ANALYZING AUTOMOBILE MARKET TRENDS USING R, POWER BI, AND TABLEAU

1. INTRODUCTION

The automobile sector is one of the most dynamic industries, where factors such as brand reputation, engine size, fuel type, year of manufacture, mileage, and price play a crucial role in influencing consumer decisions and market trends. With the availability of large volumes of car-related data, analyzing these variables has become essential for understanding pricing strategies, customer preferences, and overall market performance. Such insights not only guide potential buyers in making informed decisions but also help manufacturers and dealers in developing competitive pricing and marketing strategies.

Focuses on analyzing a comprehensive car dataset that includes details such as manufacturer, model, engine size, fuel type, year of manufacture, mileage, and price. By leveraging advanced analytical tools like R programming, Power BI, and Tableau, the study examines the relationship between these variables to uncover patterns and trends in the automobile market. R provides powerful statistical and analytical functions for data cleaning, transformation, and hypothesis testing; Power BI enables interactive dashboards that deliver real-time insights into pricing and performance; and Tableau offers intuitive visualizations that highlight correlations, variations, and market dynamics.

Furthermore, this study emphasizes the significance of data-driven decision-making in the automobile sector. Insights derived from the analysis can help in predicting car price ranges, understanding brand value, evaluating the effect of engine specifications on pricing, and identifying how fuel type and mileage influence market demand. The integration of R, Power BI, and Tableau ensures that the analysis is both statistically accurate and visually compelling, enabling actionable recommendations for consumers, dealers, and manufacturers to enhance decision-making, optimize strategies, and improve overall market competitiveness.By combining statistical rigor with modern visualization tools, this analysis not only provides clarity on the current automobile market trends but also serves as a foundation for future predictive modeling and advanced forecasting in the industry.

1.1 PROBLEM STATEMENT

The rapid growth of the used car market has created a highly competitive environment where pricing, vehicle condition, and brand reputation are key factors for success. However, with this growth comes the challenge of managing:

- Understanding how brand and manufacturer affect the resale value of cars.
- Evaluating the impact of fuel type on used car prices and buyer preferences.
- Analyzing the relationship between car specifications (engine size, mileage, year of manufacture) and market value.
- Identifying pricing trends across different years, models, and brands.
- Recognizing factors that influence consumer decisions when buying used cars.

Without proper analysis and visualization, these insights stay hidden, which can lead to missed opportunities for better pricing strategies, improved sales, and informed business decisions.

1.2 TOOL DESCRIPTION

To perform a comprehensive analysis of the used car dataset, three powerful tools were utilized: R Programming, Power BI, and Tableau. Each tool contributed uniquely, ensuring that the study achieved both statistical rigor and visually rich, interpretable insights. The combination of these tools allowed for a seamless workflow—from data preprocessing and statistical modeling to interactive visualization and business intelligence reporting.

R PROGRAMMING

R is a robust, open-source programming language specifically designed for statistical computing, data analysis, and graphical representation. It provides a wide array of built-in functions and packages, making it an ideal choice for researchers and data analysts working with diverse datasets. R is capable of handling large-scale datasets and performing advanced operations, including regression analysis, hypothesis testing, clustering, and predictive modeling. Its extensive graphical capabilities, especially through libraries such as ggplot2, dplyr, and plotly, allow the creation of high-quality, customizable visualizations that enhance data interpretation. Furthermore, R supports automation through scripting, enabling repetitive analysis tasks to be executed efficiently. The strong global community, continuous development of R packages, and its adaptability to different data types make it a versatile tool

widely adopted in academic research, industry analytics, and data science applications. For this study, R was particularly useful for performing in-depth statistical analysis and deriving predictive insights from the used car dataset.

POWER BI

Power BI, developed by Microsoft, is a leading business intelligence and analytics tool that transforms raw data into meaningful insights through interactive dashboards and reports. It supports seamless connections to multiple data sources, including Excel, SQL databases, cloud services, and APIs, allowing for efficient data integration and transformation. With its intuitive drag-and-drop interface, users can design professional dashboards without requiring extensive programming knowledge. Key features of Power BI include real-time data updates, KPI tracking, and advanced data modeling and calculations using DAX (Data Analysis Expressions). Power BI also facilitates collaboration and sharing of insights within organizations, making it an effective tool for monitoring business operations, analyzing trends, and supporting data-driven decision-making. In this project, Power BI was leveraged to create dynamic dashboards that provided stakeholders with interactive and easily interpretable visual representations of the used car market.

