

**Used Cars Price Analysis**

**ACKNOWLEDGMENT**

I want to extend my sincere gratitude to the entire team at “Analytics Space” for granting me the opportunity to work on the “**Used Cars Price Analysis**” project. I would also like to express my heartfelt thanks to my academic team, 'Data Trained,' for providing me with this enriching experience. Working on this project not only allowed me to enhance my analytical skills but also exposed me to a plethora of new insights. The invaluable suggestions and guidance from both teams played a pivotal role in the successful completion of this project, and I am truly grateful for the learning and growth it has brought to me.

**CHAPTER 1**

**INTRODUCTION**

**1.1 Project Objective**

The primary objective of the "Used Car Price Analysis" project is to analyze the factors influencing the pricing of used cars and develop insights that help in understanding market trends. This involves:

1. Collecting and preprocessing data on used cars, including attributes such as age, brand, model, mileage, location, and condition.
2. Identifying key factors that significantly impact the price of a used car.
3. Applying data visualization techniques to uncover patterns and relationships within the dataset.
4. Building predictive models using machine learning algorithms to estimate the price of a used car based on its features.
5. Providing actionable insights to buyers, sellers, and stakeholders to make informed decisions in the used car market.

This project aims to bridge the gap between market understanding and pricing strategy using data-driven techniques.

**Overview of Project**

This project involves multiple stages, including:

* **Web Scraping and Data Collection:** Extracting data on used cars, such as brand, price, emi’s, fuel, no of cars from each company. Details from sources like car Dekho
* **Data Cleaning and Preprocessing:** Preparing the dataset for analysis by handling missing values, removing duplicates, and standardizing formats.
* **Data Modelling and Analysis:** Applying statistical and machine learning techniques to uncover patterns and insights regarding college distribution, course availability, and fee variations.

**Visualization:** Presenting findings through interactive dashboards and visualizations using Tableau to facilitate decision-making for students, parents, and policymakers.

**1.2 scope**

The "Used Car Price Analysis" project focuses on analyzing factors that influence the pricing of used cars and building predictive models to estimate their market value. It includes collecting and preprocessing data, performing exploratory analysis to identify key trends, and developing machine learning models for price prediction. The project aims to assist buyers, sellers, and platforms like Cardekho in making informed decisions by offering insights into fair and competitive pricing.

This project also extends to improving pricing transparency and market efficiency while enhancing user trust on online platforms. The project provides opportunities to apply data science techniques, explore real-world datasets, and gain hands-on experience in solving industry-relevant problems. In the future, it can incorporate additional data like maintenance history and dynamic pricing for more robust models.

**1.3 Problem Statement**

The used car market is highly dynamic, with prices influenced by numerous factors such as vehicle specifications, market demand, and regional trends. Platforms like Car Dekho aim to provide accurate and fair pricing insights for buyers and sellers. However, determining the appropriate price for a used car remains challenging due to the complexity of factors affecting value, such as mileage, age, condition, fuel type, and brand reputation.

The problem is to analyze the existing data from the Car Dekho website to:

1. Understand the key factors influencing the pricing of used cars.
2. Identify trends and patterns in the market to support transparent pricing.
3. Develop a predictive model that accurately estimates the price of a used car based on its features**.**

**1.4 Project Objectives**

The **"Used Car Price Analysis"** project focuses on understanding and predicting the pricing dynamics in the used car market. The used car industry is vast and influenced by various factors such as vehicle age, brand, mileage, fuel type, location, and condition. Accurate analysis of these factors can help buyers and sellers make informed decisions, enhancing transparency and efficiency in the market.

The project involves several key steps:

1. **Data Collection**: Gathering a comprehensive dataset containing details of used cars, such as specifications, pricing, and market conditions.
2. **Data Preprocessing**: Cleaning and organizing the data to handle missing values, outliers, and inconsistencies, ensuring it is ready for analysis.
3. **Exploratory Data Analysis (EDA)**: Using visualization tools to uncover trends, correlations, and patterns that influence used car prices.
4. **Feature Engineering**: Identifying and creating key features that have the most significant impact on pricing.
5. **Model Development**: Employing machine learning algorithms to build predictive models capable of estimating used car prices based on input features.
6. **Insights and Recommendations**: Presenting actionable insights derived from the analysis to guide decision-making for buyers, sellers, and industry stakeholders.

