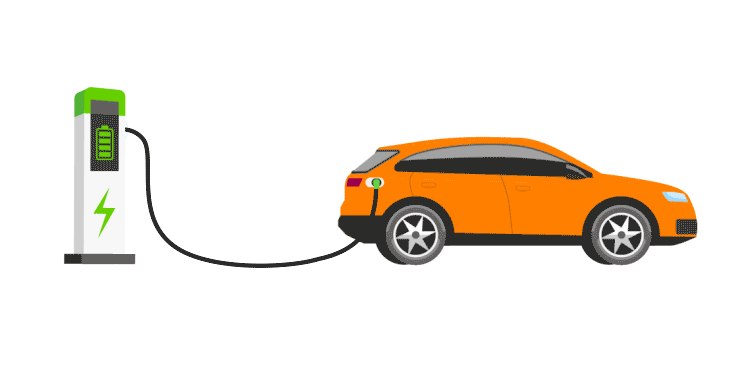
**Market Segmentation Analysis of Electric Vehicles Market in India**



**Problem Statement:**

Task is to analyse the Electric Vehicles Market in India using Segmentation analysis and come up with a feasible strategy to enter the market, targeting the segments most likely to use their product in terms of Geographic, Demographic, Psychographic, and Behavioural.

In this report we analyse the Electric Vehicles Market in India using segments such as region, price, charging facility, type of vehicles (e.g., 2 wheelers, 3 wheelers, 4 wheelers etc.), retail outlets, manufacturers, safety, plug types and much more.

**1. Introduction**

The Indian electric vehicle (EV) market is experiencing rapid growth, driven by government policies, rising fuel costs, and environmental awareness. This report analyses market segmentation to identify consumer groups most likely to adopt EVs and proposes a market entry strategy. It leverages datasets on EV sales by state and manufacturer, predictive battery life modelling, and Fermi estimation to quantify segment potential.

**Objectives**

* Segment the Indian EV market by demographic, geographic, psychographic, and behavioural factors.
* Analyse preferences for 2-wheelers, 3-wheelers, and 4-wheelers.
* Evaluate battery performance’s role in consumer choice.
* Develop a feasible market entry strategy for high-potential segments.

**Background**

India’s EV market is projected to grow significantly, with 2-wheelers dominating (~80% of sales) due to affordability and urban mobility needs. The FAME II scheme and state-level incentives have accelerated adoption, particularly in urban areas. Understanding segment preferences is critical for tailoring products and strategies.

**2. Data Description**

The analysis uses the following datasets:

* dim\_date.csv: 36 entries with columns `date`, `fiscal\_year`, `quarter` (April 2021–March 2024).
* electric\_vehicle\_sales\_by\_makers.csv: Sales by EV manufacturers, with dates converted to datetime format.
* electric\_vehicle\_sales\_by\_state.csv: State-wise EV sales, also with datetime formatting.
* Indian Automobile buying Behaviour.csv: Contains all the columns like age, salary, wife\_salary, loan to estimate the affordability of EV vehicle.
* ev\_battery\_charging\_data.csv: Contains columns like Degradation, charging time, efficiency, SoC to findout optimal temperature, predictive maintenance.
* ev\_charging\_stations\_india.csv: It has all the electric vehicles charging stations information
* ev\_blast\_dataset.csv: It contains battery\_type, external\_abuse, short\_circuits etc as its columns, used to find if the battery is in poor condition.
* ElectricCarData\_Clean.csv: It has all the information regarding the car brand, rapid charge, accelspeed, Efficiency\_WhKm.
* EV\_sales.csv: Has the sales data over the years
* Metadata.csv: It is a NASA battery dataset that contains type, ambient\_temperature, capacity, Re, Rct

**Preprocessing**

Dates were standardized using `pd.to\_datetime` with `dayfirst=True`. The `dim\_date` dataset has 36 rows and 3 columns, with no missing values (`date\_df.info()` and `date\_df.describe()`).

Null values were filled with median and outliers are handled either by replacing them with median or dropping them. State names and city names were changed to maintain consistency and readability.

Compound Annual Growth Rate (CAGR) : Used to track growth over time

* CAGR = (Ending or max value)/ (Beginning or min value) \*\* (1/ years) - 1

Penetration Rate: Used to measure market share or adoption level at a point of time

* Penetration Rate (%) = (EV Sales/ Total Vehicle Sales) \* 100

The above two are calculated to check the growth and sales rate overtime.

**Battery Life Prediction**

A neural network model predicts battery life using:

* Features: `type\_discharge` (charge/discharge), `Capacity`, `Re`, `Rct`.
* Performance: Mean Squared Error (MSE) of 4449.05.
* Example Predictions:
  + Discharge, Capacity=1.674305, Re=-4.976500e+11, Rct=1.055903e+12 → 26.15 units.
  + Charge, Capacity=20.5, Re=-2.983215e+11, Rct=1.223456e+12 → 340.93 units.

**3. Market Segmentation:**

Demographic Segmentation

* Age: Young professionals (18–35 years) favor 2-wheelers for affordability. Families (35–50 years) prefer 4-wheelers for safety.
* Income: Middle-income (INR 5–15 lakh annually) dominates 2-wheeler purchases; high-income (INR 15 lakh+) prefers 4-wheelers.
* Occupation: Urban professionals and small business owners adopt EVs for fuel savings.

Young Professionals - Age 18–35, middle-income, prefer 2-wheelers, tech-savvy.

Families - Age 35–50, high-income, prefer 4-wheelers, value safety.

Small Businesses - Delivery/transport sectors, cost-driven, prefer 3-wheelers.

Geographic Segmentation

* Urban Areas: Maharashtra, Karnataka, Delhi lead due to charging infrastructure.
* Rural Areas: Low adoption due to limited infrastructure.
* Regional Trends: Southern/western states show faster growth.

Urban (e.g., Maharashtra, Delhi) - High adoption, 2-wheelers dominant, robust infrastructure.

Rural - Low adoption, limited infrastructure, prefer low-cost 2-wheelers.

Southern/Western States - Rapid growth, supportive policies, mixed vehicle types.

Psychographic Segmentation

* Environmentally Conscious: Prioritize low emissions.
* Tech Enthusiasts: Seek smart features (e.g., connectivity).
* Cost-Conscious: Focus on affordability, prefer 2-wheelers.

Eco-Conscious - Low emissions, efficient batteries.

Tech Enthusiasts - Connectivity, range, performance.

Cost-Conscious - Affordability, prefer 2-wheelers.

Behavioural Segmentation

* Purchase Patterns: First-time buyers prefer established brands (e.g., Tata). Repeat buyers explore new entrants.
* Usage Patterns: Commuters favor 2-wheelers; commercial users choose 3-wheelers/4-wheelers.
* Loyalty: Moderate, driven by price/performance.

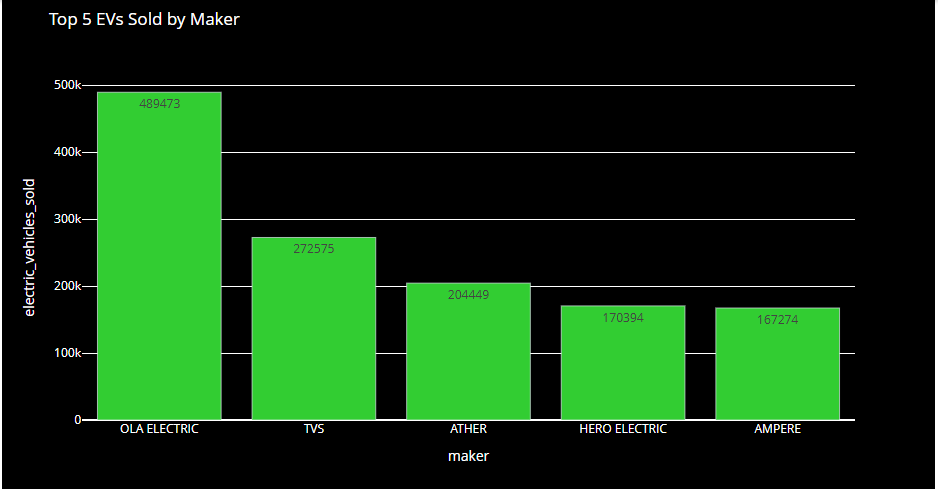
First-Time Buyers - Prefer established brands, focus on reliability.

