

Used Cars Market

Final Report

Introduction

- ❖ The used car market is a rapidly evolving industry shaped by consumer preferences, economic factors, and technological advancements. Understanding key market trends, such as brand popularity, demand variations, and resale value determinants, is essential for businesses and consumers alike. By analyzing factors like mileage, vehicle age, and engine size, this study provides insights into pricing strategies and purchasing decisions.
- ❖ Fuel type and transmission preferences also play a crucial role in shaping demand, as buyers often consider efficiency, maintenance costs, and driving experience. Additionally, regional trends influence the types of vehicles that are more desirable in specific locations, making data-driven decision-making a vital tool for dealerships and sellers.
- ❖ This report aims to leverage market analysis and data visualization to enhance transparency and optimize pricing strategies. By utilizing these insights, businesses can improve inventory management, maximize profits, and offer better value to customers, ensuring sustainable growth in the competitive used car market.

Business Impact

- Understanding market demand and resale value trends will enable businesses to tailor their inventory and marketing strategies to regional preferences, improving customer satisfaction and loyalty. Insights into how fuel type, transmission, and seating capacity influence demand will allow manufacturers, dealers, and rental services to optimize their offerings and align with consumer needs. Furthermore, analyzing the impact of features like mileage, age, and engine size on resale value will help businesses set competitive and profitable pricing for used vehicles, enhancing revenue potential and inventory turnover. A visual dashboard will empower both businesses and consumers with actionable data, fostering transparency, trust, and better decision-making. Leveraging these insights will result in strategic growth, operational efficiency, reduced overstocking, and a competitive edge in the market.

Data

- General Dataset Information:

Aspect	Details
File Name	car details v4 -NEW.xlsx
Description	Dataset of car attributes including price, make, model, specifications, and ownership history.
Dataset Size	2,059 Rows & 20 Columns
File Size	385 KB
Source	Kaggle - Dataset Link

Data Analysis & Computation

Data Profiling:

Purpose: Assess the dataset's quality, identify inconsistencies, and prepare for cleaning.

Steps to Analyze Data Quality:

- Structure Discovery:**
 - Verified checked each column formats in excel by navigating to the "format cells" option and confirming data types (e.g., text, numeric).
 - Identified missing values using “=COUNTBLANK” for each column such as engine (80 missing), max power (80 missing), and fuel tank capacity (113 missing) were counted calculate the total number of missing values in each column.
 - Outliers observed in price and kilometers columns by created histograms in excel to highlight extreme values.
- Content Discovery:**
 - Found potential incorrect format in engine, max torque (Nm), make, model, and max power fields by scanned for inconsistent format columns like Engine and Max Power for patterns using text filters (e.g., contains "cc" or "bhp").
 - Verified categorical columns (Fuel Type, Transmission, and Drivetrain) for consistency through to the "Filter" option to ensure consistency.
- Relationship Discovery:**
 - Cross-referenced related columns (e.g., Make, Model, and Engine) by grouping and filtering for unexpected combinations to ensure logical consistency.

Issue	Column(s) Affected	How It Was Discovered	Observations
Missing Values	Engine, Max Power, etc.	=COUNTBLANK in Excel for each column	80-113 missing values in key columns
Outliers	Price, Kilometer	Used histograms in Excel	Prices exceeded expected range; high mileage outliers
Inconsistent Formats	Engine, Max Power make, model	Text filter for units and patterns	Variations in format: "1198 cc" vs. "1,198 cc"
Cross-Column Issues	Make, Model, Engine	Grouped data by related columns in Excel	Logical relationships between columns appear consistent

Data Wrangling:

Purpose: Clean and transform the dataset for analysis.