The project aims to showcase the practical application of data science techniques in solving real-world problems, leveraging analytical and machine learning skills to derive meaningful conclusions.

**CHAPTER 2**

**DATA COLLECTION AND SOURCES**

**Data Source**

The primary data source for the **"Used Car Price Analysis"** project is from **Cardekho** website, a popular platform for buying and selling new and used cars in India. The website provides detailed listings of used cars, including attributes such as:

1. **Vehicle Specifications**: Brand, model, year of manufacture, fuel type, transmission, mileage, and engine capacity.
2. **Pricing Information**: Listed prices of used cars across various regions.
3. **Location Details**: City and state where the car is available for sale.
4. **Additional Features**: Vehicle condition, number of owners, insurance status, and other relevant details.

The extracted data is then stored in structured formats such as CSV or Excel, enabling easy accessibility and further analysis. To ensure reliability, the data undergoes preprocessing, including cleaning, removing duplicates, and handling missing values.

This comprehensive approach ensures the availability of high-quality datasets for analysing pricing trends and student preferences in engineering education across India.

**Web Scraping Process**

The web scraping process for this project is designed to extract detailed information about used cars price analysis from the "CarWale" website systematically.

The key Python libraries utilized in the web scraping process were:

1. BeautifulSoup: This library was used to parse the HTML content of the CarWale website. BeautifulSoup provides a simple way to navigate and search through the HTML structure, Enables easy extraction of specific data points, such as fuel type, specifications, and brand.
2. Requests: The Requests library was employed to send HTTP requests to the CarWale website and retrieve the HTML content of the pages. By using Requests, we could access multiple pages of the website programmatically, allowing for the extraction of data from a large number of cars.
3. Pandas: After extracting the data, Pandas was used to organize it into a structured format, typically a DataFrame. This allowed for easier manipulation and analysis of the data, including cleaning, filtering, and sorting.

**Steps in the Web Scraping Process:**

1. **Sending Requests**: HTTP requests were made to getmyuni.com to access data on engineering colleges across all Indian states.
2. **Parsing HTML Content**: BeautifulSoup was used to extract relevant data, identifying specific HTML tags and attributes.
3. **Extracting Data**: Parsed data was stored in structured Pandas DataFrames, ensuring consistency and accuracy.
4. **Saving Data**: The final dataset was saved in CSV format for easy analysis using data visualization tools like Tableau and Python-based analytics.

**Challenges**

Despite the effectiveness of web scraping, certain challenges were encountered:

1. Data Collection:
   * Extracting data from the Cardekho website using web scraping while ensuring compliance with legal and ethical guidelines.
   * Handling restrictions such as CAPTCHA, rate limiting, or lack of APIs for direct data access.
2. Data Quality Issues:
   * Dealing with incomplete, inconsistent, or missing data in the dataset.
   * Identifying and removing outliers or incorrect values that could skew analysis and model performance.
3. Complexity of Price Influences:
   * Accounting for intangible factors like brand reputation or market demand that are not directly measurable.
   * Handling regional variations and market-specific trends in pricing.
4. Model Development:
   * Selecting and fine-tuning machine learning algorithms to achieve high prediction accuracy.
   * Avoiding overfitting or underfitting when training models on the data.
5. Resource Limitations:
   * Managing computational resources when dealing with large datasets or complex models.
   * Ensuring the project stays within time constraints while maintaining quality.
   * **CHAPTER 3**

### **DATA PREPROCESSING**

Effective data preprocessing is a critical step in ensuring the quality and reliability of the analysis. This stage involves cleaning the raw data, transforming it into a suitable format for analysis, and importing it into Tableau for visualization. Below is a detailed explanation of the steps involved in the data preprocessing phase of this project.

