Udacity Data Science Nanodegree

Project: Wrting a Blog Post based on Findings in Airbnb Data

**Table of Contents**

* Motivation
* Installation
* File Descriptions
* Results
* Licensing, Authors, and Acknowledgements

**Motivation**

Using Seattle Airbnb dataset, this project aims to answer three business questions of interest using exploratory data analysis and , and machine learning:

Q1 : How are properties distributed among neighboorhood group?

Q2: How many times room types are reviewed? What is the average review score rating? Is there a relationship room type price and the number of review and review score ratings? What is the average highest price for room type in different neighborhoods?

Q3 : Implement linear regression model to apply ML algorithm to forecast price based on variables are selected

**Installation**

This project requires Python 3.x and the following Python libraries installed:

* NumPy
* Pandas
* matplotlib
* seaborn
* scikit-learn
* statsmodels

You will also need to have software installed to run and execute an iPython Notebook

Install Anaconda, a pre-packaged Python distribution that contains all of the necessary libraries and software for this project.

**File Description**

**Project\_Airbnb\_Seattle.ipynb**

A Jupyter Notebook with all the codes following the steps of CRISP-DM for analyzing Airbnb data in Seattle.

**Seattle/listing.csv**

The listings data of Seattle, downloaded from here

**Interacting with project**

The main findings of the code can be found at the post available here.

**Licensing, Authors, and Acknowledgements**

Credit to Airbnb for the open data. All data are downloaded from here. You can find the Licensing for the data and other descriptive information at the Kaggle link available here and here.