EV INDIA MARKET SEGMENTATION ANALYSIS

Charles Praveen Johnson

Data Pre-Processing:

Essential Libraries

Pandas: Pandas is a powerful data manipulation and analysis library. It provides data structures like DataFrame, which is used for handling and analyzing structured data.

NumPy: NumPy is a numerical computing library in Python. It provides support for large, multi-dimensional arrays and matrices, along with mathematical functions to operate on these arrays.

Matplotlib: Matplotlib is a plotting library for Python. It is used for creating various types of plots and charts to visualize data.

Seaborn: Seaborn is a statistical data visualization library based on Matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

LabelEncoder from Scikit-Learn: LabelEncoder is used to convert categorical data into numerical format, which is often required for machine learning algorithms.

VIF (Variance Inflation Factor) from Statsmodels: VIF is used to check for multicollinearity in regression analysis. It helps identify highly correlated variables in a dataset.

KMeans from Scikit-Learn: KMeans is a clustering algorithm used for partitioning a dataset into K clusters.

StandardScaler from Scikit-Learn: StandardScaler is used for standardizing features by removing the mean and scaling to unit variance. It is often applied before clustering to ensure that all features contribute equally.

Axes3D from Matplotlib: Axes3D is used for creating 3D plots. In this case, it is used to visualize clusters in a three-dimensional space.

Reading Datasets:

Dataset-1: Loading and Analyzing the Indian automobile buying behavioural characteristics. In this dataset we are trying to analyze how Demography related to buy cars, and how EDA focuses on identifying the factors influencing the car price.

Dataset-2: Loading and Analyzing the Electric vehicles sales report data across states. In this dataset we are trying to analyze how Geography related to sale vehicles, and identifying the regions for startups.

Dataset-3: Loading and Analyzing the Electric Vehicles (EV) charging stations report across states. In this dataset we are analyzing and identifying the availability of charging stations across various regions.

Data Reduction:

Some columns or variables can be dropped if they do not add value to our analysis.

For the dataset-2, dataset-3, the column "Sl. No" has been dropped, assuming they don't have any predictive power to predict the dependent variable.

Data Cleaning:

Some names of the variables are not relevant and not easy to understand. Some data may have data entry errors, and some variables may need data type conversion. We need to fix this issue in the data.

For dataset-1, dataset-2, dataset-3, we are checking data information, null values and counting the values. since there is no null values present.

For the dataset-1, all categorical columns (Marital status, Education, Profession etc..) has been label encoded to numerical type for the next step of process.

Statistics Summary:

The information gives a quick and simple description of the data. It includes Count, Mean, Standard Deviation, median, mode, minimum value, maximum value, range, standard deviation, etc. Statistics summary gives a high-level idea to identify whether the data has any outliers, data entry error, distribution of data such as the data is normally distributed or left/right skewed.

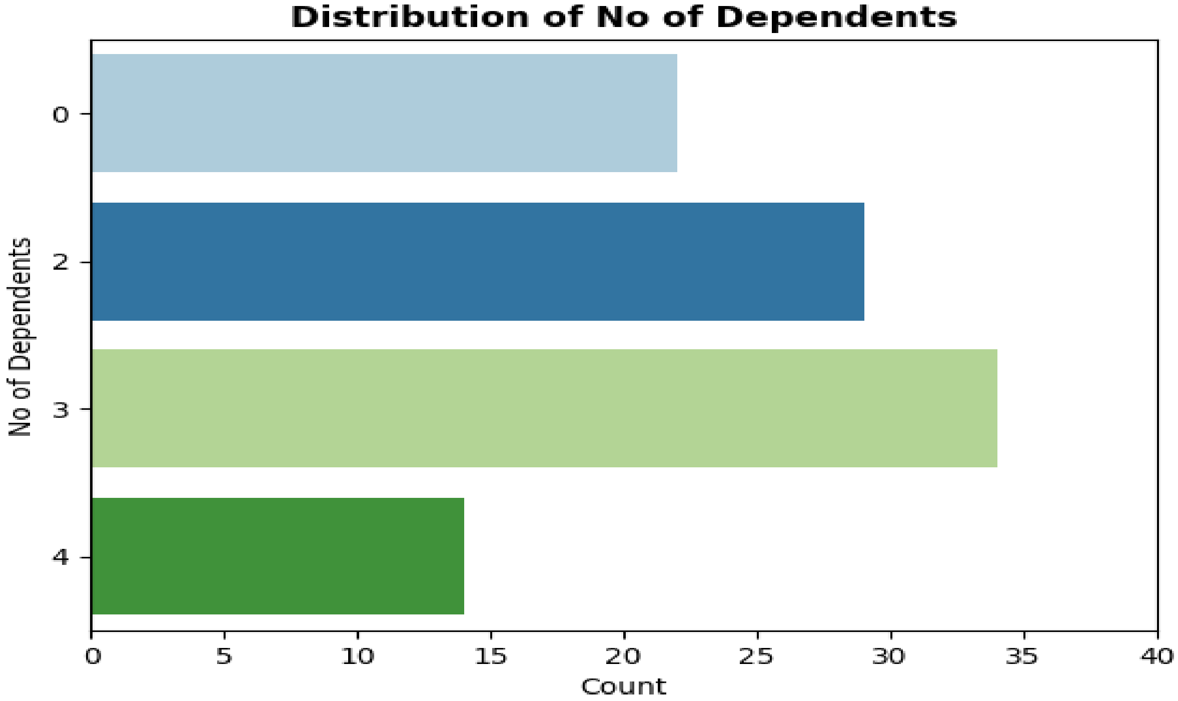
describe()– Provide a statistics summary of data belonging to numerical datatype such as int, float**.**

For the dataset-1, we are finding that min (26-51 max) age group of peoples with their minimum total salary of (200000-5200000 max).

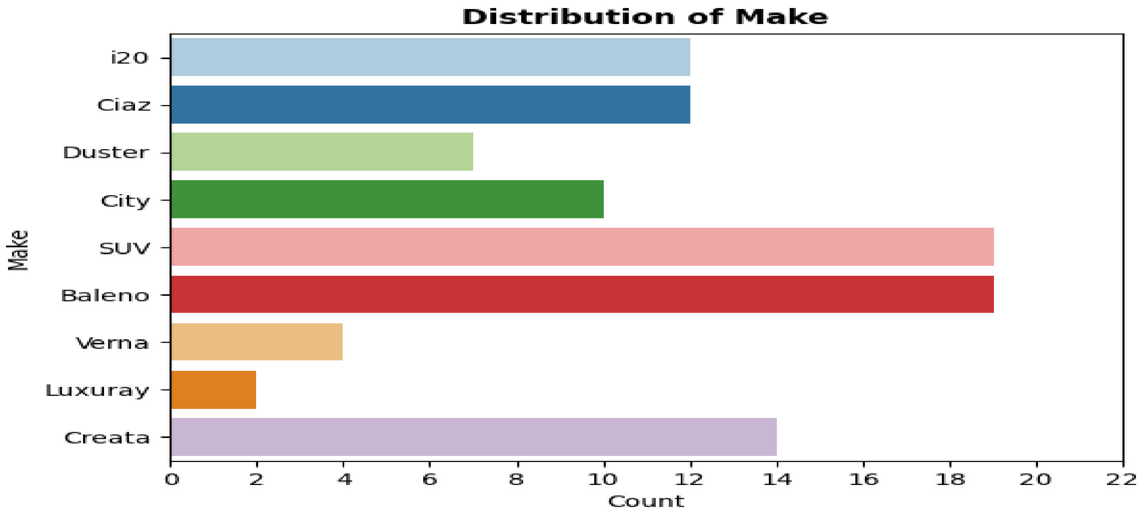
EDA Exploratory Data Analysis:

Exploratory Data Analysis (EDA) is a crucial process involving initial investigations on data. It aims to discover patterns, check assumptions using summary statistics and graphical representations, and leverage insights into the dataset to address business problems. EDA is valuable for identifying outliers, patterns, and trends in the given data, providing in-depth insights, and offering clues for imputing missing values.

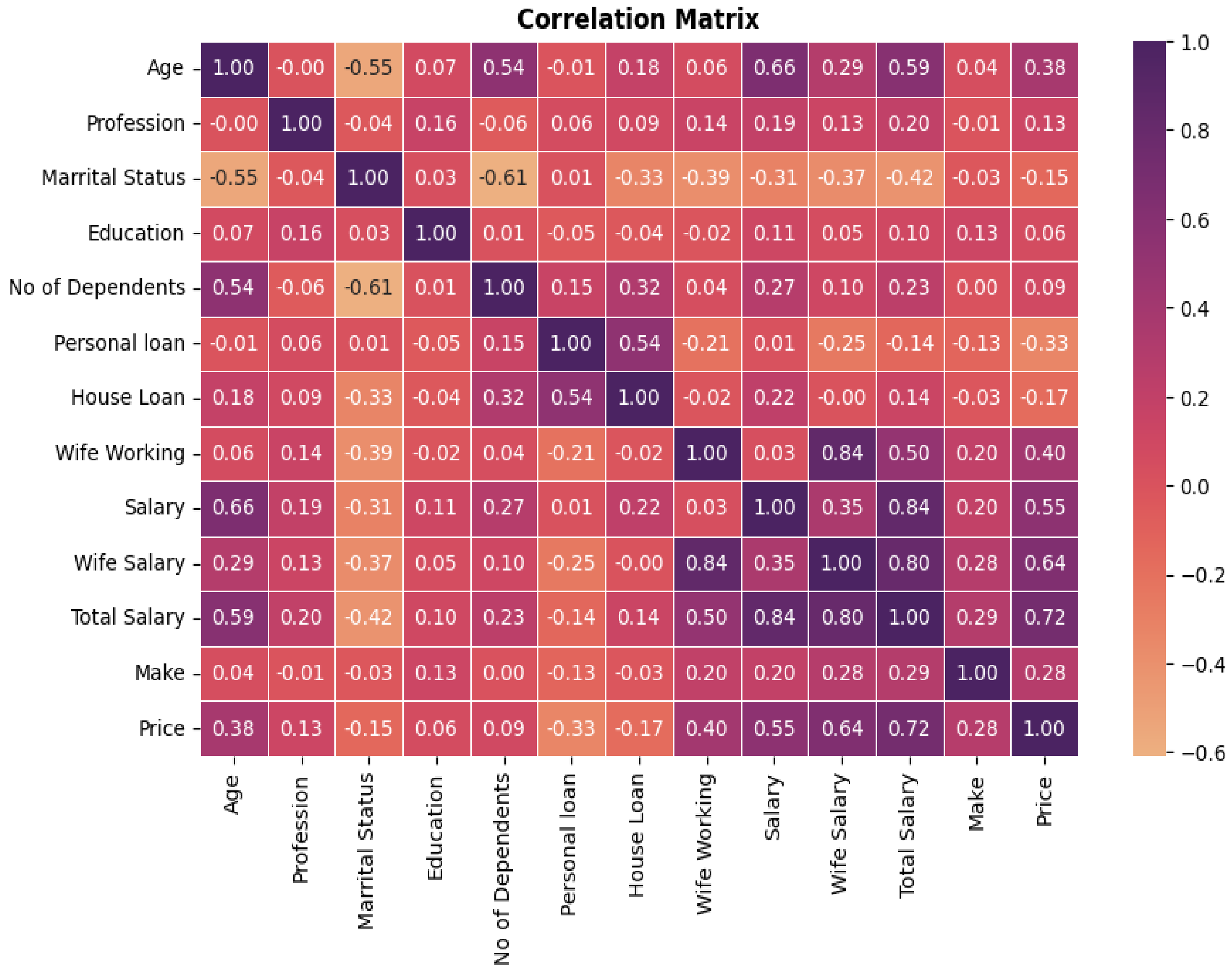
Number of Dependents:



Number of Make Available:

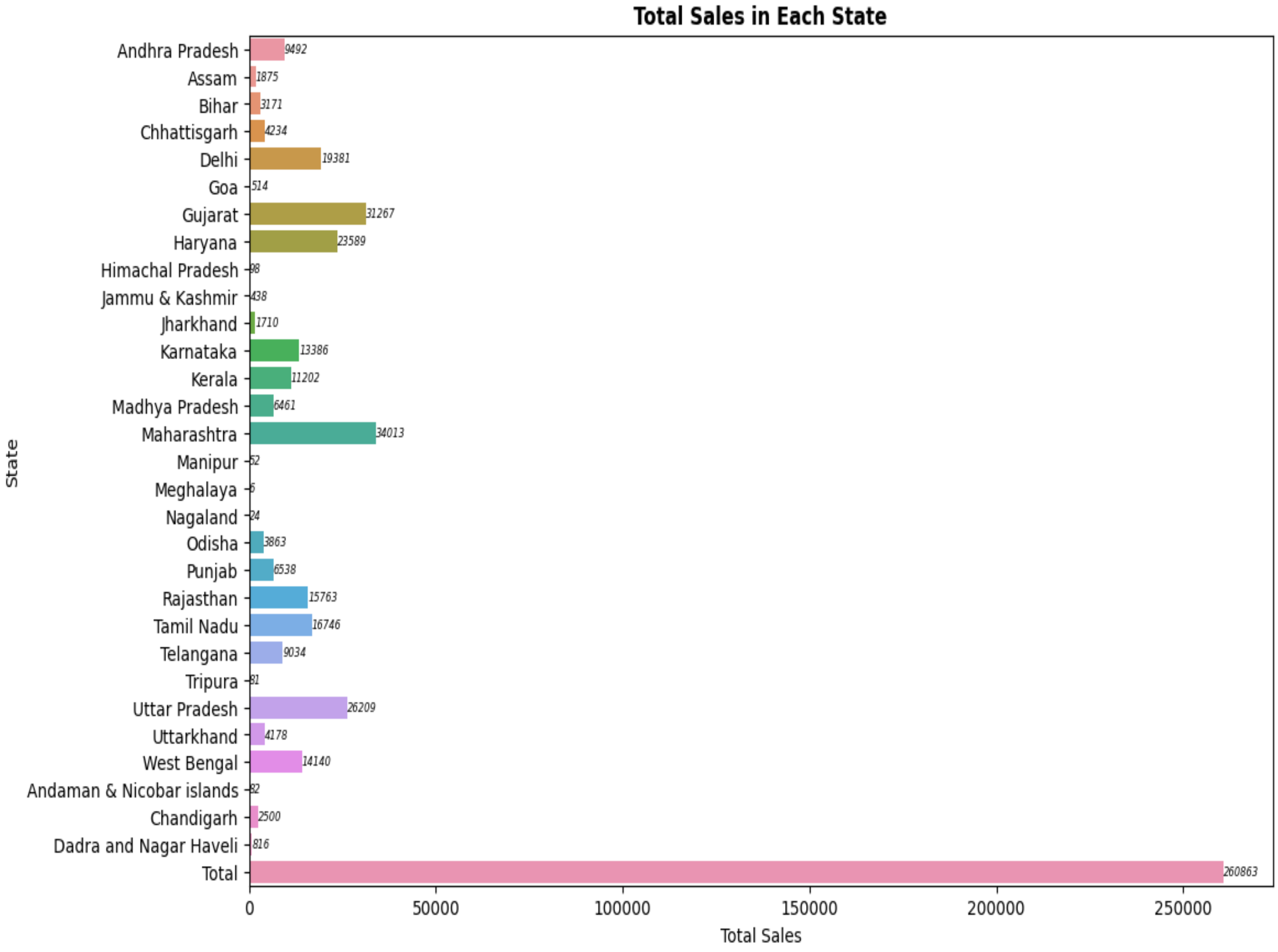


Heatmap Correlation Matrix:



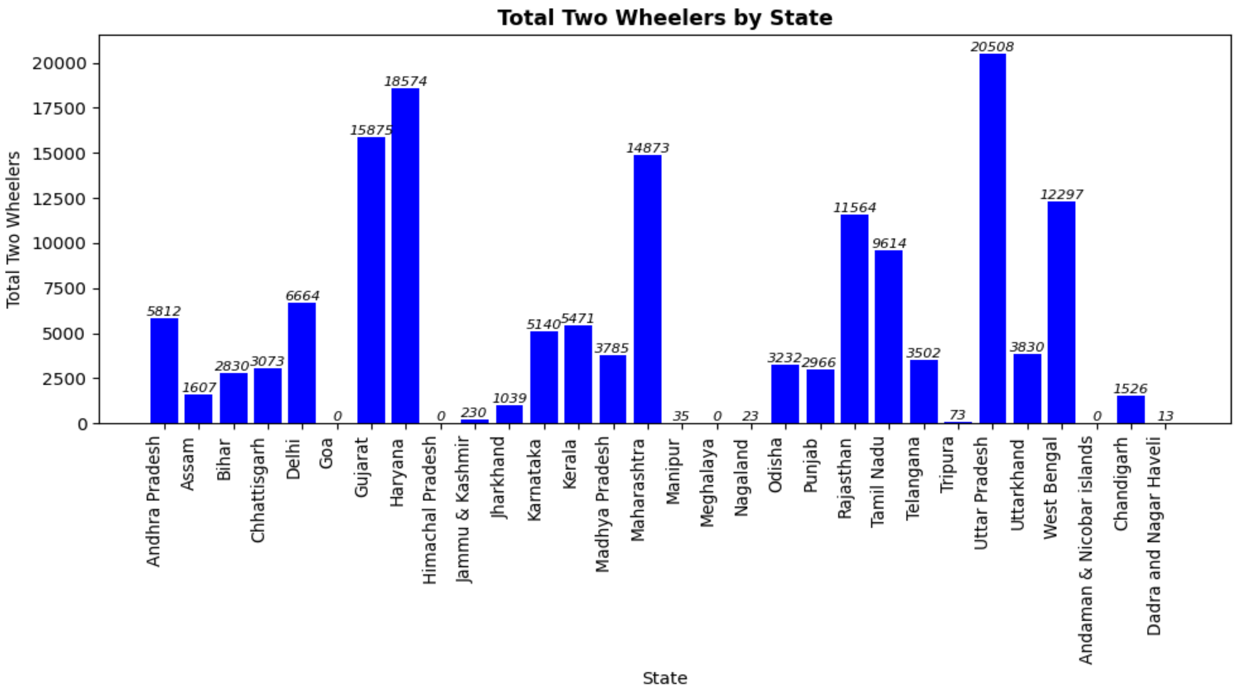
	Age	Profession	Marrital Status	Education	No of Dependents	Personal loan	House Loan	Wife Working	Salary	Wife Salary	Total Salary	Make	Price
Age	1.00	-0.00	-0.55	0.07	0.54	-0.01	0.18	0.06	0.66	0.29	0.59	0.04	0.38
Profession	-0.00	1.00	-0.04	0.16	-0.06	0.06	0.09	0.14	0.19	0.13	0.20	-0.01	0.13
Marrital Status	-0.55	-0.04	1.00	0.03	-0.61	0.01	-0.33	-0.39	-0.31	-0.37	-0.42	-0.03	-0.15
Education	0.07	0.16	0.03	1.00	0.01	-0.05	-0.04	-0.02	0.11	0.05	0.10	0.13	0.06
No of Dependents	0.54	-0.06	-0.61	0.01	1.00	0.15	0.32	0.04	0.27	0.10	0.23	0.00	0.09
Personal loan	-0.01	0.06	0.01	-0.05	0.15	1.00	0.54	-0.21	0.01	-0.25	-0.14	-0.13	-0.33
House Loan	0.18	0.09	-0.33	-0.04	0.32	0.54	1.00	-0.02	0.22	-0.00	0.14	-0.03	-0.17
Wife Working	0.06	0.14	-0.39	-0.02	0.04	-0.21	-0.02	1.00	0.03	0.84	0.50	0.20	0.40
Salary	0.66	0.19	-0.31	0.11	0.27	0.01	0.22	0.03	1.00	0.35	0.84	0.20	0.55
Wife Salary	0.29	0.13	-0.37	0.05	0.10	-0.25	-0.00	0.84	0.35	1.00	0.80	0.28	0.64
Total Salary	0.59	0.20	-0.42	0.10	0.23	-0.14	0.14	0.50	0.84	0.80	1.00	0.29	0.72
Make	0.04	-0.01	-0.03	0.13	0.00	-0.13	-0.03	0.20	0.20	0.28	0.29	1.00	0.28
Price	0.38	0.13	-0.15	0.06	0.09	-0.33	-0.17	0.40	0.55	0.64	0.72	0.28	1.00

Total Number of EV sales across states:

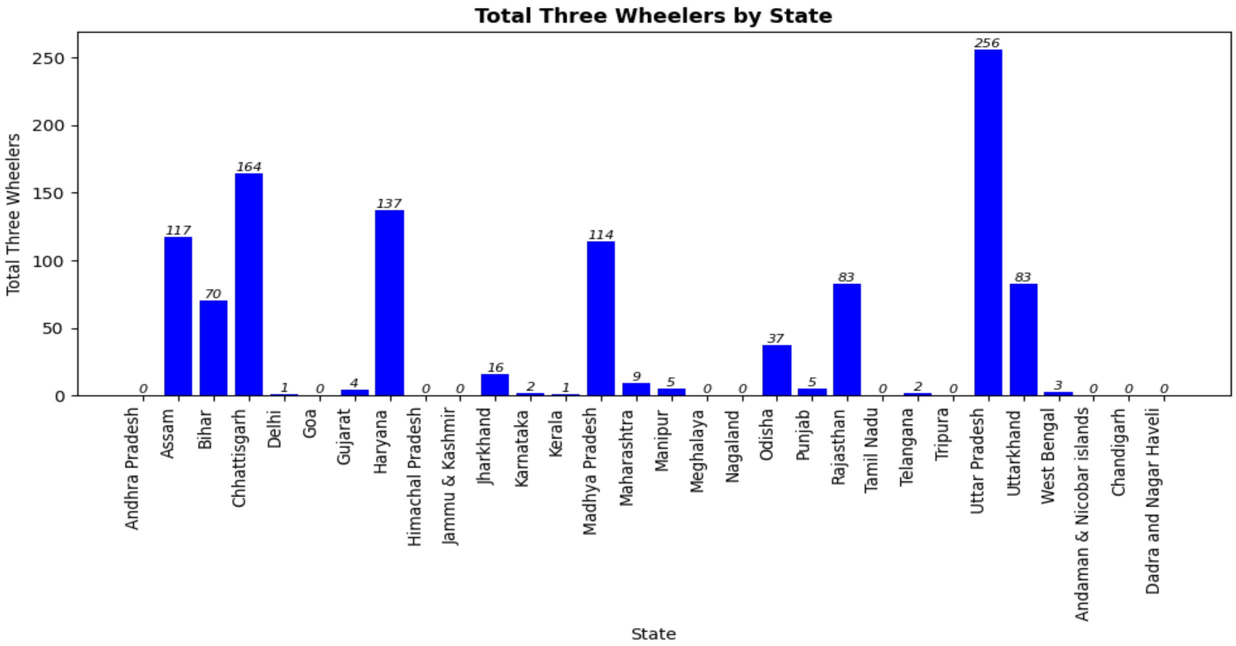


In the EV Sales Statistics data, the analysis focused on descriptive statistics, data visualization, and interpretation of trends in electric vehicle sales across different states and vehicle categories.

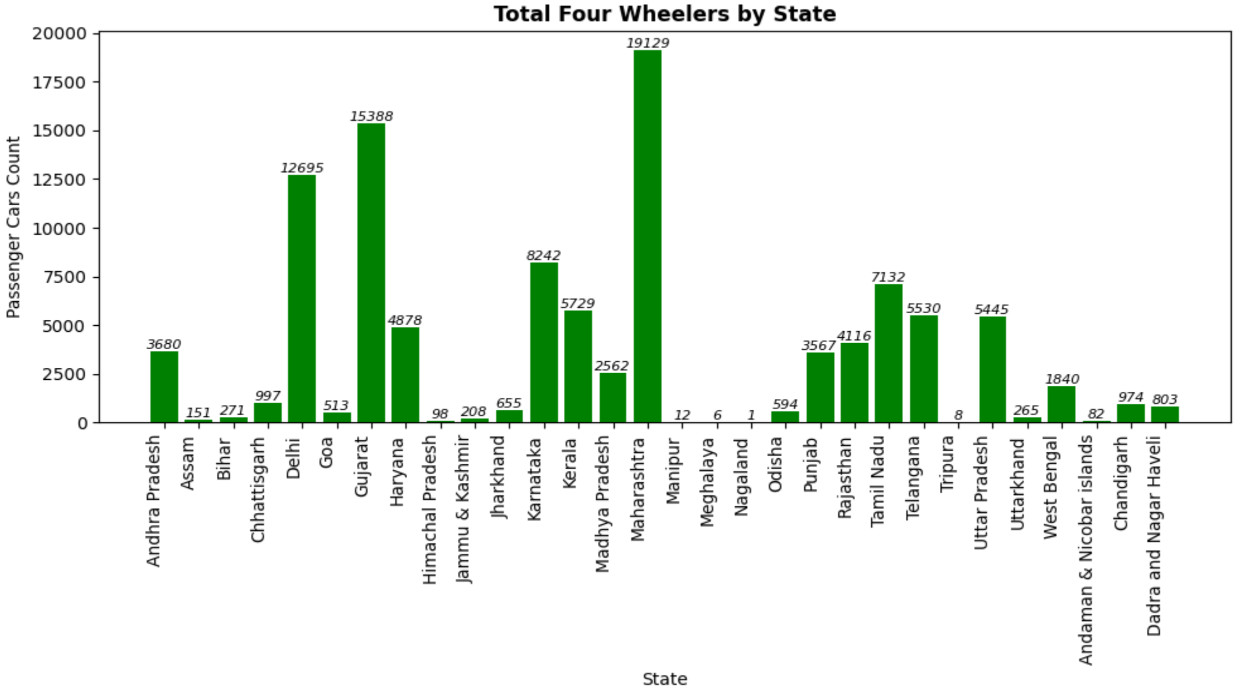
Total number of Two Wheelers:



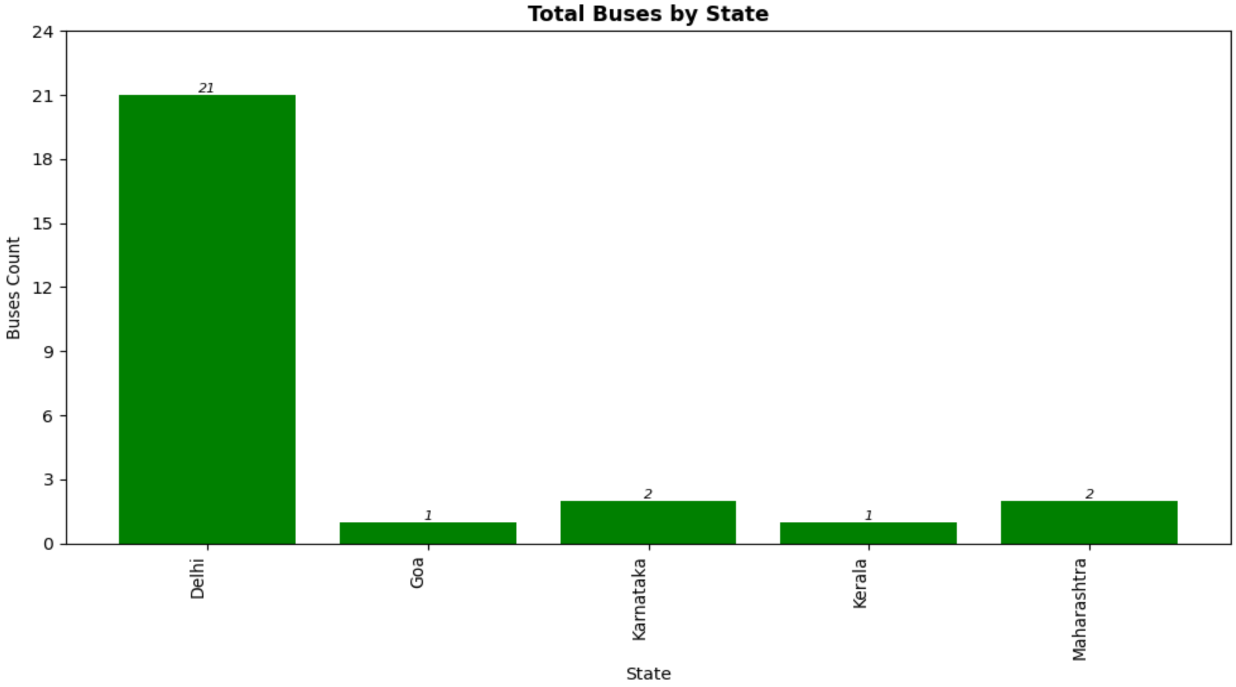
Total number of Three Wheelers:



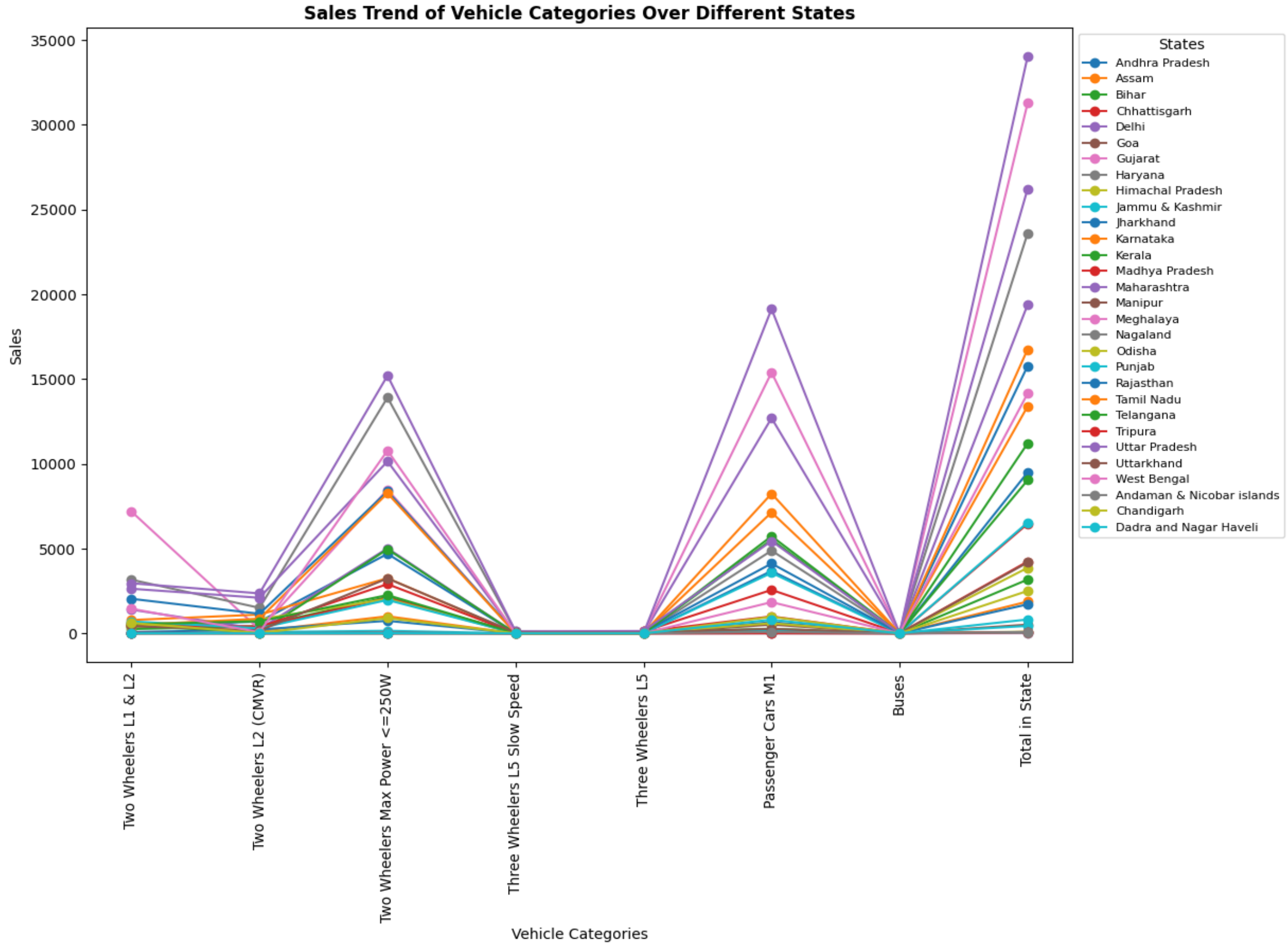
Total number of Four Wheelers:



Total number of Buses:

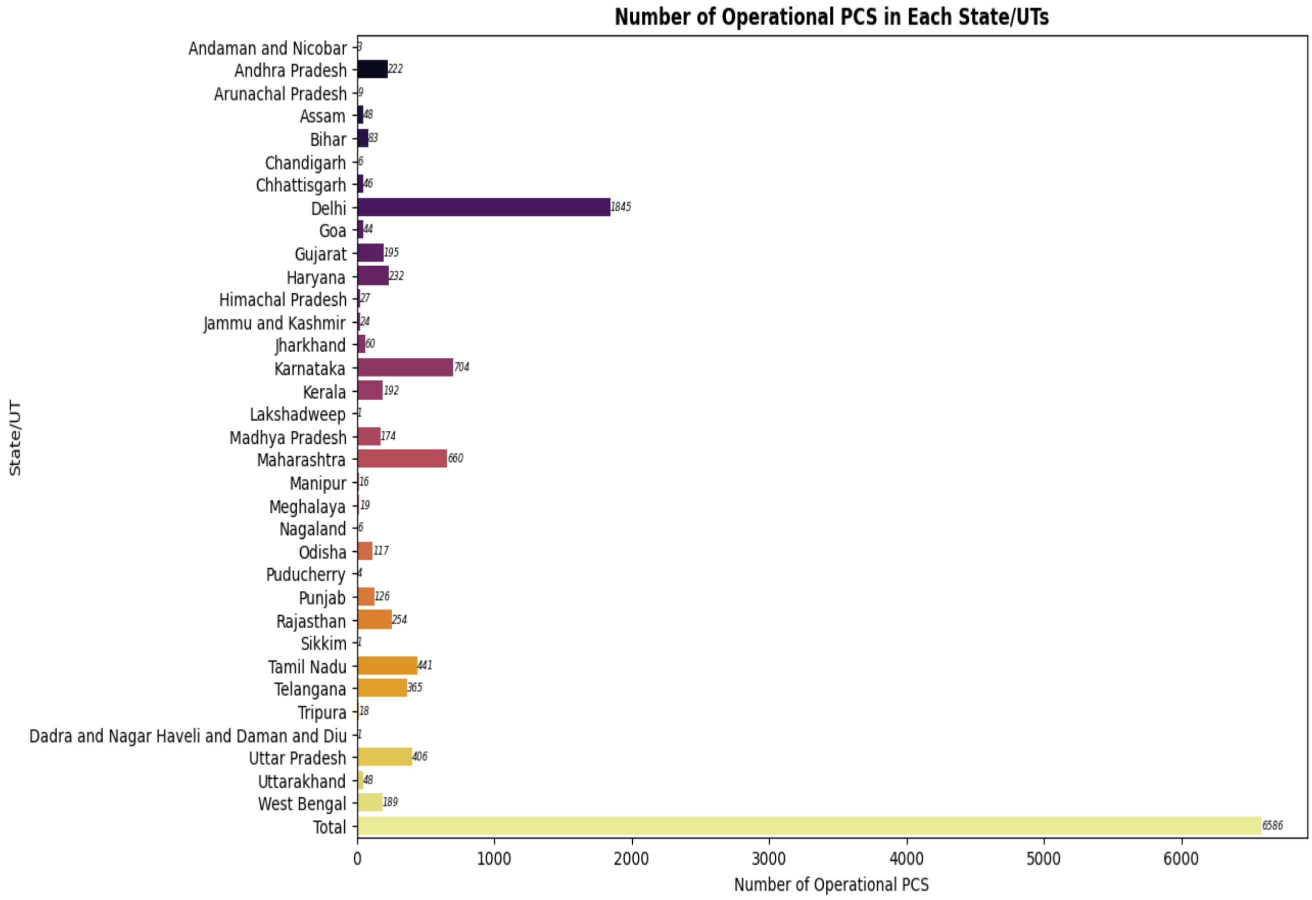


Vehicle Categories Sales Trend across states:

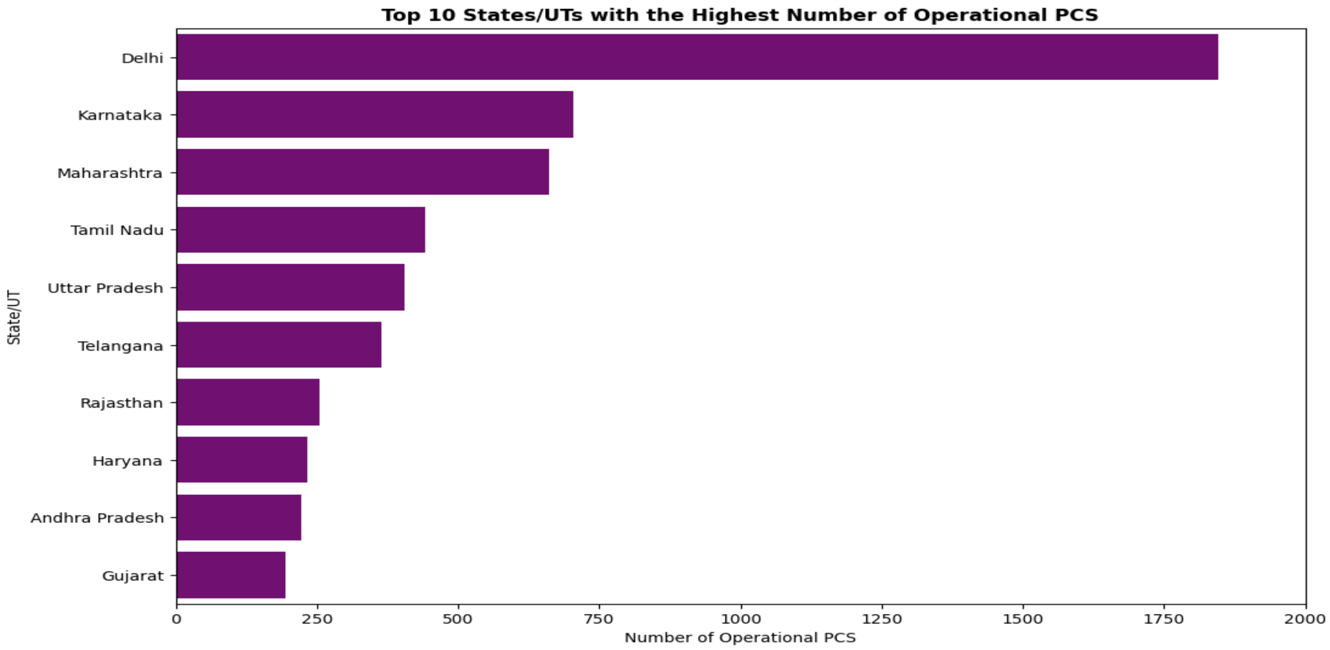


Trend analysis shows that Maharashtra, Gujarat, and Uttar Pradesh have the highest number of total EV sales. Among all states, Maharashtra stands out as the only one offering all categories of vehicles and leading in the top position.

