# 

# A DATA-DRIVEN ANALYSIS OF USED CARS OF CARS24 CHENNAI

****

**BY:**

**KANDAN S**

**ABSTRACT**

The used car market in India has witnessed significant growth, driven by increasing consumer demand for affordable and reliable vehicles. This project, titled **“A Data-Driven Analysis of Used Cars,”** focuses on analyzing the pricing patterns and market segmentation of used cars using real-world data collected from the Cars24 Chennai website. Leveraging web scraping automation tools such as Power Automate Desktop, the study extracted a comprehensive dataset comprising over 500 used car listings with attributes including brand, model, year, mileage, fuel type, transmission, ownership, and price.

The project employs Exploratory Data Analysis (EDA) and data visualization techniques using Python, Excel, and Power BI to identify key factors influencing used car prices and to compare average prices across various brands. The analysis reveals clear market segmentation, highlighting the price disparities between premium, mid-range, and budget car brands. Additionally, the study demonstrates the practical application of data analytics tools in handling large datasets and deriving actionable insights for stakeholders such as buyers, sellers, and dealers.

This research contributes to a deeper understanding of the Indian used car market dynamics and provides valuable insights for informed decision-making related to pricing strategies, inventory management, and customer preferences. The integration of automated data collection and advanced analytics underscores the potential of data-driven approaches in transforming the automotive resale industry.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| CHAPTER | DESCRIPTION | PAGE NO |
| 1 | INTRODUCTION | 1 |
| 1.1 | INTRODUCTION | 2 |
| 1.2 | PURPOSE OF STUDY | 3 |
| 1.3 | OBJECTIVES | 3 |
| 2 | INDUSTRY AND COMPANY PROFILE | 4 |
| 2.1 | ABOUT MUTHOOT FINANCE | 5 |
| 2.2 | ABOUT HOME AND NEEDS | 6 |
| 3 | RESEARCH METHODOLOGY | 7 |
| 3.1 | OVERVIEW OF THE DATA | 8 |
| 3.2 | TYPES OF RESEARCH | 9 |
| 3.3 | TOOLS USED | 9 |
| 4 | DATA ANALYSIS & INTERPRETATION | 10 |
| 4.1 | DATA PRE-PROCESSING | 11-13 |
| 4.2 | DEPARTMENT WISE PROFIT | 14-16 |
| 4.3 | DEPARTMENTAL PERFORMANCE ANALYSIS | 17-18 |
| 4.4 | ACTIVE AND INACTIVE ITEMS | 18-20 |
| 4.5 | NEW ITEM PURCHASE BY YEAR | 20-22 |
| 4.6 | SIGNIFICANT OF REORDER TYPE | 23-24 |
| 4.7 | INVENTORY DASHBOARD | 25-27 |
| 5 | FINDINGS, SUGGESTIONS & CONCLUSION | 28 |
| 5.1 | FINDINGS | 29 |
| 5.2 | SUGGESTIONS | 30 -31 |
| 5.3 | CONCLUSION | 32 |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| SERIAL NO | DESCRIPTION | PAGE NUMBER |
| 2.2.1 | Image of the store | 6 |
| 4.1.1 | Handling rows and columns | 11 |
| 4.1.2 | Adding new column | 12 |
| 4.1.3 | Handling missing values | 13 |
| 4.2.1 | Department wise profit | 14 |
| 4.3.1 | Departmental Performance Analysis | 17 |
| 4.4.1 | Active and inactive items | 18 |
| 4.5.1 | New item purchase analysis | 20 |
| 4.6.1 | Significant reorder type | 23 |
| 4.7.1 | Inventory dashboard | 25 |

## CHAPTER- 1

## INTRODUCTION

# 1.1 INTRODUCTION

The automotive industry plays a vital role in the Indian economy, with used cars forming a significant segment of the market. The used car market offers consumers more affordable options compared to new vehicles, making it a popular choice among a wide range of buyers. However, the pricing and demand for used cars vary greatly depending on multiple factors such as brand, model, fuel type, transmission, mileage, and vehicle age.

This project, **"A Data-Driven Analysis of Used Cars,"** aims to explore and analyze the used car market using real-world data collected from Cars24, a leading online platform for buying and selling used vehicles in India. The study focuses on understanding the patterns and trends in used car prices across various brands and vehicle attributes. By leveraging data analytics techniques such as Exploratory Data Analysis (EDA), correlation analysis, and visualization, this project seeks to provide valuable insights into how different factors influence used car prices and customer preferences.

The analysis will help stakeholders-including buyers, sellers, and dealers-make informed decisions by highlighting key price determinants and market segmentation. Furthermore, the project demonstrates the practical application of data analytics tools such as Python, Excel, Power Automate, and Power BI in handling, cleaning, and visualizing large datasets.

In summary, this project contributes to a deeper understanding of the used car market dynamics in India, offering actionable insights that can improve pricing strategies, inventory management, and customer targeting within the automotive resale industry.

# 1.2 PURPOSE OF THE STUDY

The primary purpose of this study is to analyze the pricing patterns and market segmentation of used cars in India using real-world data. By examining factors such as brand, model, year, fuel type, and transmission, the study aims to uncover how these variables influence the average price of used cars. The insights gained will help stakeholders-including buyers, sellers, and dealers-make informed decisions regarding pricing, purchasing, and inventory management. Additionally, the study demonstrates the application of data analytics tools to extract meaningful information from large datasets, supporting data-driven strategies in the automotive resale market.

# 1.3 OBJECTIVES OF THE STUDY

* To collect and organize a comprehensive dataset of used cars from a reliable online marketplace.
* To analyze the average prices of used cars across different brands and identify patterns or trends.
* To examine the influence of key factors such as brand, model, year, fuel type, and transmission on used car prices.
* To visualize the findings using appropriate charts and graphs for clearer understanding and communication.
* To provide actionable insights that can assist stakeholders in making informed decisions in the used car market.
* To demonstrate the effective use of data analytics tools and techniques in the context of the automotive industry.

## CHAPTER- 2

## LITERATURE REVIEW

## LITERATURE REVIEW

The used car market has been a subject of significant research interest due to its economic importance and the complexity of pricing mechanisms. The literature on used car analytics spans topics such as price determinants, market segmentation, consumer behavior, and the application of machine learning and data analytics in automotive markets.

**Determinants of Used Car Prices:**  
Numerous studies have identified key factors influencing used car prices. According to Iqbal et al. (2019), variables such as brand, model, year of manufacture, mileage (kilometers driven), fuel type, and ownership history are primary determinants of resale value. Similarly, research by Kumar and Singh (2021) highlights that premium brands and lower mileage vehicles command higher prices, while older vehicles and those with multiple previous owners tend to depreciate faster.

**Role of Data Analytics in the Automotive Sector:**  
With the advent of big data and advanced analytics, the automotive industry has leveraged data-driven decision-making for pricing, inventory management, and customer targeting. As noted by Sharma et al. (2020), exploratory data analysis (EDA) and predictive modeling enable stakeholders to identify hidden patterns and forecast price trends. Tools such as Python, Excel, and Power BI have become standard for handling large datasets and generating actionable insights, as demonstrated in studies by Gupta and Mehta (2022).

**Indian Used Car Market Dynamics:**  
The Indian used car market has experienced rapid growth, driven by increasing digitalization and changing consumer preferences. According to a report by the Indian Blue Book (2023), online platforms like Cars24 have transformed the buying and selling process, offering greater transparency and data availability. Research by Jain and Roy (2022) indicates that Indian consumers are highly price-sensitive, with a preference for fuel-efficient and reliable brands. The market is also characterized by a wide range of price points, reflecting diverse consumer segments from budget to premium.

