Electric Vehicle Market Segmentation Analysis

Pinki

11-12-2024

Data pre-processing

Required Libraries

In order to perform EDA and clustering on the collected data, the following python libraries are used:

1. Pandas: for data handling/manipulation
2. Numpy: for numerical calculation
3. Matplotlib and Seaborn: For data visualization
4. Sci-kit learn: for the k-means clustering algorithm and other algorithms

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.cluster import KMeans
```

Pulling the datasets

```
# Fetching Dataset-1
df1 = pd.read_csv('EVStats.csv')
df1.head()
```

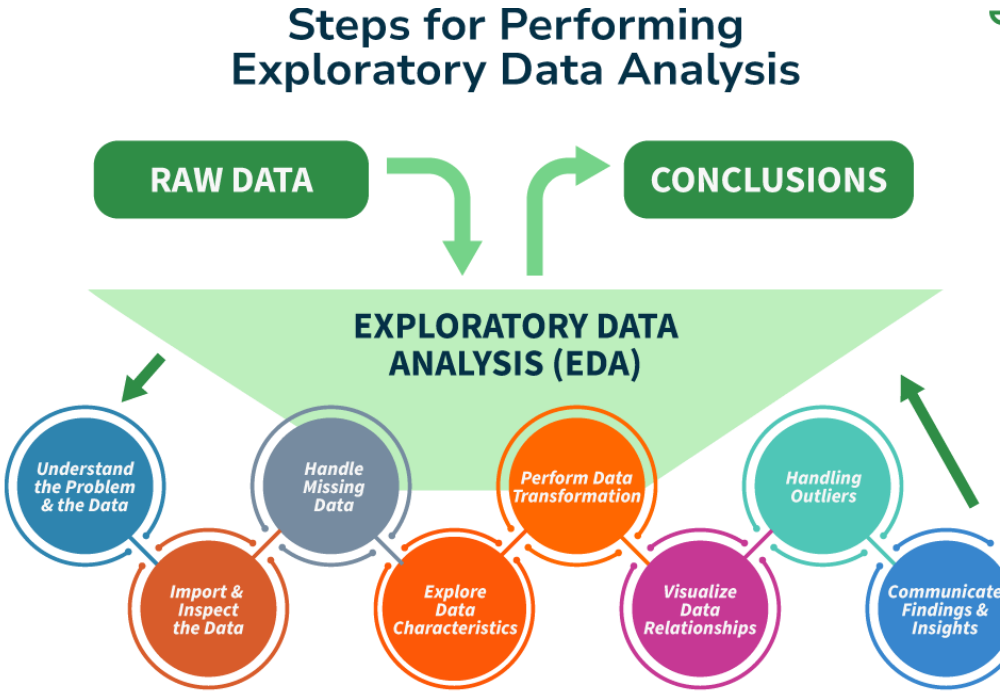
	Sl. No	State	Two Wheelers (Category L1 & L2 as per Central Motor Vehicles Rules	Two Wheelers (Category L2 (CMVR))	Two Wheelers (Max power not exceeding 250 Watts)	Three Wheelers (Category L5 slow speed as per CMVR)	Three Wheelers (Category L5 as per CMVR)	Passenger Cars (Category M1 as per CMVR)	Buses	Total in state
0	1	Meghalaya	0	0	0	0	0	6	0	6
1	2	Nagaland	0	20	3	0	0	1	0	24
2	3	Manipur	16	8	11	0	5	12	0	52
3	4	Tripura	28	9	36	0	0	8	0	81
