

Electric Vehicle Market in India Market Segmentation

i

submitted by subeena k.k

Indian Electric Car market

CAGR 34.5%
(2018-2025)

**\$71.1
Million**

2017

**\$707.4
Million**

2025



Abstract

Market segmentation is a vital approach for promoting the widespread adoption of emerging transportation technologies, such as electric vehicles (EVs), in developing markets. With EVs gaining traction as low-emission and cost-effective alternatives, their adoption is anticipated to grow significantly in the near future, sparking heightened academic interest. This study aims to identify and analyze potential buyer segments for EVs by examining psychographic, behavioral, and socio-economic factors through a structured research framework, integrating concepts of perceived benefits, attitudes, and purchase intentions.

A combination of advanced analytical techniques, including cluster analysis, multiple discriminant analysis, and Chi-square tests, was employed on survey data collected from 563 respondents via an online cross-sectional survey. The analysis revealed three distinct consumer segments: Conservatives, Indifferents, and Enthusiasts, which represent the early adopter demographics of EVs. The findings offer actionable insights for researchers and policymakers, providing strategic guidance to foster EV adoption within the context of sustainable transportation development.

This report utilizes Fermi Estimation to address the problem systematically by breaking it into manageable components for detailed analysis.

Keywords: Electric vehicles, Market segmentation, Sustainable transportation, Behavioral analysis, Adoption intention, Psychographics.

Data Collection

The data has been collected manually, and the sources used for this process are listed below :

- <https://www.kaggle.com/datasets>
- <https://data.gov.in/>
- <https://www.data.gov/>
- <https://data.worldbank.org/>
- <https://datasetsearch.research.google.com/> F

Frameworks and Libraries

- SKLearn: Simple and efficient tools for predictive data analysis
- Seaborn: Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.
- Plotly: The plotly Python library is an interactive, open-source plotting library that supports over 40 unique chart types covering a wide range of statistical, financial, geographic, scientific, and 3-dimensional use-cases.
- KElbowVisualizer: The KElbowVisualizer implements the “elbow” method to help data scientists select the optimal number of clusters by fitting the model with a range of values for k . If the line chart resembles an arm, then the “elbow” (the point of inflection on the curve) is a good indication that the underlying model fits best at that point. In the visualizer “elbow” will be annotated with a dashed line.
- Matplotlib : Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python.
- Numpy: Caffe-based Single Shot-Multibox Detector (SSD) model used to detect faces
- Pandas: pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.

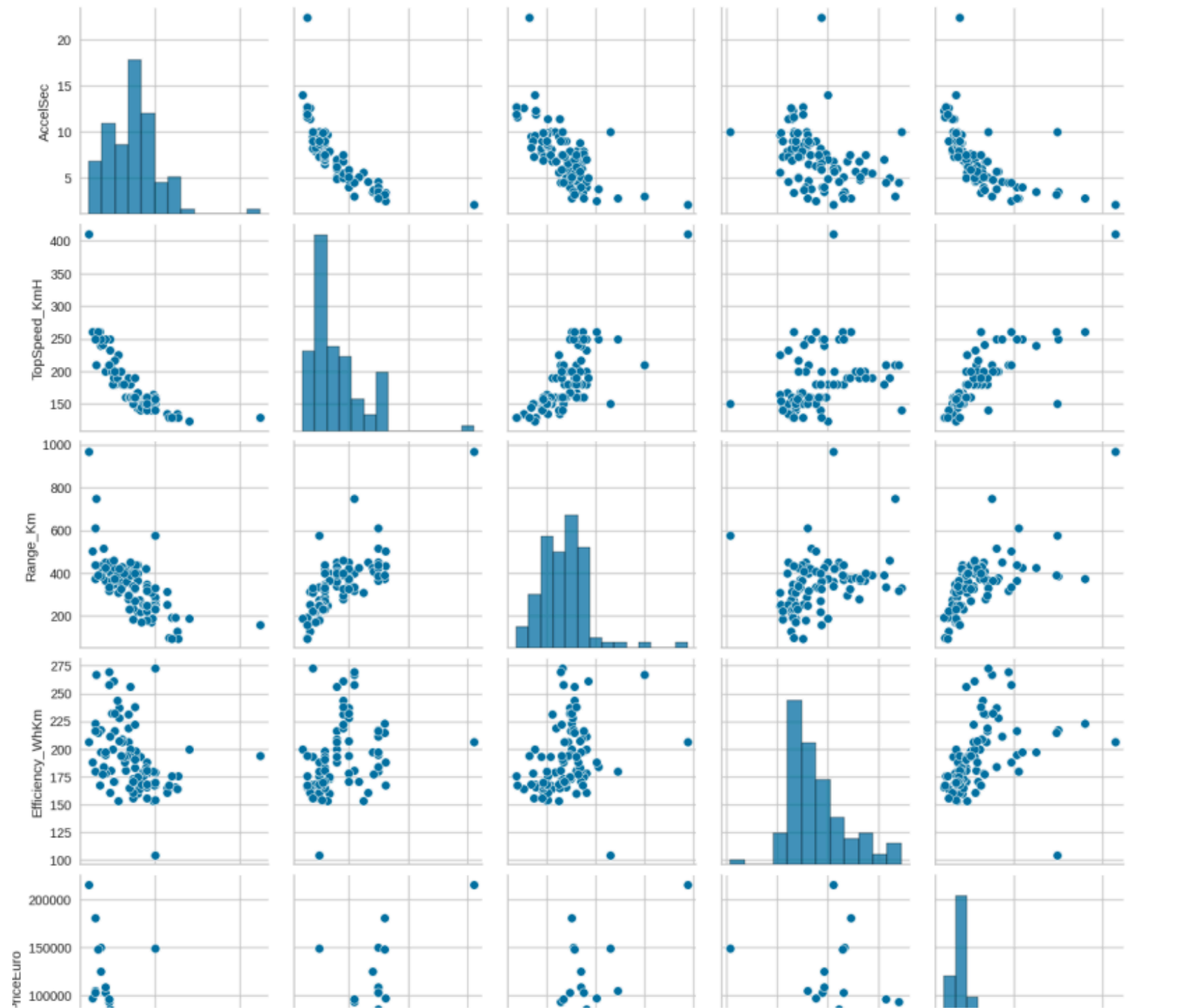
Data Cleaning

The data collected is compact and is partly used for visualization purposes and partly for clustering. Python libraries such as NumPy, Pandas, Scikit-Learn, and SciPy are used for the workflow, and the results obtained are ensured to be reproducible.