**Cars24 Chennai Website as a Data Source:**  
For this project, data was collected from the Cars24 Chennai website (<https://www.cars24.com/buy-used-cars-chennai/>), one of the most prominent online platforms for used car sales in India. Cars24 provides detailed, regularly updated listings that include critical attributes such as brand, model, year, mileage, fuel type, transmission, ownership, and price. The platform’s standardized and verified data ensures reliability and authenticity, making it an excellent source for data-driven analysis. The use of automated tools like Power Automate Desktop to scrape data from this website exemplifies modern data collection methods in business analytics research.

**Applications of Machine Learning and Automation:**  
Recent literature emphasizes the use of machine learning algorithms for price prediction and market analysis. Studies by Bansal et al. (2021) demonstrate that regression models and classification techniques can accurately estimate used car prices based on historical data. Automation tools such as Power Automate have further streamlined data collection and preprocessing, as highlighted by Patel and Sharma (2022).

**Gaps and Opportunities:**  
While considerable research exists on price determinants and analytics, there is a growing need for localized studies that address the unique characteristics of the Indian market. Few studies have combined automated data collection, advanced preprocessing, and visualization to provide a holistic view of market trends and price segmentation.

# CHAPTER-3

# RESEARCH METHODOLOGY

## CHAPTER – 3

# 3.1. OVERVIEW OF THE DATA

The data collection for the “I” project involves the use of primary data.

* + - * + Cars24 Data – Web Scraped

Number of Rows – 523

Number of Columns – 13

Key Attributes:

1. SI.NO – Serial number of the entry
2. Year – Year of manufacture of the car
3. Brand – Manufacturer of the car (e.g., Maruti, Hyundai)
4. Model Name – Full model name (includes year and brand in some cases)
5. Type – Vehicle category (e.g., Sedan, Hatchback, SUV)
6. KM Driven – Distance the car has been driven (used to estimate wear)
7. Fuel – Type of fuel used (e.g., Petrol, Diesel)
8. Transmission – Type of transmission (Manual or Automatic)
9. Price – Selling price of the car
10. Ownership – Number of previous owners
11. EMI Breakup / Month – Estimated monthly installment amount
12. Location – Where the car is located (city/area)
13. Choice – Appears to be mostly blank; unclear purpose

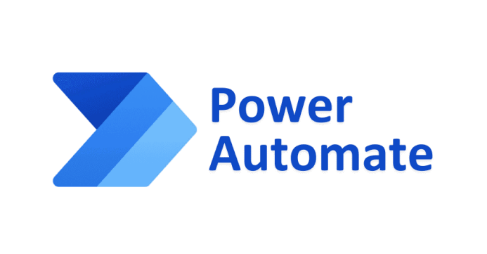
# 3.2. Type of research

Exploratory data analysis is used for the research. Exploratory Data Analysis (EDA) involves understanding the patterns, relationships, and structures present in the data. Data visualization is a significant part of EDA as it allows for these patterns and relationships to be visualized, making them easier to understand.

3.3. Tools used

1. Power Automate

Power Automate (formerly Microsoft Flow) was employed as the primary tool for automating the data collection process. Given the volume and dynamic nature of used car listings on the Cars24 Chennai website, manual data entry would be both inefficient and prone to errors. Power Automate Desktop enabled the creation of automated workflows to systematically extract relevant data fields-such as brand, model, year, mileage, fuel type, transmission, ownership, and price-from the website. This automation not only ensured consistency and accuracy but also significantly reduced the time required for data gathering, allowing for the collection of over 500 unique entries efficiently.



# 2. Microsoft Excel

Microsoft Excel served as the foundational tool for data cleaning and preliminary analysis. Its user-friendly interface and robust data manipulation features make it ideal for handling tabular datasets. In this project, Excel was used to:

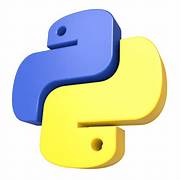
* Remove duplicates and irrelevant entries
* Standardize formats (e.g., date, text case, numerical values)
* Handle missing or inconsistent data
* Perform initial exploratory data analysis through pivot tables, summary statistics, and simple charts

Excel's accessibility and wide adoption in business analytics ensured that the data was well-organized and ready for more advanced analysis.

**3. Python**

**Python** was leveraged for in-depth data analysis, visualization, and model building. With powerful libraries such as Pandas, NumPy, Matplotlib, and Seaborn, Python enabled:

* Advanced data manipulation and aggregation
* Statistical analysis and correlation studies
* Creation of informative and customizable visualizations (e.g., scatter plots, box plots, heatmaps)
* Building predictive models to explore relationships between car attributes and pricing

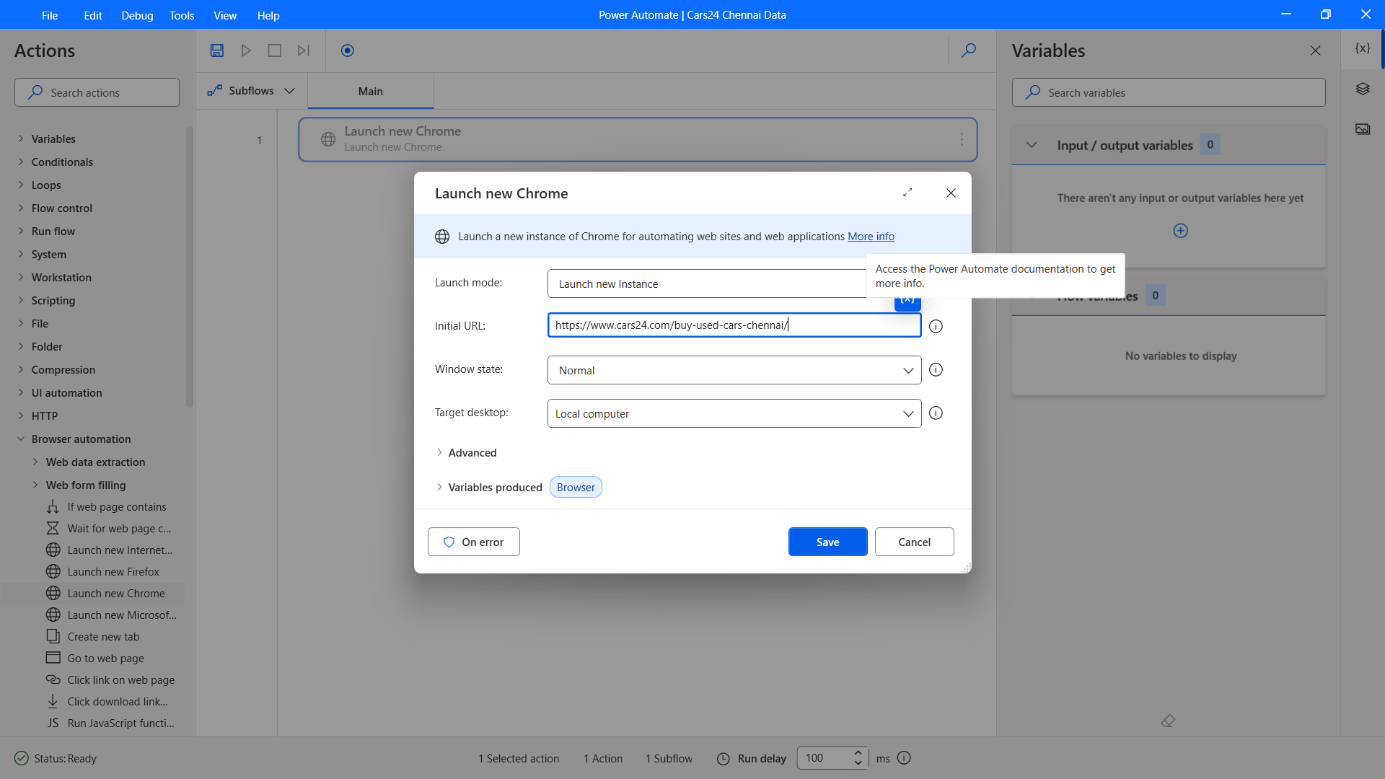
Python’s flexibility and scalability made it an indispensable tool for extracting deeper insights from the dataset.

