**Name - Aryan Kakran**

**EV Market Segmentation**

**Objective**

The objective of this report is to analyse the electric vehicle (EV) market using a dataset containing details of vehicle registrations segmented by fuel type, state, and year. The focus is on understanding trends in EV adoption, comparing it with other fuel types, and identifying key growth areas.

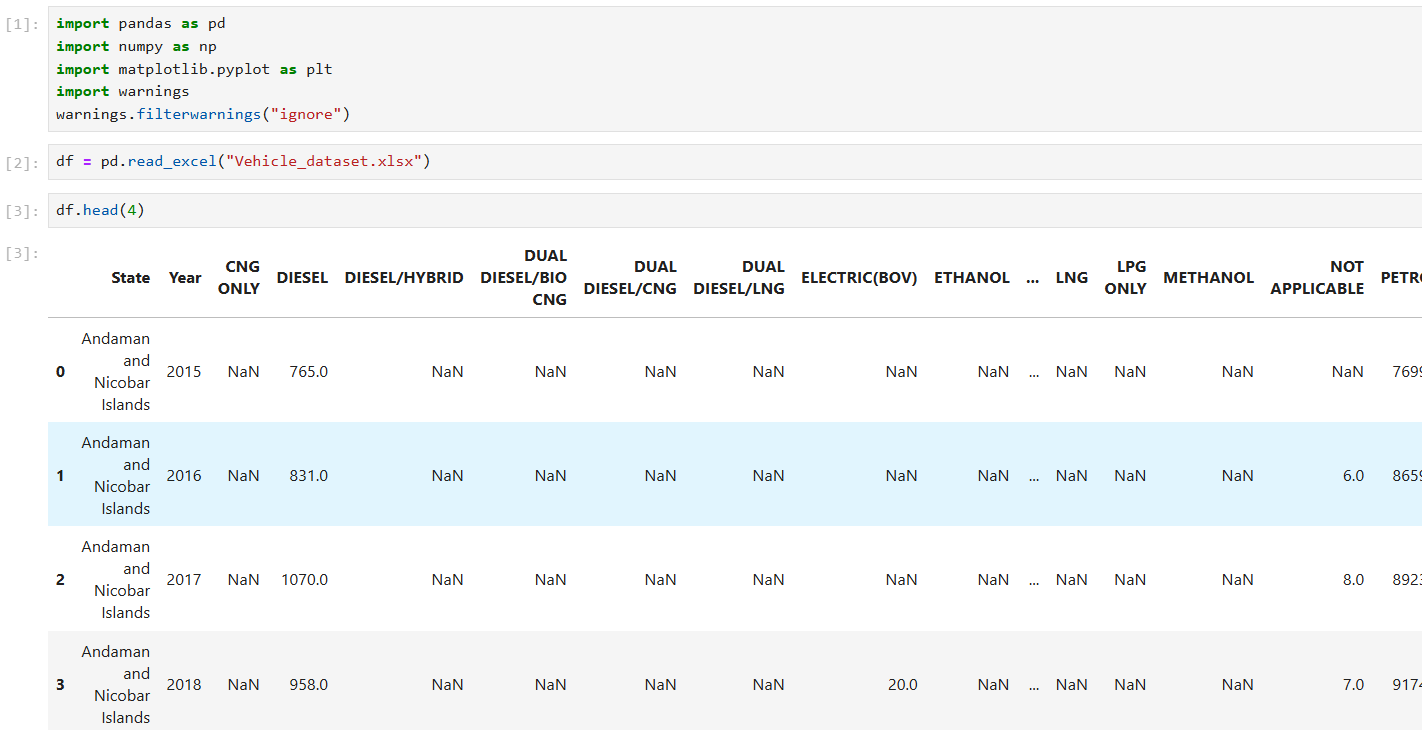
**Dataset Overview**

* **Data Structure**:
  + **Rows**: 333
  + **Columns**: 21
* **Key Features**:
  + **Fuel Types**: Includes categories such as CNG ONLY, DIESEL, ELECTRIC(BOV), PETROL, etc.
  + **States**: Records data across multiple Indian states.
  + **Years**: A data is from 2015-23.

