**Does Population Density of a State affect its sales in EV?**

**By: Keerthana J**

**Objective:**

The objective of this project is to perform market segmentation analysis using population density data and other relevant features. The goal is to divide the market into distinct segments based on similarities in population density and other factors, such as demographic or socio-economic indicators. By identifying different segments within the market, businesses can better understand their customers' needs, preferences, and behaviours, and tailor their marketing strategies and product offerings accordingly.

**Dataset:**

We have collected a dataset containing population density data for different states along with the state wise sales of the respective states.

State wise Population Dataset: <https://www.kaggle.com/datasets/diyasanthosh/statewise-distribution-of-population2011>

Sales of EV state wise dataset: <https://tn.data.gov.in/resource/stateuts-wise-current-sales-electric-vehicles-ev-country-various-segments-reply-unstarred>

**Approach:**

1. **Data Preparation:** Preprocess the dataset by handling missing values, scaling features, and encoding categorical variables as necessary.
2. **Exploratory Data Analysis (EDA):** Explore the dataset to gain insights into the distribution of population density and other features. Visualize the data using plots and charts to identify any patterns or correlations.
3. **Feature Selection:** Select the features that will be used for market segmentation analysis. These features should be relevant to the target market and likely to capture meaningful differences between segments.
4. **Clustering:** Apply clustering algorithms such as K-means, hierarchical clustering, or DBSCAN to partition the dataset into distinct segments based on similarities in population density and other features.

**Data Preprocessing**

Replacing null data:

A screenshot of a computer

Description automatically generated

Removing the sales data where sales is 0

A black rectangular object with colorful text

Description automatically generated

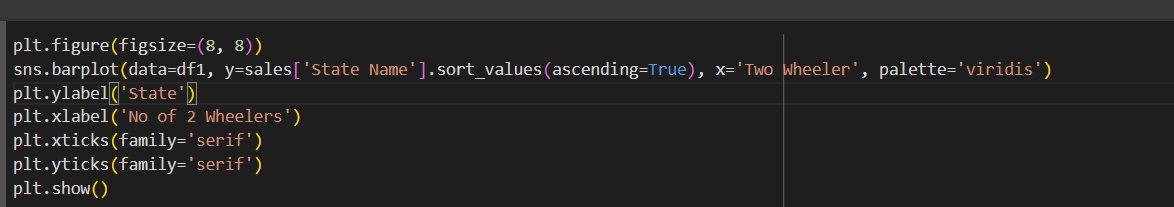
Data after preprocessing

A screenshot of a computer

Description automatically generated

**Exploratory Data analysis**

2-Wheeler Sales Data:



