

Name: Shahad Alameel

Email: Shahad.a@outlook.com

Github: ShahadAlameel

Project Proposal

Background:

The goal of this project is to apply Linear Regression on data scraped from Airbnb website, to predict the price of a stay based on the features of that stay (number of beds, rating, etc.).

Data Description:

Airbnb, Inc. is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. Based in San Francisco, California, the platform is accessible via website and mobile app.

Sample size is 4 major cities (New York, Los Angeles, Chicago, and Sad Diego) that are located in the United States.

The dataset has 1440 rows and 10 columns.

Column names are:

- **stay type:** The type of the stay and where it is located.
- **ad title:** The title of the advertise on the website listing.
- **guests:** The number of guests allowed in the stay.
- **bedrooms:** The number of bedrooms.
- **beds:** The number of beds.
- **baths:** The number of baths.
- **rating:** The rating of the stay in a scale from 0 to 5.
- **number of reviews:** The number of comments and reviews of the stay.
- **price:** The price of the stay per night.
- **city:** The city where the stay is located.

Tools:

- Technologies: Python, Jupyter notebook.
- Libraries: Numpy, Pandas, Pandas profiling, Matplotlib, Seaborn, Sklearn, BeautifulSoup, patsy, selenium.