**Data Preprocessing: Required Libraries**

To conduct EDA and clustering on the collected data, the following Python libraries are utilized**: -**

1. Pandas: For data handling and manipulation.
2. Matplotlib & Seaborn: For creating insightful visualizations.
3. Scikit-learn: For clustering (e.g., k-means) and preprocessing tasks like scaling and normalization.



**Exploratory Data Analysis**

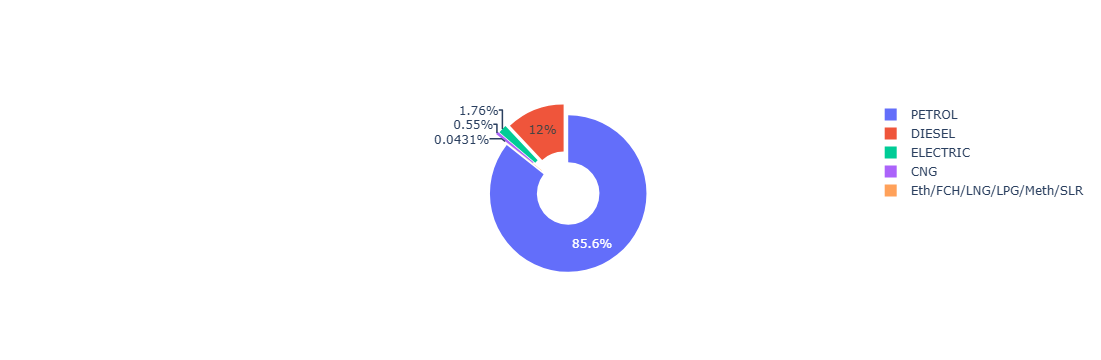
Exploratory Data Analysis, commonly referred to as EDA, is a critical step in the data science workflow. It involves uncovering valuable insights within the dataset using summary statistics and visualizations. The core aspects of this technique are highlighted in the image below.