A graph of different colored bars

Description automatically generated

**Karnataka** has the highest 2-wheeler EVs sold.

3-Wheeler sales data

A computer screen with text on it

Description automatically generated

A graph with different colored bars

Description automatically generated

**Uttar Pradesh** has the highest 3-wheeler EVs sold.

4-Wheeler’s sales data

A screen shot of a computer

Description automatically generated

A graph with different colored bars

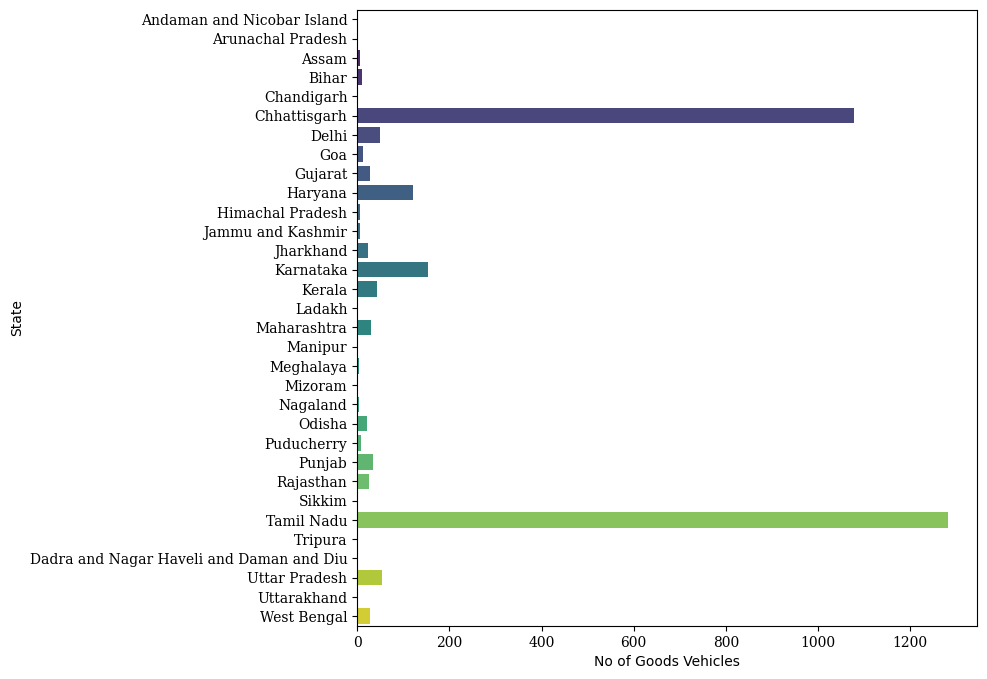
Description automatically generated

**Karnataka** has the highest 4-wheeler EVs sold.

Goods Vehicle sales data

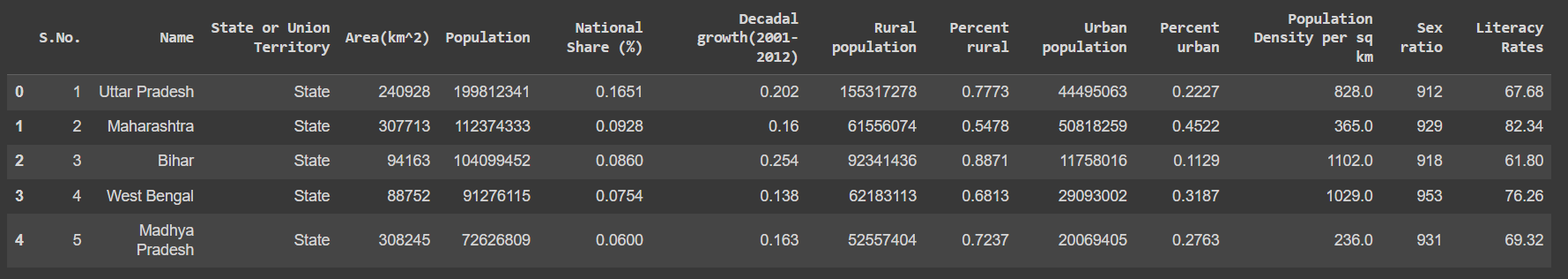
A screen shot of a computer

Description automatically generated



**Tamil Nadu** has the highest Goods EVs sold.

**Pre-Processing – Population Dataset**

****

**A black background with orange text

Description automatically generated**