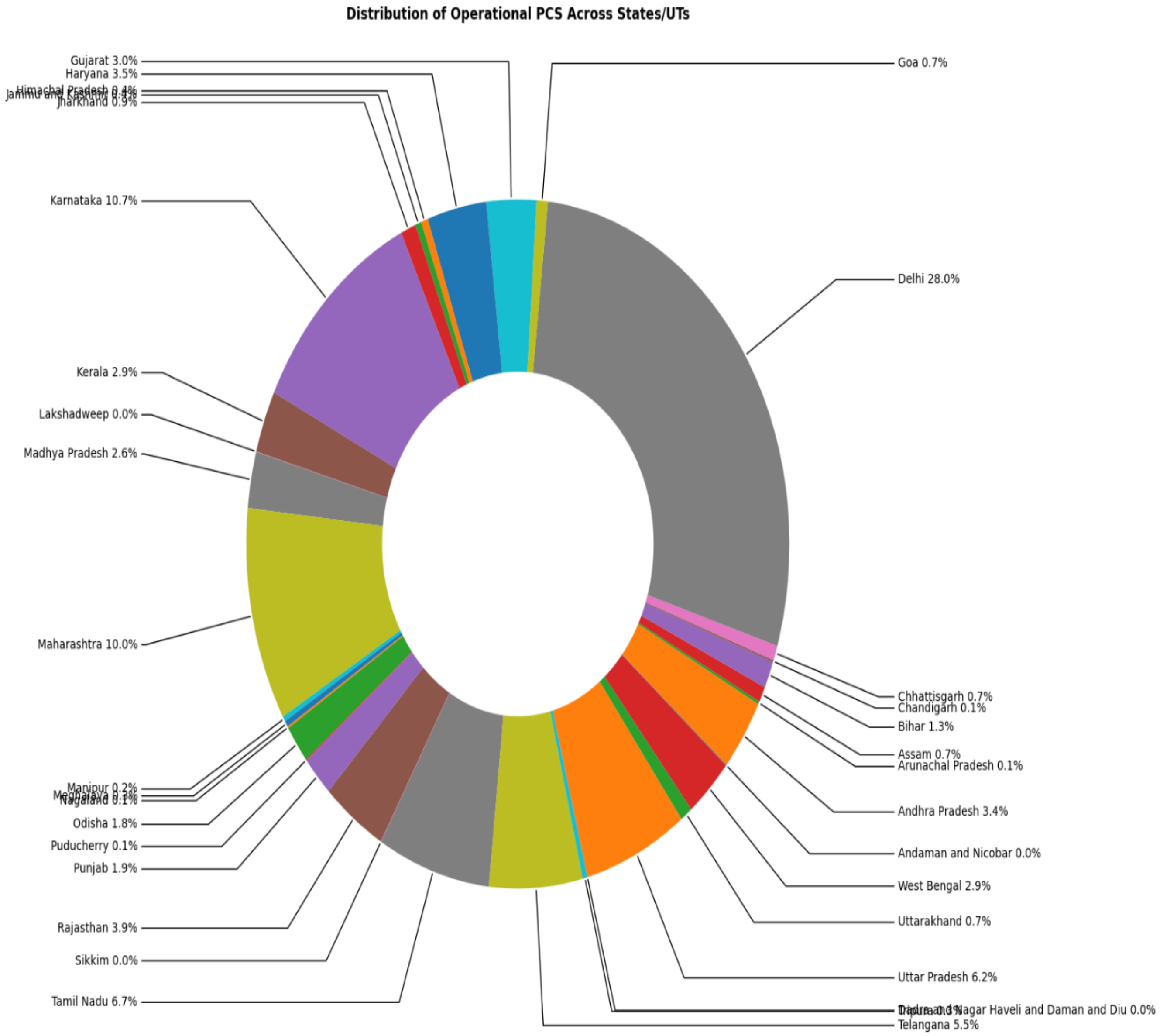
Total Number of charging stations across states:



Top 10 Highest number of charging stations states:



Distribution of EV Charging stations across States/UTs:

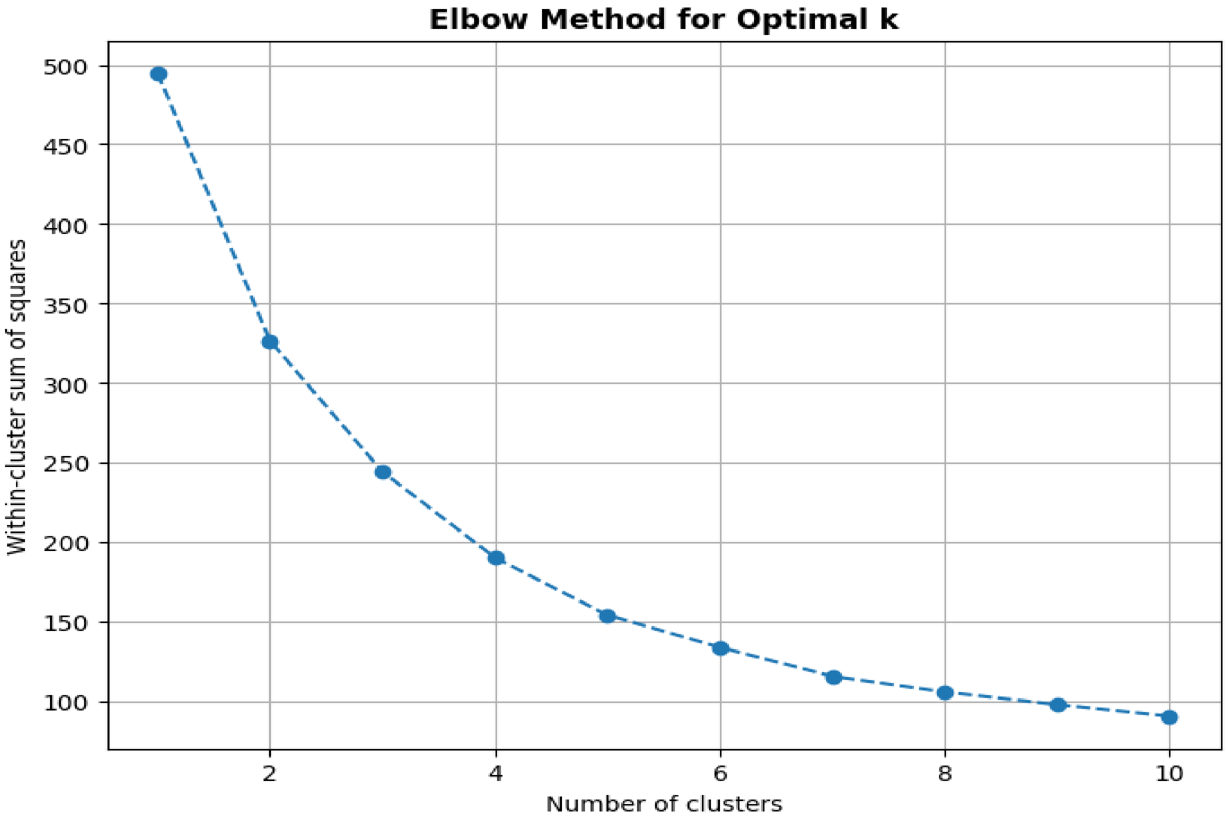


From the pie-chart, state Delhi(28%), Karnataka(10.7%) & Maharashtra(10%) has the highest distributions of EV charging stations shares. Lakshadweep, Sikkim, Dadra and Nagar Haveli and Daman and Diu, Andaman and Nicobar, Puducherry has the least share in India.

Elbow Method:

The elbow method is a technique used to determine the optimal number of clusters in a dataset for a clustering algorithm, such as KMeans. The basic idea is to run the clustering algorithm for a range of values of k (number of clusters) and plot the within-cluster sum of squares (WCSS) or inertia for each k. WCSS is the sum of squared distances between each point in a cluster and the centroid of that cluster.

As the number of clusters increases, WCSS tends to decrease because smaller clusters can better fit the data. However, after a certain point, adding more clusters does not significantly reduce the WCSS, leading to an "elbow" in the plot. The point at which the reduction in WCSS slows down is considered the optimal number of clusters.



Identification of Optimal K: The main goal is to find a balance between having a low WCSS (indicating tight and well-separated clusters) and avoiding too many clusters, which may not provide meaningful insights.