Exploratory Data Analysis:

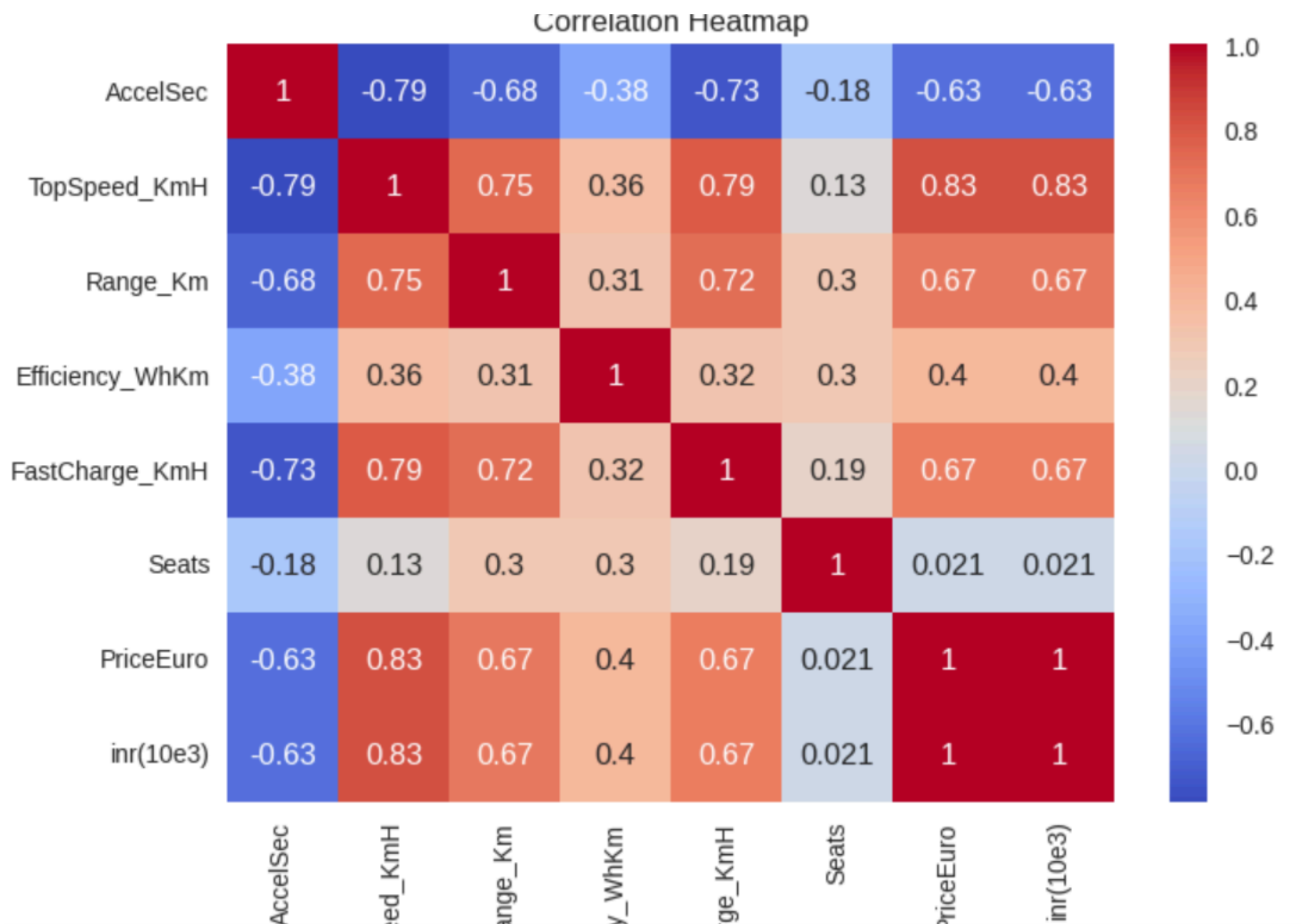
Exploratory Data Analysis is a crucial step in understanding your dataset and uncovering insights before applying statistical models or machine learning algorithms. Here are some key steps and notes about the EDA process.