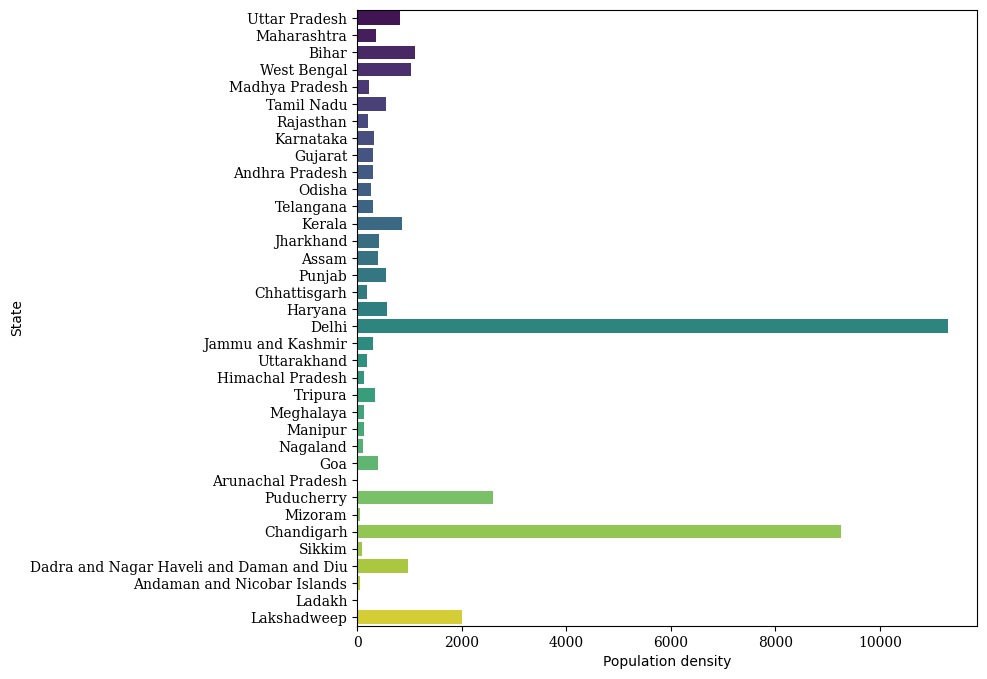
**A screenshot of a computer

Description automatically generated**

**Exploratory data Analysis- Population Dataset**

**A screen shot of a computer code

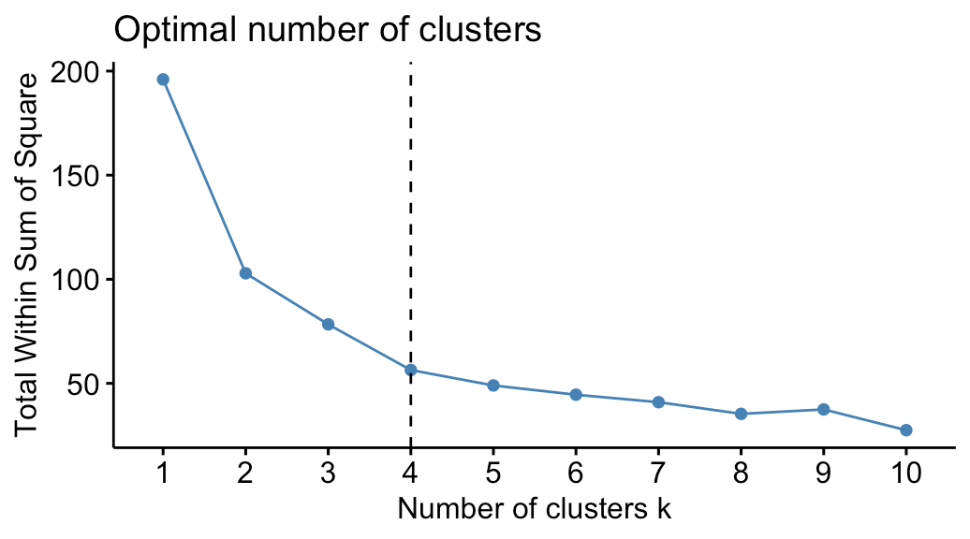
Description automatically generated**



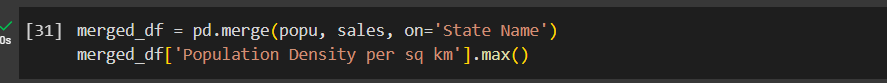
**Segment Extraction**

In the context of our market segmentation analysis project, "Segment Extraction" refers to the process of extracting actionable insights and information from the segmented data. Once we have divided the market into distinct segments based on population density and other relevant factors, segment extraction becomes crucial for understanding the characteristics of each segment and making informed business decisions.