TABLEAU

Tableau is a leading platform for data visualization and interactive analytics, renowned for its ability to transform complex data into visually compelling and easily interpretable dashboards. It connects effortlessly to a variety of data sources, enabling users to explore data in depth through interactive visualizations, filters, and dynamic dashboards. Tableau emphasizes visual storytelling, helping users uncover trends, patterns, and actionable insights without requiring extensive programming skills. Its efficiency in handling large datasets and its sophisticated visualization options make it particularly suitable for enterprise-level analysis. Tableau also allows for advanced features such as calculated fields, parameter-driven analysis, and scenario modeling, enhancing analytical depth. In the context of this study, Tableau was instrumental in creating interactive dashboards that highlighted key metrics such as car prices, fuel types, manufacturers, and trends across different regions, making insights more accessible and decision-friendly for end-users.

2. DOMAIN SPECIFICATION – USED CAR MARKET

The domain of this work is the used car market within the automobile industry, a rapidly growing segment driven by demand for affordable and reliable vehicles. Technology-driven platforms and online marketplaces have transformed buying and selling processes by enabling accurate vehicle evaluation, data-driven pricing, and seamless interactions. By analyzing historical and transactional data including vehicle specifications, pricing trends, brand performance, and buyer preferences stakeholders can uncover insights to optimize pricing, improve sales strategies, and enhance customer satisfaction, ultimately increasing transparency, efficiency, and competitiveness in the market.

Key Applications

1. Buyer Behavior Analysis

Understanding buyer behavior is crucial in the used car market. By analyzing factors such as preferred brands, fuel types, price ranges, mileage patterns, and geographic preferences, sellers can identify trends in consumer demand. These insights allow dealerships to predict future demand, tailor marketing strategies, and personalize offers for different buyer segments. For instance, data may reveal that younger buyers prefer compact, fuel-efficient vehicles, while families prioritize SUVs or multi-purpose vehicles. Leveraging such insights helps sellers enhance the overall buying experience, increase conversion rates, and retain customers for repeat sales.

2. Vehicle Valuation and Brand Performance

Analyzing historical sales data enables the evaluation of vehicle performance and brand value. By examining variables such as price, year of manufacture, engine size, mileage, and fuel type, stakeholders can determine which vehicles and brands are in high demand and which have lower resale value. This information is crucial for dealers and sellers to adopt optimal pricing strategies, manage inventory effectively, and make informed purchasing decisions. Furthermore, understanding brand performance helps businesses identify market opportunities and potential gaps in supply.

3. Pricing Trend Analysis

Historical sales and transactional data provide insights into pricing trends over time, across different models, and within various regions. Statistical analysis and visualization tools

can highlight patterns such as seasonal price fluctuations, depreciation rates, and market demand shifts. These insights allow sellers to set competitive yet profitable pricing strategies, ensuring that vehicles are priced accurately according to market conditions while maximizing revenue.

4. Market Segmentation Insights

Market segmentation in the used car industry involves analyzing the relationship between vehicle specifications, fuel type, and customer preferences. Segmentation allows sellers to categorize buyers into distinct groups based on demographics, buying behavior, and vehicle choices. This targeted approach enables effective marketing campaigns and better allocation of resources. For example, electric or hybrid vehicles may be marketed more aggressively to environmentally conscious buyers, while premium car models can target high-income customer segments.

5. Business Intelligence Dashboards

Interactive dashboards using tools like Power BI or Tableau can visually represent key performance indicators (KPIs) such as average resale value, brand popularity, sales distribution, and inventory status. These dashboards provide real-time insights to stakeholders, enabling quick, data-driven decisions. The use of dashboards enhances operational efficiency by presenting complex datasets in an intuitive and interactive format, helping managers monitor sales performance and adjust strategies proactively.

6. Inventory and Sales Optimization

By analyzing demand patterns, vehicle performance, and customer preferences, dealerships can optimize inventory management. Data insights can guide which vehicles to stock, in what quantity, and in which location, minimizing overstock and understock situations. Effective inventory optimization ensures that high-demand vehicles are readily available, improving sales potential, reducing holding costs, and maximizing profitability.

7. Customer Satisfaction and Trust Enhancement

Customer trust and satisfaction are key in the used car market. Analyzing vehicle quality, pricing, and feedback helps ensure fairness and transparency. This leads to higher satisfaction, repeat sales, and a stronger reputation.

3. DATASET DESCRIPTION

The car dataset contains seven attributes Manufacturer, Model, Engine Size, Fuel Type, Year of Manufacture, Mileage, and Price with over 5,000 records. It enables multi-dimensional analysis of numerical and categorical variables, making it ideal for statistical analysis in R and interactive dashboards in Power BI and Tableau to study price trends, correlations, and model popularity.

