     - ◆ http://data.beta.nyc/dataset/inside-airbnb-data/resource/9d64399b-36d6-40a9-b0bb-f26ae0d9c53f
   - ➔ Another data source layer that can be taken into consideration is the seasonality of pricing depending on holiday seasons
     - ◆ http://data.beta.nyc/dataset/inside-airbnb-data/resource/ce0cbf46-83f9-414a-8a1d-7fd5321d83ca
   - ➔ For advanced analysis, you can also consider the review of the housing listed and perform text analysis to gauge customer satisfaction with the property
     - ◆ http://data.beta.nyc/dataset/inside-airbnb-data/resource/8115833e-8a0e-4af6-8aed-4d96a0ae0b73
1. Setting Up
   - ➔ Decide the computing resource to process the data and gather insights (ie Python, R, etc)

- ➔ Look through data to identify missing values or clean up columns to be ready for analysis
- ➔ Remove rows with lots of missing values or unreliable data

3. Exploratory Data Analysis/Visualization

- ➔ Check for correlations between price and basic features of the home
- ➔ If there are missing values, decide whether to fill these values with 0, 1, global mean, global median, etc.
- ➔ Visualize the target variable to understand skewness and identify transformation that might be necessary
- ➔ Run visualization on each feature against the target variable to identify transformations that can form better linear relationships between each feature and target variable
- ➔ Visualize the supply of Airbnb homes by location and be sure to reflect the price of the average home in each region
- ➔ Visualize high demand dates by exploring relative spikes in prices in each property
    - ◆ Find the difference in pricing of a home against the average pricing of this home over the year to identify high demand dates

**4. Implement Baselines:**

Implement the following baseline techniques

- Ordinary Least Squares Regression:
    - Feature Extraction: Decide on features of a listing that you think might have most intuitive relationship with the pricing
    - Incorporate the most popular holidays (4th July, Christmas, etc) to add an indicator variable to understand the premium of specific holidays
    - Take into account the location of the listing as categorical variables

**5. Proposal:**

Propose methodologies and ideas to be implemented, tested and interpreted for your final project to extend across the baseline.

- Consider clustering similar houses based on size or location to add that as a feature to create a multi-level model to improve accuracy.
- Explore other regressions, such as RandomForestRegressor, Ridge, and Lasso Regressions and tune parameters
- Extract other features that you may think can introduce more predictive power

Whichever methodology you choose, be sure to explain your logic and interpret the results from you finding in a clear and coherent format with appropriate visualizations.