

Project Description Document

Name: Al-Qasem AbuKashef

❖ PROBLEM STATEMENT:

The used car market has numerous obstacles in establishing proper pricing and recognizing buyer preferences. Pricing an automobile is a complex procedure because aspects such as age, usage, make, model, and condition determine its worth. Furthermore, buyers frequently lack sufficient knowledge about the car's history and upkeep, resulting in low trust between the parties. The market also necessitates ongoing research of trends such as demand for new fuels and technologies, as well as enhanced inventory management, in order to meet consumer demands without increasing unsold vehicles. This problem provides an excellent opportunity to create accurate pricing models, examine market trends, and improve the purchasing and selling experience.

❖ POSSIBLE IMPACT OF YOUR ANALYSIS:

The goal of the analysis is to have a significant impact on the used car market by improving efficiency and transparency. The analysis can help provide accurate models for pricing cars based on factors such as age, model, and fuel type, ensuring fair prices for both the seller and the buyer. It also increases trust by providing detailed information about the condition and history of the cars, helping buyers make informed decisions. Additionally, understanding market trends, such as the rise in demand for electric or economical cars, helps companies manage their inventory more effectively and reduce the costs associated with unsold inventory. This analysis not only improves the buying and selling processes but also helps companies to enhance their competitiveness, supporting market growth and sustainability.

❖ DATASET:

Source: car details v4, [here](#)

Description of Variables:

The dataset contains detailed information about vehicles in the used car market. Below are the key variables:

1. **Make:** The brand/manufacturer of the car (e.g., Honda, Toyota).
2. **Model:** The specific model and variant of the car.
3. **Price:** Selling price of the car.
4. **Year:** Year of manufacture of the car.
5. **Kilometer:** Distance driven by the car (in kilometers).
6. **Fuel Type:** The type of fuel the car uses (e.g., Petrol, Diesel).
7. **Transmission:** Transmission type (e.g., Manual, Automatic).
8. **Location:** The location where the car is listed.
9. **Color:** Exterior color of the car.
10. **Owner:** Ownership details (e.g., First owner, Second owner).
11. **Seller Type:** Whether the seller is an individual or a corporate entity.
12. **Engine:** Engine capacity of the car (in cc).
13. **Max Power:** Maximum power output of the engine.
14. **Max Torque:** Maximum torque output .
15. **Drivetrain:** Drive type (e.g., FWD, RWD).
16. **Length:** Car length (in mm).

This dataset offers comprehensive attributes that can be used to analyze and predict car prices, explore market trends, and assess vehicle preferences across locations.

Project Scoping Document

❖ Business Problem:

The used car market faces significant challenges that impact buyers, sellers, and businesses, making this issue critical to address. Pricing inefficiencies, driven by the absence of accurate, data-driven models, often lead to overpricing or underpricing of vehicles, causing sellers to lose potential buyers. As consumer preferences shift toward electric and fuel-efficient vehicles, businesses must adapt to avoid unsold inventory and missed opportunities. Poor inventory management further exacerbates costs and limits sales potential. Ultimately, these challenges reduce customer satisfaction and hinder market efficiency.

❖ Business Impact:

The analysis will provide significant benefits by improving efficiency, transparency, and profitability in the market. Accurate, data-driven pricing models will ensure vehicles are fairly priced, maximizing revenue for sellers while offering competitive deals to buyers. Enhanced transparency through detailed insights into vehicle specifications and history will build trust and confidence, encouraging more informed purchase decisions. Businesses will benefit from improved inventory management by aligning stock with market demand, reducing excess inventory and associated costs. Additionally, predicting fast-selling vehicles will streamline sales processes and shorten the time vehicles spend on the market. The analysis will also help businesses adapt to emerging trends, such as the increasing demand for electric and fuel-efficient vehicles, ensuring long-term competitiveness. Ultimately, these improvements will lead to a better customer experience, higher satisfaction, and sustainable growth in the used car industry.

❖ Dataset(s):

Dataset Description

The dataset being used is the ["Vehicle Dataset"](#), specifically the "Car Details v4.csv" file. This dataset includes 20 variables, providing comprehensive information about used cars, including their specifications, performance, and resale value. Here's an overview of the key variables:

1. **Make:** Manufacturer of the car (categorical).
2. **Model:** Specific model of the car (categorical).
3. **Price:** Resale price of the car in the given currency (numerical).
4. **Year:** Year of manufacture (numerical).
5. **Kilometer:** Distance traveled by the car in kilometers (numerical).
6. **Fuel Type:** Type of fuel used (categorical: Petrol/Diesel/CNG).
7. **Transmission:** Transmission type (categorical: Manual/Automatic).
8. **Location:** City or region where the car is listed (categorical).
9. **Color:** Exterior color of the car (categorical).
10. **Owner:** Ownership history (categorical: First, Second, etc.).
11. **Seller Type:** Type of seller (categorical: Dealer/Individual).
12. **Engine:** Engine displacement in cc (numerical).
13. **Max Power:** Maximum power output of the car (numerical).
14. **Max Torque:** Maximum torque output of the car (numerical).
15. **Drivetrain:** Drivetrain type (categorical: FWD/RWD/AWD).
16. **Length:** Length of the car in mm (numerical).
17. **Width:** Width of the car in mm (numerical).
18. **Height:** Height of the car in mm (numerical).
19. **Seating Capacity:** Number of seats in the car (numerical).
20. **Fuel Tank Capacity:** Fuel tank capacity in liters (numerical).

Strengths:

- **Rich Variety of Features:** Includes both technical specifications and market-related variables, allowing for detailed analysis.