Commercial User - High usage, prefer 3-wheelers/4-wheelers, cost-driven.

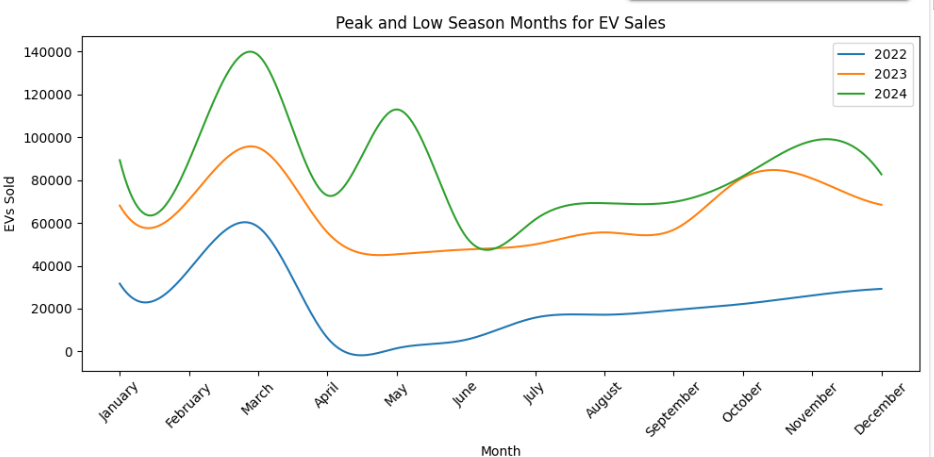
Repeat Buyers - Open to new brands, prioritize performance.

**4. Exploratory Data Analysis**

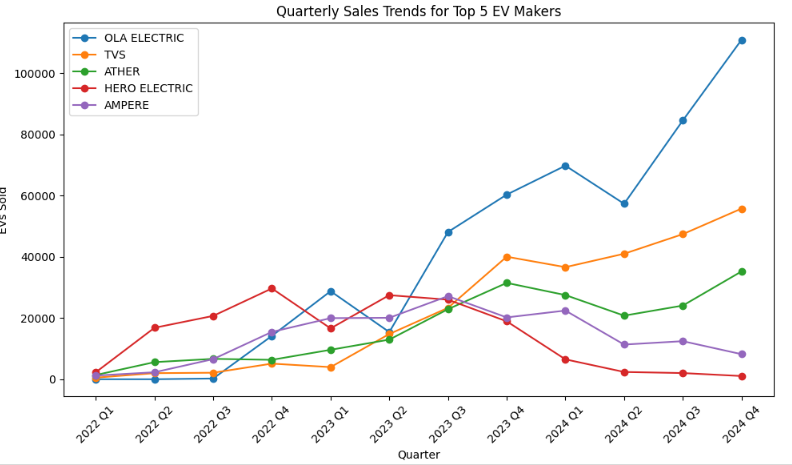
**a. Makers Data**



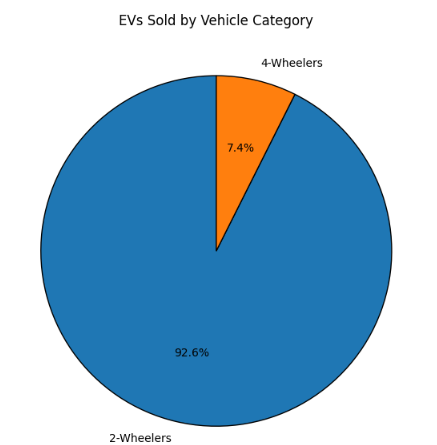
The above includes both 2 wheeler and 4 wheeler makers. From the above, it is clear that OLA ELECTRIC are sold the most Electric Vehicles.



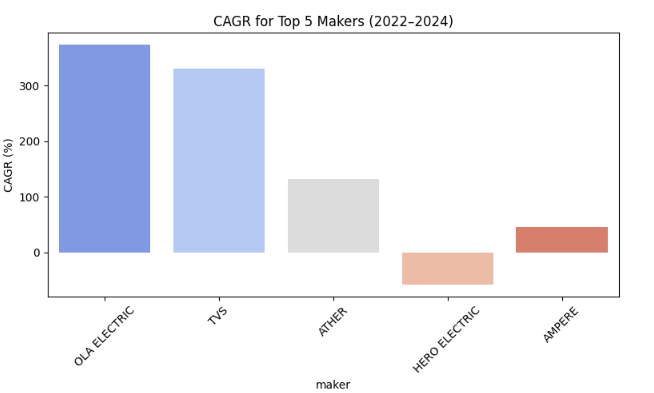
Sales are high in the beginning of the years and the same pattern is being seen across all the years. It might be because of lower sale prices or the showrooms trying to get the new models etc.



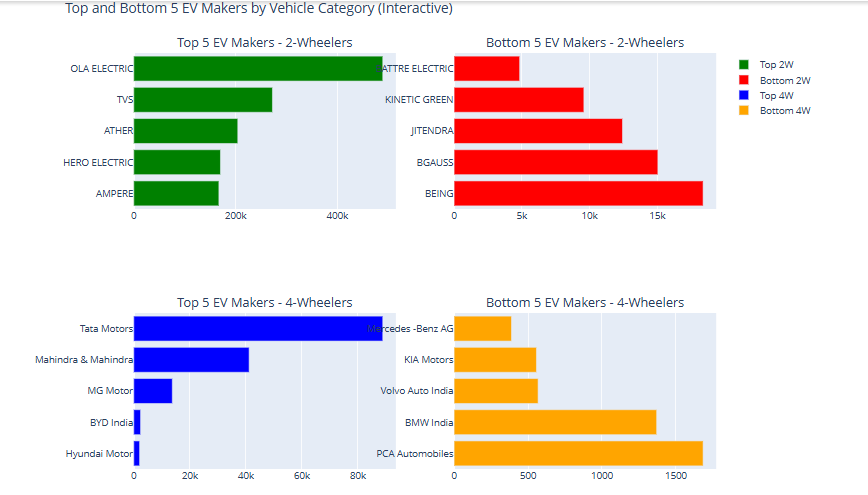
For the top 3 sellers, the sales are high in the 4th quarter.



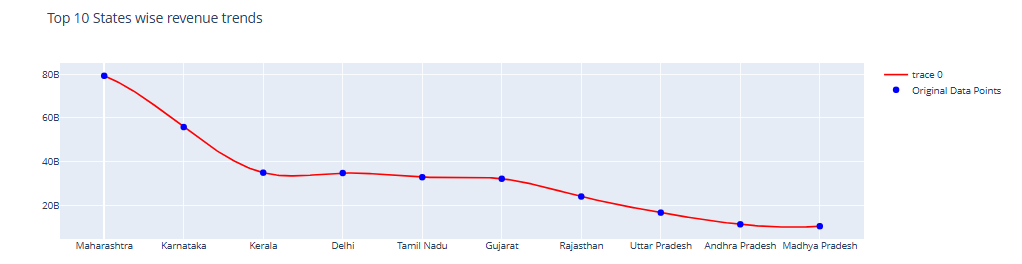
Around 93% of Indian population prefer 2 wheelers.