4	5	Andaman & Nicobar islands	0	0	0	0	0	82	0	82

```
# Fetching Dataset-2
df2 = pd.read_csv('behavioural_segment_data.csv')
df2.head()
```

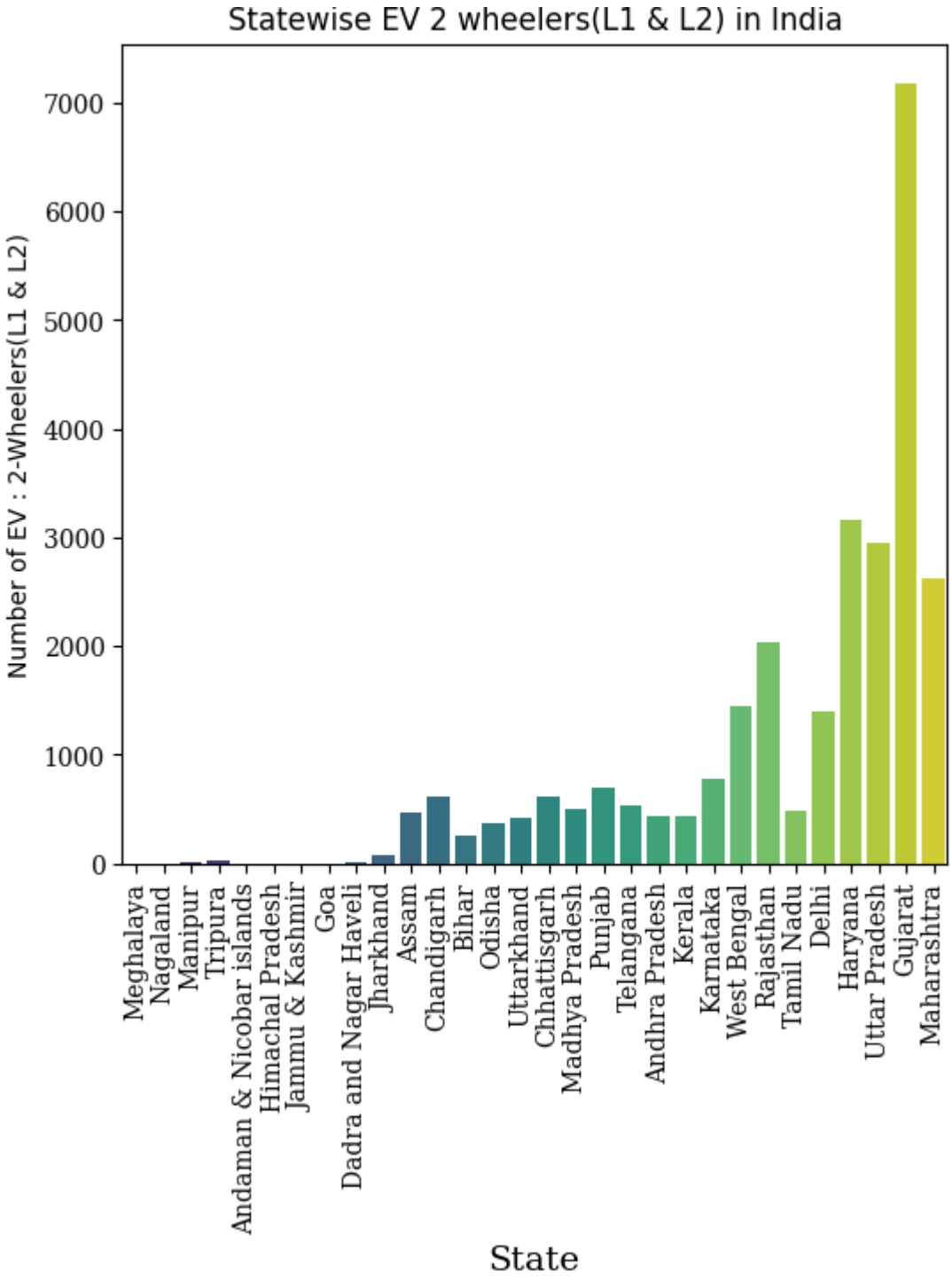
	Age	Profession	Marrital Status	Education	No of Dependents	Personal loan	Total Salary	Price
0	27	Salaried	Single	Post Graduate	0	Yes	800000	800000
1	35	Salaried	Married	Post Graduate	2	Yes	2000000	1000000
2	45	Business	Married	Graduate	4	Yes	1800000	1200000
3	41	Business	Married	Post Graduate	3	No	2200000	1200000
4	31	Salaried	Married	Post Graduate	2	Yes	2600000	1600000

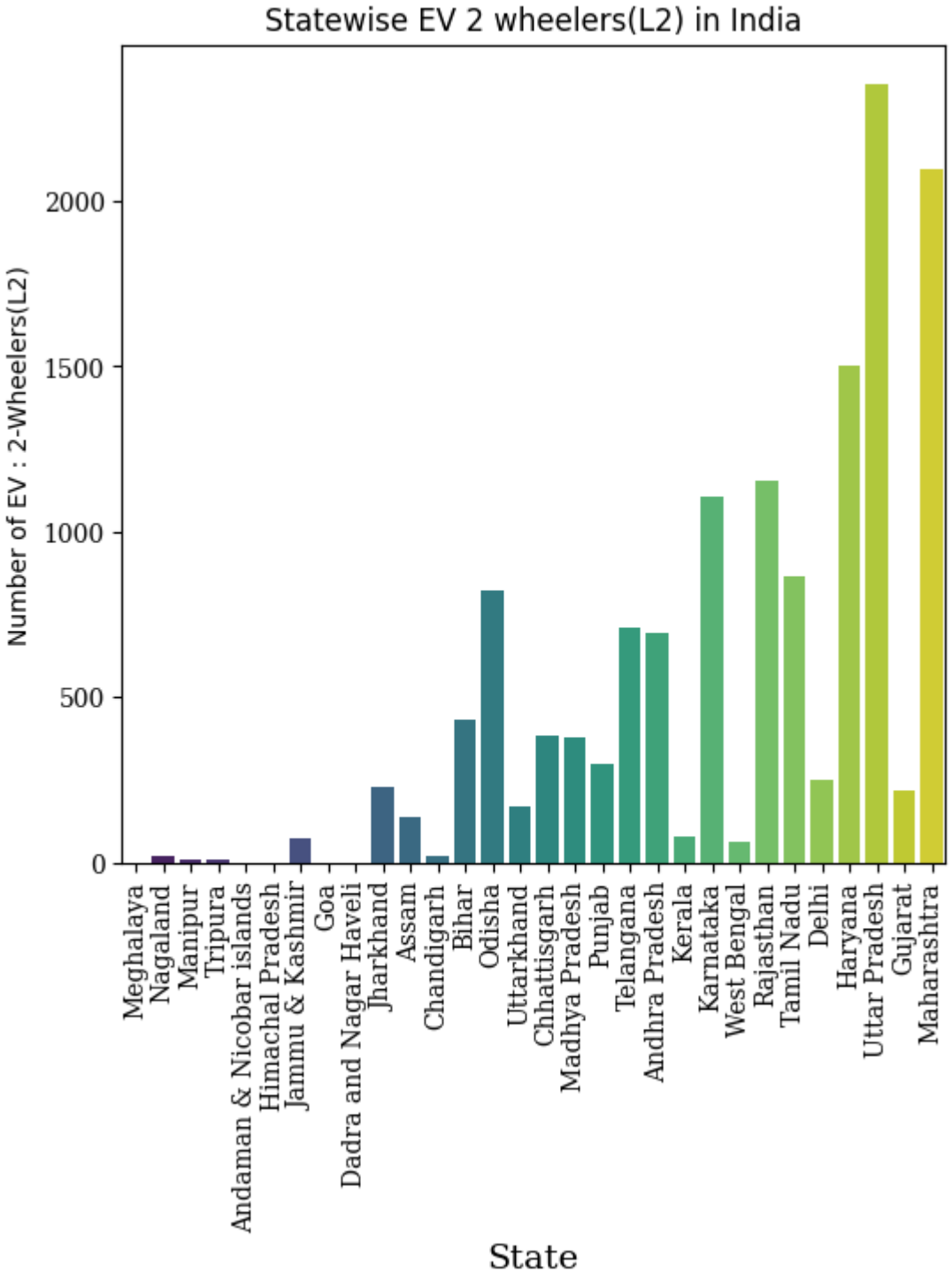
Exploratory Data Analysis

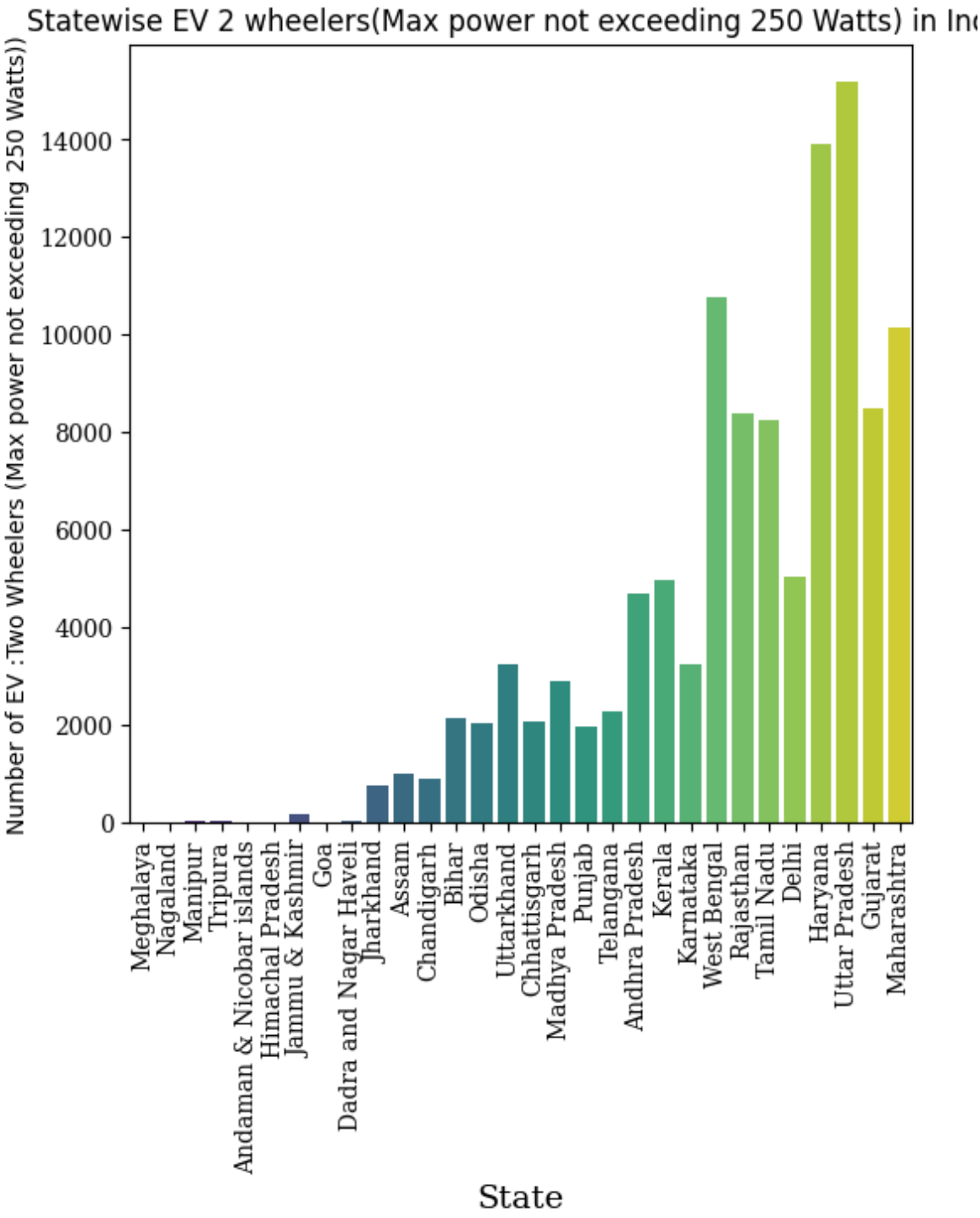
It is popularly abbreviated as EDA, is the major step in the data science pipeline. It is the process of getting insights of data with the help of visual representations and statistical summary.

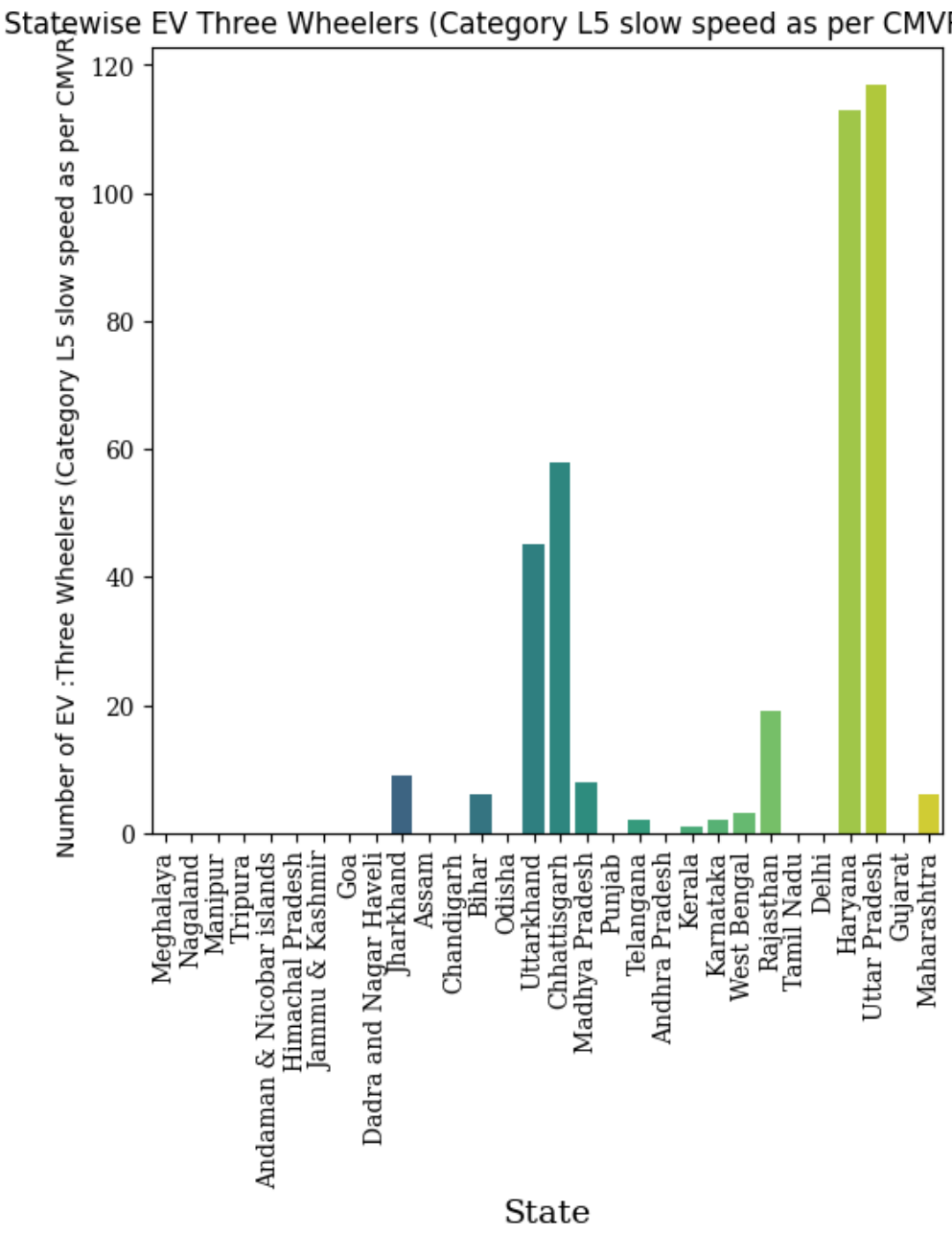


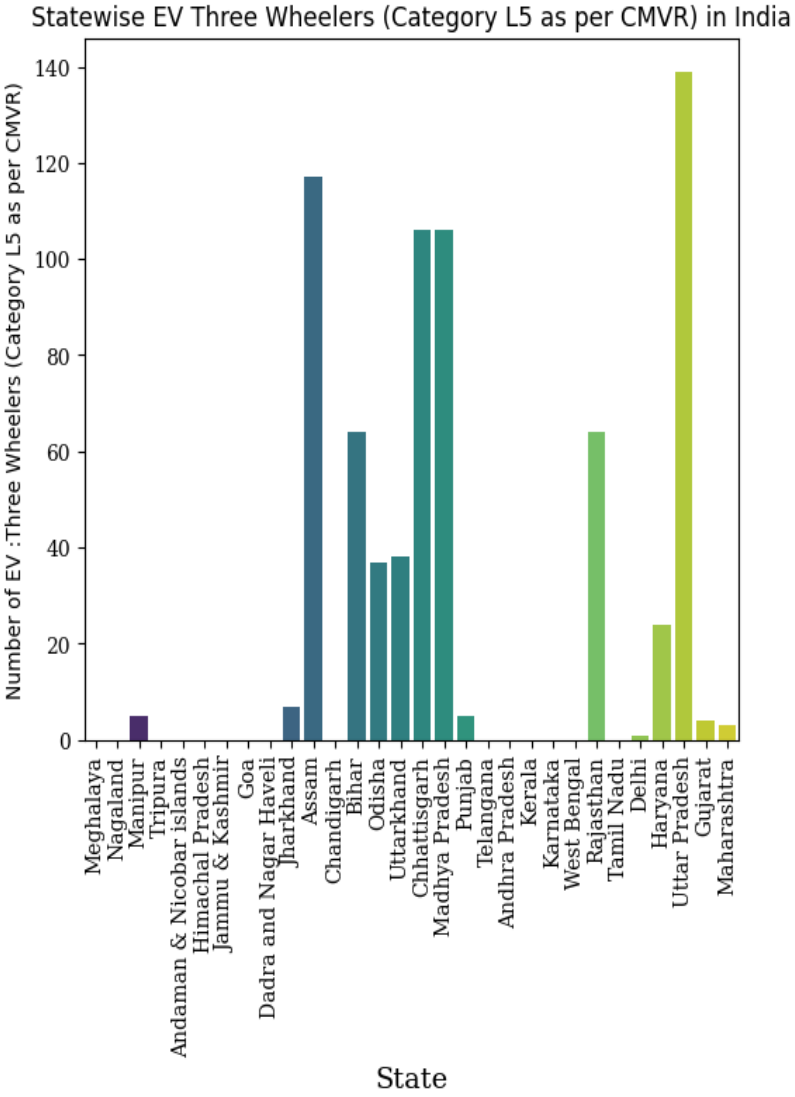
Implementing EDA on the datasets

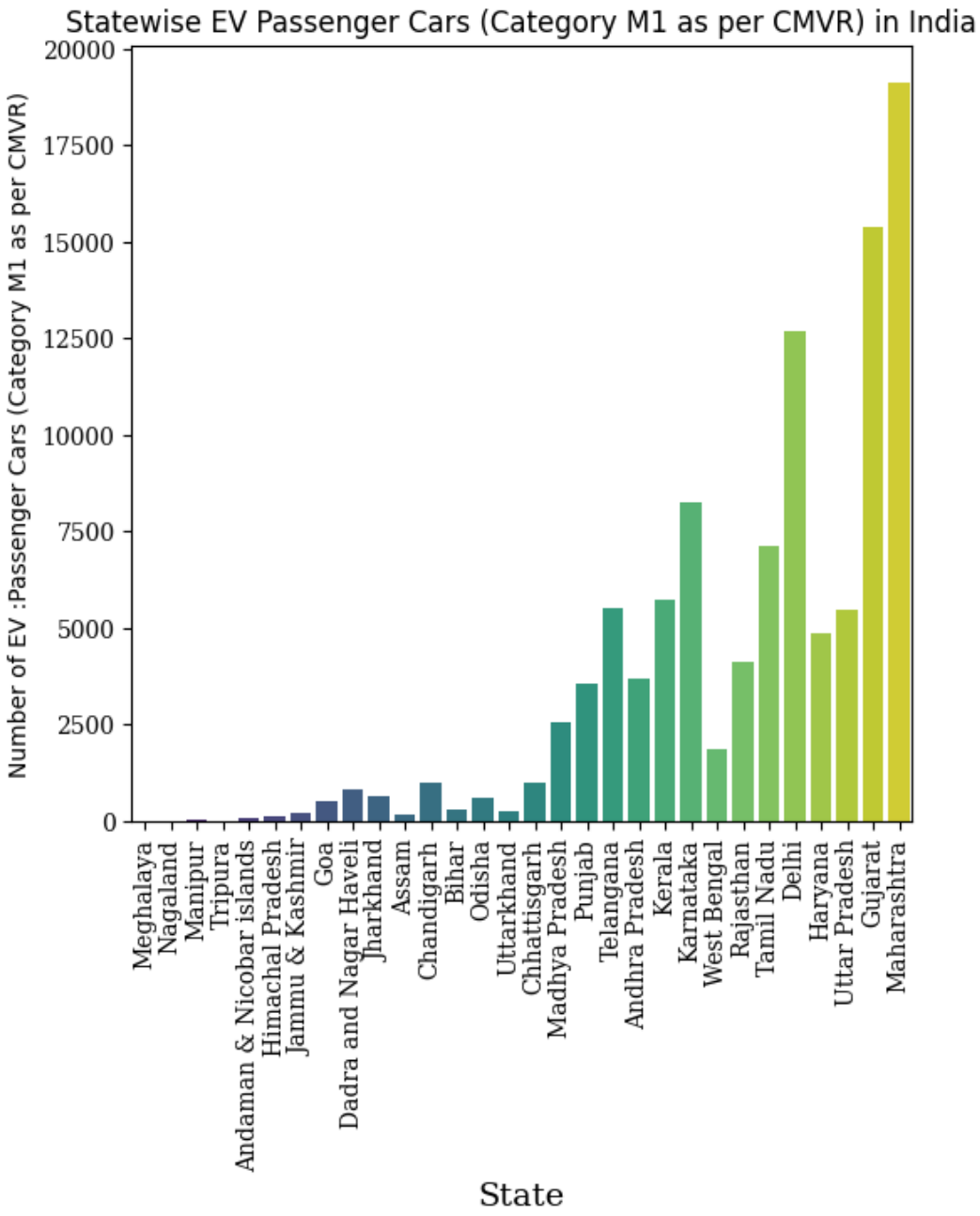

Statewise EV 2 wheelers(L1 & L2) in India
