**Data 602 – Final Project Proposal**

**By: Mubashira Qari**

**Project Title:** **Airbnb Data Analysis and Price Prediction Using Machine Learning**

**Describing Dataset:** This is an Airbnb listing dataset. This dataset is acquired from the Airbnb website. It has details about the host, location, type of listing, and reviews, along with more interesting features in different columns. It consists of (18337, 18) rows and columns and needs thorough data cleaning. The analysis will provide interesting insights and guidelines on choosing Airbnb for your next vacation trip.The metrics for the columns in the dataset are as below:

id float64 Airbnb's unique identifier for the host/user

name object name of the listing

host\_id int64 unique identifier for the host

host\_name object name of the host

neighbourhood\_group float64 next town zip code

neighbourhood int64 Field that can be filled using the lat & Long

latitude float64 latitude

longitude float64 longitude

room\_type object type of rental space

price int64 price

minimum\_nights int64 minimum nights available

number\_of\_reviews int64 count of reviews

last\_review object date of last review

reviews\_per\_month float64 per month reviews

calculated\_host\_listings\_count int64 total host's listing count

availability\_365 int64 Availability for the number of days

number\_of\_reviews\_ltm int64 Count of reviews in last 12 months

license permits for listing

**Data Source Link:**

Data Source: <http://insideairbnb.com/get-the-data/>

Data Dictionary: <https://docs.google.com/spreadsheets/d/1iWCNJcSutYqpULSQHlNyGInUvHg2BoUGoNRIGa6Szc4/edit#gid=1322284596>

**Justification for Dataset Selection:**

Whenever it comes to traveling, there is always a search for a good rating Airbnb listing. There are many important factors that need to consider. For example, the room type, comfort rating, activity, top hosts, rent, and many others. And paying close attention to your selection adds a lot to the enjoyment of the trip. For this reason, I decided to choose the Airbnb dataset.Now there are couple of questions related to electric vehicle that require analysis and those are my research questions.

**Research Questions & Objectives:**

The research questions for the electric vehicle’s dataset are the following:

1- Which listing has the best reviews?

2- What is the availability in days?

3- What is the type of room?

4- What is the correlation between the variables.

5- What is the price of the listing?

5- What are the description of the listing?

6- Which factors affect the price of the listing?

7- Upload the dataset in the SQL database for further analysis.

8- Predict the price of Airbnb listing using machine learning.

**Libraries Used for Project Implementation:**

* Python Pandas
* Python NumPy
* Python Matplotlib
* SQL Database
* Python sklearn
* Python seaborn

**EDA and Summary Statistics:**

Below are the images of exploratory data analysis and summary statistics:

A picture containing table

Description automatically generated

A picture containing graphical user interface

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated with medium confidence

Chart, scatter chart

Description automatically generated

Chart, bar chart

Description automatically generated