#### **Data Cleaning**

The raw data collected through web scraping contained several inconsistencies and errors that needed to be addressed before proceeding with the analysis. Data cleaning is essential to remove noise and inaccuracies, ensuring that the dataset is accurate, consistent, and ready for analysis. The following steps were taken to clean the data:

1. **Removal of Duplicates:**
   * **Identification of Duplicates:** The first step in the data cleaning process was to identify and remove duplicate entries. Duplicate data can occur when the same hotel is listed multiple times, possibly due to variations in room types or slight differences in hotel name formatting.
   * **Removal Process:** Using the Pandas library in Python, duplicate entries were detected based on unique identifiers such as hotel names and addresses. These duplicates were then removed, leaving only one entry per hotel. This step ensured that each hotel was represented only once in the dataset, preventing any skewing of the analysis.
2. **Handling Missing Values:**
   * **Identifying Missing Data:** Missing values can significantly impact the results of the analysis if not handled properly. In this project, missing data was primarily found in columns related to customer reviews and ratings, where some entries lacked complete information.
3. **Correction of Errors:**

**Data Accuracy:** During the data collection process, some errors were identified, such as incorrect price formatting or misaligned

* + columns. For example, some prices were mistakenly captured in different currencies or formats.
  + **Standardization:** These errors were corrected by standardizing the data. Prices were converted to a consistent currency format (Indian Rupees), and columns were realigned to ensure that all data points were correctly labelled and categorized.

1. **Outlier Detection:**
   * **Identifying Outliers:** Outliers, such as extremely high or low prices, were identified as they could distort the analysis. These outliers were detected using statistical methods like the interquartile range (IQR) and Z-score analysis.
   * **Handling Outliers:** Depending on the nature of the outlier, some were retained if they provided meaningful insights, while others were adjusted or removed if they were determined to be errors or anomalies.

**CHAPTER 4**

**DATA ANALYSIS AND VISUALIZATION**

**4.1 Introduction**

This chapter presents the analysis and visualization of used cars in India, focusing on key aspects such as brand, price, emi’s, fuel, no of car’s from each company. The study provides insights into state-wise and city-wise Car price variations based on various cars brands, Best cars with good quality with less price.

**4.2 Engineering Education Landscape: State-Wise Distribution**

The first visualization presents an overview of used cars across India. It includes filters for brands and locations, enabling users to analyze specific brands and regions. The primary metrics showcased in the dashboard include:

**4.2.1 Key Metrics**

* **Average price:** The average price across the analyzed cars is ₹1.25M
* **Minimum Price:** The minimum price ₹75.00k
* **Maximum Price:** The maximum price ₹24.50M
* **Top 10 Cars by Price:**
  + 2024Mercedes-Benz Maybach S-ClassS580 BSVI with price of 24,500,000.00
  + 2021Mercedes-Benz AMG GLE 53Coupe BSVI with price of 10,500,000.00
  + 2023Audi Q7Technology Audi with price of 8,300,000.00.

**4.2.2 Count of Cars by Brand Distribution**

A bar chart in the dashboard visualizes the number of cars each brand holds.

* Maruti has the highest number of used cars, followed by Hyundai, kia, and

Tata

**4.2.3 Fuel and Transmission Distribution**

* Donut Chart: Distribution of cars by Fuel Type (Petrol, Diesel, CNG, Electric)
* Pie Chart: Distribution by Transmission (Manual vs Automatic)

These visuals help users understand the market share of fuel options and transmission preferences in the used car segment.

**4.3 Price Analysis Dashboard**

**4.3.1 Price Comparison by Brand**

This section includes a bar chart comparing the average price of cars across brands. Premium brands like BMW, Mercedes-Benz, and Audi show significantly higher average prices than budget-friendly brands like Maruti and Tata.

**4.3.2 Price vs Year**

A line chart shows the relationship between car age (year of manufacture) and average price. As expected, newer cars have higher prices, while older models show a declining trend due to depreciation.

**4.3.2 Average of Price by Fuel Type and Transmission**

This visualization provides insights into how the average price of used cars varies based on fuel type and transmission. By segmenting the dataset in this way, users can identify which combinations are typically more expensive and which offer better affordability.

**4.4 Conclusion**

This dashboards presents a comprehensive view of the Indian used car market:

* Users can filter by **brand, price range, and location**
* KPIs help benchmark average and extreme prices
* Visuals support exploratory analysis to identify undervalued or overpriced listings
* Power BI enhances interactivity and insights, making the data easily understandable for both technical and non-technical users

The dashboard serves as a valuable tool for data-driven decision-making in buying or selling used cars.