Number of EV : 2-Wheelers(L1 & L2)
0
1000
2000
3000
4000
5000
6000
7000
Meghalaya
Nagaland
Manipur
Tripura
Andaman & Nicobar islands
Himachal Pradesh
Jammu & Kashmir
Goa
Dadra and Nagar Haveli
Jharkhand
Assam
Chandigarh
Bihar
Odisha
Uttarkhand
Chhattisgarh
Madhya Pradesh
Punjab
Telangana
Andhra Pradesh
Kerala
Karnataka
West Bengal
Rajasthan
Tamil Nadu
Delhi
Haryana
Uttar Pradesh
Gujarat
Maharashtra
State

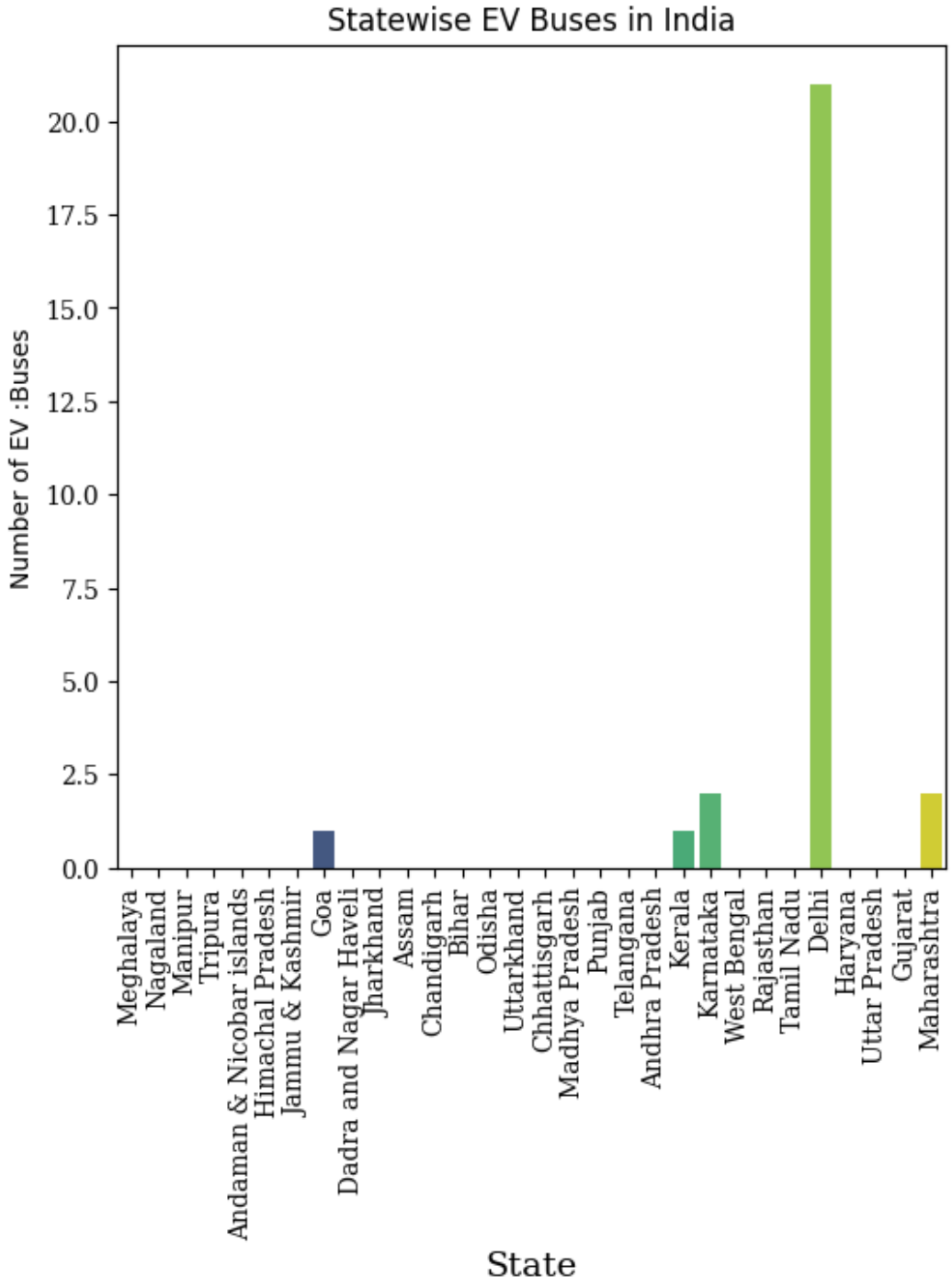

Statewise EV 2 wheelers(L2) in India
Number of EV : 2-Wheelers(L2)
0
500
1000
1500
2000
Meghalaya
Nagaland
Manipur
Tripura
Andaman & Nicobar islands
Himachal Pradesh
Jammu & Kashmir
Goa
Dadra and Nagar Haveli
Jharkhand
Assam
Chandigarh
Bihar
Odisha
Uttarkhand
Chhattisgarh
Madhya Pradesh
Punjab
Telangana
Andhra Pradesh
Kerala
Karnataka
West Bengal
Rajasthan
Tamil Nadu
Delhi
Haryana
Uttar Pradesh
Gujarat
Maharashtra
State

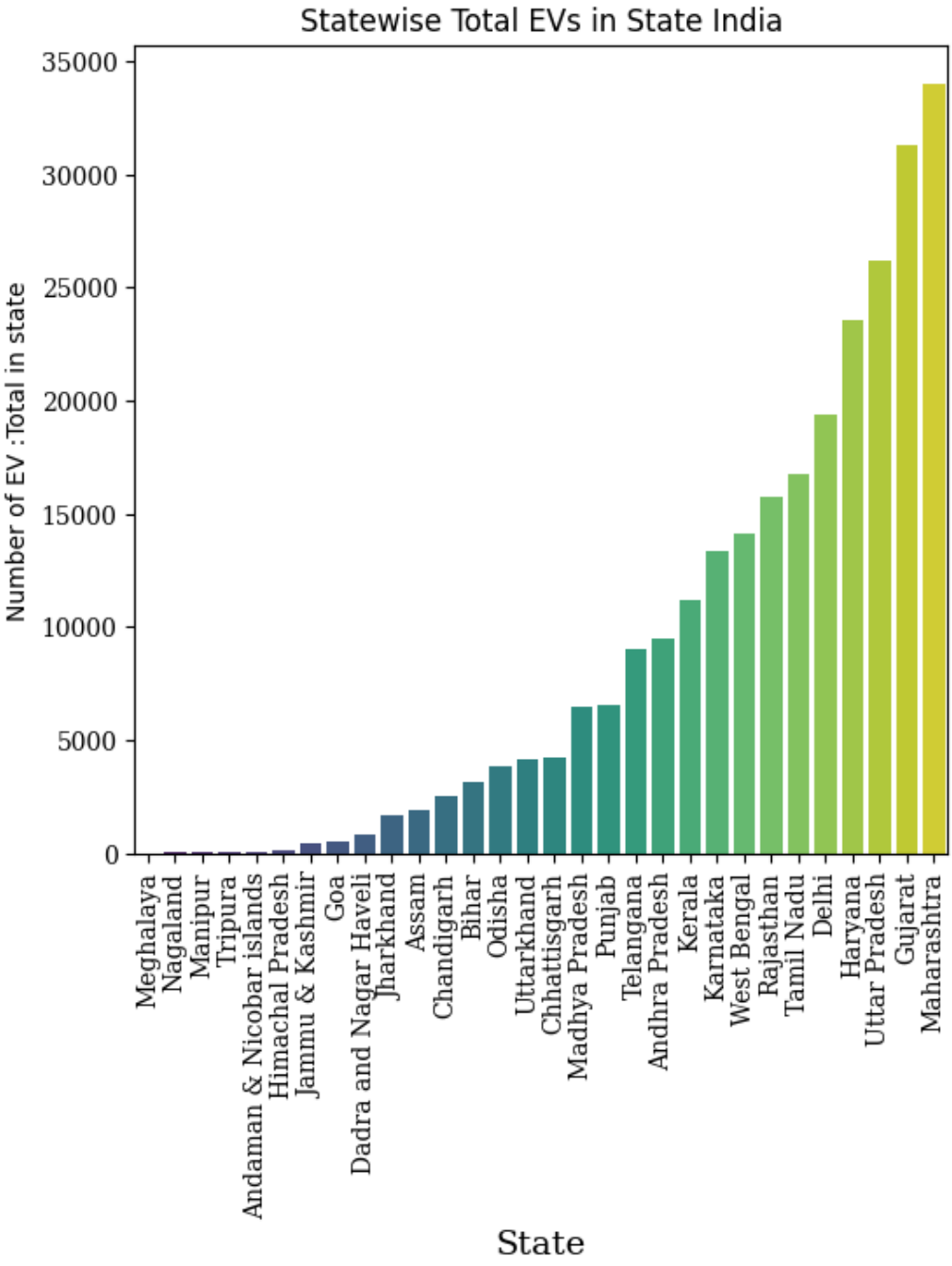

Statewise EV 2 wheelers(Max power not exceeding 250 Watts) in In
Number of EV :Two Wheelers (Max power not exceeding 250 Watts))
0
2000
