An Exploratory Analysis of Airbnb's Data to understand the Rental Landscape in San Francisco.

Aim:

- The Aim of this Project is to perform Exploratory Data Analysis and Visualization with the help of Machine Learning Techniques on Airbnb's Data to gain insights in San Francisco Area.
- All the Results and Visualizations will be displayed on an Interactive Web application.

Technologies:

- **Languages:** Python, Dash, Plotly, HTML, CSS
- **Web Framework :** Flask

Following are the questions that I aim to answer through this Project:

1. How do prices of listings vary by location?
2. How does the demand for Airbnb rentals fluctuate across the year and over years?
3. Are the demand and prices of the rentals correlated?
4. What are the diff erent types of properties in San Francisco?
5. Do they vary by neighborhood?
6. What localities in San Francisco do guests rate highly?
7. What makes a host Super host?
8. Do regular hosts and super hosts have diff erent cancellation and booking policies?
9. Are there any common themes that can be identified from the free-text section of the reviews? What aspects of the rental experience do people like and what aspects do they abhor?
10. To predict the price for each San Francisco neighborhood using listing descriptions?
11. What are the busiest times of the year to visit San Francisco? By how much do prices spike?
12. To uncover trends in reviews of Airbnb visitors to San Francisco?
13. Can we determine a spot-on daily price for a new accommodation that fits into its specific market environment and competitors in San Francisco?
    - The question focuses on the accommodation features and decisions a new host can make about initial presentation, i.e. posting a picture of him- or herself on the website, determining a minimum length of stay, offering instant bookings etc. A machine-learning algorithm will be applied to try to get an answer.

Description of Data:

The data is sourced from the Inside Airbnb website http://insideairbnb.com/get-the-data.html which hosts publicly available data from the Airbnb site.

The dataset comprises of following main tables:

Listings - Detailed listings data-showing 99 attributes for each of the listings. Some of the attributes used in the analysis are price (continuous), longitude (continuous), latitude (continuous), listing type (categorical), is_superhost (categorical), neighborhood (categorical), ratings (continuous) among others.

Reviews - Detailed reviews given by the guests with six attributes. Key attributes include date (date time), listing_id (discrete),

Reviewer_id - (discrete) and comment (textual).

Calendar - Provides details about booking for the next year by listing. Four attributes in total including **listing_id** - (discrete), date (date time), available (categorical) and price (continuous).

Neighbourhoods.csv - Neighborhood list for geo filter. Sourced from city or open source GIS files.

Methodology:

1. Literature Review
2. Obtaining and Viewing the Data ¶
3. Preprocessing the Data
    - 2.1. Deciding which columns to work with
    - 2.2. Cleaning Price Columns
    - 2.3. Dealing with Missing Values
    - 2.4. Feature Engineering
4. Exploratory Data Analysis (EDA)
5. Modeling the Data + Implementation of Machine Learning Techniques
    - 4.1. Preparing Target and Features
    - 4.2. Splitting and Scaling the Data
    - 4.3. Training an XGBoost Regressor
    - 4.4. Cross Validation
6. Interpreting the Results
7. Developing an Interactive Web application to display the Results.