**End Term Project**

Airbnb is a home-sharing platform that allows homeowners 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. Although Airbnb and other sites provide some general guidance, there are currently no free and accurate services that help hosts price their properties using a wide range of data points.

Paid third party pricing software is available, but generally, you are required to put in your own expected average nightly price (‘base price’), and the algorithm will vary the daily price around that base price on each day depending on the day of the week, seasonality, how far away the date is, and other factors.

Airbnb pricing is important to get right, particularly in big cities like New York where there is lots of competition and even small differences in prices can make a big difference. It is also difficult 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.

This project aims to solve this problem by using the predictive modeling technique (Multiple Linear Regression). But before that, the data has to be made tidy, cleaned, understood, and explored.

**Do the following.**

1. Check the missing values of all the variables. Remove the missing values of categorical variables by removing the entire observation. Replace the missing values of numeric variables with the mean.
2. Check the Distribution of all numeric and categorical variables and interpret the results. (Bar plot for Categorical variables and histogram for Numeric Variables)
3. Create a cluster bar plot of variables ‘room\_type’ and ‘city’ and interpret it.
4. Is there any influence of ‘cancellation\_policy’ on the ‘Price’?
5. What is the mean Price of property with different ‘room\_type’ for a different city?
6. Is there any relationship (influence) of Price and ‘Host\_has\_profile\_pic’?
7. Is it true that properties with ‘cleaning fee’ has higher price?
8. Create histograms of Prices of different ‘room\_type’ and interpret.
9. Which ‘city’ has the maximum Price for ‘Private room’?
10. Is there any relationship between ‘review\_score\_rating’ and ‘Price’?
11. Is there any relationship between ‘review\_score\_rating’ and ‘cleaning fee’?
12. Create a model of predicting the price of property with respect to all the independent variables?