SAMPLE DATASET

	A	B	C	D	E	F	G
1	Manufacturer	Model	Engine size	Fuel type	Year of manufacture	Mileage	Price
2	Ford	Fiesta	1	Petrol	2002	127300	3074
3	Porsche	718 Cayman	4	Petrol	2016	57850	49704
4	Ford	Mondeo	1.6	Diesel	2014	39190	24072
5	Toyota	RAV4	1.8	Hybrid	1988	210814	1705
6	VW	Polo	1	Petrol	2006	127869	4101
7	Ford	Focus	1.4	Petrol	2018	33603	29204
8	Ford	Mondeo	1.8	Diesel	2010	86686	14350
9	Toyota	Prius	1.4	Hybrid	2015	30663	30297
10	VW	Polo	1.2	Petrol	2012	73470	9977
11	Ford	Focus	2	Diesel	1992	262514	1049
12	VW	Golf	2	Diesel	2014	83047	17173
13	BMW	Z4	2	Petrol	1990	293666	719
14	VW	Golf	1.2	Diesel	2007	92697	7792
15	Toyota	RAV4	2.2	Petrol	2007	79393	16026
16	Toyota	Yaris	1.4	Petrol	1998	97286	4046
17	VW	Golf	1.6	Diesel	1989	222390	933
18	Toyota	RAV4	2.4	Hybrid	2003	117425	11667
19	Toyota	Yaris	1.2	Petrol	1992	245990	720
20	Toyota	RAV4	2	Hybrid	2018	28381	52671
21	VW	Polo	1.2	Petrol	1998	155038	2118
22	VW	Golf	1.2	Hybrid	1987	121744	1890
23	Ford	Mondeo	1.6	Diesel	1996	77584	5667
24	Toyota	Prius	1	Hybrid	2003	115291	6512
25	Toyota	Prius	1	Hybrid	1990	238571	961

car dataset

Figure 3: Sample Dataset

These attributes are as follows:

Manufacturer – Name of the car manufacturer (e.g., Maruti, Hyundai, Toyota).

Model – Specific model name of the car (e.g., Swift, Creta, Innova).

Engine Size – Engine capacity of the car in liters (e.g., 1.2L, 2.0L)

Fuel Type – Type of fuel the car uses (e.g., Petrol, Diesel, CNG, Electric).

Year of Manufacture – The year in which the car was manufactured.

Mileage – Distance the car can travel per unit of fuel (e.g., km/l).

Price – Market price of the car (in INR).

4. ANALYSIS

4.1 Charts and Visualizations in R

The car dataset was analyzed using scan, correlation, histograms, boxplots, and chi-square tests.

1. Correlation Matrix

Correlation analysis of numeric variables Price, Mileage, Engine Size, and Year of Manufacture shows that Price is positively correlated with Engine Size and Year of Manufacture, and negatively correlated with Mileag**e**. This indicates that newer cars with larger engines cost more, while higher fuel efficiency tends to lower price. These insights highlight key factors affecting car pricing and help understand relationships between vehicle attributes in the used car market.

2. Histograms.

The histograms of numeric variables highlighted the distribution of cars across different price ranges, engine sizes, and mileages. These plots revealed clusters where most cars are concentrated, as well as outliers representing extremely high-priced or high-capacity vehicles. Boxplots further illustrated price variations across categorical features such as Manufacturer and Fuel Type, showing which brands or fuel types have more consistent pricing and which exhibit greater variability.

3. Boxplots

Boxplots were used to visualize price variations across categorical features such as Manufacturer and Fuel Type. These plots highlighted which brands or fuel types maintain more consistent pricing and which exhibit greater variability, helping to identify patterns, outliers, and differences in market positioning among various car brands and fuel categories.

4. Chi-Square Tests

Chi-square tests were performed between categorical variables, including Manufacturer vs. Fuel Type and Model vs. Fuel Type, revealing statistically significant associations. The results indicate that certain manufacturers or specific models predominantly use particular fuel types, providing valuable insights into market segmentation

4.2 Charts and Visualizations in PowerBI

1. Column Chart

The Bar chart illustrates the average price of cars by manufacturer, offering a straightforward comparison of pricing patterns across different brands. This visualization makes it clear which manufacturers are positioned in the premium segment, such as Porsche and BMW, with significantly higher average prices, and which brands cater to the budget or mid-range segment, such as Toyota and VW. By visualizing brand pricing in this way, the chart provides insights into market segmentation, affordability, and brand value perception.

2. Pie Chart

The pie chart depicts the distribution of car models by fuel type, showing how Petrol, Diesel, and Hybrid options contribute to the dataset. It provides a quick, high-level overview of customer and market preferences, revealing whether traditional fuel options dominate or whether eco-friendly technologies like Hybrids are gaining ground. This type of visualization is useful for understanding energy trends, environmental impact considerations, and shifts in consumer demand.