- **Comprehensive Data:** Covers essential features like engine size, fuel type, and drivetrain, relevant to both buyers and sellers.

Weaknesses:

- **Data Imbalance:** An uneven distribution of data, such as fuel types or seller types, could potentially skew the model.
- **Potential Outliers:** Variables like price and kilometers driven might have extreme values affecting the results.
- **Data Cleaning Needs:** Missing values and inconsistent formats in columns like "Max Power" and "Max Torque" will require preprocessing.

❖ Dashboard:

Key Metrics

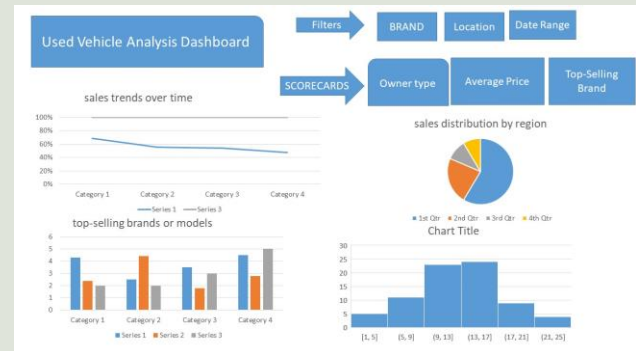
- The dashboard will display key summary statistics to provide a high-level overview of the dataset:
 - **Total Number of Cars:**
Metric: Count of all car entries in the dataset.
 - **Average Resale Price:**
Metric: Mean resale price (price column).
 - **Median Car Age:**
Metric: Median age of cars.

Filters

- **Transmission Type:**
Dropdown or toggle button to filter between Manual and Automatic transmission types.
- **Seller Type:**
Dropdown to filter data by seller type (Dealer or Individual).
- **Car Age Range:**
Slider to filter cars based on their age.
- **Price Range:**
Slider to set a minimum and maximum price range for resale prices.
- **Kilometers Driven:**
Slider to filter cars based on mileage (e.g., from low to high kilometers driven).
- **Visualizations:**
 - Bar charts for distributions (e.g., fuel type, transmission type, seller type).
 - Line charts for trends in price based on car age or kilometers driven.

Tools:

- Excel for basic data analysis, cleaning, and visualization.
- Sql for querying and managing large datasets, Perform data aggregation, filtering, and transformations before importing it into other tools.
- Tableau for dynamic and user-friendly visualizations



❖ Methods

What are your variables?

- **Price:** The resale price of the car.
- **Age:** The age of the car.
- **Kilometer:** Total distance traveled by the car in kilometers.
- **Fuel Type:** The type of fuel used (e.g., Petrol, Diesel, CNG).
- **Transmission:** Transmission type (Manual or Automatic).
- **Owner:** Ownership history (e.g., First owner, Second owner).
- **Seller Type:** Whether the seller is a Dealer or an Individual.
- **Engine:** Engine displacement in cc.
- **Max Power:** Maximum power output of the car.
- **Max Torque:** Maximum torque of the car.
- **Vehicle Dimensions:** Attributes like length, width, and height, if applicable.

How do you intend to analyze your data?

1. **Data Cleaning and Preprocessing:**
 - Handle missing or inconsistent values.
 - Correct data types for categorical and numerical features.
2. **Exploratory Data Analysis (EDA):**
 - Use **statistical summaries** and **visualizations** to understand data distributions and identify correlations.
 - Apply **scatter plots** to explore relationships.
3. **Hypothesis Testing:**
 - Use statistical tests to confirm significant relationships between variables and resale price.
4. **Visualization and Insights:**
 - Present findings through a **dashboard** featuring:
 - Trends by fuel type, or transmission.
 - Inventory analysis (e.g., fast vs. slow-moving cars).

❖ Milestones

1. Data Understanding

Goal: Gain familiarity with the dataset to identify its structure, contents, and potential use cases.

- Define key variables such as resale price, present price, fuel type, and car age.
- Perform exploratory data analysis (EDA) to uncover initial patterns, distributions, and relationships between variables

2. Data Cleaning

Goal: Prepare the data for accurate analysis by addressing inconsistencies and inaccuracies.

- Handle missing values by applying appropriate imputation techniques or filtering.
- Ensure uniformity in formats, especially for numerical and categorical columns like "Max Power" or "Fuel Type."
- Remove outliers and standardize units for reliable comparisons.

3. Data Visualization

Goal: Transform raw data into meaningful visual insights.

- Create visualizations to identify key trends, such as resale price variations over time, fuel type distribution, and mileage patterns.
- Use visual tools to highlight regional differences, pricing bands, and other actionable insights.

4. Dashboard Development

Goal: Develop an interactive and user-friendly dashboard for real-time data exploration.

- Include filters for fuel type, transmission, seller type, price range, and car age.
- Display key metrics like total cars, average resale price, most common fuel type, and price range.
- Use tools like Tableau or Google Looker Studio for dynamic visuals and real-time interactivity.

5. Final Reporting

Goal: Summarize findings and present actionable recommendations.

- Prepare a detailed report outlining insights, trends, and correlations identified during the analysis.
- Highlight potential business strategies, such as pricing models, inventory management, and target audience segmentation.

❖ Timeline:

Week	Tasks

Week 1	Exploratory Data Analysis (EDA): Gain a comprehensive understanding of the dataset. EDA report with visualizations and initial insights.
Week 2	Data Preparation and Modeling: Prepare the dataset for analysis and develop models if needed. Trained and validated models, along with performance metrics.
Week 3	Datafolio + dashboard: Extract actionable insights and present findings. Dashboard showcasing insights and recommendations
Week 4	Final Report: A polished report for dashboard to present findings.