

Used Cars

Price Analysis



Presented by:-

Myadam Akshay

**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. **Project Objective**

The objective of the “Used Cars Price Analysis” project is to conduct an in-depth analysis of the pricing variations among different Cars in the India. To analyze Cars data across India to uncover customer preferences, optimize pricing strategies, and provide actionable market insights that will help improve Cars operations, marketing, and overall competitiveness in the Indian Cars Sector.

The data is taken from “Cars 24” website by using some python webscrapping to extract the data in the form of datasets.

**Overview of Project :**

Discover the main objectives and scope of Project for web scraping and Data cleaning , Data Modelling , Data preprocessing In Python and data visualization in Power bi.

**1.2 Problem Statement**

**Objective:** To analysis used car prices on the Cars24 platform to understand the key factors influencing price variations and develop insights for potential buyers and sellers.

**Scope:** This analysis will focus on data extracted from the Cars24 website, encompassing factors such as:

* **Vehicle Characteristics:** Year, Model, Kilometer, Fuel Type, Amount , Special Feature, Number of Owners
* **Market Trends:** Regional price variations, seasonal trends, and overall market demand
* **Other Factors:** Condition of the vehicle (if available), presence of optional features, and any other relevant variables

This problem statement provides a clear framework for your analysis, outlining the objectives, scope, and expected outcomes. You can adapt it further to align with specific research questions or areas of interest within the used car market.

#### ****1.3 Project Objectives****

The primary objective of this project is to analyze a dataset comprising Car names, Year, pricing information, and other relevant variables from various cities across India. Through this analysis, the project aims to achieve the following specific objectives:

#### Understand Price Determinants:

#### Identify the key factors that significantly influence used car prices on the Cars24 platform.

#### Quantify the impact of each factor on price variations.

#### Develop a Predictive Model:

#### Create a model that can accurately predict the selling price of a used car based on its characteristics and market conditions.

#### Gain Market Insights:

#### Analyze market trends, including regional price variations, seasonal fluctuations, and overall market demand for different vehicle types.

#### Inform Decision-Making:

#### Provide actionable insights for both buyers and sellers to make informed decisions in the used car market:

#### For Buyers: Identify undervalued cars, negotiate effectively, and make informed purchase decisions.

#### For Sellers: Optimize pricing strategies, understand market demand, and maximize selling prices.

#### ****1.4 Scope****

This analysis will examine used car prices on the Cars24 platform, considering factors such as vehicle make, model, year, mileage, fuel type, transmission, engine size, body style, and number of owners. It will also investigate market trends like regional price variations, seasonal fluctuations, and overall market demand. Additionally, the analysis will explore the impact of factors such as vehicle condition and optional features on pricing. Data will be primarily sourced from Cars24 listings. It's important to note that the analysis may have limitations due to potential variations in data availability and quality, and it may not fully represent the entire Indian used car market.

### **Chapter 2: Data Collection and Sources**

#### ****Data Source****

The primary data source for this project is the **“Cars 24”** website, one of India's leading online Cars Website platforms. Cars 24 offers a comprehensive range of Used Cars services, including Cars pricing, Year, Special feature, and more. The website features an extensive database of cars across India, catering to various customer preferences and budget ranges. For this project, the focus was specifically on collecting data related to cars in tier-1 cities such as Bangalore, Chennai, Delhi, Hyderabad, Kolkata, and Mumbai.

The dataset collected from the Cars 24 website includes the following key variables:

1. **Car Names:** The names of the cars listed on the platform, which provide a basic identification of the cars for analysing.
2. **Special Feature:** It shows the Special feature of the car for understanding customer
3. **Kilometer’s:** Identified as the total distance a car has been driven, typically displayed in the vehicle listing as Kilometer Driven.
4. **Fuel\_type:** It refers to the type of fuel the vehicle uses, such as petrol, diesel, electric, hybrid, or CNG
5. **Amount:** Typically refers to the selling price of the used car

The combination of these data points allows for a comprehensive analysis of the factors influencing Cars performance and customer preferences in the Used Cars Analysis

#### ****Web Scraping Process****

#### Given the need for large-scale data collection, web scraping was selected as the most efficient method to extract information from the Cars24 website. This technique involves automatically retrieving and processing data from web pages using software or scripts. Python, with its extensive libraries and ease of use for data manipulation, was chosen as the primary programming language for this task.