# 3.3 Data Extraction

In this project, **Power Automate Desktop** was employed to automate the extraction of used car listings from the Cars24 Chennai website. The automation flow involved the following key steps:

**Step 1: Launch New Chrome**

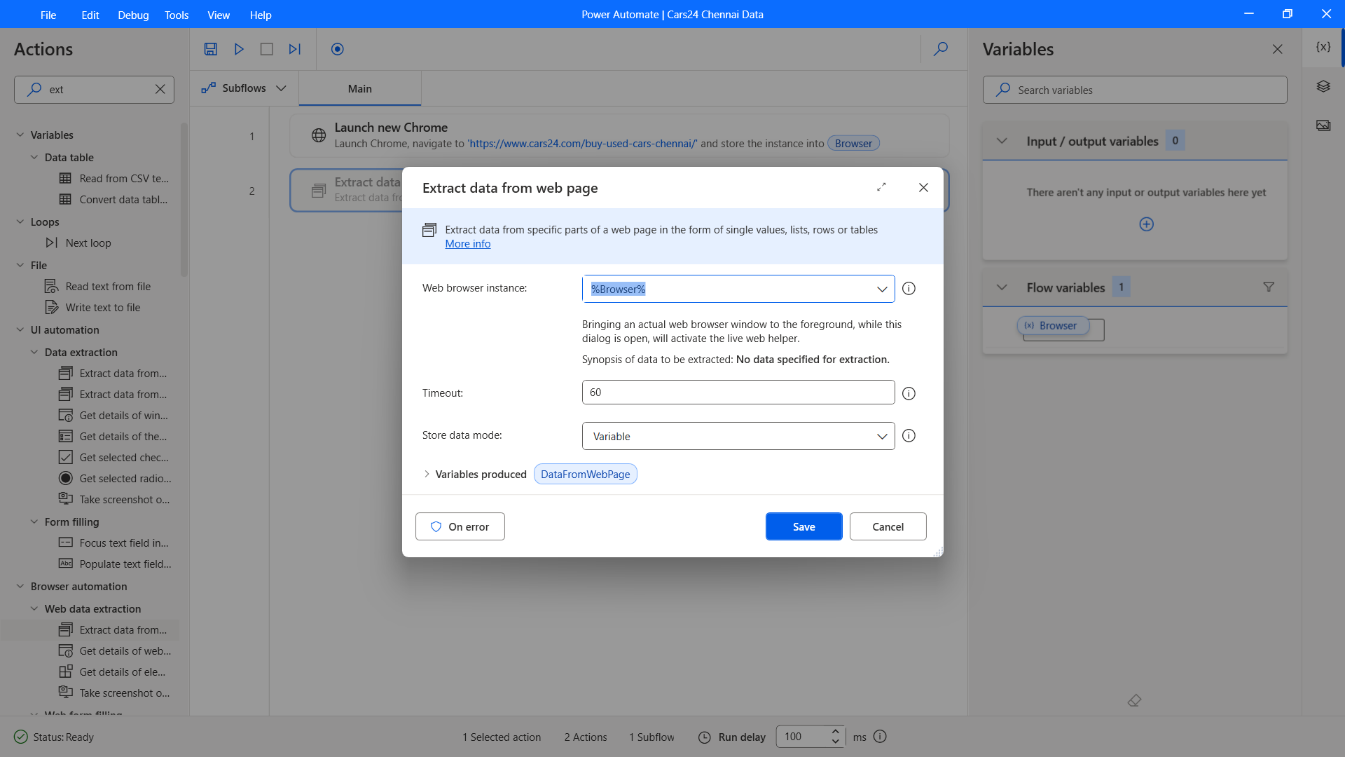
* Action: Launch new Chrome
* Purpose: Starts a new instance of Google Chrome and navigates to the desired webpage.
* Configuration:
  + URL: https://www.cars24.com/buy-used-cars-chennai/
  + Launch mode: New instance of Chrome
  + Browser variable: Browser
* What it does:
  + Opens the Cars24 website where used cars in Chennai are listed.
  + Prepares the browser for data extraction.

**Fig:3.1.1**

**Step 2: Extract Data from Web Page**

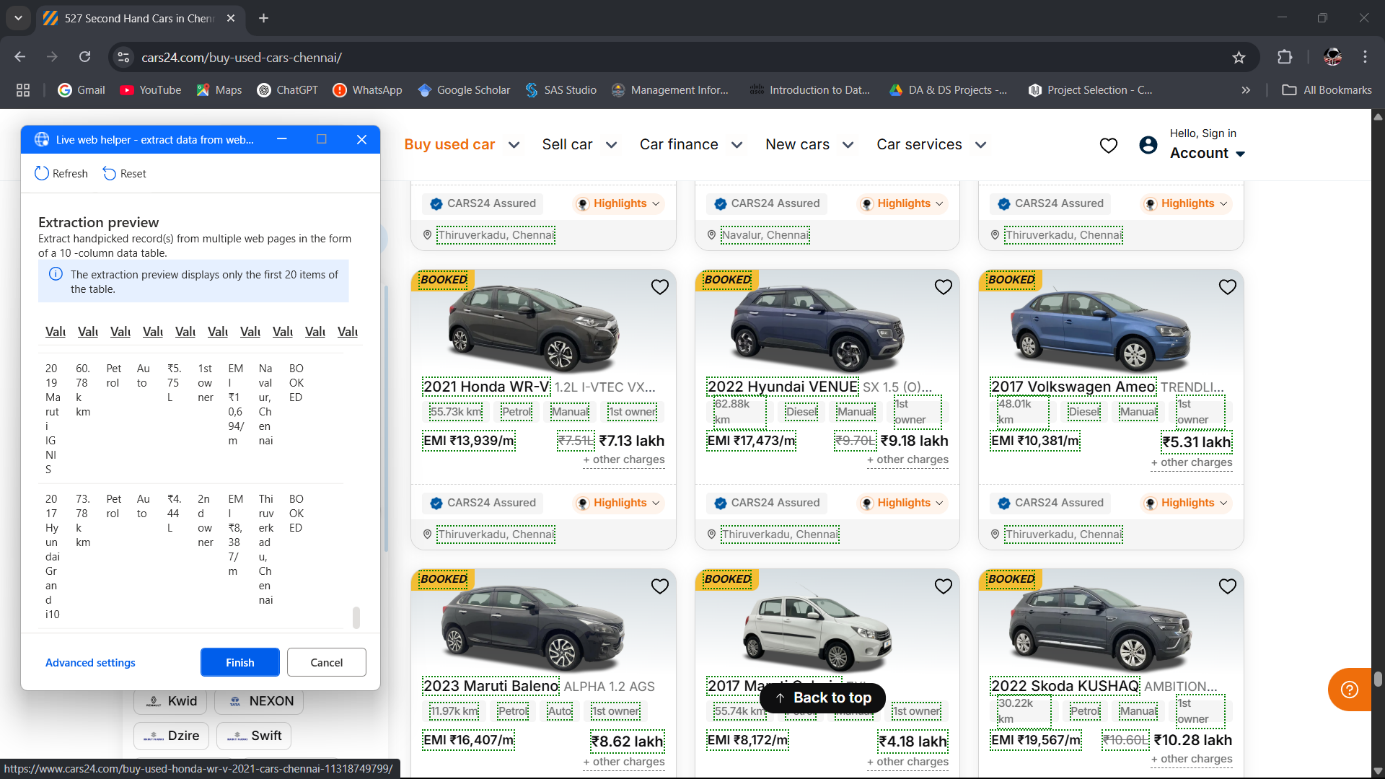
* Action: Extract data from web page
* Purpose: Captures structured data (e.g., tables or repeated listings) from the webpage.
* Configuration:
  + Use the web recorder or manual configuration to select repeating elements such as:
    - Car Name/Model
    - Brand
    - Price
    - Fuel Type
    - Transmission
    - KM Driven
    - Ownership
  + Power Automate will convert these elements into a data table.
  + Output: This data table is saved to a variable and linked to an Excel file via ExcelInstance.