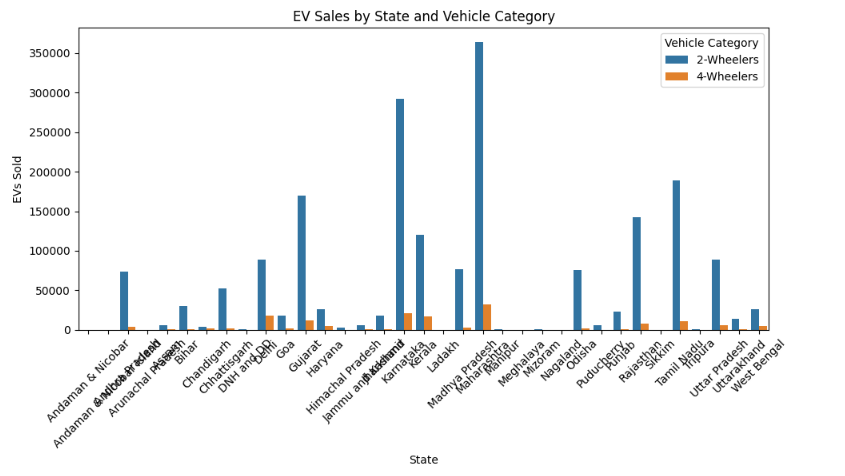


The above chart is Compound Annual Growth Rate and Hero Electric is going in negative direction, whereas Ola Electric has the highest CAGR.

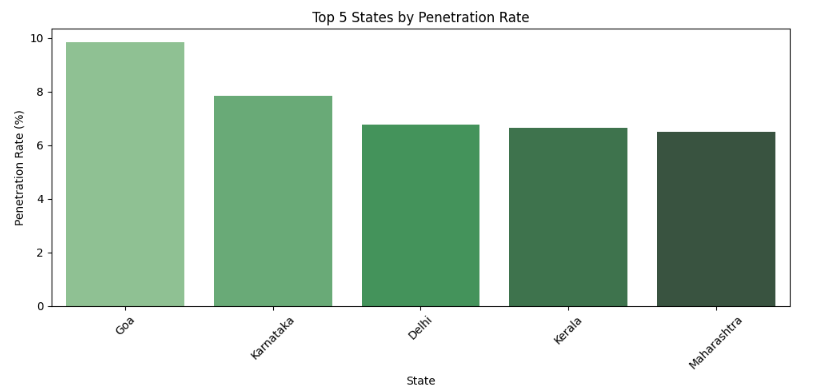


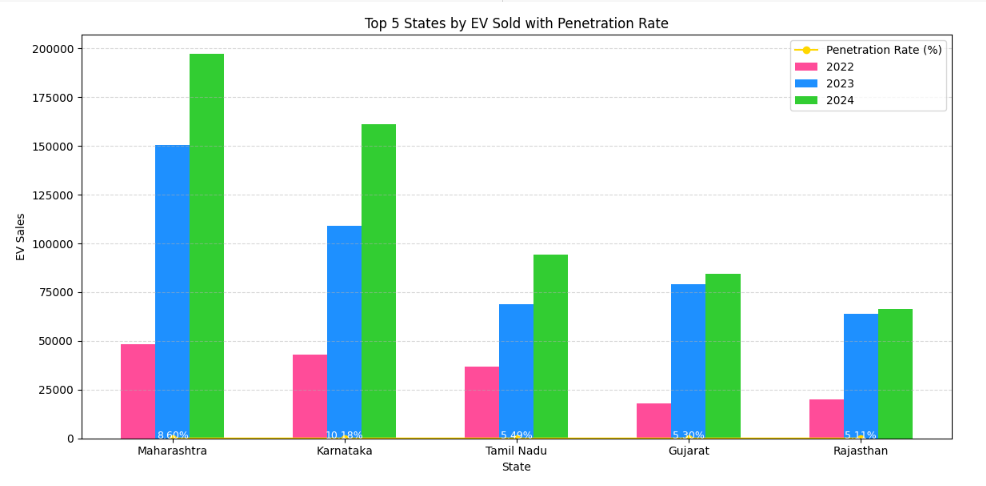
**b. State wise Sales**



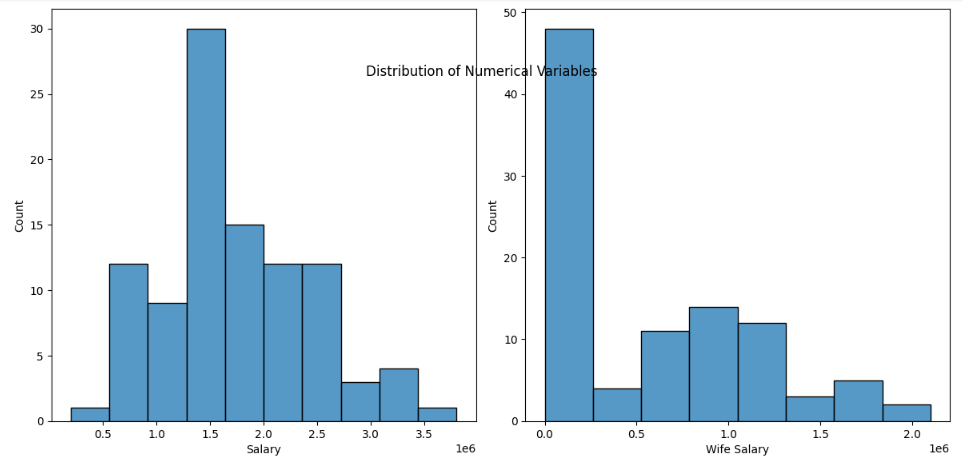


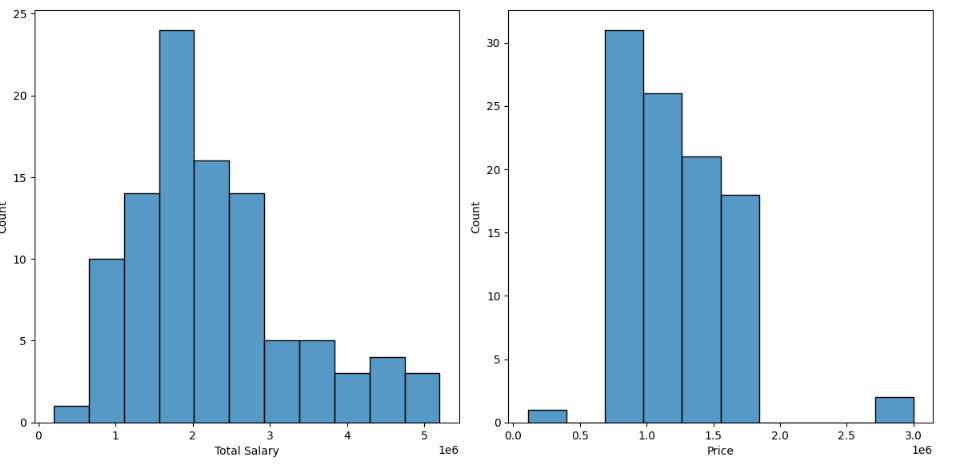
2 wheelers are sold highest in almost all the states.