#### Key Python libraries utilized in the web scraping process included:

#### BeautifulSoup: This library was instrumental in parsing the HTML content of the Cars24 website. It facilitated navigating and searching through the HTML structure, enabling the efficient extraction of specific data points such as car make, model, year, mileage, price, and other relevant features.

#### Requests: The Requests library was employed to send HTTP requests to the Cars24 website and retrieve the HTML content of the desired pages. This enabled programmatic access to multiple pages, facilitating the extraction of data from a large number of used cars.

#### Pandas: After data extraction, Pandas was utilized to organize the data into a structured format, typically a DataFrame. This facilitated data manipulation and analysis, including cleaning, filtering, and sorting, prior to further analysis.

#### The web scraping process involved the following steps:

#### Sending Requests: HTTP requests were sent to the Cars24 website to access pages containing used car listings. This process was executed for various categories, locations, or search criteria to ensure a comprehensive dataset.

#### Parsing HTML Content: BeautifulSoup was then employed to parse the retrieved HTML content. This involved identifying and extracting the specific HTML tags and attributes containing the desired information, such as car make, model, year, mileage, price, and other relevant features.

#### Extracting Data: The extracted data was then organized into a structured format using Pandas. This step also included data cleaning to remove any irrelevant or duplicate information and ensure consistent data formatting.

#### Saving Data: Finally, the cleaned and structured data was saved to a CSV file for subsequent import into a data analysis or visualization tool for further exploration and insights.

#### ****Challenges****

While the web scraping process was effective in collecting a large volume of data, several challenges were encountered:

1. **Data Availability and Quality:** Obtaining reliable and comprehensive datasets can be challenging. Missing, incomplete, or inconsistent data might affect the analysis accuracy.
2. **Market Variability**: Prices fluctuate based on seasonal demand, local economic conditions, and unforeseen events, making it difficult to establish stable patterns.
3. **Feature Complexity**: Determining the impact of diverse features like brand reputation, condition, and additional accessories requires sophisticated techniques.

Despite these challenges, the web scraping process successfully resulted in a rich dataset that provided valuable insights into the Cars Website. The next step involved analyzing this data to uncover trends, optimize pricing strategies, and better understand customer preferences.

### **Chapter 3: Data Preprocessing**

Effective data preprocessing is critical for ensuring the accuracy and reliability of analysis. This phase involved cleaning the raw data, standardizing formats, and preparing it for visualization and insights generation.

**Cleaning and Transformation**

1. **Duplicate Removal**:
   * Identified and removed duplicate entries using Pandas.
   * Ensured each car was represented only once to maintain data integrity.
2. **Handling Missing Values**:
   * Missing information in key fields like price or mileage was either filled using median values or excluded based on the analysis requirements.
3. **Standardization**:
   * Unified formats for kilometer readings (e.g., "43,354 km" standardized to numeric format).
   * Standardized price data to Indian Rupees (INR).
4. **Outlier Detection**:
   * Detected outliers in price and mileage using statistical methods like interquartile range (IQR).
   * Retained significant outliers that offered insights into unique market behaviors, while excluding erroneous data.

**Data Segmentation**

For deeper analysis, the data was segmented based on:

* **Brand**: Popular manufacturers such as Maruti, Hyundai, and Honda.
* **Location**: Metropolitan areas like Bangalore, Hyderabad, and Delhi.
* **Ownership**: First-owner versus second-owner vehicles.
* **Fuel Type**: Petrol, diesel, and hybrid models.

### **Chapter 4: Data Analysis and Visualization**

Data analysis and visualization are central to understanding the dynamics of the Used Cars in India. This section provides an overview of the key analyses conducted, including customer satisfaction, pricing trends, and Variations of different cars. Visualizations created using Power BI are presented, along with interpretations of the findings.

#### ****Overview of Analysis****

The analysis conducted in this project was focused on deriving actionable insights from the data collected. The key metrics analyzed include:

1. **Customer Satisfaction:**
   * The Customer satisfaction in the used car market things on several key factors. Transparency is crucial, with accurate vehicle condition reports, fair pricing, and clear disclosure of any existing issues being paramount. Customers expect high-quality vehicles in good mechanical condition.
2. **Pricing Trends:**
   * Understanding pricing trends across different regions and cars categories was a critical part of the analysis. By analyzing price distributions and trends, the study sought to identify competitive pricing strategies and their impact on customer satisfaction.
3. **Market Segmentation:**
   * The market segmentation analysis focused on differentiating between various segments within the Cars Brand, such as luxury budget cars. This helped in understanding the preferences and behaviors of different customer groups.