4000
6000
8000
10000
12000
14000
Meghalaya
Nagaland
Manipur
Tripura
Andaman & Nicobar islands
Himachal Pradesh
Jammu & Kashmir
Goa
Dadra and Nagar Haveli
Jharkhand
Assam
Chandigarh
Bihar
Odisha
Uttarkhand
Chhattisgarh
Madhya Pradesh
Punjab
Telangana
Andhra Pradesh
Kerala
Karnataka
West Bengal
Rajasthan
Tamil Nadu
Delhi
Haryana
Uttar Pradesh
Gujarat
Maharashtra
State

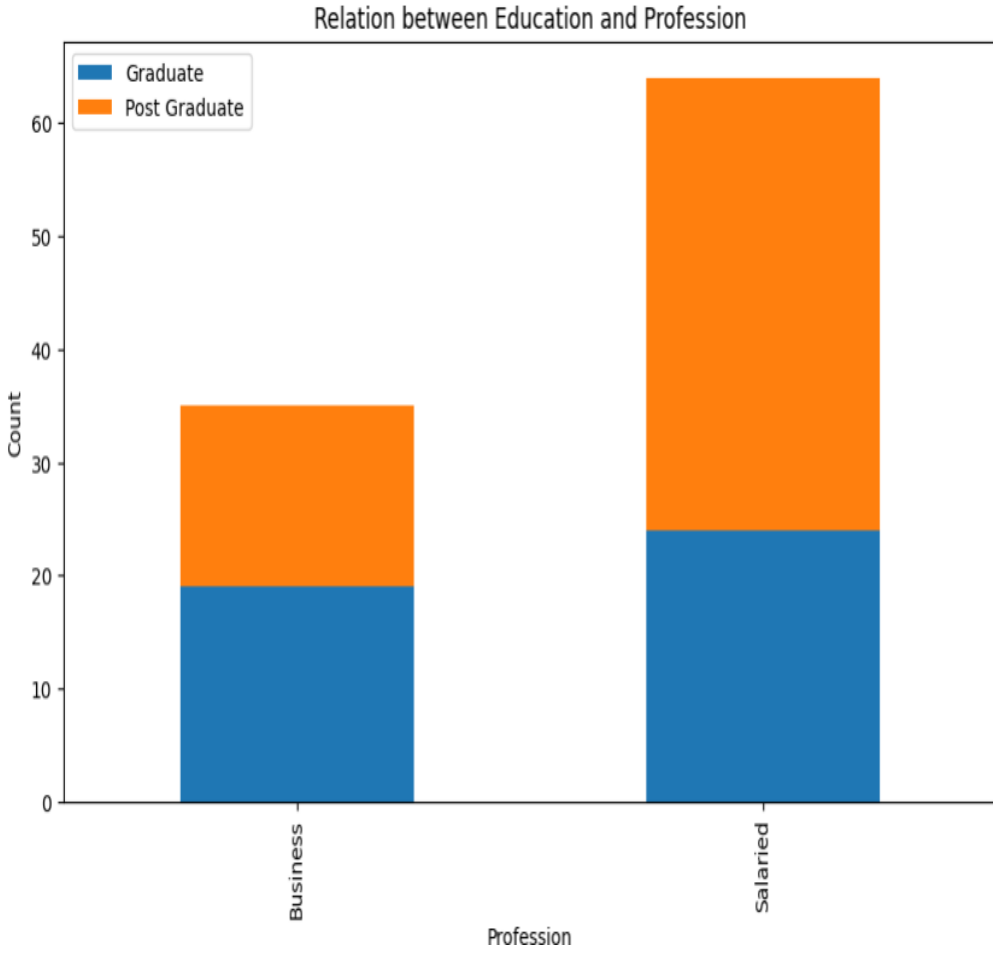

Statewise EV Three Wheelers (Category L5 slow speed as per CMVR) in India
Number of EV :Three Wheelers (Category L5 slow speed as per CMVR)
0
20
40
60
80
100
120
Meghalaya
Nagaland
Manipur
Tripura
Andaman & Nicobar islands
Himachal Pradesh
Jammu & Kashmir
Goa
Dadra and Nagar Haveli
Jharkhand
Assam
Chandigarh
Bihar
Odisha
Uttarkhand
Chhattisgarh
Madhya Pradesh
Punjab
Telangana
Andhra Pradesh
Kerala
Karnataka
West Bengal
Rajasthan
Tamil Nadu
Delhi
Haryana
Uttar Pradesh
Gujarat
Maharashtra
State

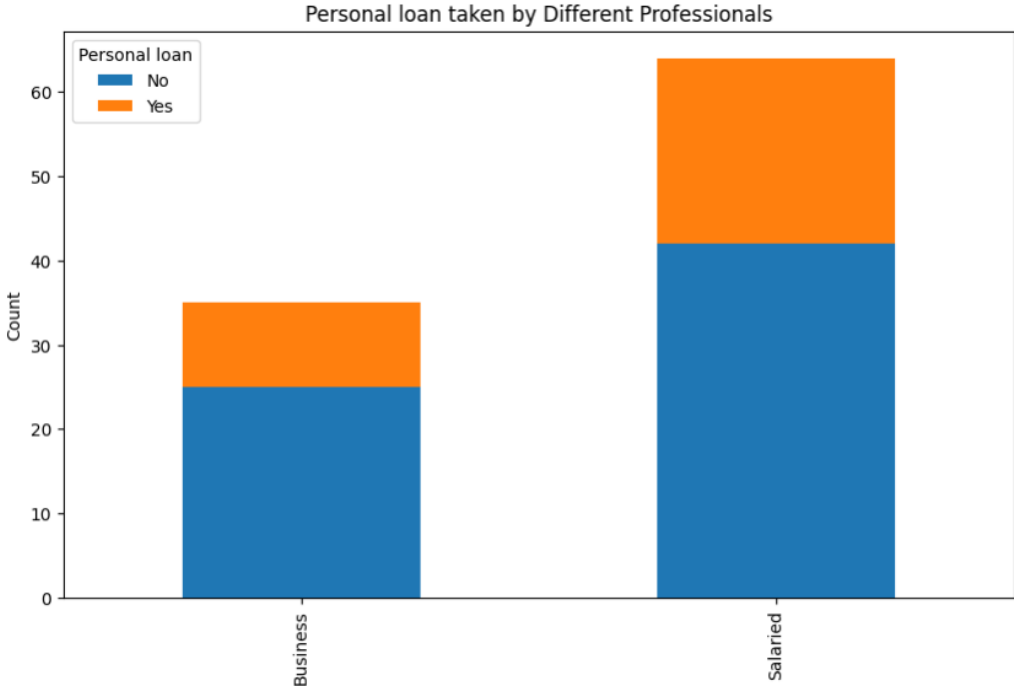

Statewise EV Three Wheelers (Category L5 as per CMVR) in India
Number of EV :Three Wheelers (Category L5 as per CMVR)