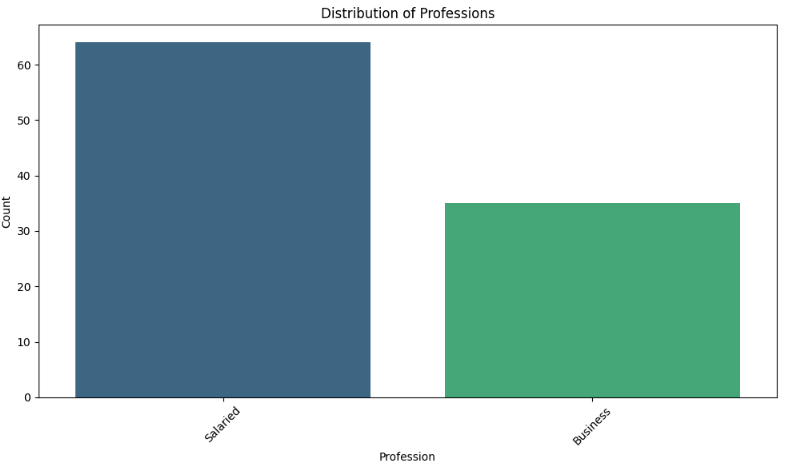


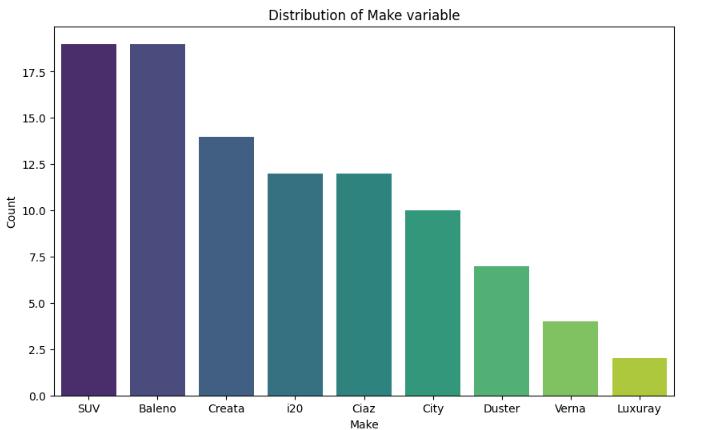


**c. Consumer Behaviour**

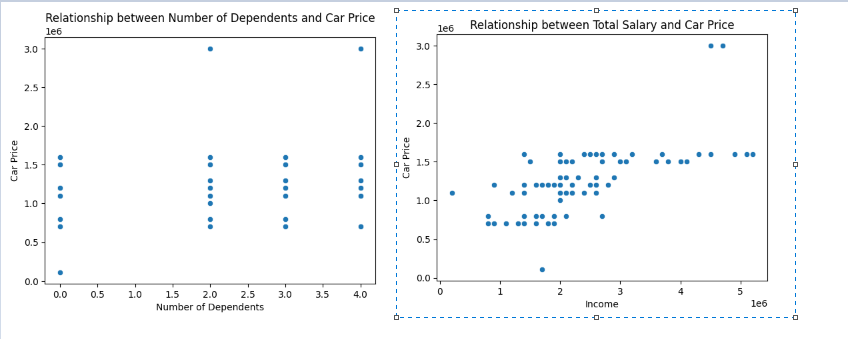




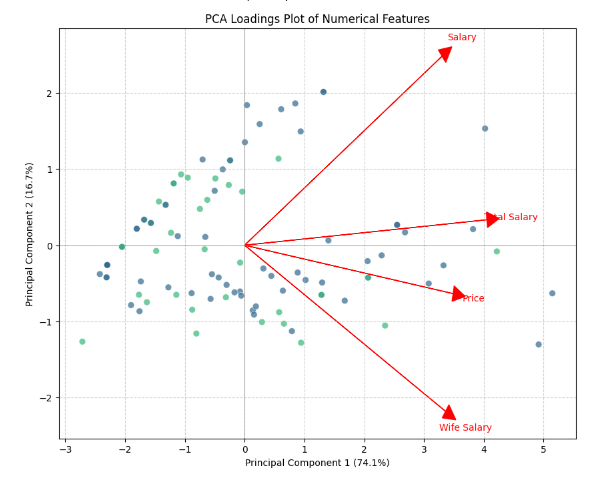




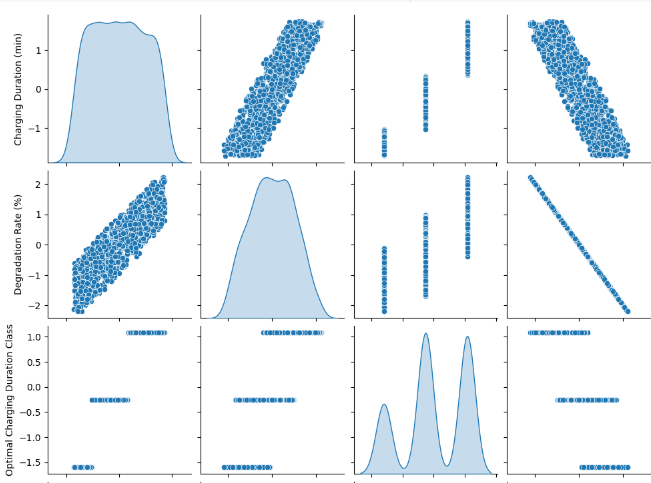
SUVs are sold a lot as it might be family friendly as it is spatial and has more comfort that small cars with same number of seats.

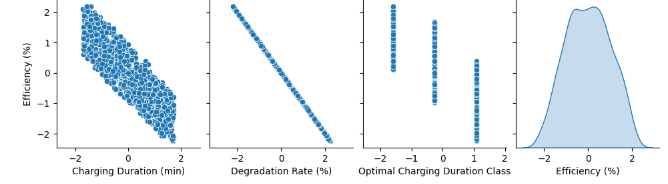


There is no proper relationship between number of dependents and car price whereas a positive relationship is seen between total salary and car price.



**d. Battery:**





Relationships observed:

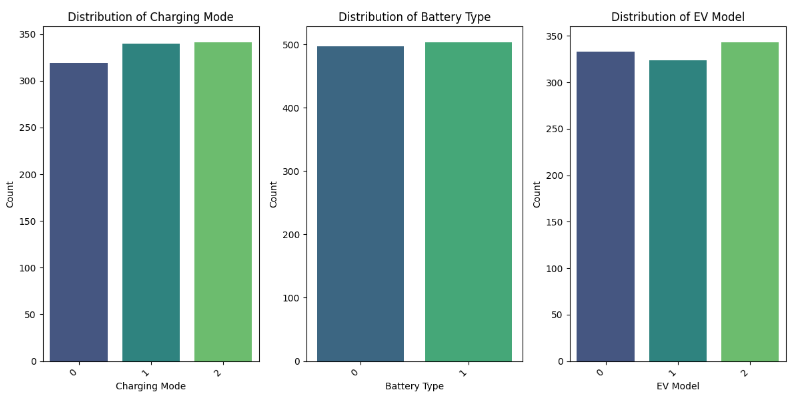
* Charging Duration (min) and Efficiency (%): Strong negative linear correlation: as duration increases, efficiency decreases. Longer charging durations may reduce battery efficiency.
* Charging Duration vs Degradation Rate: Positive linear correlation. Longer charging durations correlate with increased degradation rate.
* Efficiency vs Degradation Rate: Strong negative correlation (almost perfect line). As degradation rate increases, efficiency decreases sharply. Indicates that degradation directly affects efficiency.

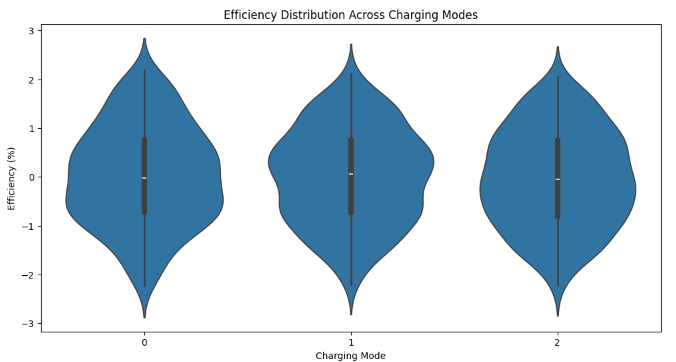
Practical Insights:

* Shorter durations usually correlate with better efficiency and less degradation.
* High degradation directly reduces efficiency, indicating the importance of minimizing degradation.
* Battery efficiency and charging duration show clear inverse relationship.