A diagram of a data analysis

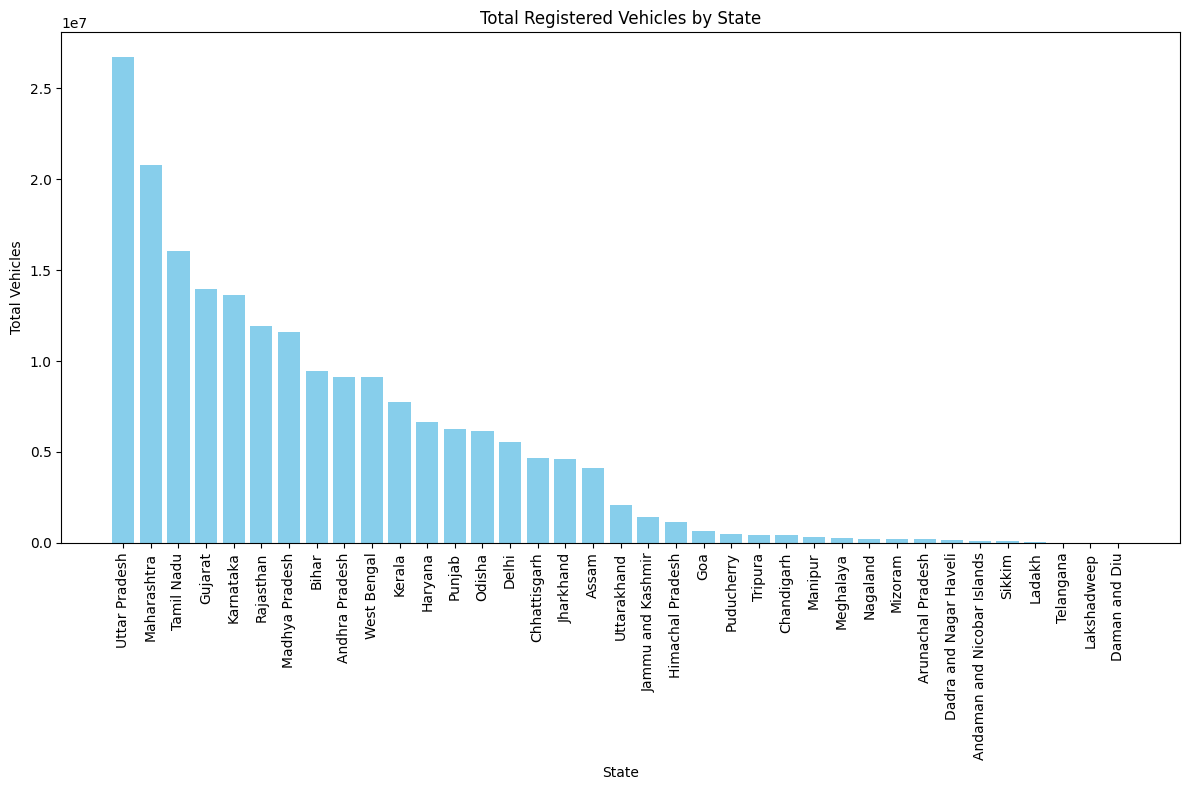
Description automatically generated

**Implementing EDA on the datasets**

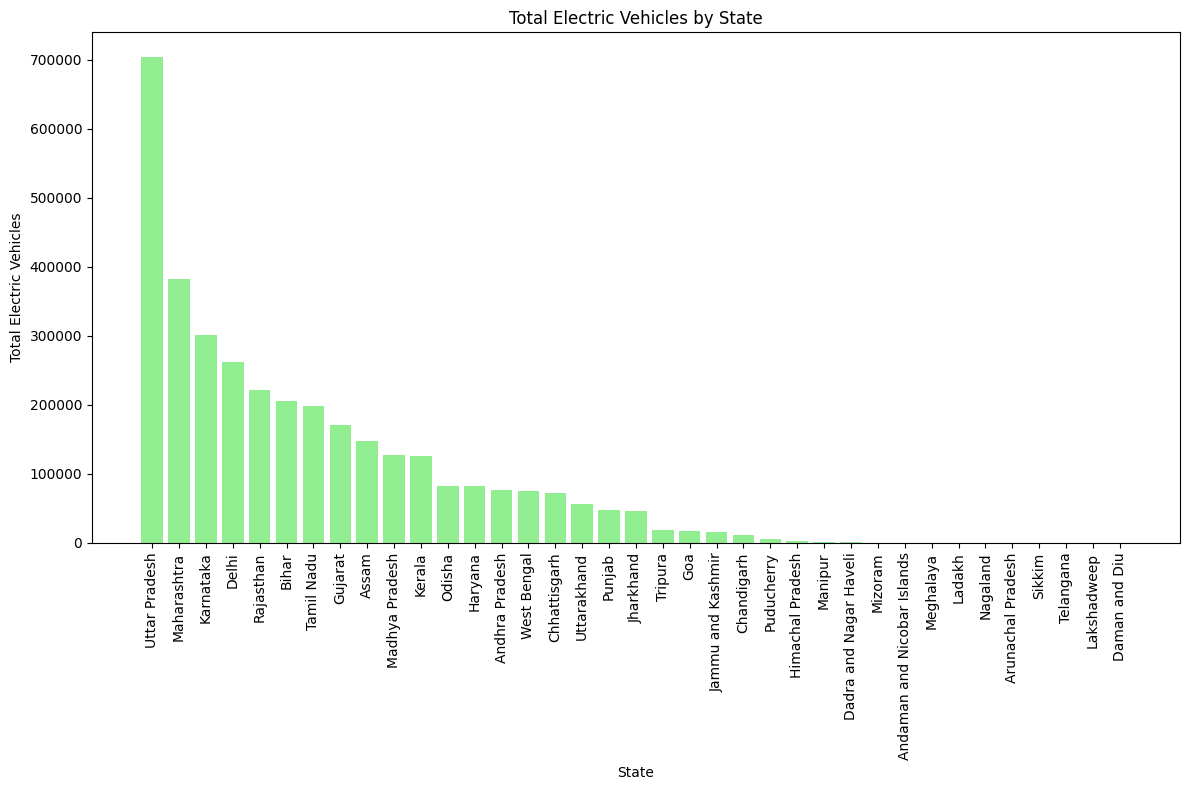
Pie chart of percentage of vehicles by fuels:

****

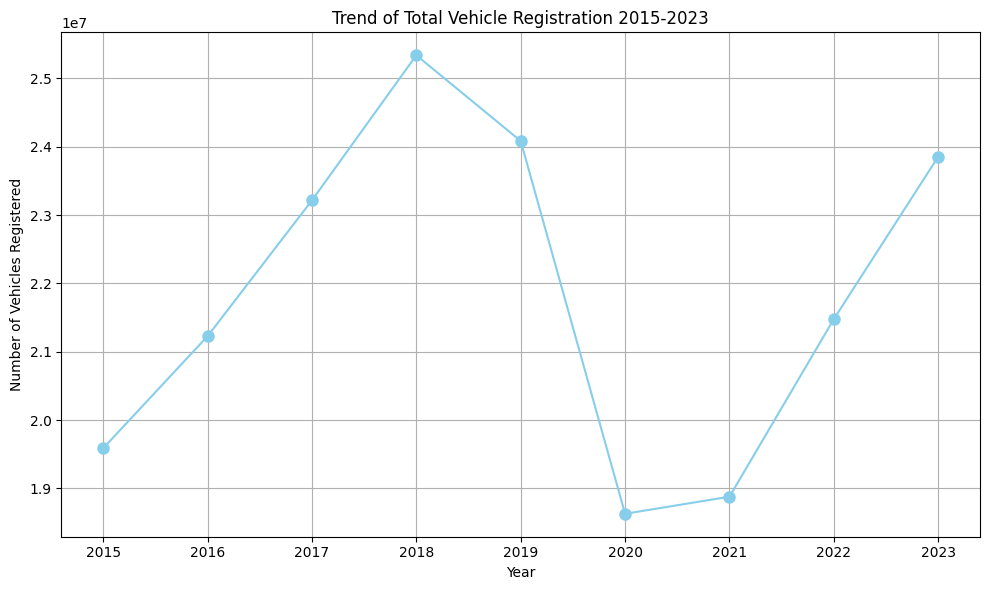
Total Registered Vehicles by State:



Total Electric Vehicles by State:



Trend of Total Vehicle Registration 2015-2023:



Trend of Electric Vehicle Registration 2015-2023:

A graph with a line going up

Description automatically generated

State-wise Vehicle Registration with fuel type:

A graph with different colored bars

Description automatically generated with medium confidence

State-wise Electric Vehicle Registration Year-wise Stack:

A graph of different colored lines

Description automatically generated

**Segmentation Approaches**

**Clustering**

Clustering is an unsupervised machine learning technique used to group similar data points into distinct clusters. The primary goal of clustering is to identify and separate data points with similar characteristics into different clusters. Common clustering algorithms include:

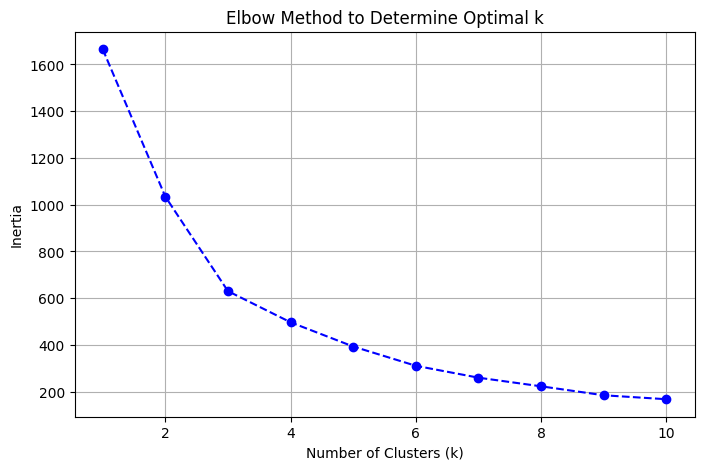
* K-Means Clustering
* Hierarchical Clustering
* Density-Based Clustering

**K-Means Clustering**

K-Means is a popular unsupervised algorithm that groups an unlabelled dataset into distinct clusters, ensuring each data point belongs to only one cluster. The "K" represents the number of clusters to be formed.

**Elbow Method**

The Elbow method is a way of determining the optimal number of clusters (k) in K-Means Clustering. It is based on calculating the Within Cluster Sum of Squared Errors (WCSS) for a different number of clusters (k) and selecting the k for which change in WCSS first starts to diminish. When you plot its graph, at one point the line starts to run parallel to the X-axis and that point, known as the Elbow Point, is considered as the best value for the k.



**Clustering Result: -**

A graph with colored dots

Description automatically generated

Github Link :