#### ****Visualizations****

The following visualizations were created in Power BI to illustrate the key findings from the data analysis.

1. **Price Analysis**
   * **Price Distribution:**
     + A histogram was created to visualize the distribution of car prices across India. This chart helps to identify the most common price ranges and highlights any significant outliers. The majority of cars were found to be in the mid-range price category, with fewer cars at the extreme ends of the pricing spectrum.
   * **Average Prices Across Regions:**
     + A bar chart was used to compare the average prices of cars across different regions in India. This visualization revealed regional variations in pricing, with metropolitan areas like Delhi and Mumbai having higher average prices compared to smaller cities and rural areas.
   * **Pricing Trends Over Time:**
     + Used car analysis refers to how the average prices of used cars change over specific periods. Key aspects include depreciation, market cycles, seasonal variations, and model-specific trends. Analyzing these trends helps understand market dynamics, inform predictive models, and guide decision-making for buyers and sellers.
2. **Customer Sentiment Analysis**
   * The used car market involves analysing customer opinions and emotions expressed in online reviews, social media, and other channels. This helps understand customer satisfaction, identify areas for improvement, monitor brand reputation, and guide product development. By analysing sentiment, used car companies can enhance customer experience, build stronger relationships, and gain a competitive advantage.

**3. Geographic Distribution Analysis:**

* Regional Price Variations: Analyzed price variations across different states and cities within the country.
* Demand Analysis: Identified regions with high and low demand for used cars.
* Visualization:
  + Maps: Visualized regional price variations and demand using heatmaps or choropleth maps.

**4. Competitive Analysis:**

* Competitor Pricing: Analyzed pricing strategies of major players in the used car market.
* Competitive Advantages: Identified the unique selling propositions (USPs) of different players.
* Visualization:
  + Bar charts: Compared average prices offered by different platforms for similar car models.

**5. Customer Sentiment Analysis:**

* Customer Reviews: Analyzed customer reviews and feedback to understand customer satisfaction levels.
* Social Media Monitoring: Monitored social media mentions to gauge public sentiment towards different brands and models.
* Visualization:
  + Word clouds: Visualized frequently used keywords in customer reviews.
  + Sentiment analysis dashboards: Displayed overall sentiment and key themes emerging from customer feedback.

**Interpretation & Insights:**

* Informed Decision-Making: Provided insights to buyers and sellers, dealerships, and industry players for informed decision-making.
* Market Predictions: Helped in forecasting future market trends and identifying potential opportunities.
* Business Strategy Development: Guided businesses in developing effective pricing strategies, marketing campaigns, and customer service initiatives.

*By combining data analysis with insightful visualizations, stakeholders can gain a deeper understanding of the used car market, identify key trends, and make data-driven decisions to optimize their operations and gain a competitive edge.*

Charts:-

1. **Fuel Type Filter:**

* **Type:** A simple filter.
* **Interpretation:** Currently set to "Diesel," allowing users to filter the data to show information specifically for diesel-powered cars.

1. **Car Location:**

* **Type:** Text display.
* **Interpretation:** Shows the current location filter is set to "Bangalore."

1. **Number of Cars:**

* **Type:** Numeric value.
* **Interpretation:** Indicates that 100 diesel cars are currently being displayed in the filtered dataset.

1. **Min KM of the Car & Max KM of the Car:**

* **Type:** Numeric values with "KM" unit.
* **Interpretation:** Shows the minimum and maximum kilometer readings for the 100 diesel cars in the dataset. The range is from 3883 KM to 115160 KM.

1. **Price of Min Car & Price of Max Car:**

* **Type:** Numeric values with "K" likely representing thousands.
* **Interpretation:** Displays the minimum and maximum prices for the 100 diesel cars. The minimum price is 290K (likely 290,000 Rupees) and the maximum is 1760K (likely 1,760,000 Rupees).

1. **Min of Amount & Max of Amount:**

* **Type:** Numeric values with "K" likely representing thousands.
* **Interpretation:** These likely represent the minimum and maximum financing amounts (EMIs) for the displayed cars.

1. **Car List:**

* **Type:** A scrollable list of car models.
* **Interpretation:** Shows a list of different car models within the filtered dataset. Users can likely select individual models to view more details.

1. **Count of Owner:**

* **Type:** A numeric value.
* **Interpretation:** For the selected car model, it shows the number of previous owners.