3. Scatter Plot

The scatter plot compares average mileage, average price, and engine size across fuel types. Each bubble represents a combination, enabling viewers to spot trade-offs between efficiency, engine performance, and cost. For instance, Petrol cars may show higher mileage but varying prices, while Diesel and Hybrid vehicles cluster differently, reflecting their unique positioning in the market. This chart is particularly valuable for analyzing the relationship between performance metrics and affordability, allowing stakeholders to balance fuel efficiency against ownership costs**.**

4. Line Chart

The line chart visualizes the trend of car prices across different engine sizes, showing a consistent upward trajectory. This indicates that as engine capacity increases, car prices also rise, reflecting higher production costs, better performance, and advanced features in larger vehicles. The chart provides a clear narrative about how engine capacity influences pricing strategies, helping to identify whether certain brands follow or deviate from the general trend.

5. Treemap

The treemap represents the count of models by manufacturer, where the size of each block corresponds to the number of models a brand offers. Larger blocks, such as those for Ford, Toyota, and VW, indicate strong market presence and diverse product offerings, while smaller blocks, such as Porsche, highlight niche or luxury segments. This visualization allows for an immediate understanding of market dominance, product variety, and competitive positioning among manufacturers.

4.3 Charts in Tableau

1. Clustered Bar Chart (Fuel Type Details)

A clustered bar chart was used to display the distribution of Fuel Types across different manufacturers. For example, selecting a fuel type like Hybrid shows details such as Manufacturer (e.g., Toyota), Price, and count of cars in the dataset. This chart allows comparison of multiple categories side by side, making it easy to identify which manufacturers offer which fuel types and understand the relative volume of cars for each category.

2. Bar Charts (Manufacturer-Specific Analysis)

Separate bar charts were created for individual manufacturers such as BMW, Ford, Porsche, and Volkswagen, showing Fuel Type, Average Price, and Total Listings for each brand. This helps compare manufacturers, understand fuel type distribution, and analyze pricing patterns.

3. Line / Curve Chart (Car Price by Brand)

A line chart was used to visualize Price trends across different brands, highlighting how average car prices vary between manufacturers and identifying brands with higher or lower market values.

4. Heatmap (Car Price by Engine Size)

A heatmap was created to analyze Price in relation to Engine Size, where color intensity represents the concentration of cars in specific price ranges and engine capacities, revealing patterns in the used car market.