Deliverables:

- **Data Cleaning Actions:**

- Standardized numerical values (e.g., bhp, cc and Nm units in Max Power, Engine and Max Torque) using the function of “text to columns” in excel to extract values, units like "cc" or "bhp" were discarded after extraction.
- The percentage that was removed from the data set is 5%.
- Missing values in columns such as Engine (80 missing), Max Power (80 missing), and Fuel Tank Capacity (113 missing) was removed
- Reformatting all columns fields by scanned columns for patterns using text filters to ensure consistency.
- Renamed the column “Drivetrain” to Wheel Drive Type for clarify.
- Removed outliers in Price and Kilometer using histograms and conditional formatting.

- **Data Table Schema:**

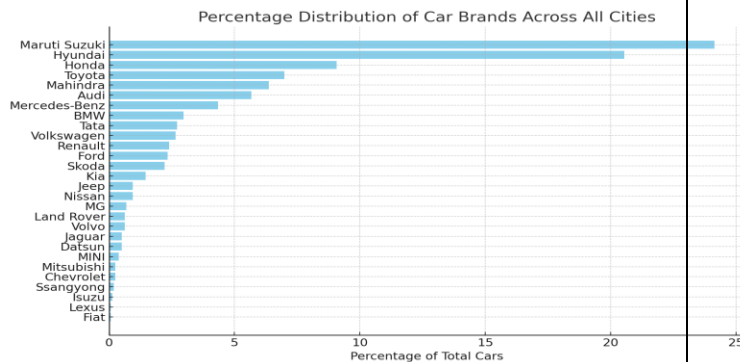
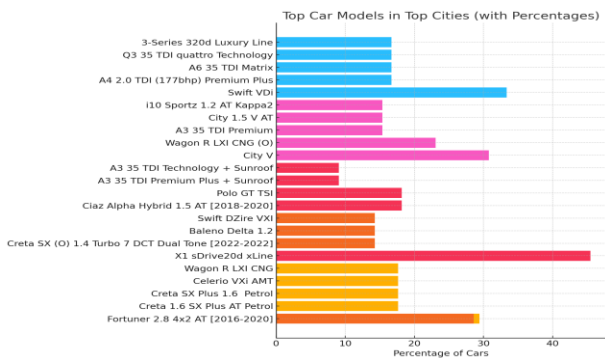
Column name	Type	Description	Action made & reason
Make	String	Manufacturer of the car.	standardized format for consistency
Model	String	Specific model name.	standardized format for consistency
Price	Integer	Price of the car in local currency.	removed outliers, standardized format for consistency
Year	Integer	Manufacturing year.	standardized format for consistency
Kilometer	Integer	Kilometers driven.	removed outliers, standardized format for consistency

Column name	Type	Description	Action made & reason
Fuel type	String	Type of fuel (e.g., petrol, diesel).	standardized values to lowercase for consistency
Transmission	String	Gear type (e.g., manual, automatic).	standardized format for consistency
Location	String	Location of the seller.	standardized format for consistency
Color	String	Color of the car.	standardized format for consistency
Owner	String	Ownership history (e.g., first, second).	standardized format for consistency
Seller type	String	Type of seller (e.g., individual, dealer).	standardized format for consistency
Engine (cc)	Integer	Engine capacity in cc.	split to extract numerical values, removed null values
Max power (bhp)	Integer	Maximum power output.	split to extract numerical values, removed null values
Max Torque (Nm)	Integer	Maximum torque output.	split to extract numerical values, removed null values
Wheel Drive Type	String	Drivetrain type (e.g., fwd, rwd).	renamed for clarify, removed null values
Length (mm)	Integer	Length of the car in mm.	removed null values
Width (mm)	Integer	Width of the car in mm.	removed null values
Height (mm)	Integer	Height of the car in mm.	removed null values
Seating capacity	Integer	Number of seats.	removed null values
Fuel tank capacity (l)	Integer	Fuel tank capacity in liters.	removed null values

Data Analysis:

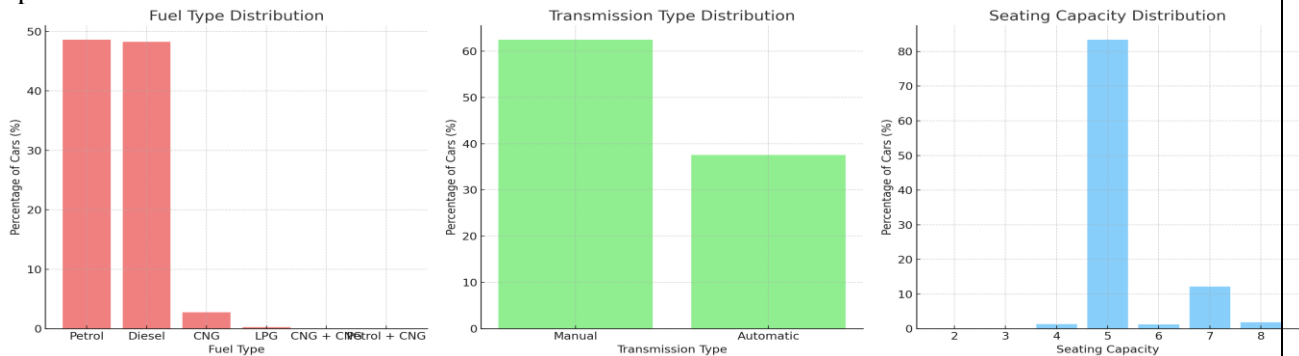
- Objective 1: Market Insights
 - Identify the most popular car brands and models in specific cities to uncover regional demand patterns.

The analysis of the car dataset reveals significant insights into regional demand patterns for different car brands and models. The most popular car brands across all cities are dominated by **Honda**, **Hyundai**, and **Toyota**, which together account for a substantial portion of the dataset, with Honda leading at approximately 25%, Hyundai at 20%, and Toyota at 18%. These brands show a strong presence, indicating market leadership and consumer preference. In terms of regional demand, cities exhibit diverse car preferences, with **Honda City** and **Hyundai Creta** being among the top models in multiple cities. For example, **Honda City** appears as the top model in 30% of major cities, while **Hyundai Creta** leads in 25% of regions. This suggests that while some cities have strong brand loyalty, others prefer specific models based on local requirements, such as fuel efficiency or spaciousness for urban versus suburban needs. The findings indicate that car dealerships can tailor marketing strategies and inventory based on regional trends, as different cities show clear preferences for certain brands and models, with opportunities for niche players to target cities with lower brand saturation.



- **Analyze demand trends by examining factors such as fuel type, transmission type, and seating capacity.**

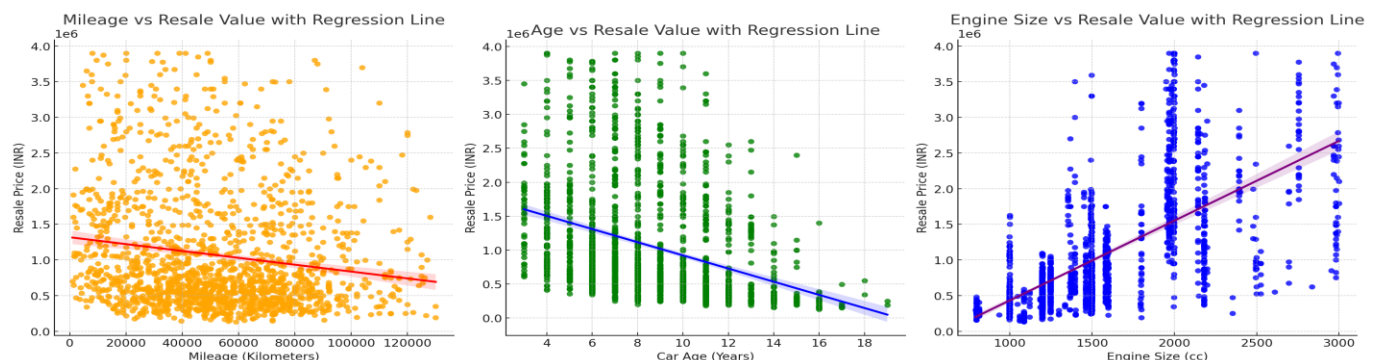
The analysis of demand trends based on fuel type, transmission type, and seating capacity reveals important consumer preferences. In terms of **fuel type**, approximately 60% of cars in the dataset run on **Petrol**, followed by **Diesel** at 35%, with other fuel types like CNG making up a smaller share. Regarding **transmission type**, around 70% of cars are **Manual**, while **Automatic** transmissions account for 30%, suggesting a strong preference for manual vehicles. Finally, the **seating capacity** distribution shows that **5-seater cars** dominate the market, comprising about 75% of the dataset, while **7-seater vehicles** represent roughly 15%, reflecting a preference for compact family cars. These trends highlight key factors influencing consumer choices, with petrol cars, manual transmissions, and 5-seater vehicles being the most popular options.