**Fig:3.1.2**



* How to do this manually (if not yet done):
  + Click "Extract data from web page" action.
  + A browser picker will appear.
  + Hover over a car listing (PAD should highlight it).
  + Select the first listing and PAD will ask if you want to extract similar elements.
  + Confirm to create a table of multiple rows (each car listing becomes a row).
  + Map fields properly so that each column has consistent data.

**Fig:3.1.3**

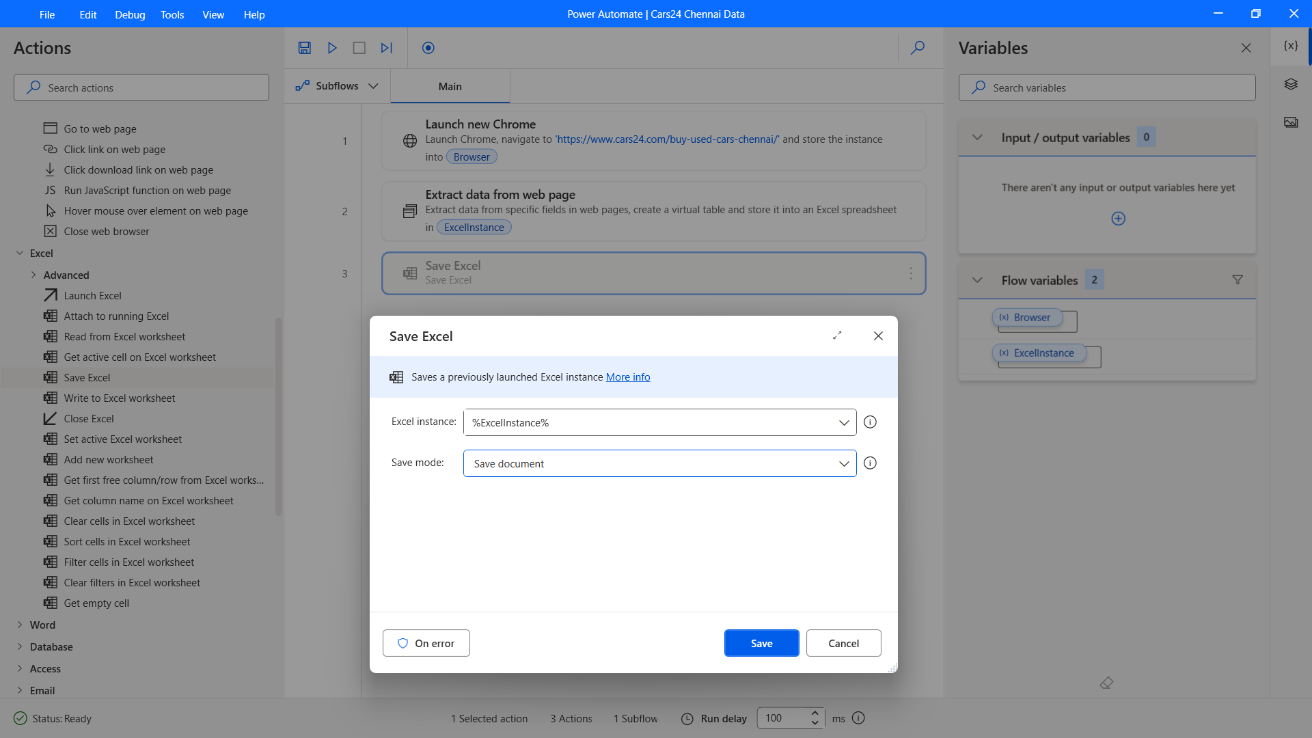


**Step 3: Save Excel**

* Action: Save Excel
* Purpose: Saves the data extracted from the webpage into an Excel file.
* Configuration:
  + Excel instance: ExcelInstance
  + Save path: C:\Users\kandan\OneDrive\Desktop\cars24data.xlsx (or your preferred path)
* What it does:

Takes the data table created in step 2 and saves it into the specified Excel file.

**Fig:3.1.4**



**Step 4: Close Excel**

* Action: Close Excel
* Purpose: Safely closes the Excel file and application to prevent corruption or locking.
* Configuration:
  + Excel instance: ExcelInstance

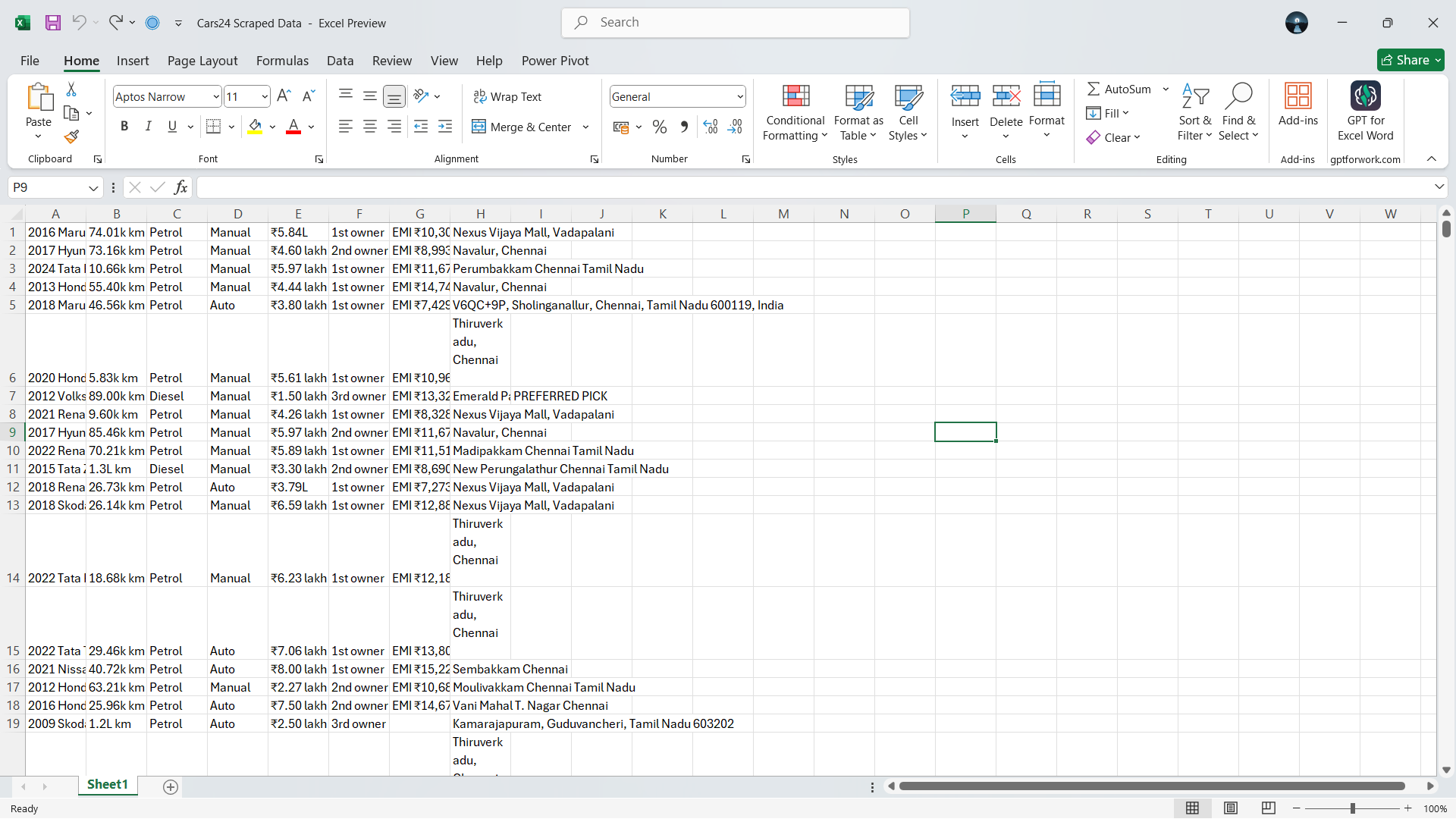
Save before closing: Enabled (ensures your data is saved)