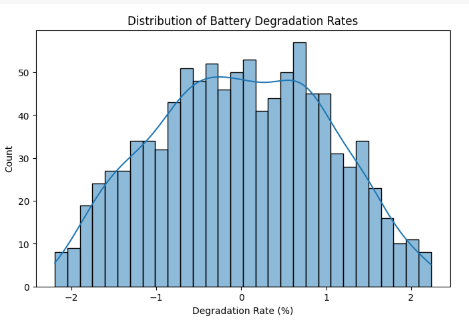
Actionable insights from this analysis:

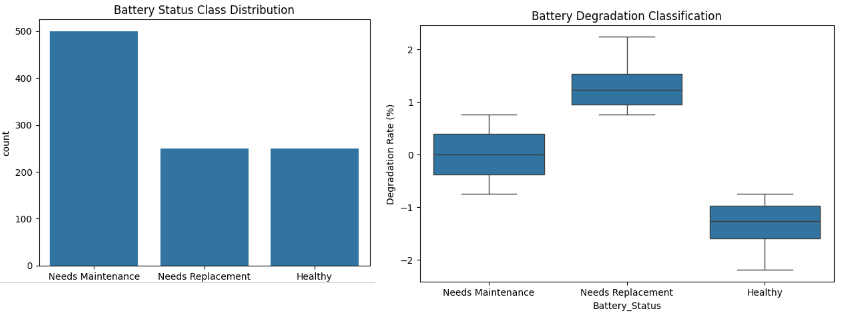
* Optimal charging practices should prioritize shorter charging times to maintain efficiency.
* Strategies should consider balancing charging duration and acceptable degradation rates.

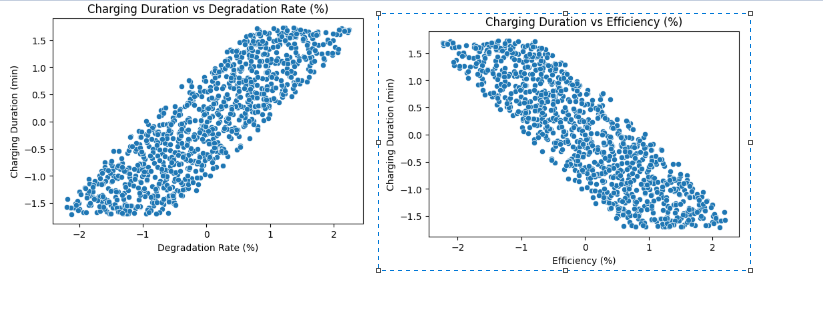




This shows that it is not a significant factor.







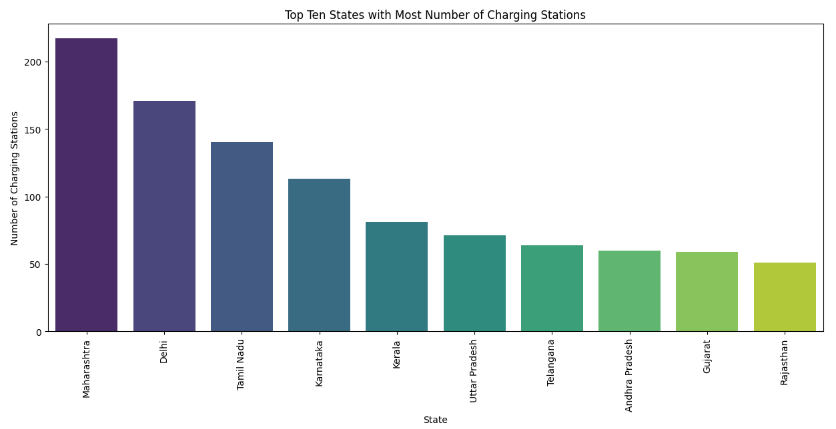
Degradation Rate vs. Charging Duration

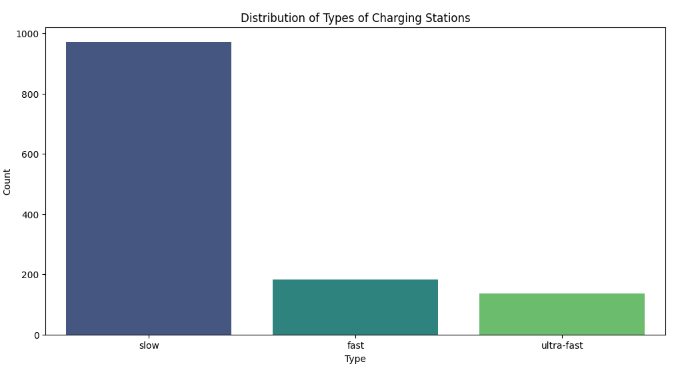
* Strong positive correlation → As degradation increases, charging time increases.
* Likely cause: Older/degraded batteries may charge slower due to efficiency loss.

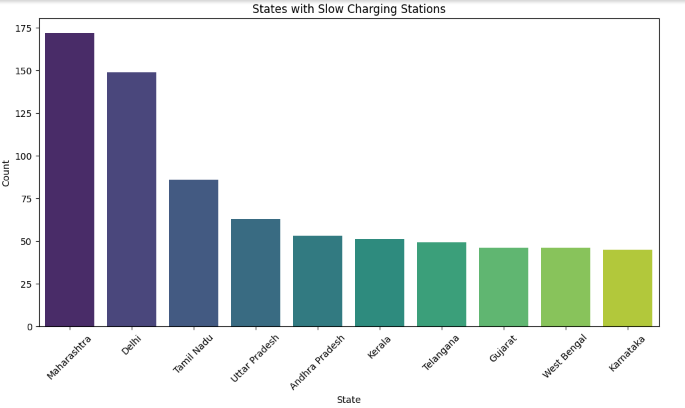
Efficiency vs. Charging Duration

* Strong negative correlation → More efficient batteries charge faster.
* Likely cause: Lower efficiency means more energy is lost as heat, extending charging time.

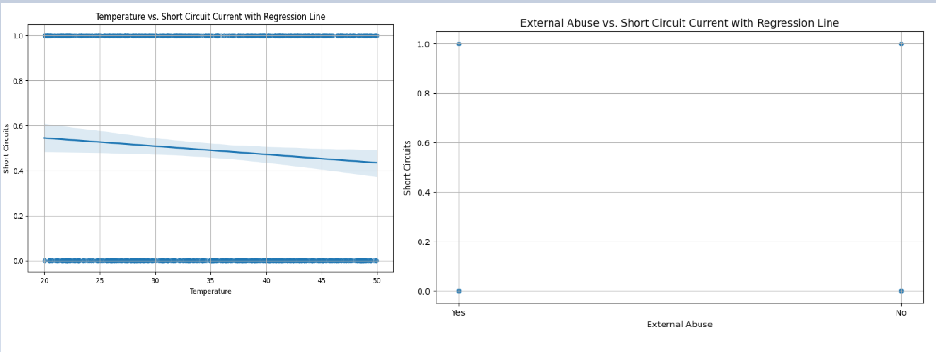
**e. Charging Stations:**

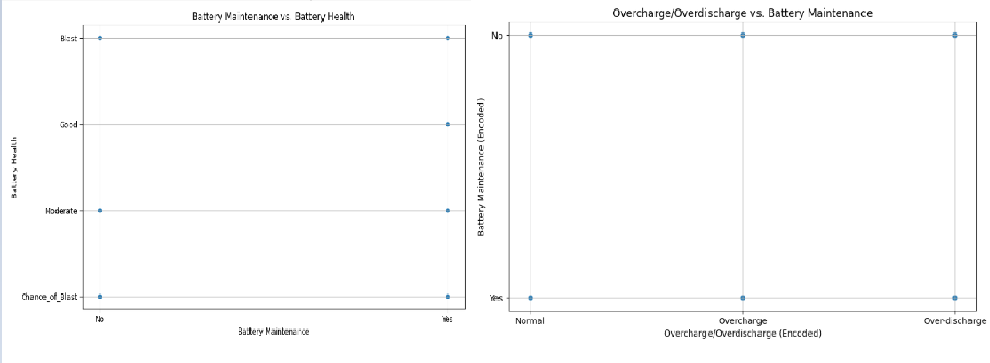
  
This makes absolute sense as the number of electric vehicles are sold highest in Maharastra and Delhi from state wise sales data. So, more number of charging stations should be present in these states.