5. IMPLEMENATION AND EVALUATION

POWERBI

5.1 Used Car Market Analysis Dashboard (Power BI)

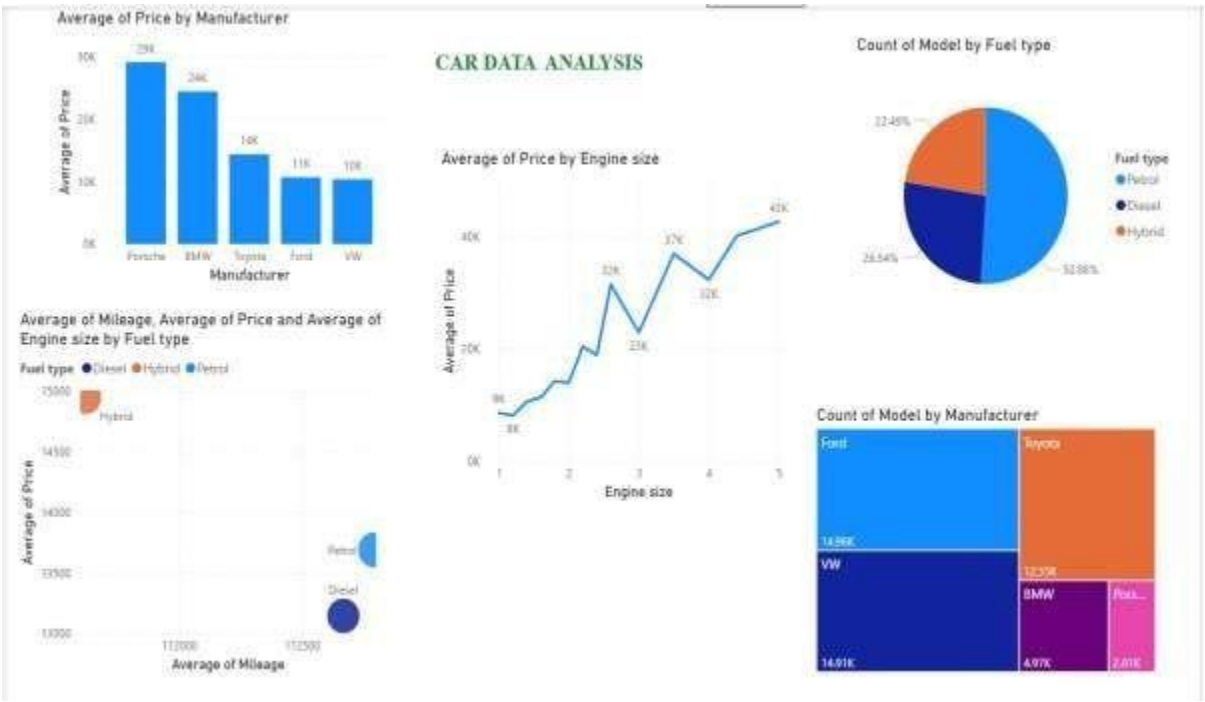

Figure 5.1 Used Car Analysis Dashboard

Figure 5.1 provides an overview of the car dataset through key visualizations. The column chart highlights price differences across manufacturers, while the pie chart shows fuel type distribution with Petrol dominating. The scatter plot compares mileage, price, and engine size across fuel types. The line chart illustrates how prices rise with larger engine sizes, and the treemap reveals the model distribution, with Ford, Toyota, and VW contributing the most.

1. Average Price of a Car by Manufacturer (Bar Chart)

A bar chart was used to visualize the average car price across different manufacturers. This provided an easy comparison of premium brands with higher average prices versus budget-friendly brands, highlighting pricing trends in the market.

2. Count of Models (Pie Chart)

A pie chart illustrated the distribution of car models among manufacturers. The size of each slice represented the proportion of models contributed by each brand, helping to identify market leaders with diverse offerings compared to smaller contributors.

3. Average Mileage (Cards Plot)

Cards were used to display the average mileage of cars in the dataset. This simple KPI visualization provided a quick snapshot of efficiency levels across vehicles without complex comparisons.

4. Average Price Trend (Line Chart)

A line chart was created to show average car price variations across different attributes. This helped in identifying pricing patterns over categories such as fuel type or manufacturer, revealing upward or downward trends.

5. Count of Models by Manufacturer (Heatmap)

A heatmap was used to analyze the number of car models offered by each manufacturer. Color intensity highlighted brands with the highest variety of models, making it easy to distinguish dominant players from smaller ones at a glance.

TABLEAU

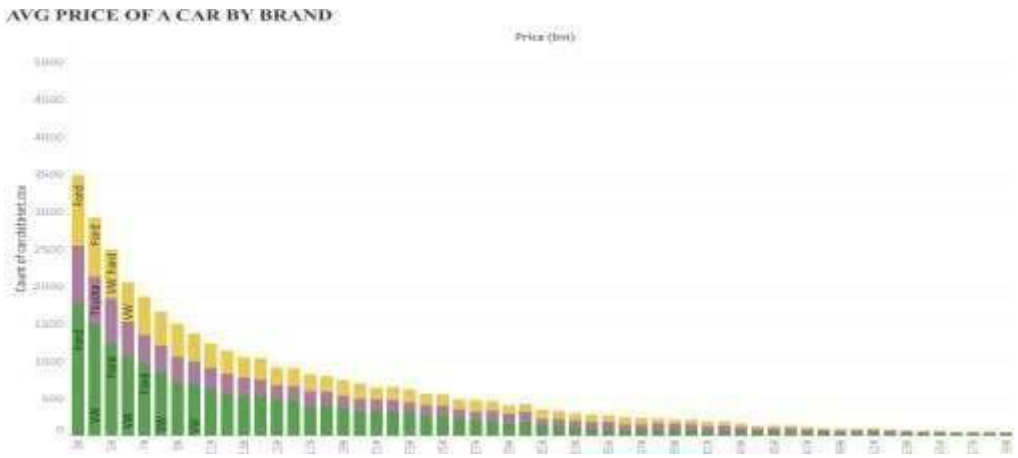

Figure 5.2 Average Price of a car by Brand

Figure 5.2 shows the average car price across different brands. Premium manufacturers such as Porsche and BMW record significantly higher average prices, while brands like Toyota, Ford, and VW represent more affordable options. This visualization highlights the price segmentation among manufacturers, making it easy to distinguish luxury brands from mid-range and budget ones.