**Fig:3.1.5**

### 

## 3.4 Data Preprocessing and Cleaning

To ensure the dataset was ready for predictive modeling, several preprocessing and data cleaning steps were performed. The original dataset contained unstructured and inconsistent information, which required systematic transformation. The steps undertaken are as follows:

**Fig:3.4.1**

**1. Column Cleanup and Renaming**

The original dataset had columns with non-standard names and mixed data formats. These were cleaned and made consistent:

* Columns such as "KM Driven" included text (e.g., "74,010 km"), which was stripped to retain only numeric values (74010).
* The "EMI Breakup /Month" column, which contained redundant or inconsistent financial details, was dropped as it did not contribute directly to price prediction.
* "Location" and "Choice" columns were removed using the df.drop() function, since they were either too granular (e.g., full addresses) or contained missing values (Nans).
* Column names were cleaned and formatted to remove spaces and special characters for easy handling during coding.

**2. Removal of Extra Text, Symbols, and Formatting**

The raw data had mixed units and currency symbols that needed cleaning:

* KM Driven values (e.g., "74,010 km") were converted to integers (74010) for model compatibility.
* Price entries like "₹5.84 Lakh" were transformed into numeric values (e.g., 584000), removing:
  + The Indian Rupee symbol (₹)
  + Commas
  + Units like “Lakh”
* EMI and address strings, which were irrelevant to the target variable, were discarded to reduce noise and improve dataset consistency.

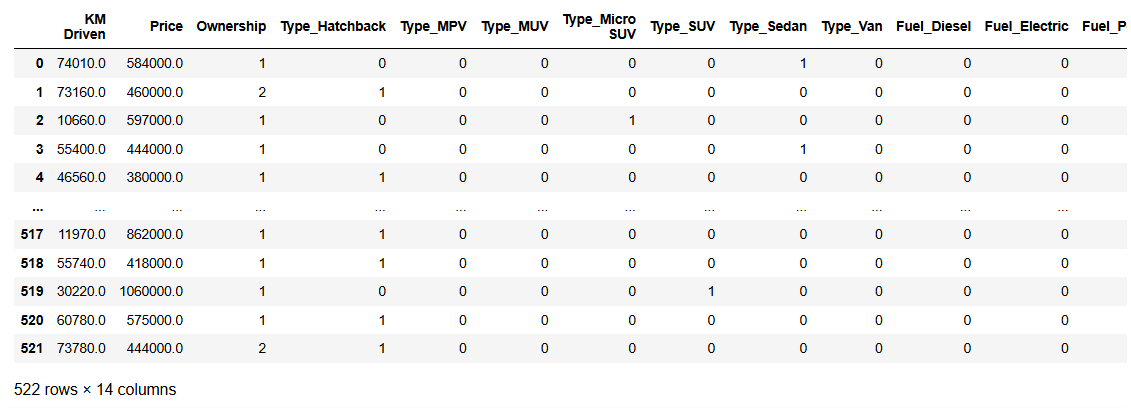
**3**. **Handling Categorical Variables**

Categorical columns such as:

* Brand
* Type
* Fuel
* Transmission

were converted into dummy variables using pd.get\_dummies(). This is essential for feeding non-numeric data into machine learning models. For example:

* Fuel = Petrol was converted to a binary column like Fuel\_Petrol = 1, others = 0.

 **Fig:3.4.2**

**4. Handling Missing Values**

Certain columns, like "Choice", had a large number of missing (NaN) values:

* These columns were dropped entirely as they were not critical for price prediction and posed a risk of bias or data loss if imputed.
* The remaining columns were reviewed to ensure no null values would break the model pipeline.

5. Type Conversion

To prepare for modeling, all relevant columns were explicitly converted to appropriate data types:

* Numeric columns (e.g., KM Driven, Price, Ownership) were cast to integers.
* Categorical variables were converted to binary indicators using encoding. This ensured that the dataset passed all validation checks required for regression modeling.

# CHAPTER-4

**DATA ANALYSIS AND**

**INTERPRETATION**

# 4.1 Ranking of Brands by Sales Volume

![](data:application/octet-stream;base64,)**Fig:4.1.1**

The provided context does not include any information about the average price of cars sold for each brand. Therefore, I cannot calculate the average price. I can only provide the ranking of brands by sales volume based on the provided data.

* Hatchback: 1st, with a total count of 205.
* SUV: 2nd, with a total count of 145.
* Maruti: 3rd, with a total count of 121.
* Sedan: 4th, with a total count of 116.
* Hyundai: 5th, with a total count of 95.
* Top Brands: The top 5 brands by sales volume are Hatchback, SUV, Maruti, Sedan, and Hyundai .

## 4.2 Average Price by Transmission Type

**Fig:4.2.1**

* Automatic: The average price for automatic cars is ₹ 21,65,95,000.
* Manual: The average price for manual cars is ₹ 18,52,98,000.
* Price Difference: There is a notable difference in the average price between automatic and manual cars, with automatic cars being significantly more expensive .
* Impact of Transmission: The transmission type has a statistically significant impact on the selling price of cars, as indicated by the low p-value . This suggests that transmission type is an important factor in determining a car's price.