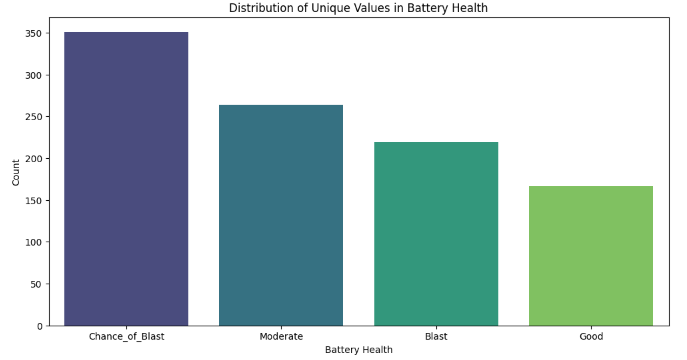


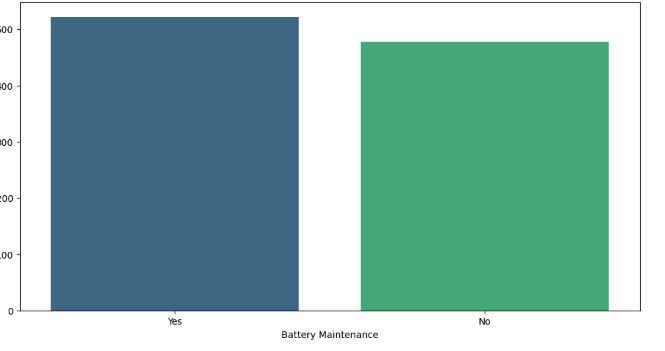


**f. Battery Blast (Synthetic dataset from Kaggle)**



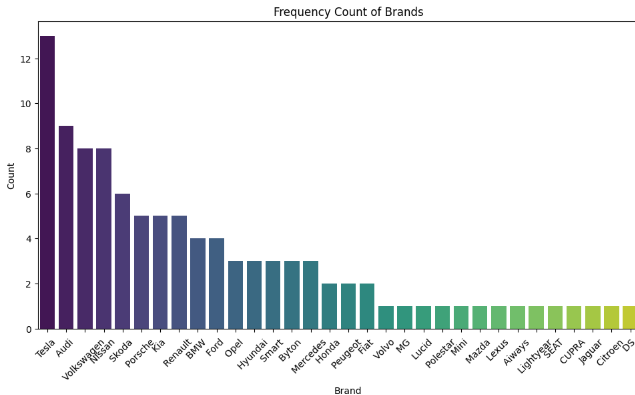


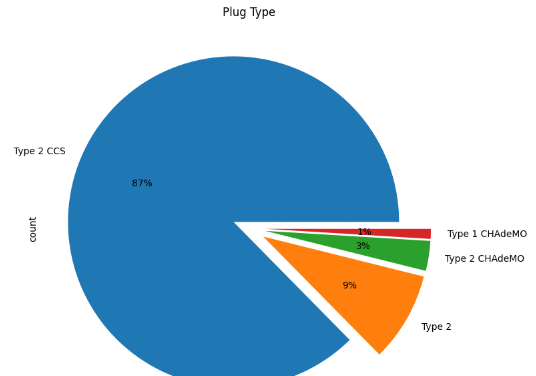


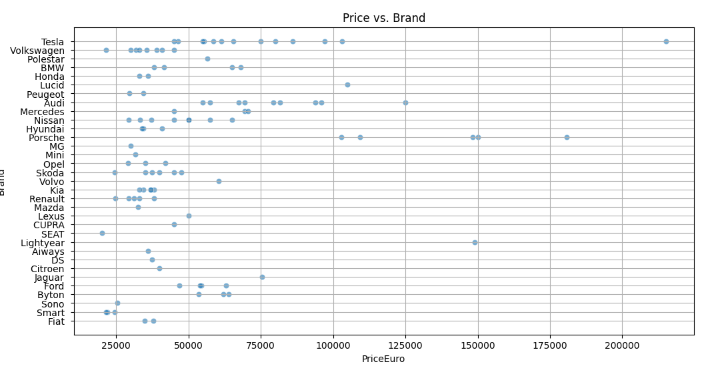


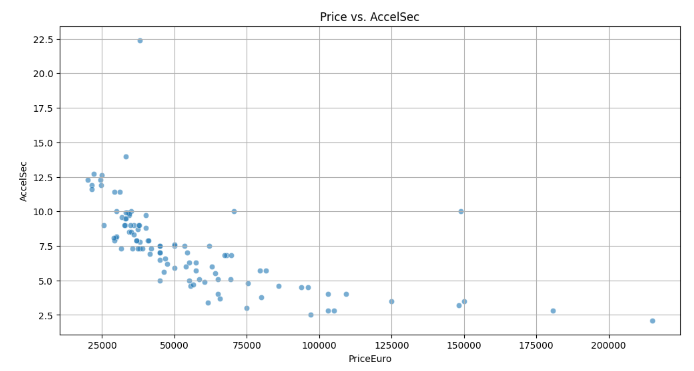
Since this is a synthetic dataset, no much relationship is found among the variables.

**g. Vehicle Specifications:**





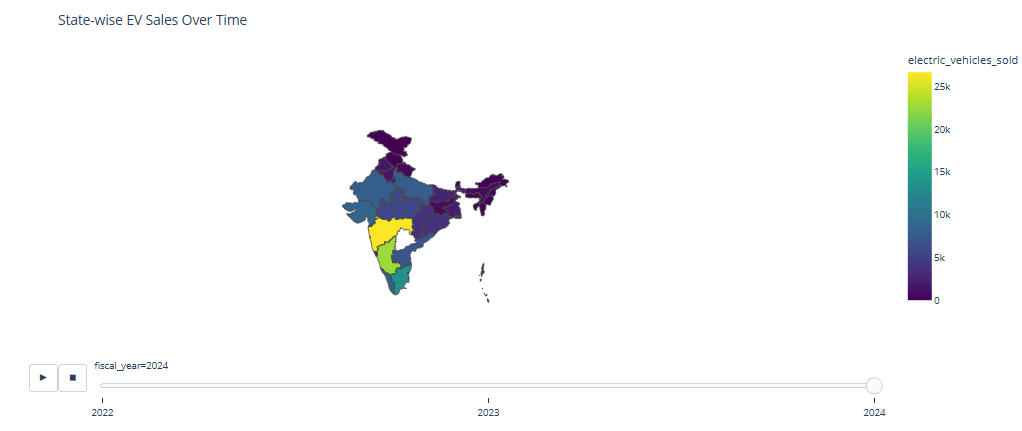




As the price increases, the AccelSec is decreasing. They might be negatively correlated.



It is clear that there is a linear relationship between price and efficiency.



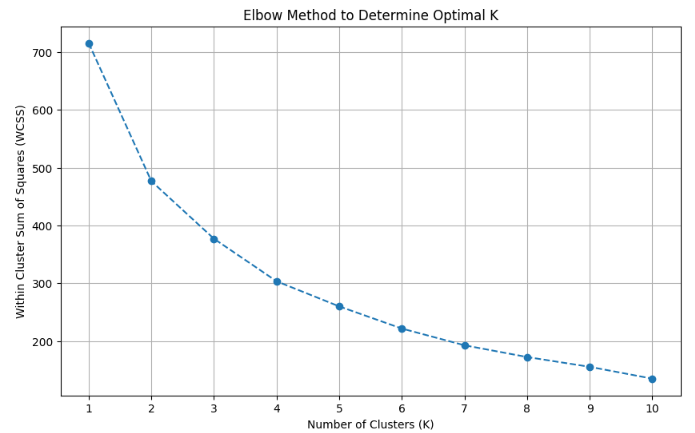
From the above map, it is clear that over the time, the sales of the electric vehicles increased.