Correlation Between Numerical Variables

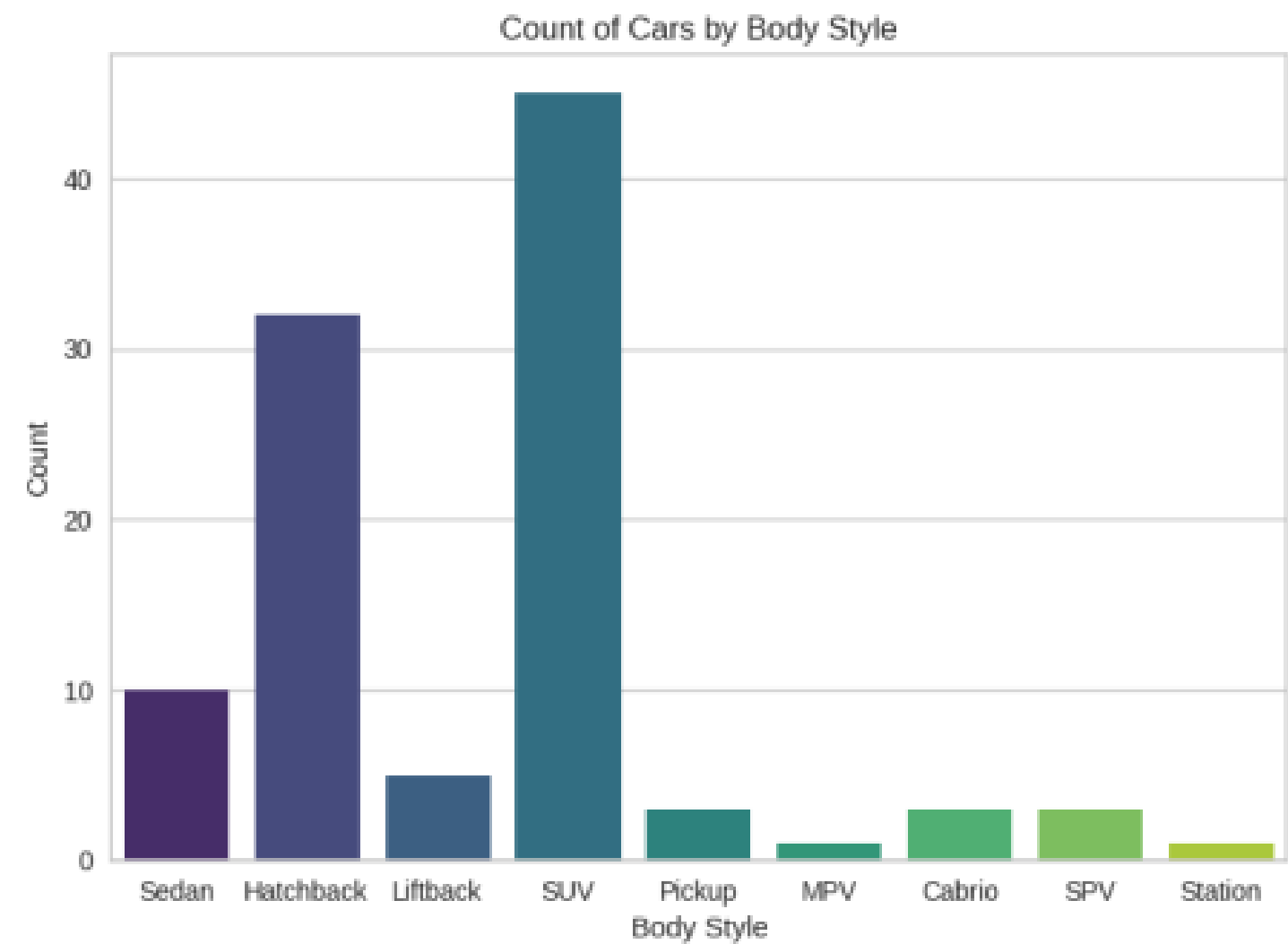


Correlation Matrix: A correlation matrix is simply a table that displays the correlation between each pair of variables. It is best used in variables that demonstrate a linear relationship between each other. Coefficients for different variables. The matrix depicts the correlation between all the possible pairs of values through the heatmap in the below figure. The relationship between two variables is usually considered strong when their correlation coefficient value is larger than 0.7.