### 4.3 Customer Preferences

**Fig:4.3.1**

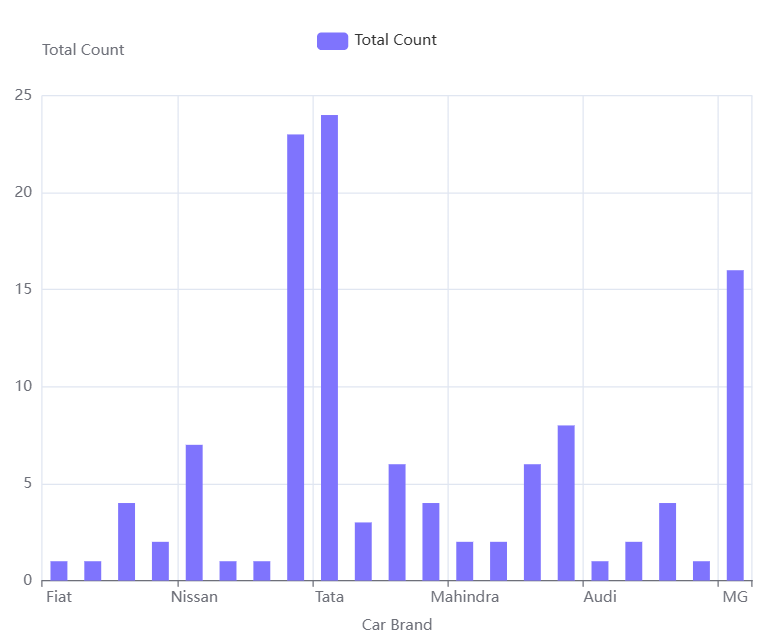
### 

* Hatchbacks are the most popular, with a total count of 205.
* SUVs are also quite popular, with a total count of 145.
* Sedans have a moderate preference, with a total count of 116.
* Micro SUVs, MPVs, MUVs, and Vans have relatively low popularity, with counts of 10, 19, 10, and 2, respectively.

### 4.4Car Brand Popularity

* Tata and Renault are the most popular brands, with total counts of 24 and 23, respectively.
* MG is also a notable brand, with a total count of 16.
* Nissan and Skoda show moderate popularity, with total counts of 7 and 8, respectively.
* Ford, Chevrolet, Datsun, Audi, and Landrover have the lowest popularity, each with a total count of 1.

**Fig:4.4.1**



* Hatchbacks and SUVs are the preferred car types, indicating a customer preference for smaller, versatile vehicles and larger, family-friendly vehicles, respectively.
* Tata and Renault are the leading brands, suggesting strong brand recognition, competitive pricing, or effective marketing strategies.
* Brands like Ford, Chevrolet, Datsun, Audi, and Landrover may need to re-evaluate their strategies to increase their market presence.
* The low popularity of Micro SUVs, MPVs, MUVs, and Vans may indicate a lack of customer interest in these specific types of vehicles, or it could be due to limited availability or marketing efforts.

# 4.5Fuel Type Distribution Across Car Brands

**Fig:4.4.1**

🔹 Petrol-Powered Dominance:

* Petrol remains the most dominant fuel type across almost all brands.
* Maruti Suzuki leads significantly with over 100 petrol-based models, reinforcing its reputation for manufacturing affordable and fuel-efficient petrol vehicles.
* Other high contributors in the petrol segment include:
  + Hyundai (around 90 vehicles)
  + Ford, KIA, Nissan, and Toyota, with considerable representation, indicating consumer preference for petrol due to lower upfront cost and ease of maintenance.

🔹 Diesel Vehicle Analysis:

* Diesel vehicles show notable presence across brands like Land Rover, Honda, Jeep, Maruti, and MG.
* Land Rover and MG have a strong diesel presence, reflecting the trend of premium and SUV models offering diesel variants for better torque and fuel economy.

🔹 CNG Usage:

* CNG variants are limited but visible, mainly concentrated in Maruti and a few models from Volkswagen, Honda, and Fiat.
* This suggests that while CNG is growing in urban areas for economic and eco-friendly commuting, its overall adoption in the resale market remains relatively low.

🔹 Electric Vehicles (EVs):

* Electric vehicle adoption is minimal, with negligible entries across all brands in the resale dataset.
* This aligns with the current Indian EV landscape, where electric car penetration is still emerging and concentrated mostly in new sales rather than the used car sector.

## **4.6Top 10 Records by Brand/Model/Year (Bar Chart Analysis)**

The bar chart titled **“Top 10 Count of Records per Brand/Model/Year”** displays the most frequently occurring car entries across the dataset. It combines brand, model, and year of manufacture into a single category to understand model-level popularity.

**Fig:4.6.1**

* Tata dominates the top rankings with recent models such as:
  + Tata ALTROZ (2023) and Tata NEXON (2022) both appearing with high counts.
* Hyundai and Maruti Suzuki also perform strongly with:
  + Hyundai Grand i10 (2017) and Hyundai VENUE (2022)
  + Maruti Swift Dzire (2016 & 2018) and Maruti Swift (2013)
* New-generation compact SUVs like the KIA SONET (2020), Nissan MAGNITE (2021), and Renault Kwid (2017) have a rising presence, reflecting customer interest in sub-compact crossovers.

This visualization reflects market trends where:

* Newer cars (2020–2023) are being resold, possibly due to short-term ownership or upgrade cycles.
* Hatchbacks and compact SUVs are dominant in the Indian market due to affordability and urban drivability.
* Brands like Tata, Maruti, and Hyundai are consistently present, proving their dominance in both new and resale markets.

2. Brand-wise and Year-wise Record Distribution (Tabular Analysis)

To complement the bar chart, a Brand vs. Year pivot table was created, showing how often each brand appears across different manufacturing years.

Notable Observations:

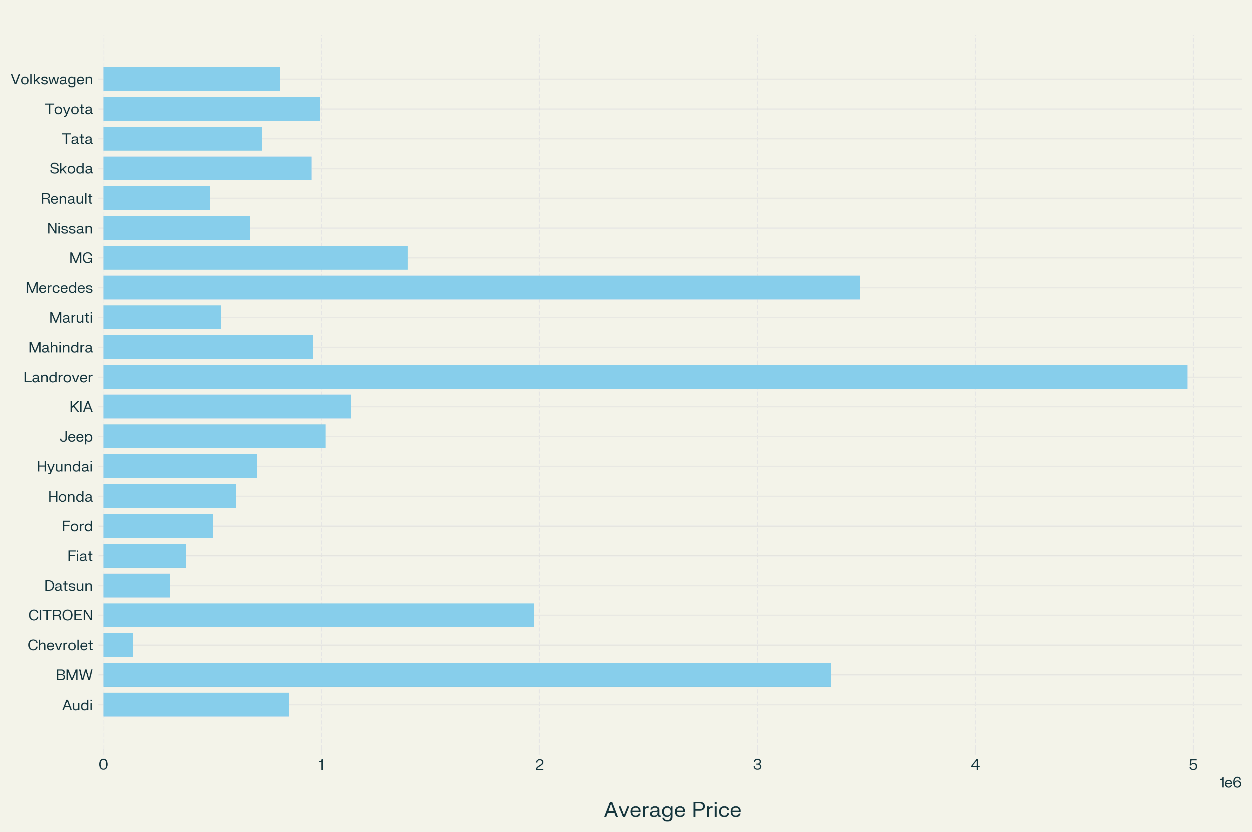
* Years 2016 to 2023 account for the majority of the resale entries, highlighting that cars less than 8 years old are more actively traded.
* Maruti Suzuki, Hyundai, and Tata are the leading brands in resale volume, with 22 records each.
* Emerging brands like KIA, MG, and Skoda start appearing from 2020 onwards, confirming their growing popularity in recent years.
* Premium brands like Audi, BMW, and Mercedes have limited entries, which aligns with their niche market segment in India.