**5. Segment Extraction**

**a. K-Means**

K-Means Clustering is one of the most popular Unsupervised Machine Learning Algorithms Used for Solving Classification Problems. K Means segregates the unlabeled data into various groups, called clusters, based on having similar features, common patterns. Suppose we have N number of Unlabeled Multivariate Datasets of various features like water-availability, price, city etc. from our dataset. The technique to segregate Datasets into various groups, on the basis of having similar features and characteristics, is called Clustering. The groups being formed are known as Clusters. Clustering is being used in Unsupervised Learning Algorithms in Machine Learning as it can segregate multivariate data into various groups, without any supervisor, on the basis of a common pattern hidden inside the datasets.

In the Elbow method, we are actually varying the number of clusters (K) from 1 – 10. For each value of K, we are calculating WCSS (Within-Cluster Sum of Square). WCSS is the sum of squared distance between each point and the centroid in a cluster. When we plot the WCSS with the K value, the plot looks like an Elbow. As the number of clusters increases, the WCSS value will start to decrease. WCSS value is largest when K = 1. When we analyze the graph, we can see that the graph will rapidly change at a point and thus creating an elbow shape. From this point, the graph starts to move almost parallel to the X-axis. The K value corresponding to this point is the optimal K value or an optimal number of clusters.

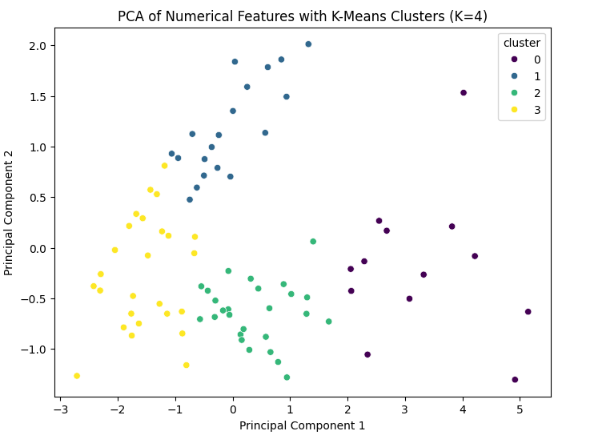


**Behavioural Segmentation:**

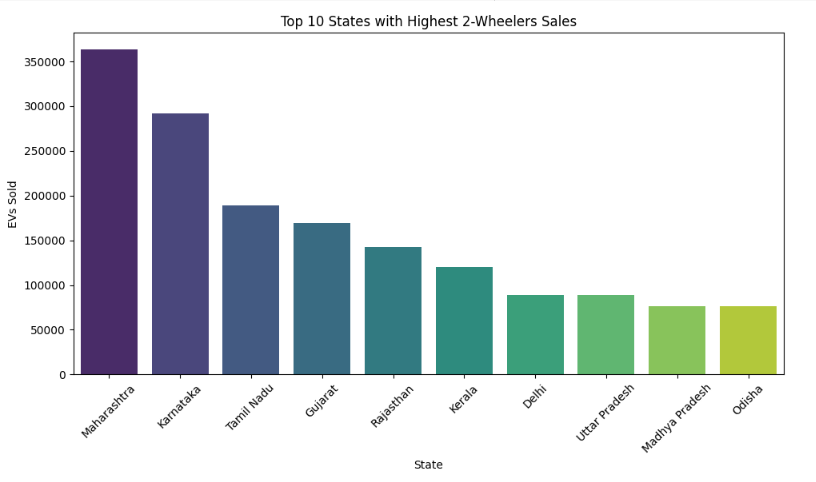
Segmenting Consumers based on Behavior and Attributes (Clustering): If you want to group consumers with similar buying behaviors or attributes, clustering models are suitable.

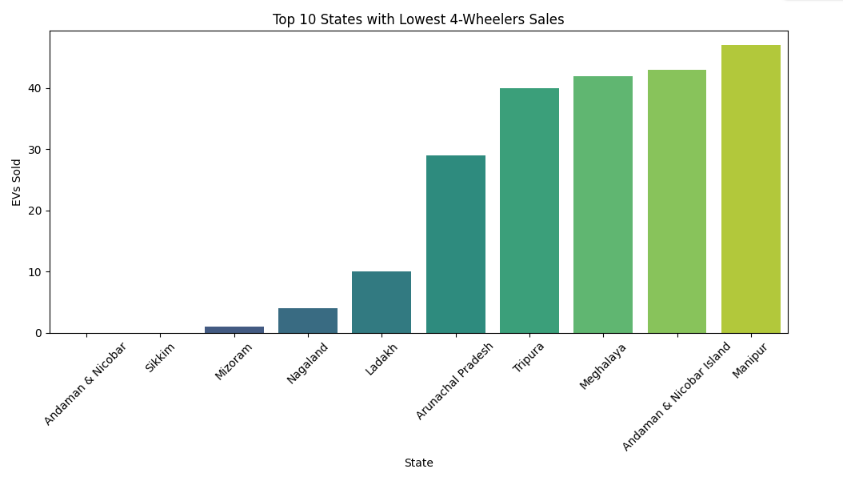
* K-Means Clustering: A popular algorithm for partitioning data into a predefined number of clusters.
* DBSCAN: Can find clusters of varying shapes and handle noise.

In this notebook, K-Means has been used.



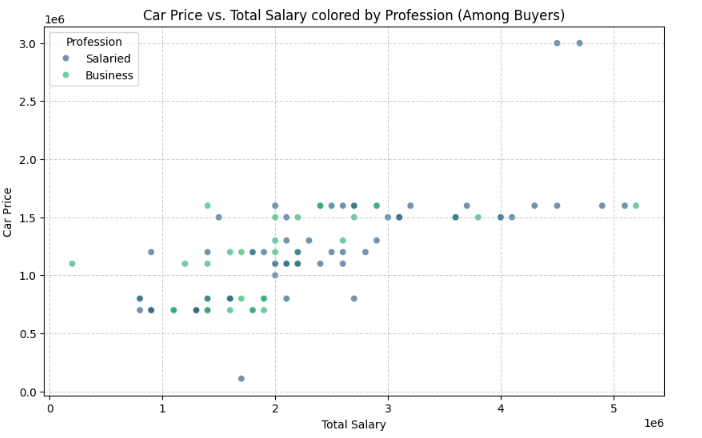
**Geographic Segmentation:** Segmenting the market based on geography. This mainly includes characteristics of the market based on the location.