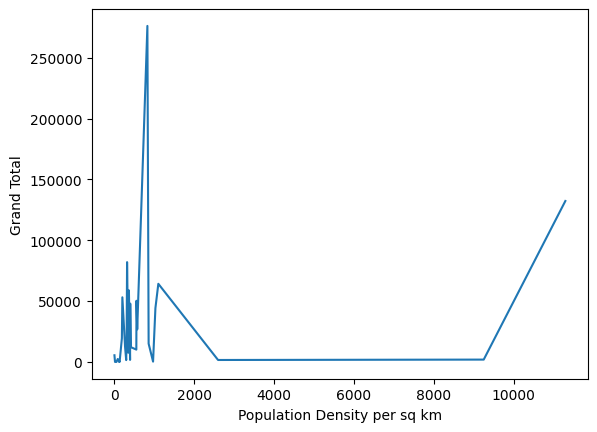
K-means clustering is a popular unsupervised machine learning algorithm used for partitioning a dataset into a predetermined number of clusters. The algorithm iteratively assigns each data point to the nearest cluster centroid and then recalculates the centroids based on the mean of the data points assigned to each cluster. This process continues until the centroids converge or a specified number of iterations is reached. K-means is an efficient and scalable algorithm suitable for large datasets, making it widely used in various applications such as market segmentation, image compression, and anomaly detection. However, K-means requires the number of clusters (k) to be specified a priori, which can be a limitation in some cases. Additionally, K-means is sensitive to the initial placement of centroids and may converge to a local optimum, making multiple initializations with different random seeds necessary to find a more stable solution. Despite these limitations, K-means remains a versatile and widely used clustering algorithm due to its simplicity, effectiveness, and computational efficiency.

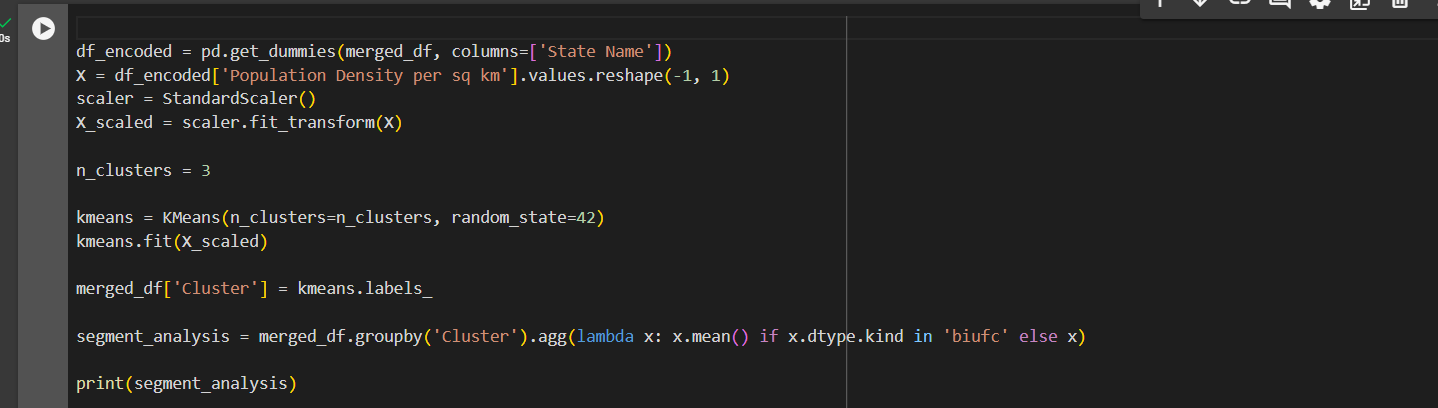


**Segmentation**

****

sns.lineplot(data=merged\_df,x="Population Density per sq km",y="Grand Total",estimator='mean', errorbar=None)



****

**A screenshot of a computer screen

Description automatically generated**

import matplotlib.pyplot as plt

import seaborn as sns

plt.figure(figsize=(10, 6))

sns.barplot(data=segment\_analysis.reset\_index(),x='Cluster', y='Population Density per sq km', palette='viridis')

plt.xlabel('Cluster')

plt.ylabel('Mean Feature1')

plt.title('Cluster Profiles - Mean Feature1')

plt.show()

A graph of a cluster

Description automatically generated with medium confidence

**Conclusion**

In conclusion, our analysis indicates that the sales of electric vehicles (EVs) do not solely depend on population density. While population density is a relevant factor that may influence consumer behaviour and market dynamics, other variables also play significant roles in determining EV sales. Our findings suggest that factors such as demographic characteristics, economic conditions, government incentives, infrastructure development, and consumer preferences may have a more substantial impact on EV sales than population density alone. Therefore, a comprehensive understanding of the market requires considering a wide range of factors beyond population density. Future research and analysis should explore these additional variables to develop more accurate models and strategies for predicting and promoting EV sales effectively.