0
20
40
60
80
100
120
140
Meghalaya
Nagaland
Manipur
Tripura
Andaman & Nicobar islands
Himachal Pradesh
Jammu & Kashmir
Goa
Dadra and Nagar Haveli
Jharkhand
Assam
Chandigarh
Bihar
Odisha
Uttarkhand
Chhattisgarh
Madhya Pradesh
Punjab
Telangana
Andhra Pradesh
Kerala
Karnataka
West Bengal
Rajasthan
Tamil Nadu
Delhi
Haryana
Uttar Pradesh
Gujarat
Maharashtra
State


Statewise EV Passenger Cars (Category M1 as per CMVR) in India
Number of EV :Passenger Cars (Category M1 as per CMVR)
20000
17500
15000
12500
10000
7500
5000
2500
0
Meghalaya
Nagaland
Manipur
Tripura
Andaman & Nicobar islands
Himachal Pradesh
Jammu & Kashmir
Goa
Dadra and Nagar Haveli
Jharkhand
Assam
Chandigarh
Bihar
Odisha
Uttarkhand
Chhattisgarh
Madhya Pradesh
Punjab
Telangana
Andhra Pradesh
Kerala
Karnataka
West Bengal
Rajasthan
Tamil Nadu
Delhi
Haryana
Uttar Pradesh
Gujarat
Maharashtra
State


Statewise EV Buses in India
Number of EV :Buses
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Meghalaya
Nagaland
Manipur
Tripura
Andaman & Nicobar islands
Himachal Pradesh
Jammu & Kashmir