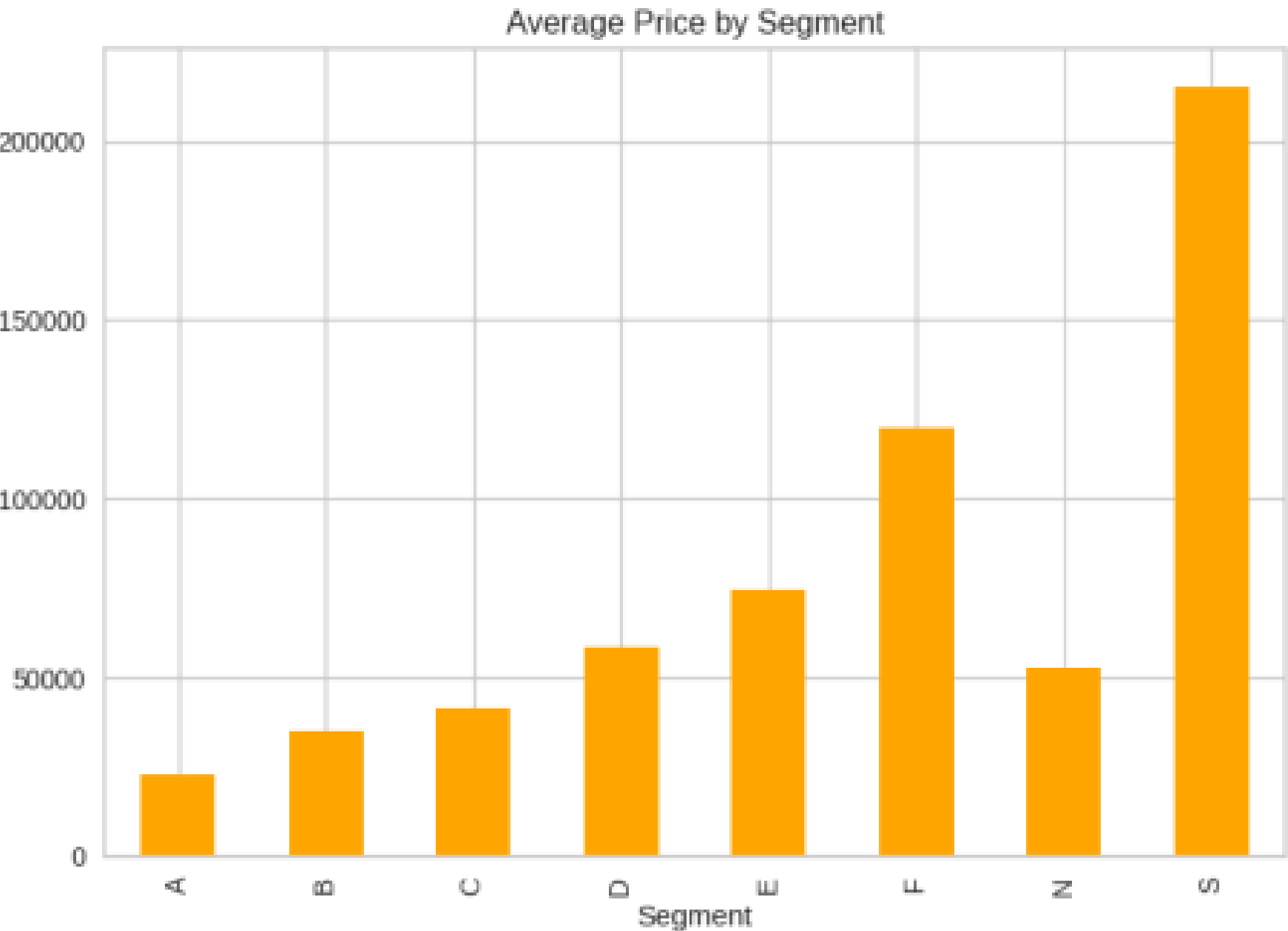
Heatmap for correlation



Countplot for BodyStyle

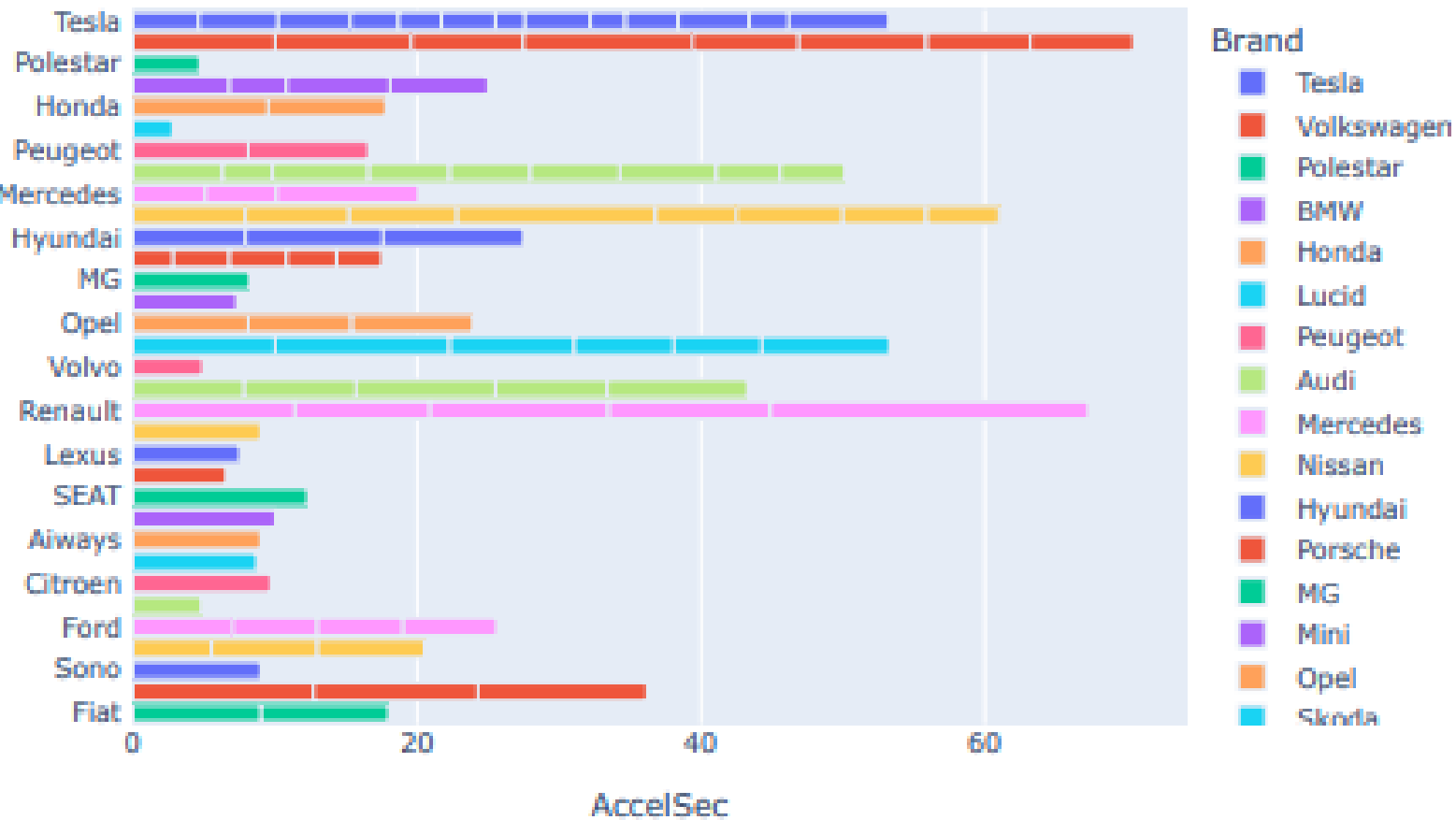


Bar chart for average Price by Segment