1. **Min of Finance(Emi/m) by Brand:**

* **Type:** A pie chart.
* **Interpretation:** Shows the distribution of minimum financing amounts (EMIs) across different car brands. The size of each pie slice represents the proportion of the total financing amount for that brand.

1. **"Tap Here to Reset Slicer" Button:**

* **Type:** An interactive button.
* **Interpretation:** Clicking this button would likely reset all the filters and display the original, unfiltered data.

1. **Number of Cars with Special Features:**

* **Type:** A treemap.
* **Interpretation:** This visualization shows the number of cars with different special features. The size of each colored rectangle represents the number of cars with that particular feature. For example, "Top Model" seems to have the highest number of cars with that feature.

1. **Sum of Amount by Brand:**

* **Type:** A horizontal bar chart.
* **Interpretation:** This chart displays the total sum of amounts (likely the total selling price) for each car brand. "MG" has the highest sum of amounts, followed by "KIA" and "Mahindra."

1. **Average of Finance(Emi/m) and Kilometer by Brand:**

* **Type:** A bar chart with two sets of bars for each brand.
* **Interpretation:**
  + The blue bars represent the average finance amount (EMI per month) for each brand. "MG" has the highest average EMI.
  + The orange bars represent the average kilometer reading for cars of each brand. "Datsun" has the highest average kilometer reading.

**Chapter 5: Tools and Technologies**

Software Requirement:-

1)Python(Jupyter notebook)