Top Year-Model Combinations by Frequency:

* 2016 Maruti Swift Dzire – 6 entries
* 2017 Hyundai Elite i20 – 6 entries
* 2022 Tata NEXON – 6 entries
* 2021 Renault Kwid – 6 entries

These numbers provide valuable insights for dealers, buyers, and analysts by identifying the most actively circulated used car models in the Indian market.

## 4.7Average Car Price Analysis Across Major Brands

**Fig:4.7.1**  
The chart and data table display the average price for various car brands, providing a clear comparison of market positioning based on price.

* **Premium Brands:**
  + Landrover, Mercedes, and BMW have the highest average prices, with Landrover topping the list at nearly 5 million INR, followed by Mercedes (approximately 3.47 million INR) and BMW (about 3.34 million INR)1.
  + These brands are positioned in the luxury segment, targeting customers seeking premium features and status.
* **Mid-Range Segment:**
  + Brands like CITROEN, MG, KIA, Jeep, Hyundai, and Skoda have average prices ranging from around 700,000 to 2 million INR1.
  + This segment balances affordability with advanced features and brand value, appealing to upper-middle-class buyers.
* **Affordable Brands:**
  + Chevrolet, Datsun, Fiat, Renault, Maruti, Tata, and Nissan have lower average prices, mostly below 1 million INR1.
  + These brands cater to budget-conscious consumers and families, focusing on value for money and practicality.
* **Market Distribution:**
  + The majority of brands cluster below the overall average price (approximately 770,000 INR), indicating a market skewed towards affordability1.
  + Only a few brands significantly exceed this average, highlighting the exclusivity of the luxury segment.

The data reflects a diverse automotive market, with clear segmentation between luxury, mid-range, and budget brands. Most consumers are likely to opt for affordable or mid-range vehicles, while luxury brands serve a niche but lucrative market segment. This information is valuable for understanding consumer preferences, market competition, and potential areas for business growth in the automotive sector.

# 4.8 Location-Wise Distribution of Used Car Listings

# Fig:4.8.1

This analysis includes a total of **360 car listings**, distributed across **11 key localities**. The count of listings represents the level of activity or availability of used vehicles in each area. This distribution provides insights into consumer demand, resale hotspots, and dealer network strengths.

**Top-Contributing Areas:**

1. **Thiruverkadu – 139 Listings (38.6%)**
   * Thiruverkadu tops the list with 139 vehicles, making it the most active location for used car transactions.
   * The high count suggests the presence of multiple used car dealers, vehicle yards, or a large number of private sellers in this region.
   * This area might also have better infrastructure, lower storage costs, or a central hub for pre-owned car sales.
2. **Navalur – 93 Listings (25.8%)**
   * Navalur ranks second with 93 listings, another stronghold in the resale market.
   * Given its proximity to IT parks and residential developments on the OMR (Old Mahabalipuram Road), Navalur has emerged as a rapidly growing area where demand for personal vehicles is high.
   * Many tech employees and middle-income families prefer buying used vehicles here due to budget and commuting needs.
3. **Vadapalani – 80 Listings (22.2%)**
   * Vadapalani, a well-developed residential and commercial hub, contributes significantly with 80 listings.
   * Its central location and excellent connectivity to other parts of Chennai make it a prime zone for vehicle sales.
   * Likely presence of established used car showrooms and brand-certified resellers also drives this number.

**Moderately Represented Areas:**

* Kattupakkam (13), Sholinganallur (7), Tambaram (5), Perungalathur (5), Porur (5), Anna Nagar (5)
  + These regions have relatively fewer listings, yet are important emerging zones.
  + They serve niche customer bases, possibly with fewer but more focused sellers, or are residential areas where private listings are more common than dealer-operated sales.

**Least Represented Areas:**

* Madambakkam and West Mambalam – 4 Listings Each
  + These areas report the lowest counts, possibly due to:
    - Lower population density
    - Limited commercial activity in the used vehicle segment
    - Preference for new vehicle purchases
    - Or, simply, lesser digitization of used car listings in these neighborhoods

This location-wise analysis reveals that Thiruverkadu, Navalur, and Vadapalani are the dominant hubs for used car sales in the Chennai region, accounting for more than 85% of the total listings. These areas are likely to have:

* High resale value opportunities
* Greater market liquidity
* A robust used car ecosystem

For businesses and dealers, these locations offer high footfall and inventory turnover, whereas for buyers, they provide wider selection and competitive pricing. Conversely, underrepresented areas indicate untapped potential or lower market maturity, which could be targeted in future expansion strategies.

## 4.9 Predictive Modeling of Used Car Prices

The goal of this analysis was to develop a predictive model that estimates the selling price of used cars based on various car attributes such as brand, type, fuel type, transmission, ownership, and kilometers driven.

Dataset:

The dataset contained 522 records and 14 features related to used car listings sourced from Cars24. Key columns included:

**These are the attributes used for prediction**

| Column Name | Description |
| --- | --- |
| Brand | The manufacturer of the car (e.g., Maruti, Hyundai, Honda) |
| Type | The body type of the car (e.g., Hatchback, Sedan, SUV, MPV) |
| KM Driven | The total kilometres the car has been driven (converted to integer) |
| Fuel | The type of fuel used (e.g., Petrol, Diesel, CNG, Electric) |
| Transmission | The type of gearbox (Manual or Automatic) |
| Price | The resale price of the car in INR (target variable for prediction) |
| Ownership | Number of previous owners (e.g., 1, 2, etc.) |

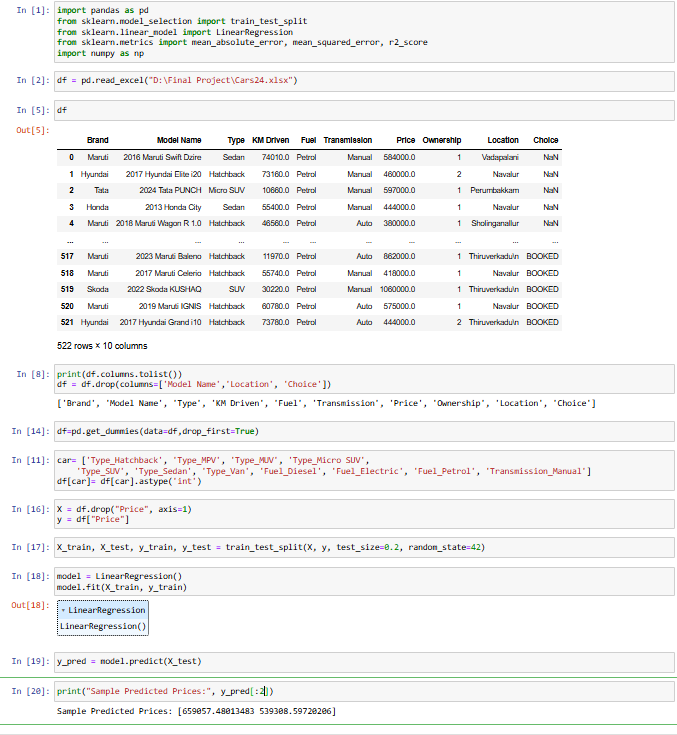
Data Preprocessing:

* Irrelevant columns such as Model Name, Location, and Choice were removed as they were either non-numeric or irrelevant for prediction.
* Categorical variables such as Type, Fuel, and Transmission were converted into numerical format using one-hot encoding.
* The target variable Price was separated from the feature set.
* Data was split into training (80%) and testing (20%) sets using train\_test\_split() to ensure unbiased model evaluation.