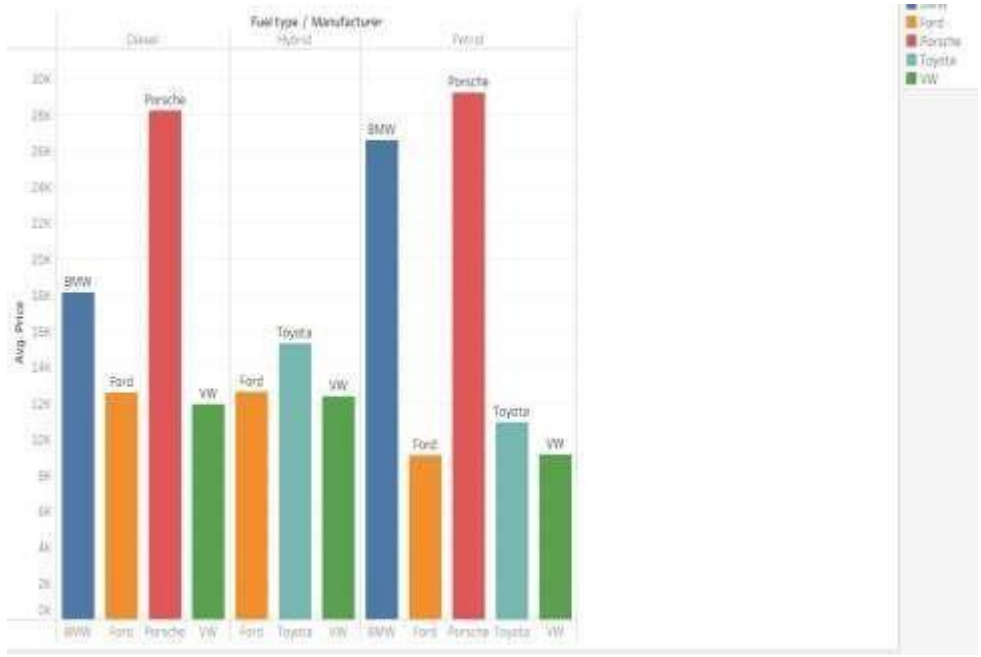

Figure 5.3 Average Car Price by Brand and Fuel Type

Figure 5.3 This visualization Chart compares the average price of cars across different brands, with each brand further divided by fuel type such as petrol, diesel, and hybrid. It helps identify how fuel type influences the pricing of models within the same brand, and also highlights price differences between brands. This allows for a better understanding of brand positioning and fuel-type cost variations in the automobile market.

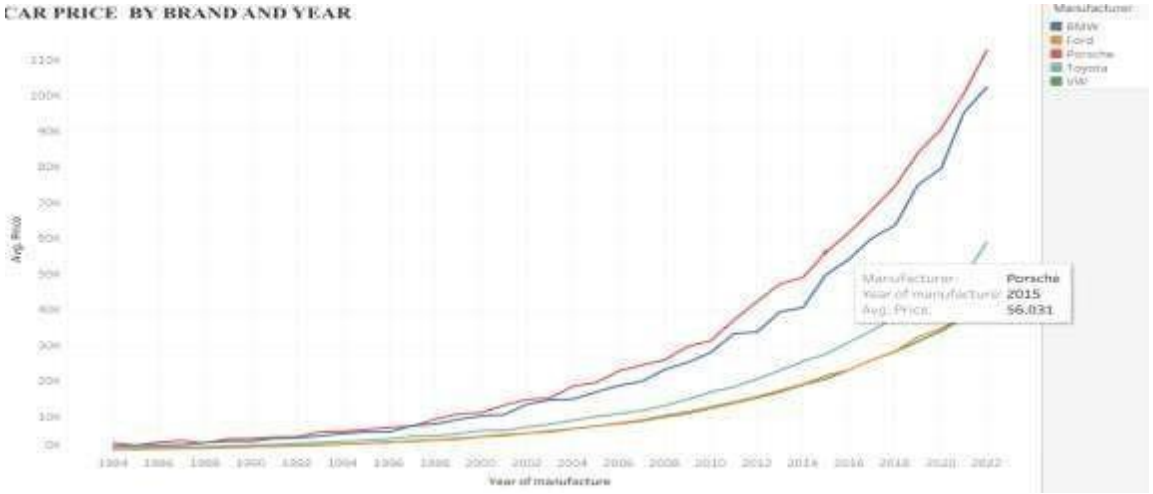

Figure 5.4 Car Price by Brand and Year

Figure 5.4 This chart shows how car prices vary across brands over different years, highlighting trends and shifts in brand pricing patterns. It provides insights into long-term brand positioning and market price evolution.