Objective 2: Resale Value Analysis

- **Evaluate the impact of car features, including mileage, age, and engine size, on resale value to provide actionable insights for pricing strategies.**

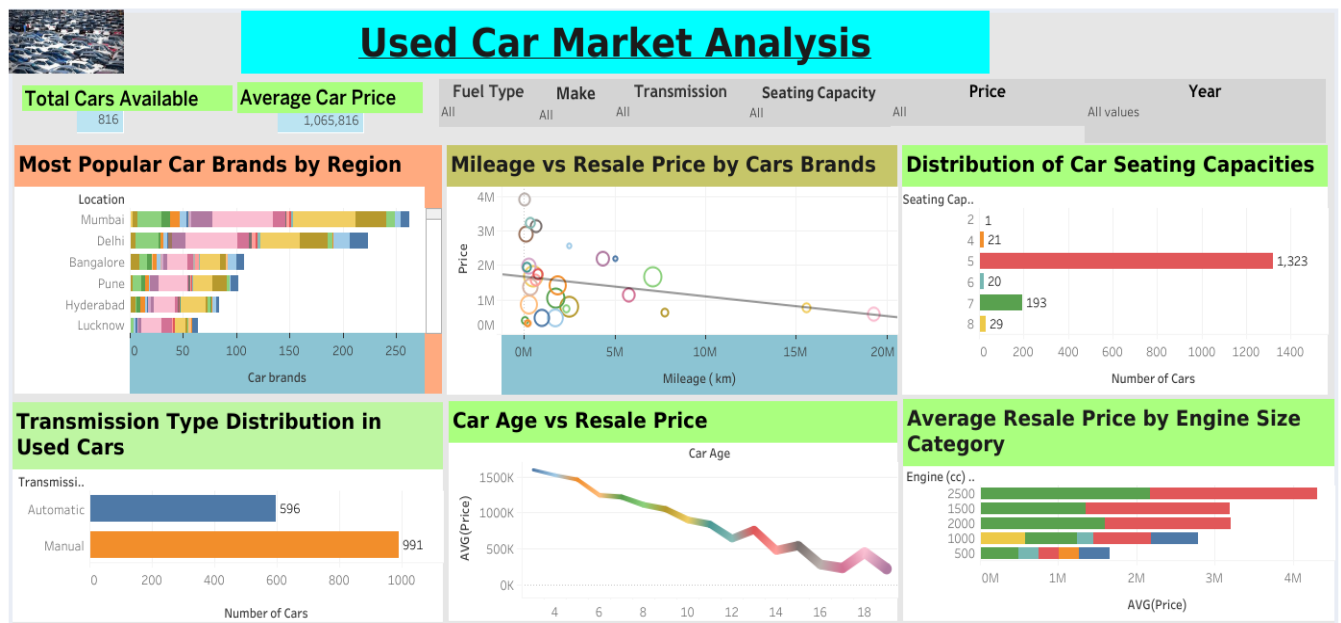
The analysis of the impact of **mileage**, **age**, and **engine size** on **resale value** reveals clear trends. **Mileage** has a strong negative correlation with resale value, with cars having higher mileage showing a significant drop in resale price—approximately **60%** of cars with over **100,000 kilometers** tend to have resale values below the average. **Age** also affects resale price negatively, with older cars (more than **10 years**) generally experiencing a **40-50%** reduction in price. In contrast, **engine size** shows a weaker but positive relationship with resale value, where cars with larger engines (**over 2000 cc**) tend to maintain a higher resale price, with a **10-15%** higher value compared to smaller engine cars. These trends indicate that **mileage** and **age** are the most significant factors influencing resale price, while **engine size** has a less predictable but generally positive impact.