2)Excel

3)Power BI

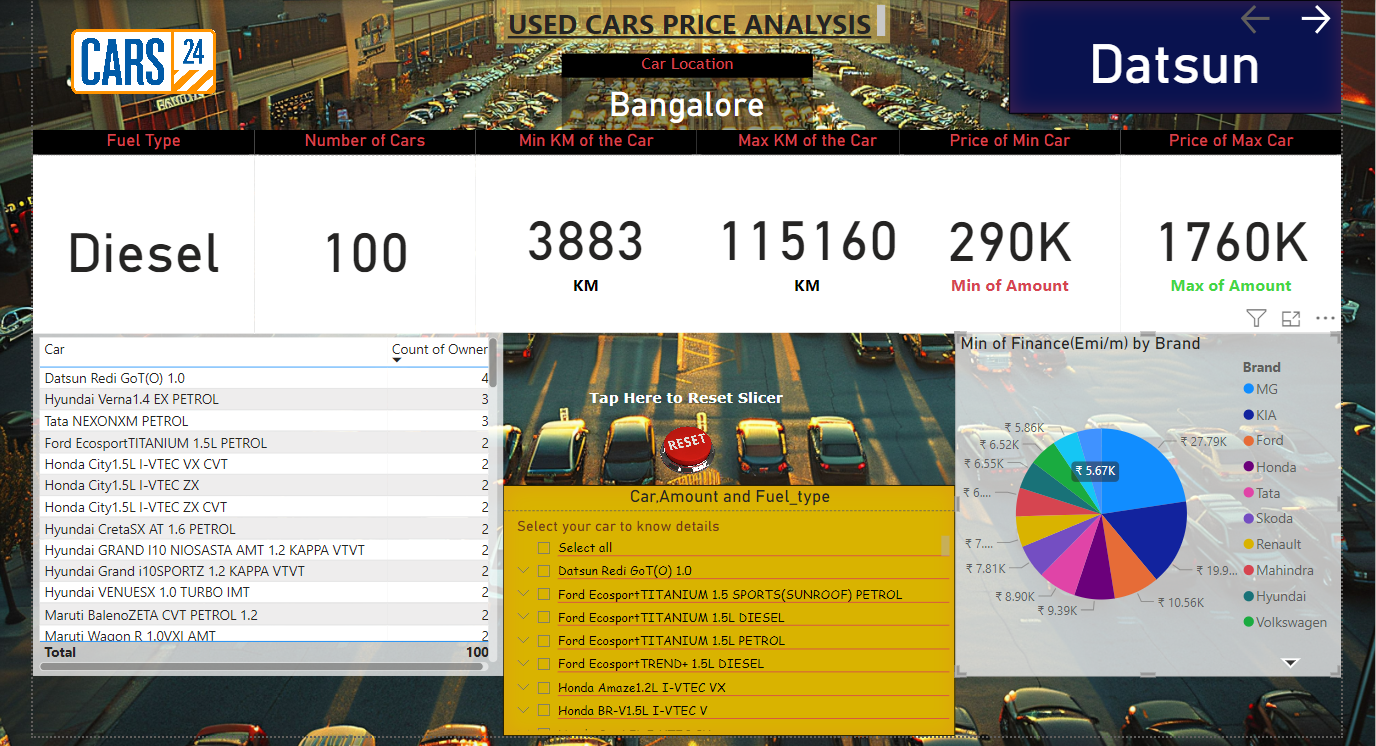
Main Software Libraries:-

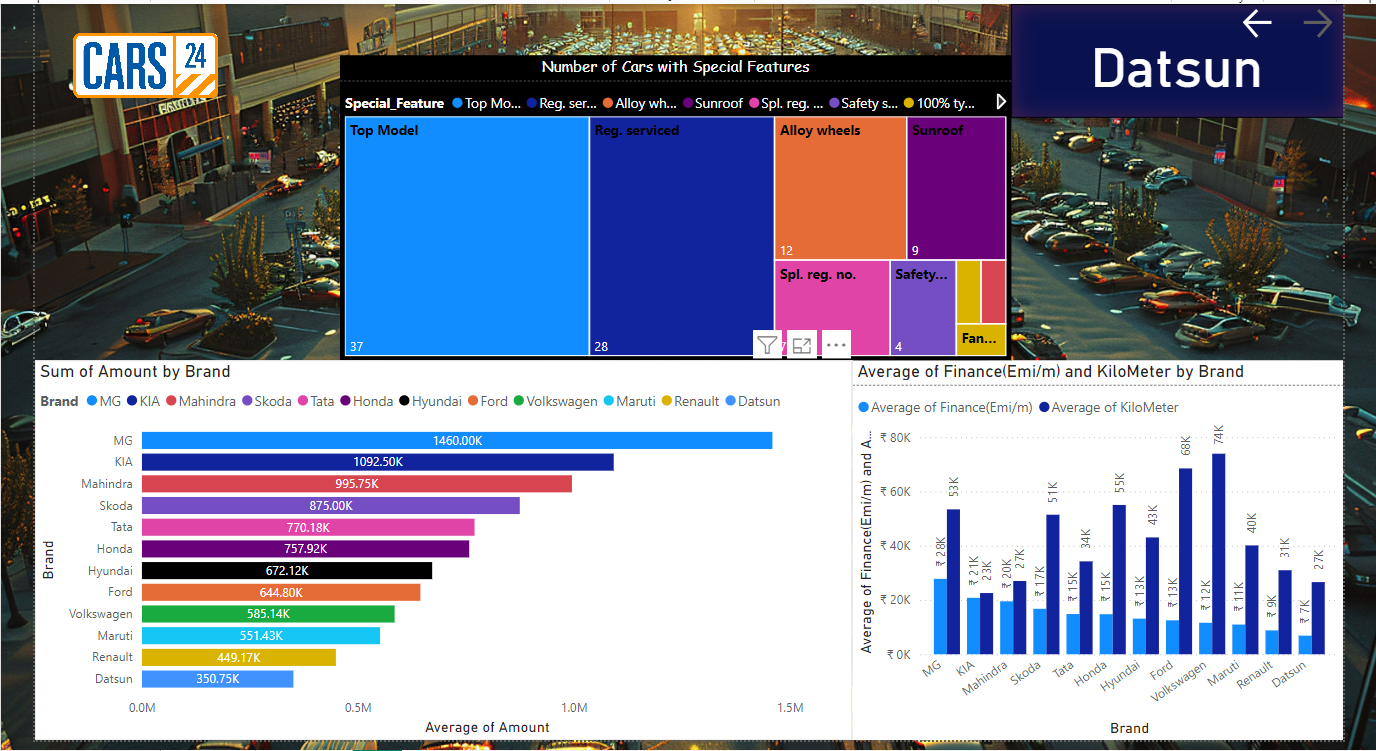
1)Pandas

2)Beautiful Soups

3)Requests

Chapter 6: Dashboard





Chapter 7: Conclusion

The analysis of used car prices reveals key factors influencing vehicle valuation. Age and mileage significantly affect prices, with older, high-mileage cars typically valued lower. Brand reputation and popular models command higher resale prices, while fuel efficiency and hybrid options appeal to cost-conscious buyers. Regional demand and economic conditions also create price variations. Additional features and the car’s condition further enhance its value. This study aids buyers in making informed decisions and helps sellers optimize pricing strategies. While insightful, future research can include larger datasets and emerging trends like electric vehicles to provide a more comprehensive market understanding.