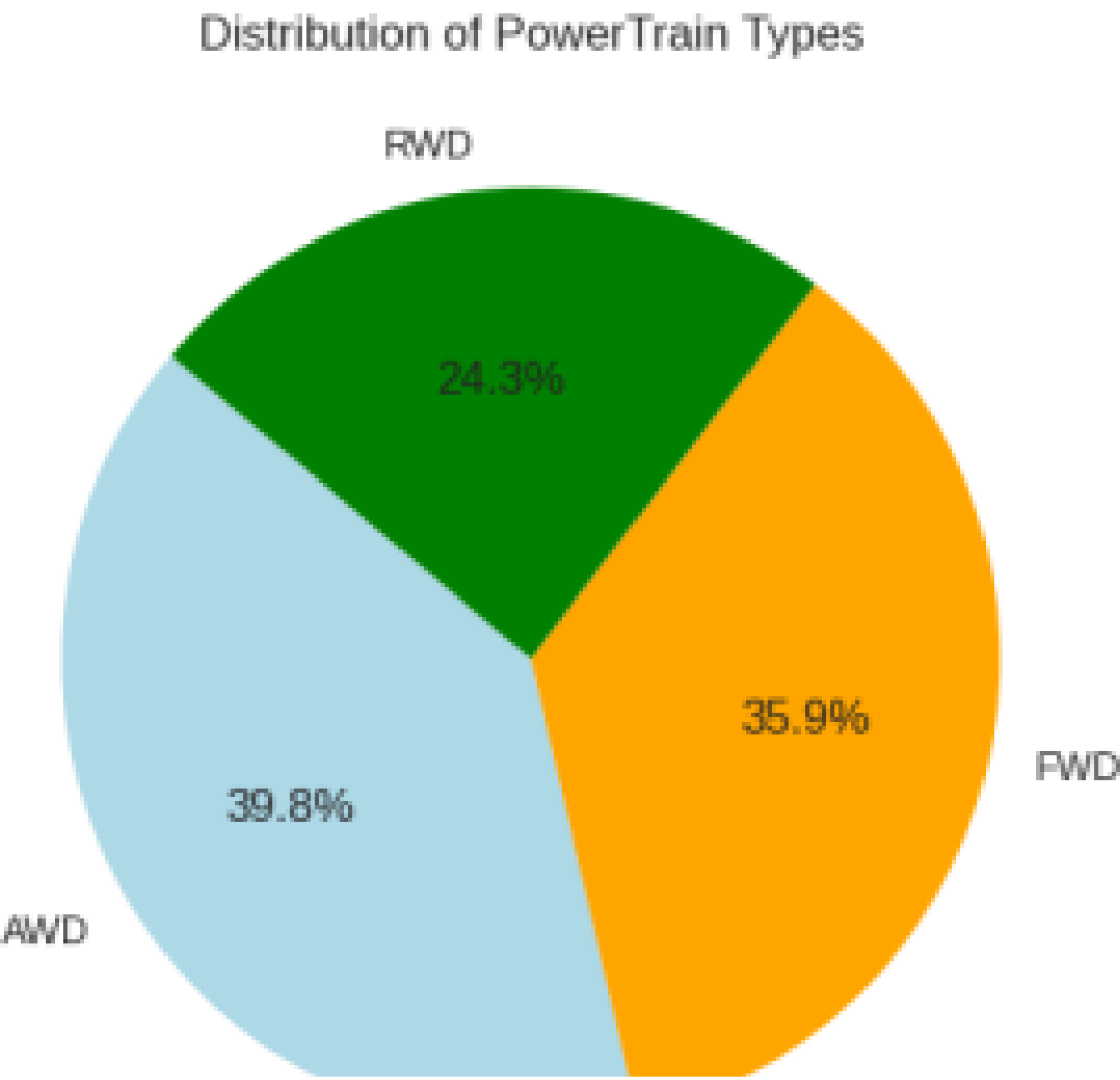


Which car has fastest accelaration?

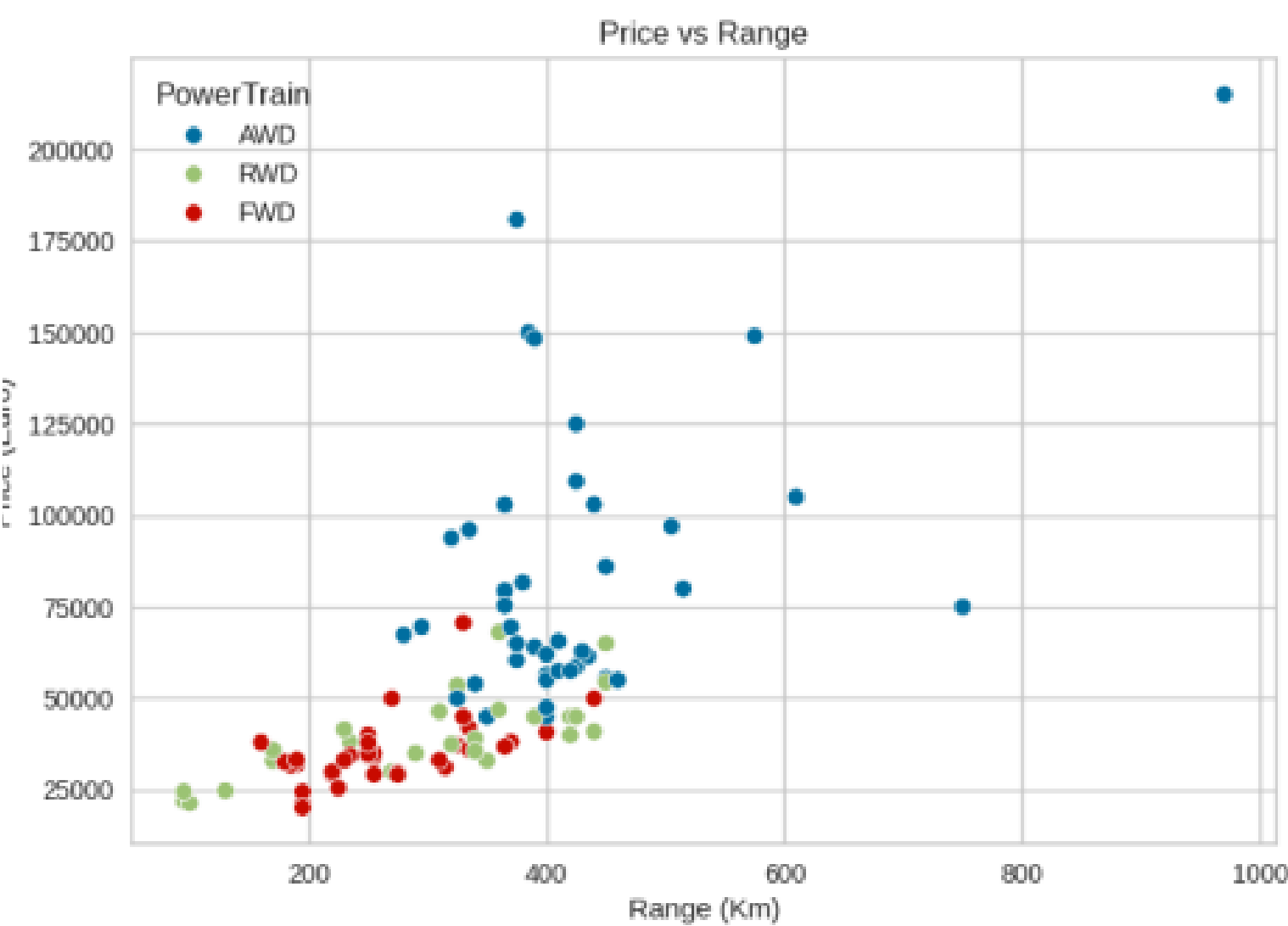
Which car has fastest accelaration?