Goa
Dadra and Nagar Haveli
Jharkhand
Assam
Chandigarh
Bihar
Odisha
Uttarkhand
Chhattisgarh
Madhya Pradesh
Punjab
Telangana
Andhra Pradesh
Kerala
Karnataka
West Bengal
Rajasthan
Tamil Nadu
Delhi
Haryana
Uttar Pradesh
Gujarat
Maharashtra
State

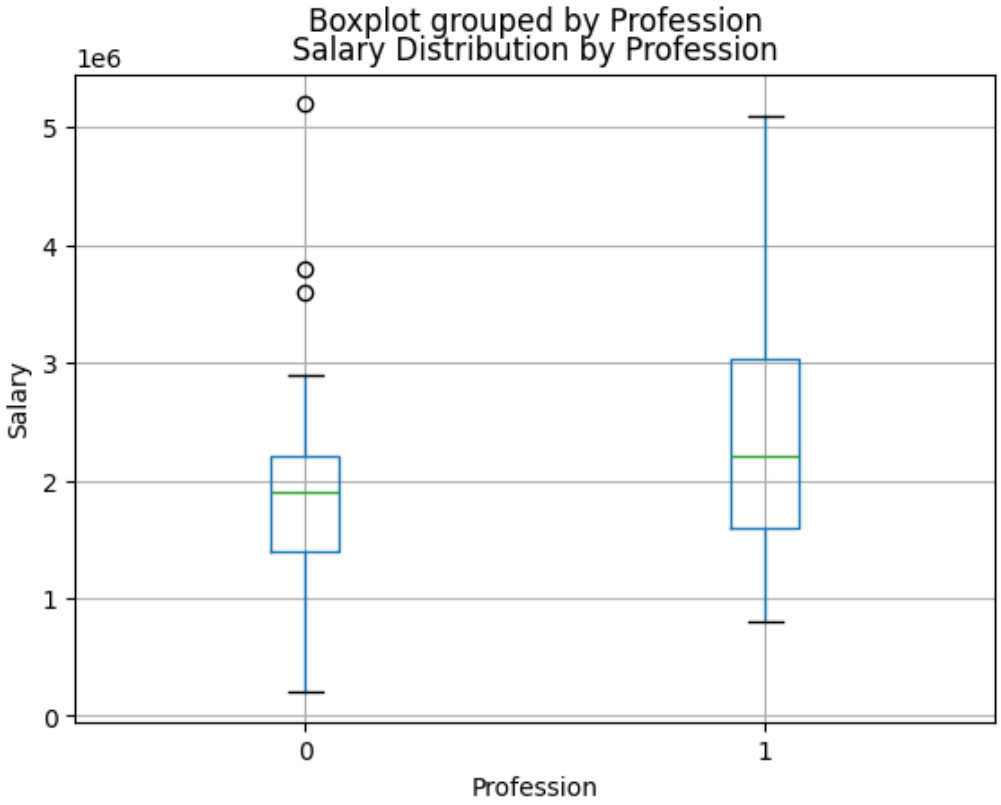

Statewise Total EVs in State India
Number of EV :Total in state
35000
30000
25000
20000
15000
10000
5000
0
Meghalaya
Nagaland
Manipur
Tripura
Andaman & Nicobar islands
Himachal Pradesh
Jammu & Kashmir
Goa
Dadra and Nagar Haveli
Jharkhand
Assam
Chandigarh
Bihar
Odisha
Uttarkhand
Chhattisgarh
Madhya Pradesh
Punjab
Telangana
Andhra Pradesh
Kerala
Karnataka
West Bengal
Rajasthan
Tamil Nadu
Delhi
Haryana
Uttar Pradesh
Gujarat
Maharashtra
State


Relation between Education and Profession
Graduate
Post Graduate
Count
0
10
20
30
40
50
60
Business
Salaried
Profession


Personal loan taken by Different Professionals
Personal loan
No
Yes
Count
0
10
20
30
40
50
60
Business
Salaried

Profession

<Figure size 1000x600 with 0 Axes>



Segmentation Approaches

Clustering