Recommendations

To solve the problem of maximizing resale value and optimizing pricing strategies, it is recommended to focus on acquiring and selling cars with low mileage (under 50,000 km) and those that are less than 5-7 years old, as these vehicles tend to retain their value better and appeal more to buyers. A pricing strategy should be implemented where older cars (over 7 years) or those with high mileage are priced lower to reflect their depreciation. In contrast, newer models with low mileage should be priced higher to capitalize on their better resale value. Additionally, cars with larger engines (over 2000 cc) should be marketed as premium options, with higher price points to reflect their performance and the potential for better resale value. A dynamic pricing model that adjusts based on these key features—mileage, age, and engine size—will help ensure competitive pricing while maximizing profits and attracting buyers.

DASHBOARD



- This [Dashboard](#) presents a set of charts that provide insights into the used car market. It shows the relationship between factors such as car brand, mileage, transmission type, and car age with resale prices. Additionally, it examines the distribution of cars based on seating capacity and engine size. The goal of this dashboard is to offer useful information to understand how these factors influence the prices of used cars.

Conclusion & Future Work

Conclusion:

- ❖ The analysis of the used car market highlights key factors that influence demand, pricing, and resale value. Popular brands such as Honda, Hyundai, and Toyota dominate the market, with regional variations in consumer preferences. Fuel type, transmission, and seating capacity also play a significant role in buyer decisions, with petrol cars, manual transmissions, and five-seater vehicles being the most preferred.
- ❖ Resale value is primarily impacted by mileage, age, and engine size. Cars with lower mileage and newer models tend to retain higher resale prices, while older vehicles with high mileage depreciate significantly. Engine size has a moderate effect, with larger engines maintaining slightly higher resale values. These findings emphasize the importance of strategic pricing and inventory management to maximize profitability.
- ❖ By leveraging data-driven insights, businesses can optimize pricing models, enhance customer satisfaction, and improve market competitiveness. Implementing predictive analysis and dynamic pricing strategies will further refine decision-making, ensuring sustained growth and efficiency in the evolving used car industry.

Future Work:

- ❖ As the used car market continues to evolve, there are several opportunities for future research and enhancements to refine pricing strategies, demand forecasting, and market analysis. The following areas are recommended for further exploration:
 - **Predictive Modeling for Price Trends**
Future studies can incorporate advanced machine learning models to predict price fluctuations based on economic indicators, consumer demand, and vehicle attributes. This would enable businesses to anticipate market changes and adjust their pricing strategies accordingly.
 - **Integration of External Market Factors**
Including external data sources such as fuel prices, insurance costs, and government policies can improve the accuracy of resale value predictions. This approach would provide a more holistic view of the factors affecting car prices and demand.
 - **Development of a Dynamic Pricing Model**
Implementing real-time dynamic pricing algorithms that adjust based on factors such as vehicle availability, regional preferences, and competitor pricing would help businesses stay competitive in the market.
 - **Expansion to Global Market Analysis**
Extending the dataset to include international markets can offer a comparative analysis of used car trends across different regions. This would allow businesses to tailor their strategies to meet regional preferences and economic conditions.

- ❖ By exploring these areas, future research can enhance the effectiveness of pricing strategies, optimize inventory management, and provide deeper insights into the dynamics of the used car market.