**Data 602 – Final Project Proposal**

**By: Mubashira Qari**

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

**Describing Dataset:** This dataset is about electric vehicles. It has 103 row and 14 columns.

This dataset requires detailed data cleaning before performing the data manipulation steps.

The metrics for the columns in the dataset are as below:

Vehicle Battery capacity (in KWH)

Acceleration (0-100) in Seconds

Top Speed in Km/hr

Range of Vehicle in km

The efficiency of Vehicle in Wh/km

Fast charge speed in km/hr

Drive Configuration

Number of seats

Price in Germany (Euro)

Price in the UK (Pound)

**Data Source Link:**

This dataset is acquired from Kaggle website, and the link is provided below:

<https://www.kaggle.com/datasets/geoffnel/evs-one-electric-vehicle-dataset?select=ElectricCarData_Norm.csv>

<https://www.kaggle.com/datasets/kkhandekar/cheapest-electric-cars>

**Justification for Dataset Selection:**

There are multiple reasons for choosing this dataset.

First, I eagerly wanted to do the research about the electric vehicles because I am interested in buying one for myself. In terms of carbon emission, Electric cars are good for the environment by preventing greenhouse gas emission.

The operating and maintenance costs are lot lower for electric vehicles and that makes them more economical and viable in the long run.

Initially, one of the biggest challenges is range anxiety and uncomfortable feeling of not finding the charging stations within the battery range. These concerns are being taken care by the long-range batteries hitting the market and GPS apps, providing the location of the nearest charging station. And people are more likely to buy electric vehicles.

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 car has the fastest charging battery capacity?

2- Which vehicle has the highest efficiency?

3- Does a difference in power train effect the range, top speed, efficiency?

4- What is the correlation between the variables.

5- Which vehicle has the most range?

5- Which manufacturer has the greatest number of vehicles?

6- Which factors affect the price of the vehicle?

7- Merge the two datasets.

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

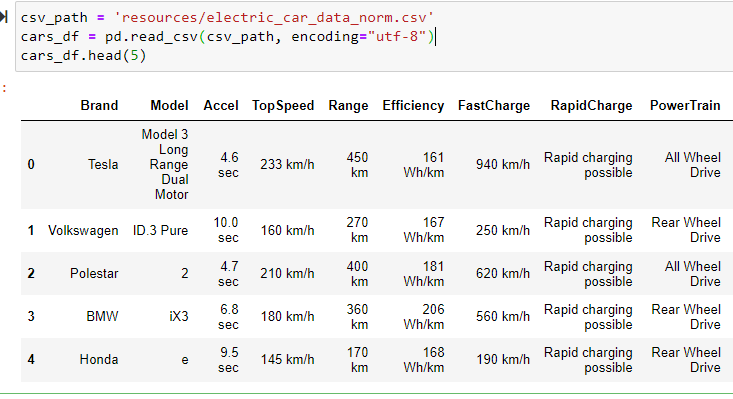
9- Predict the price of electric vehicles using machine learning.

**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:



Table

Description automatically generated

Text, table

Description automatically generated with medium confidence

Graphical user interface, text, application

Description automatically generated

Text

Description automatically generated

Table

Description automatically generated