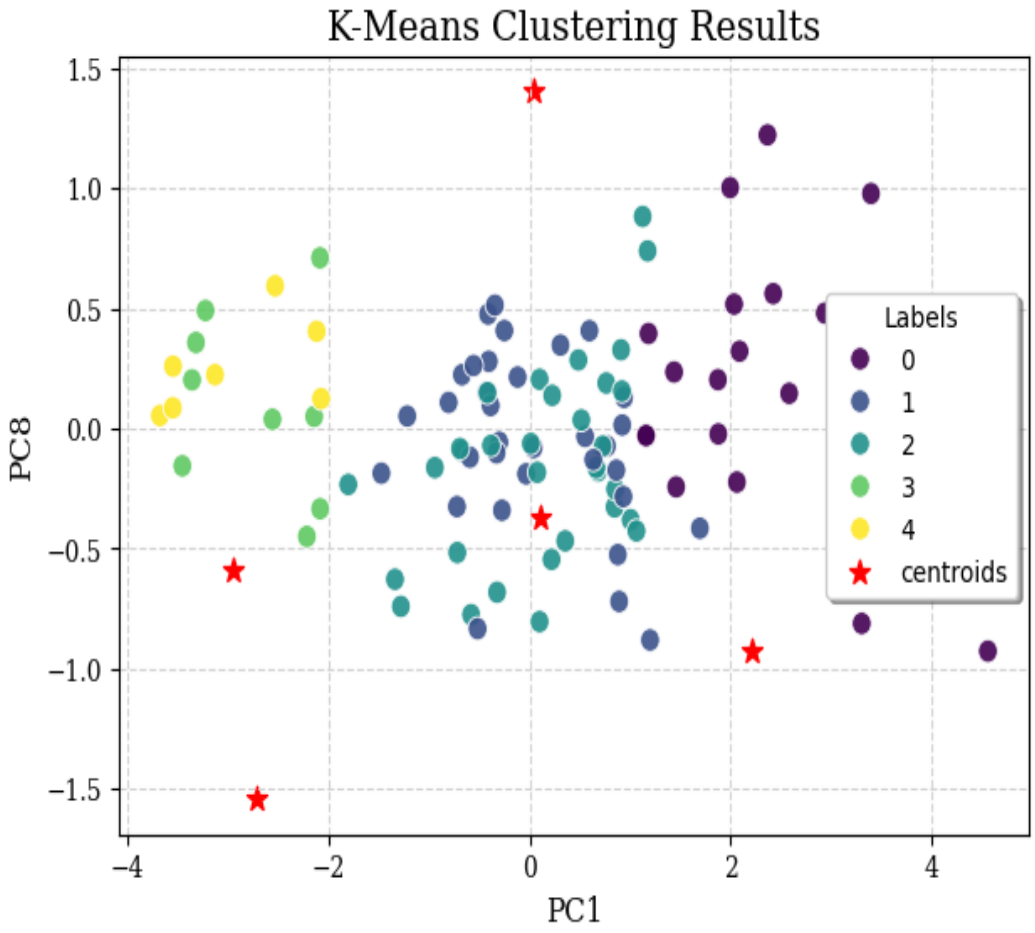
Clustering is an unsupervised machine learning technique of grouping similar data points into clusters. The objective of this technique is to segregate data points with similar traits and place them into different clusters. There are several algorithms to perform clustering, density-based clustering etc.

K-Means Clustering

K-Means Clustering is an unsupervised learning algorithm whose job is to group the unlabeled dataset into different clusters where each data point belongs to only one cluster.

Here, K is the number of clusters that need to be created in the process. The algorithm finds its applicability into a variety of use cases including market segmentation, image segmentation, image segmentation, image compression, document clustering etc. The below image is the results of clustering of one the datasets.

```
plt.show()
```



The K-Means Algorithm works the following ways:

1. Specify the number of clusters, i.e. K
2. Select k random points in the dataset. These points will be the centroids (centers) of each of the K clusters.
3. Assign each data point in the dataset to one of the K centroids, based on its distance from each of the centroids.
4. Consider this clustering to be correct and reassign the centroids to the mean of these clusters.
5. Repeat step 3 if any of the points clusters, go to step 4 Else Go to step 6.
6. Calculate the variance of each of the clusters.
7. Repeat this clustering ‘n’ number of times until the sum of variance of each cluster is minimum.

Principle Component Analysis

Principle component analysis is a linear dimensionality-reduction technique that is used to reduce the dimensionality of large data sets by transforming a large data sets by transforming a large set of variables into a smaller one while preserving most of the information present in the large set.

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 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 k (as 5 in the below figure).

```
plt.tick_params(axis= both , direction= inout , length=6, color= purple , grid_color= lightgray , gr
plt.show()
```