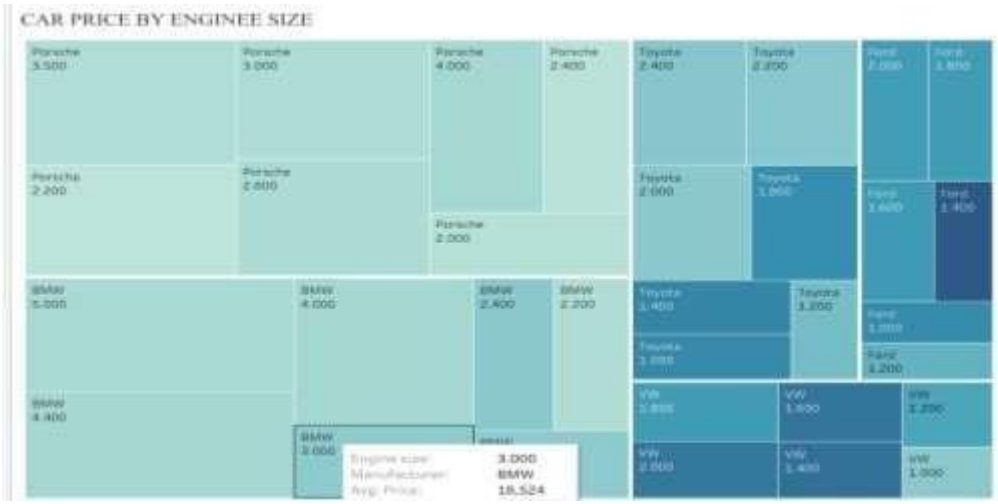

Figure 5.5 Car Price by Engine Size

Figure 5.5 This chart shows the variation in car prices with respect to engine size. Larger engine capacities generally correspond to higher prices, reflecting performance and technology costs. The visualization highlights clear price gaps between small, mid, and large engines. It provides insights into consumer affordability and market segmentation based on engine capacity.

R Studio

```
> corrplot(corr_matrix, method = "color", type = "upper",
+          tl.col = "black", tl.srt = 45,
+          addCoef.col = "black", number.cex = 0.7)
> summary(df)
 Manufacturer          Model              Engine.size     Fuel.type
 Length:50000       Length:50000        Min.   :1.000   Length:50000
 Class :character   Class :character    1st Qu.:1.400   Class :character
 Mode  :character   Mode  :character    Median :1.600   Mode  :character
                                        Mean   :1.773
                                        3rd Qu.:2.000
                                        Max.   :5.000
 Year.of.manufacture     Mileage            Price
 Min.   :1984        Min.   :   630   Min.   :    76
 1st Qu.:1996        1st Qu.: 54352   1st Qu.:  3061
 Median :2004        Median :100988   Median :  7972
 Mean   :2004        Mean   :112497   Mean   : 13829
 3rd Qu.:2012        3rd Qu.:158601   3rd Qu.: 19027
 Max.   :2022        Max.   :453537   Max.   :168081
```

Figure 5.6 Summary Statistics of Car Dataset

Figure 5.6 represents the summaryand descriptive statistics of the car dataset with 50,000 records**,** covering manufacturer, model, year, mileage, engine size, fuel type, and price. The year of manufacture ranges from 1984 to 2022 (median 2004), showing most cars are modern. Mileage spans 630–453,537 km (median 100,988), indicating varied usage. Engine sizes fall between 1.0–5.0 liters (mean 1.77), reflecting mostly small to mid-sized engines. Prices range from ₹76 to ₹168,081**,** with a median of ₹7,972 and mean of ₹13,829**,** highlighting both budget and premium cars.

```
        Pearson's Chi-squared test

data:  table(df$Manufacturer, df$Fuel.type)
X-squared = 31399, df = 8, p-value < 2.2e-16

> # Model vs Fuel Type
> chisq.test(table(df$Model, df$`Fuel.type`))

        Pearson's Chi-squared test

data:  table(df$Model, df$Fuel.type)
X-squared = 54712, df = 28, p-value < 2.2e-16
```

Figure 5.7 The Chi-square test results for the car dataset

Figure 5.7 shows the Chi-square test results for the car dataset. A significant association was found between Manufacturer and Fuel Type (χ^2 = 31,399, df = 8, p < 2.2e-16) and between Model and Fuel Type (χ^2 = 54,712, df = 28, p < 2.2e-16). This indicates that both manufacturer and model strongly influence fuel type selection.

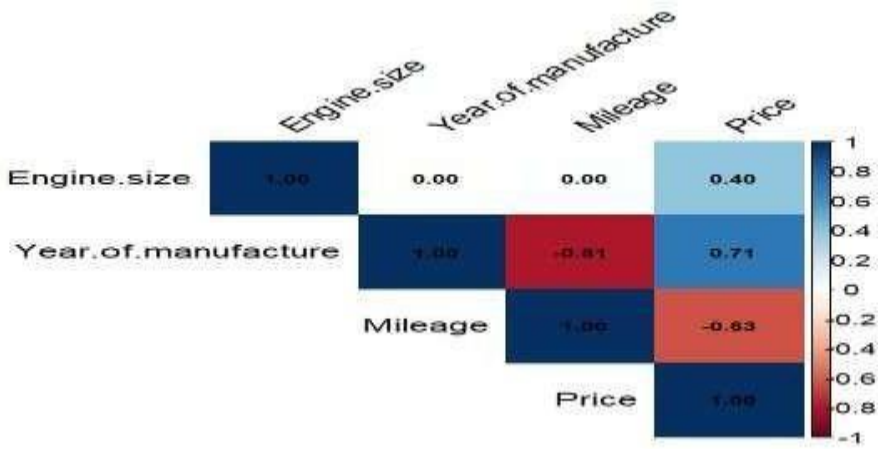

Figure 5.8 : Correlation Matrix of Numerical Variables

Figure 5.8 displays the correlation matrix of numerical variables in the car dataset. A strong negative correlation is observed between Year of Manufacture and Mileage (-0.81) and

between Mileage and Price (-0.63), while a strong positive correlation exists between Year of Manufacture and Price (0.71). Engine size shows a moderate positive correlation with price (0.40).