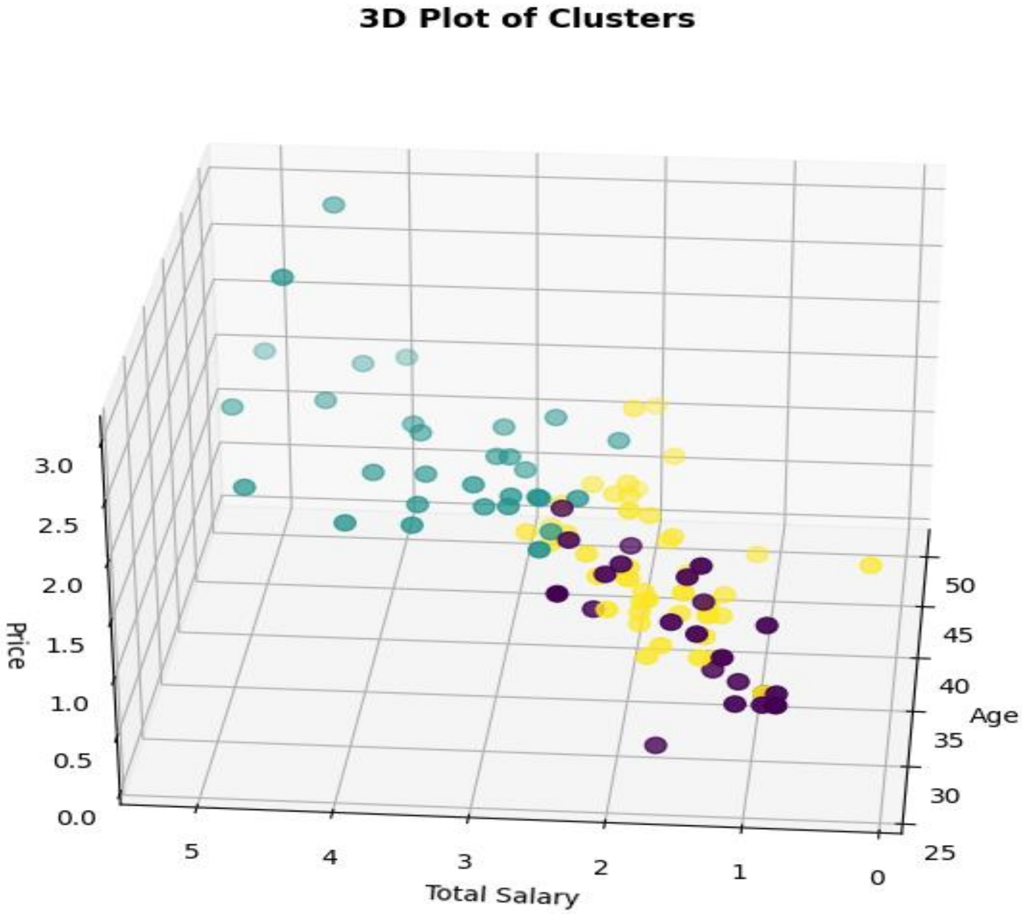
Trade-off: The elbow point represents a trade-off between the goodness of fit (low WCSS) and model simplicity (fewer clusters). Beyond the elbow, the improvement in clustering quality is not significant compared to the increase in complexity.

Avoiding Overfitting: Adding too many clusters can lead to overfitting, where the model captures noise in the data rather than meaningful patterns.

Interpretability: A moderate number of clusters is often more interpretable and useful for decision-making than a large number of clusters.

KMeans Clustering:

Now, regarding the 3D k-means clustering, the elbow method helps you determine the optimal number of clusters for the 3D space defined by standardizing the features 'Age', 'No of Dependents', 'Salary', 'Wife Salary', and 'Price' before applying the elbow method. The “fit_transform” method ensures that the scaling parameters (mean and standard deviation) are calculated from the data and then applied to standardize the features. The elbow method is not directly related to the 3D visualization itself but guides you in choosing the appropriate number of clusters for the subsequent 3D k-means clustering.



After determining the optimal number of clusters(optimal_k = 3), we proceed to apply k-means clustering in 3D space and visualize the clusters using a 3D scatter plot. The 3D plot allows you to observe how the data points are distributed in the three-dimensional feature space based on their assigned clusters. This visualization helps in understanding the patterns and relationships within the data in a more complex space than a 2D plot.

In our case, we are using the elbow method applied to determine the optimal number of clusters for the KMeans algorithm using the 'Age,' 'Total Salary,' and 'Price' features in a three-dimensional space. The 3D scatter plot visualizes the clusters in this space, providing insights into the structure of the data.

Conclusion:

The conclusions and insights gained from the research and analysis work are as follows:

Indian Automobile Buying Behavior:

- Three distinct clusters were identified, each exhibiting unique characteristics in terms of demographics, financial attributes, and preferences.
- Insights into the average purchase price, marital status, profession, education, and age group preferences for different clusters were provided.
- Based on the insights from the buying behavior analysis, it is evident that people can afford Electric Vehicle (EV) car brands available in the Indian market, considering their price range.

EV Sales Statistics:

- Two-wheelers dominate EV sales across states, with substantial contributions from three-wheelers and passenger cars.
- Maharashtra, Gujarat, and Uttar Pradesh emerge as hotspots for EV sales, presenting lucrative opportunities for startups to target these states.
- Maharashtra, in particular, stands out as a leader in EV adoption, offering a market for all categories of vehicles.

EV Charging Stations:

- Some states have a higher number of Public Charging Stations (PCS), indicating better charging infrastructure.
- The availability of charging stations varies across states. Startups could focus on regions with a lower density of charging stations to bridge the infrastructure gap.
- The top 10 states with the highest number of charging stations present potential markets for startups looking to invest in and expand EV charging infrastructure.

Identifying optimal market segments involves considering variables that strongly influence consumer behavior and preferences. The target variables for creating optimal market segments in the Electric Vehicle (EV) automobile domain could includes:

Income Level: Income is a key determinant of purchasing power and influences the affordability of different car models.

Lifestyle Preferences: Lifestyle choices, such as urban or rural residence, commuting patterns, and environmental consciousness, impact the type of vehicles preferred.

Demographic Factors: Age, marital status, and number of dependents play a role in determining the size and type of vehicle suitable for a household.

Brand Loyalty and Past Purchase Behavior: Analyzing brand preferences and past purchase behavior helps identify consumers likely to remain loyal to specific brands or models.

Government Initiatives and Policies: Keep an eye on government initiatives and policies supporting the EV industry. Understanding and aligning with these policies can provide startups with a competitive advantage and facilitate smoother market entry.

Sustainable and Innovative Solutions: Given the increasing focus on sustainability, startups in the EV space should emphasize environmentally friendly and energy-efficient solutions. Innovations in battery technology, charging infrastructure, and vehicle design can set startups apart in a competitive market.

The Electric Vehicle sector presents significant opportunities for startups to thrive and contribute to the transition towards sustainable transportation. By strategically addressing regional variations, focusing on specific vehicle categories, investing in charging infrastructure, and understanding consumer preferences, startups can position themselves for success in this dynamic and rapidly evolving market.