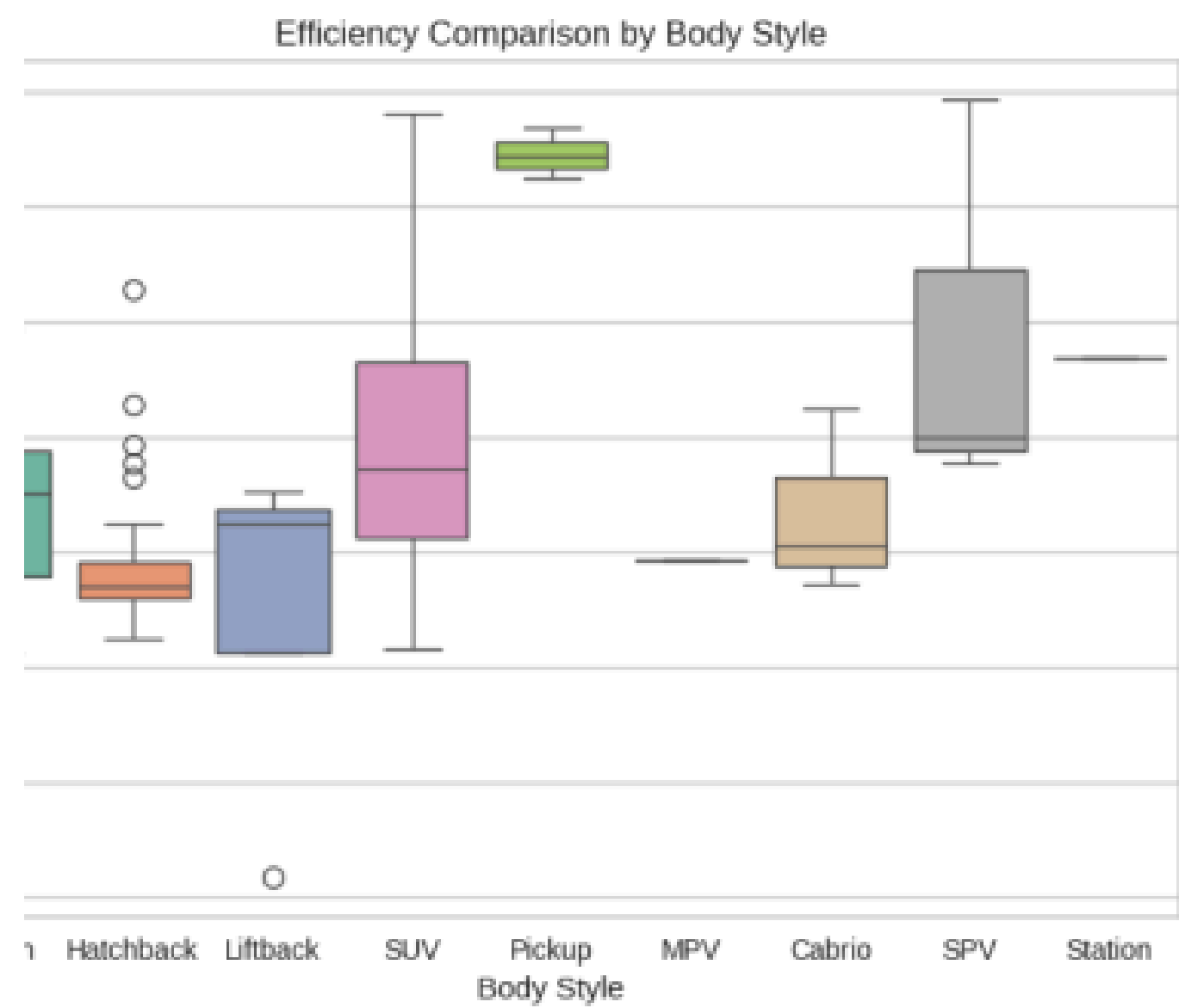
Distribution of PowerTrain Types



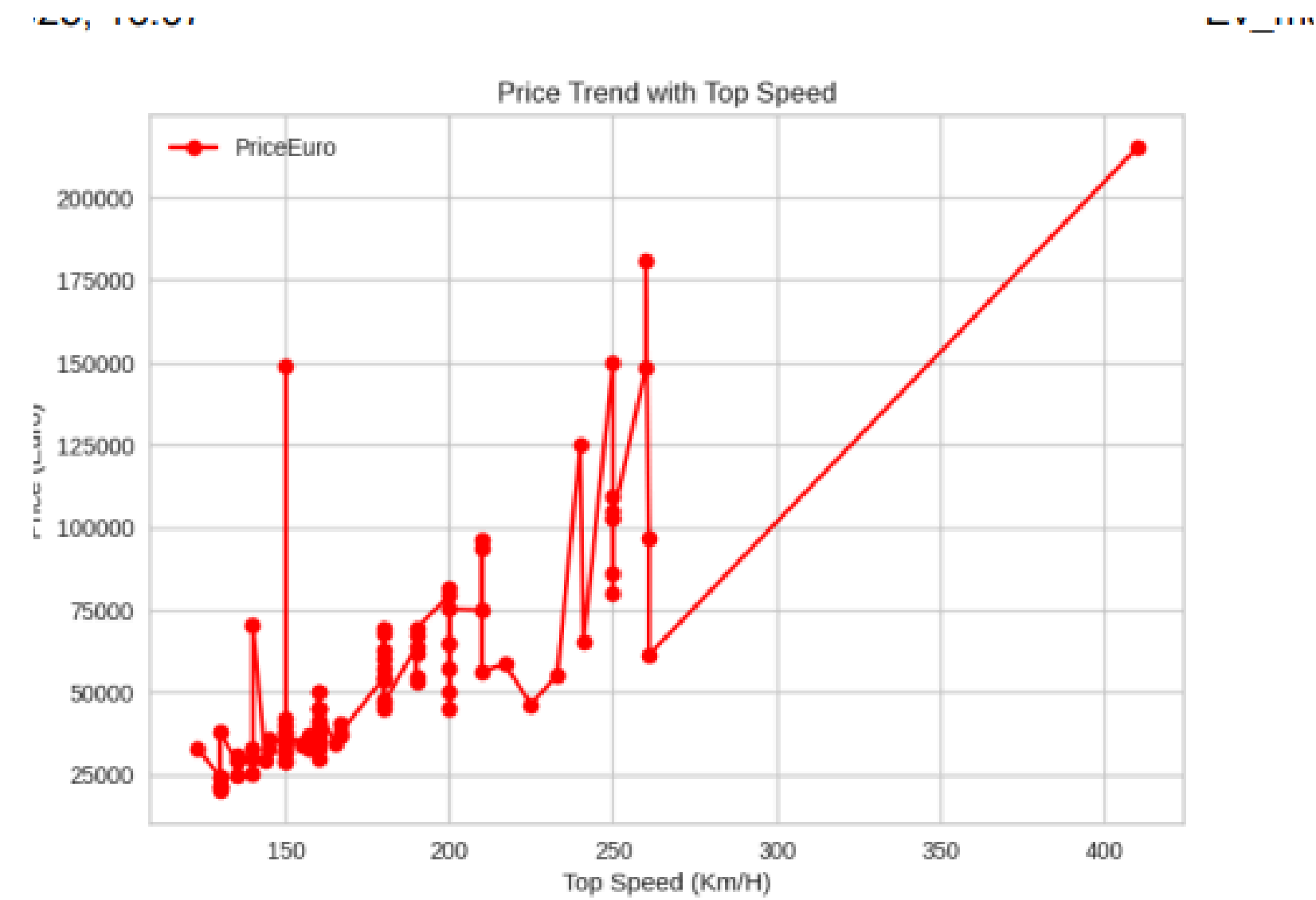
Scatter plot for Range vs Price



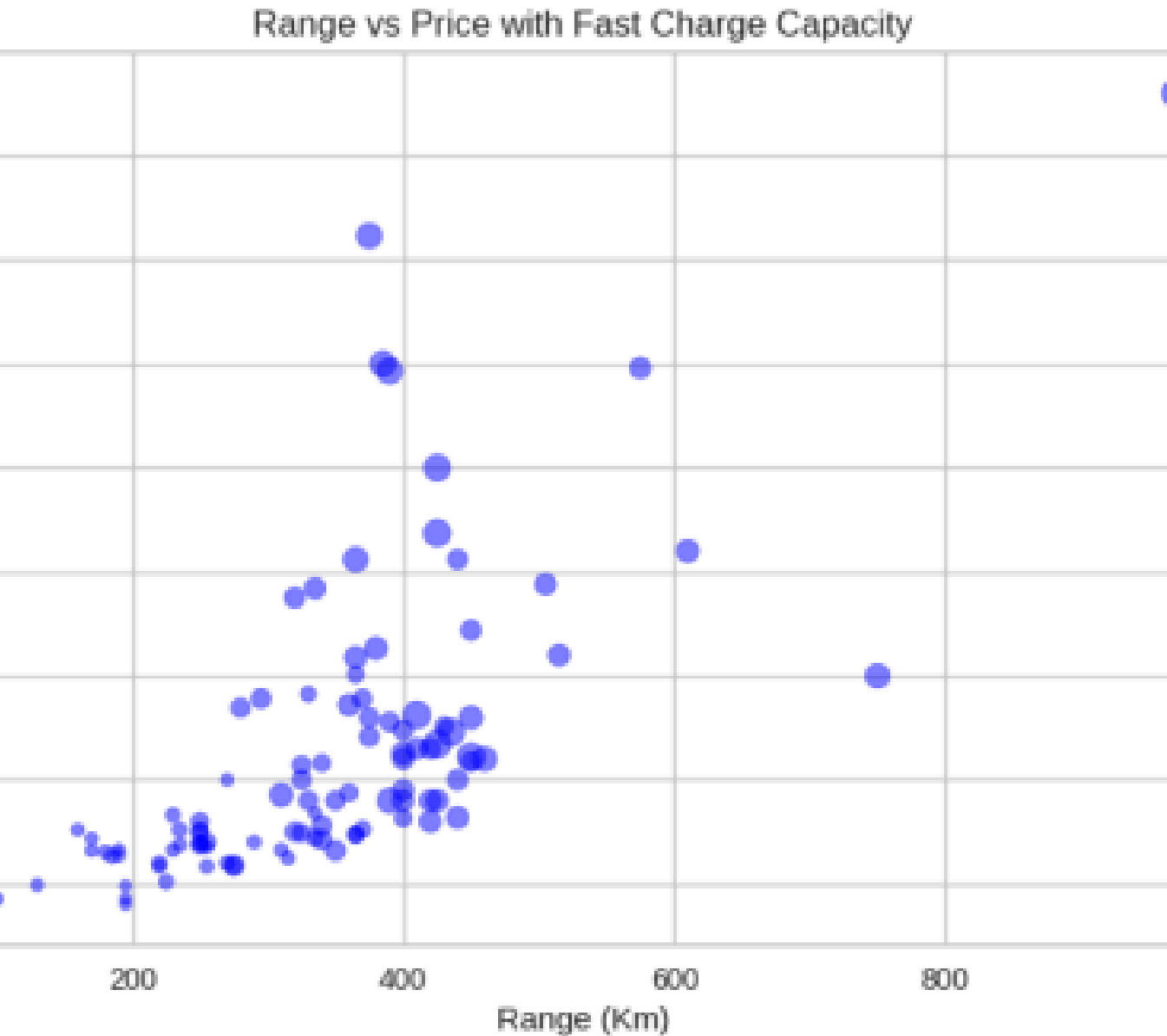
Boxplot for Efficiency by BodyStyle



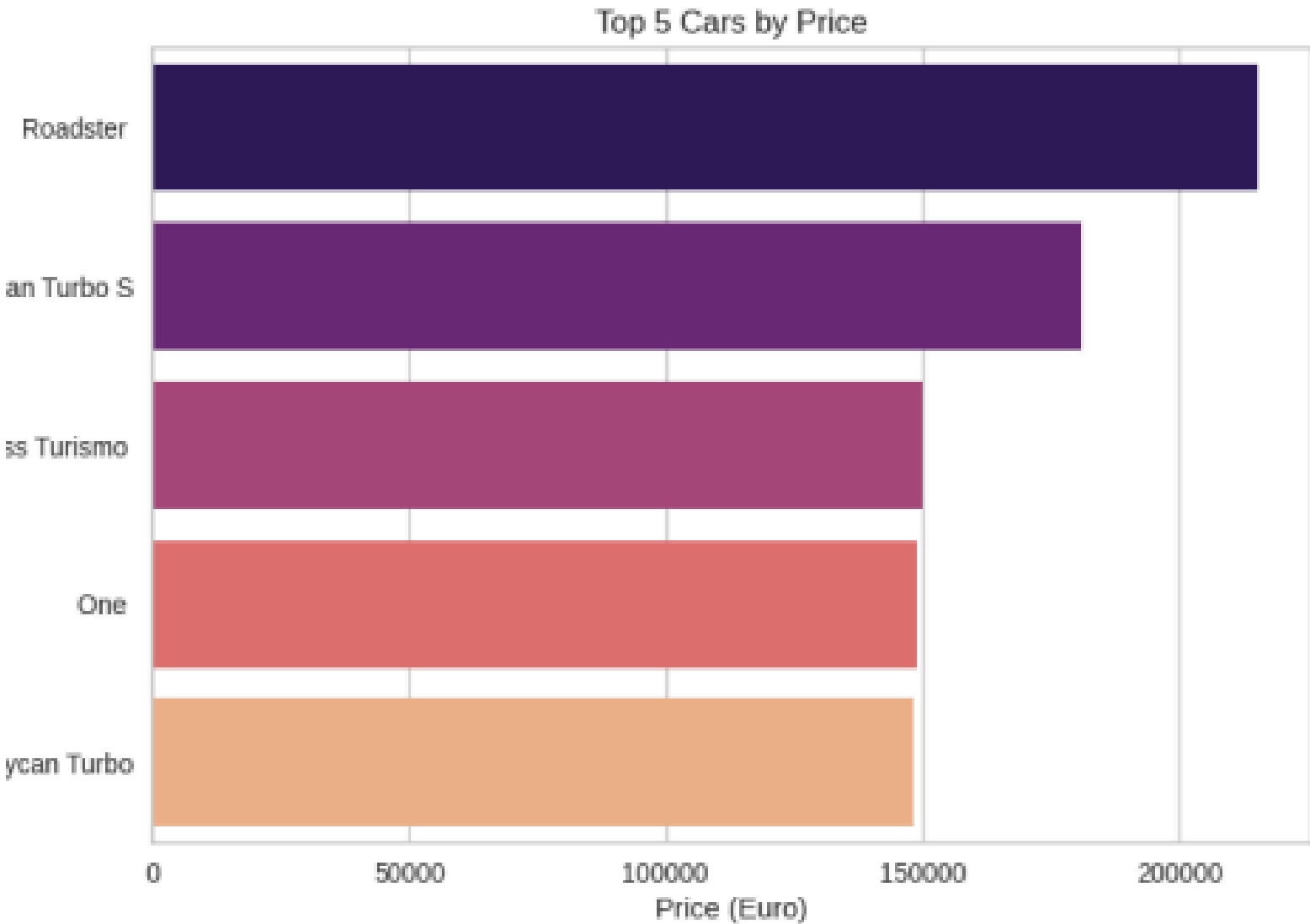
Efficiency Comparison by Body Style



Price Trend with Top Speed



Range vs Price with Fast Charge Capacity



Prediction of Prices for Most Used Cars

Linear regression is a supervised machine learning algorithm that performs regression tasks by predicting target values based on independent variables. This model is widely employed to understand relationships between variables and to make forecasts.

In this analysis, we use a linear regression model to predict the prices of various electric cars offered by different companies. The independent variables (XXX) include various features of the cars, while the dependent variable (yyy) represents the prices to be predicted.

The dataset is split into training and testing sets with a 40:60 ratio, meaning 40% of the data is used to train the model. The model is trained using the command:

```
LinearRegression().fit(X_train, y_train),
```

which fits the training data into the regression model.

Conclusion

Based on the insights from the exploratory data analysis and visualizations:

1. **Body Style Preferences:** Sedans and SUVs dominate the market, while MPVs have the least demand, indicating a strong consumer preference for style and versatility.
2. **Segment Pricing Insights:** Segment S (premium) vehicles are the most expensive, while Segment A (economy) vehicles are the least expensive, reflecting a wide range of consumer affordability and preferences.
3. **PowerTrain Distribution:** AWD is the most preferred drivetrain type, followed by FWD and RWD, indicating a consumer preference for versatility and performance.
4. **Top Brands by Price and Performance:** Brands like Volkswagen stand out for acceleration performance, while premium brands dominate the high-price range, catering to different consumer priorities such as speed, luxury, and innovation.
5. **Price Segmentation:** The luxury segments are dominated by high-performing brands, while affordable segments cater to practicality and efficiency, showcasing a diverse market landscape.

Suggestions

1. Focus on Popular Body Styles: Manufacturers should invest in developing Sedans and SUVs to meet the growing demand, while evaluating the potential for increasing MPV popularity through marketing or feature enhancements.
2. Strategize Pricing: Balance pricing strategies to attract both premium buyers (Segment S) and budget-conscious customers (Segment A) by offering a mix of features and affordability.
3. Innovate Drivetrain Options: Given the preference for AWD vehicles, manufacturers could expand their AWD lineup while maintaining options for FWD and RWD for diverse consumer needs.
4. Brand-Specific Strategies: Premium brands should continue to focus on innovation and luxury, while other brands may focus on affordability and practical performance to capture wider market shares.
5. Enhance Consumer Awareness: Highlight the unique selling points (e.g., performance, efficiency, and technology) of each model through targeted campaigns to influence buyer preferences and drive sales growth.