Figure 5.9 Histogram of Car Prices

Figure 5.9 presents the histogram of car prices. Most vehicles are concentrated in the lower price range, with a steep decline in frequency as price increases. This right-skewed distribution highlights that affordable cars dominate the dataset, while high-priced cars are relatively rare.

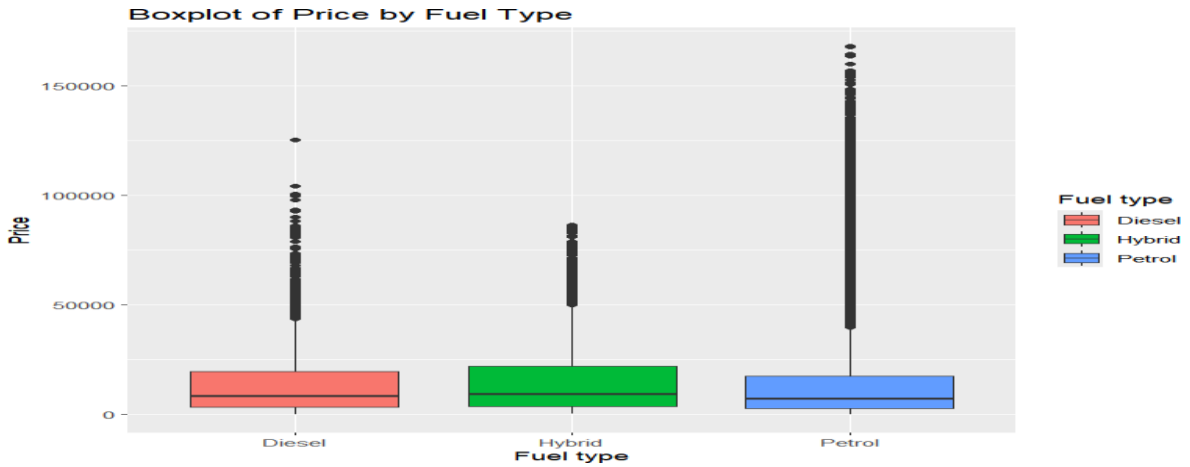

Figure 5.10 Boxplot of Price by Fuel Type

Figure 5.10 The boxplot shows price variations across fuel types, with petrol and hybrid cars exhibiting wider spreads and higher outliers. Diesel cars have lower median prices and a narrower range. This indicates fuel type significantly influences car price distribution.

6. CONCLUSION

The Car dataset analysis of the used car dataset provided meaningful insights into various aspects of the automobile market, including brand performance, fuel type preferences, pricing patterns, mileage distribution, and year-wise market trends. Using R, statistical analysis techniques such as descriptive statistics, correlation analysis, regression modeling, and hypothesis testing were applied to uncover relationships between factors like engine size, mileage, fuel type, and car prices. These results helped in understanding the factors that influence customer purchasing decisions and brand positioning. Power BI contributed to building interactive dashboards with KPI cards, bar charts, pie charts, line charts, and heatmaps, which offered a high-level overview of the dataset and supported quick comparisons across brands and fuel types. On the other hand, Tableau was used to design more advanced visualizations such as dual-axis charts, trend analysis, and treemaps, which provided deeper insights into pricing variations, brand popularity, and year-wise performance. Overall, this project demonstrated how the integration of statistical analysis and visualization tools can transform raw automobile data into actionable insights, enabling data-driven decision-making for manufacturers, dealers, and customers.

Future Enhancements

This car data Work successfully explored key dimensions of the used car market, several future enhancements could add further value. The dataset can be expanded to include additional attributes such as customer demographics, loan/finance details, and insurance costs to gain a more holistic view of buyer behavior. Incorporating machine learning and predictive modeling could improve price prediction, resale value estimation, and demand forecasting for specific brands and fuel types. Real-time data integration with Power BI or Tableau would allow live monitoring of market trends, enabling car dealers and customers to make informed decisions instantly. Furthermore, applying natural language processing (NLP) on customer reviews and online listings could complement numerical data with qualitative insights, giving a well-rounded understanding of buyer sentiment. Finally, optimization models for pricing strategies, promotional offers, and inventory management would enhance efficiency and strengthen the competitive edge in the rapidly growing used car market**.**