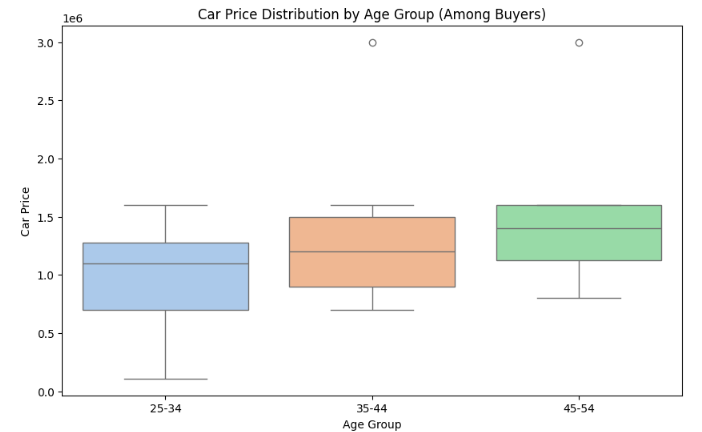


**Demographic Factors**

Major demographic factors to consider – age, income, number of dependents

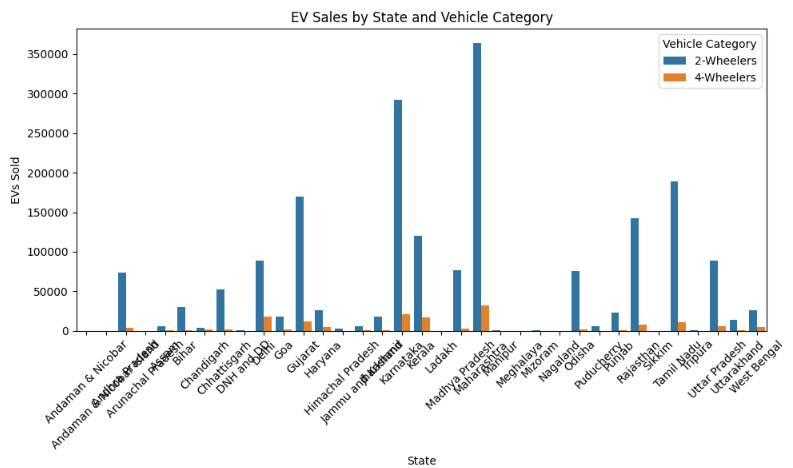


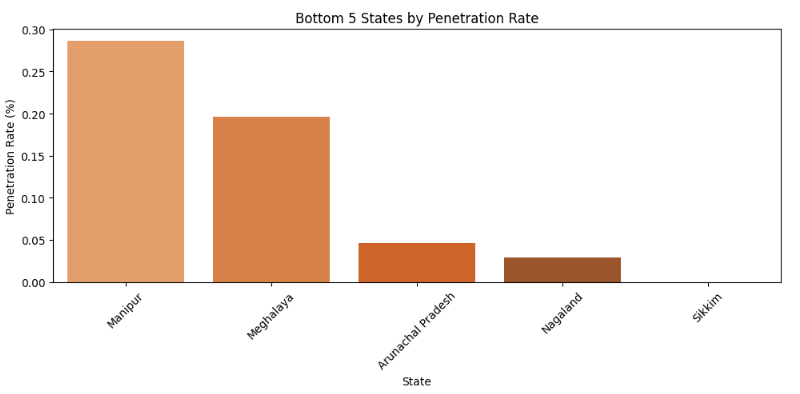
As the total salary increases, the car price is higher.



Age group 45-54 can afford higher car price as they might be well settled in their careers or occupations and earn well than the younger age groups.

**Geographic Factors**





**b. Prophet**

Prophet is an open-source time series forecasting library developed by Meta AI, designed for business and economic data with trends, seasonality, and holidays. It’s built in Python (and R), user-friendly, and robust to missing data, outliers, and trend shifts, making it ideal for forecasting EV sales trends in your datasets.

**Mechanics**:

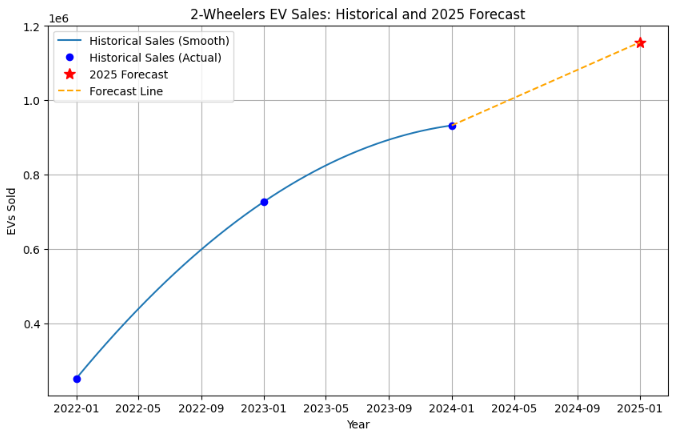
1. **Input**: Requires a DataFrame with columns ds (date, datetime format) and y (value to forecast, e.g., sales).
2. **Fitting**: Uses Bayesian inference or maximum a posteriori (MAP) estimation to fit the model, automatically detecting changepoints and seasonality.
3. **Forecasting**: Generates future predictions with uncertainty intervals (e.g., 80% confidence bounds).
4. **Components**: Outputs trend, seasonality, and holiday effects separately for analysis.
5. **Robustness**: Handles missing data, outliers, and non-uniform time intervals effectively.

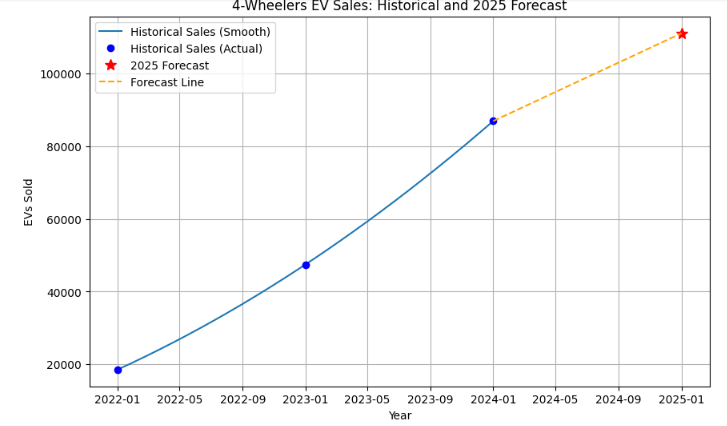
**Key Features**:

* Automatic changepoint detection for trend shifts.
* Customizable seasonality (yearly, monthly, weekly).
* Holiday effects for specific events.
* Scalable for large datasets with fast computation.

**EV Sales Forecasting (2025+)**

* Goal: Predict future EV sales by state/category
* Model: Time Series (e.g., ARIMA, Prophet) or Regression (e.g., XGBoost)
* Features: state, fiscal\_year, quarter, vehicle\_category





**6. Target Segments**

Based on the analysis, the target segment can be narrowed down to EVs having:

* Demographic factors such as age, income, number of dependents
* Behavioural factors such as good Acceleration and viable Price range
* Geographic factors such as States which are more market friendly

In conclusion, the target segment should comprise of EVs having Acceleration of 7.5-10 sec, High in Comfort and Value for Money ratings, have a Price range of 20-30 Lakhs, and be focused mainly on States such as Maharashtra, Karnataka, Tamil Nadu and Rajasthan.

**7. Customizing the market mix**



The marketing mix, comprising the 4Ps (Product, Place, Promotion, and Price), is a critical framework for driving business growth in the automotive industry by aligning strategies with the target market as outlined in the marketing plan. The company aims to maximize sales and strengthen its market presence through a robust competitive position. However, strategic decision-makers must maintain flexibility in their approaches to adapt to evolving market dynamics. The automotive sector presents significant growth opportunities, particularly for products incorporating advanced computing technologies. Concurrently, the company faces external threats in its business environment. By leveraging SWOT analysis, managers can strategically adjust the 4Ps to capitalize on opportunities and mitigate threats, ensuring sustained competitiveness and market relevance.

**Product Mix**

The product component of the marketing mix encompasses the company's offerings, with each product line representing a distinct group of outputs or products. The complete set of product lines constitutes the product mix, reflecting a strategic level of business diversification. In the context of the automotive industry, the company provides a diverse portfolio, including various brands, types, and models of electric vehicles (EVs), tailored to meet the needs of target market segments such as young professionals, urban families, and commercial operators.

* Automobiles
* Automobile parts
* Commercial vehicles
* Financial services

**Prices and Pricing Strategies**

The pricing component of the marketing mix focuses on establishing optimal price points and ranges for the company’s electric vehicle (EV) products to align with market expectations and business objectives. Strategic pricing shapes the perceived value of brands and models, significantly influencing sales performance in price-sensitive markets such as India, where affordability drives consumer adoption across diverse segments. The pricing strategies for its automotive products are as follows:

* Market-oriented pricing strategy
* Premium pricing strategy

**Promotional Mix**

Promotional activities are considered in this aspect of marketing mix of 4Ps. These activities are also known as marketing communications tactics. The combination of these tactics is called a promotional mix or marketing communications mix the following promotional activities are used, arranged according to significance in the automotive business:

* Advertising (primary)
* Direct marketing
* Personal selling
* Sales promotion
* Public relations

**Place/Distribution**

In this aspect of marketing mix or 4Ps, the virtual or physical locations of transactions are considered. Such locations are significant because they enable the company to reach target customers in specific markets, while also allowing customers to access information and products available from the automotive business. The following places are used in the distribution of products and services:

* Official websites
* Dealerships
* Automotive shows and exhibits