**CHAPTER 5**

**TOOLS AND TECHNOLOGIES**

**Software Requirement: -**

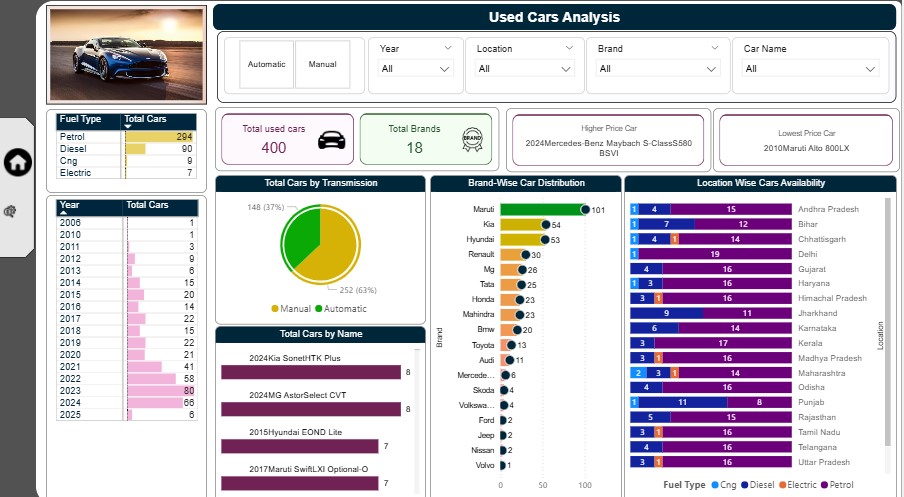
* Python (Jupyter notebook)
* Power BI
* Excel

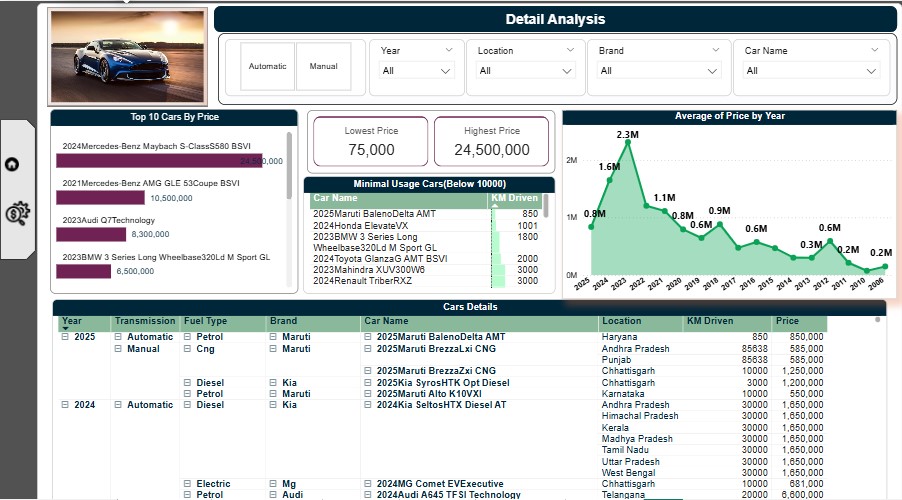
**Main Software Libraries: -**

* Pandas
* NumPy
* Beautiful Soups
* Requests
* CSV

**CHAPTER 6**

**DASHBOARDS**





**CHAPTER 7**

**CONCLUSION**

The "Price Analysis for Used Cars" project on the Cardekho platform highlights key factors affecting used car prices, such as car age, brand, fuel type, transmission, and location. It was found that older cars and those with higher mileage generally have lower prices, while premium brands and specific models tend to retain higher value. Diesel cars and those with automatic transmission also command higher prices. Regional variations in prices were observed, with urban areas typically having higher car prices. The project provides valuable insights for both buyers and sellers, helping them make informed decisions on pricing and purchases. Future enhancements, like incorporating maintenance history and real-time data, could further refine the pricing model. Overall, this analysis allows Cardekho to improve its pricing algorithms, creating a more transparent and efficient used car market.