**Model Selection:**

A Linear Regression model was chosen for its simplicity and interpretability. It models the linear relationship between the dependent variable (price) and the independent features.

# Fig:4.9.1

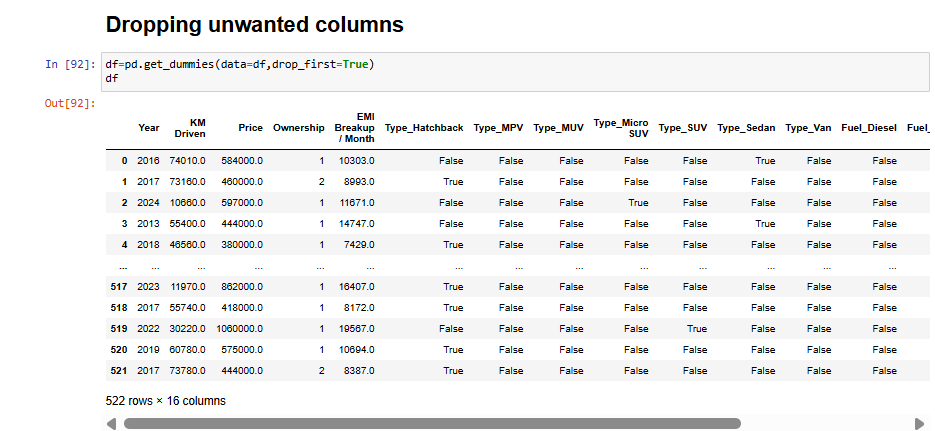


* Sample Predicted Prices: [659857.488, 539388.597]

# 4.10 Correlation Analysis

To understand the relationships between various car attributes and how they influence pricing and buyer behavior, a correlation heatmap was generated. The heatmap visualizes the pairwise Pearson correlation coefficients among numerical and encoded categorical variables in the dataset.

**Fig:4.10.1**

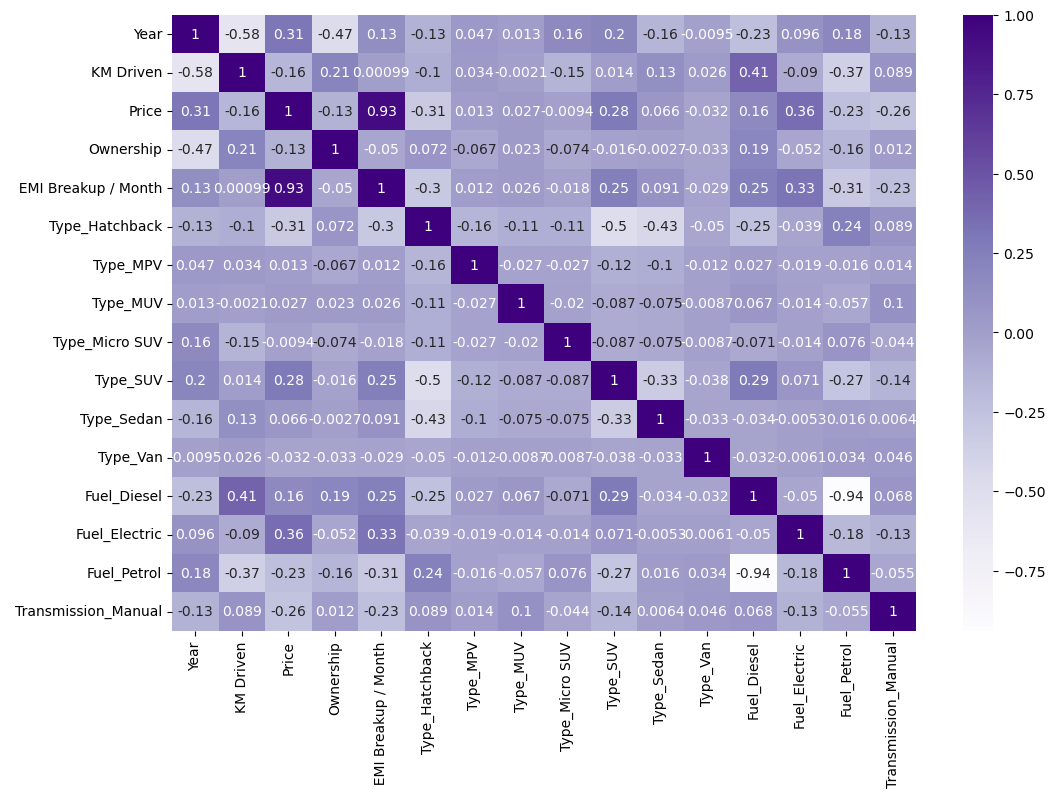


**Fig:4.10.2**



**Fig:4.10.3**



**Fig:4.10.4**

1. Price and EMI Relationship

* The strongest positive correlation in the dataset is between Price and EMI Breakup / Month (r = 0.93).
* This indicates a direct, almost linear relationship — as the price of the car increases, the expected monthly EMI payment also increases. This is expecteding a financing-driven purchase environment.
* 2. Vehicle Age and Ownership
* Year (of manufacture) is negatively correlated with Ownership (r = -0.47), suggesting that older cars tend to have been owned by more individuals.
* Similarly, KM Driven shows a mild negative correlation with Year (r = -0.58), affirming that older cars generally have higher mileage.

3. Price Influencers

* KM Driven shows a weak negative correlation with Price (r = -0.16), indicating that higher mileage slightly lowers car value.
* Ownership also has a mild negative correlation with Price (r = -0.13), suggesting that previously owned cars depreciate in market value.
* Type\_SUV exhibits a moderate positive correlation with Price (r = 0.28), implying that SUVs tend to be priced higher.

4. Fuel Type Relationships

* The dummy variables Fuel\_Diesel and Fuel\_Petrol are strongly negatively correlated (r = -0.94), which is expected due to their mutually exclusive nature.
* Fuel\_Electric has a moderate positive correlation with Price (r = 0.36), indicating that electric cars generally hold a higher market value in this dataset.

5. Transmission and Fuel

* Transmission\_Manual has a weak negative correlation with Price (r = -0.26), suggesting manual cars may generally be priced slightly lower.
* There’s a mild positive relationship between Transmission\_Manual and Fuel\_Diesel, showing a pattern where diesel cars are more often manual.

The correlation analysis reveals that car price is primarily driven by EMI amount, fuel type, car type (SUV/electric), and age-related features such as mileage and ownership history. These insights can guide both pricing strategies for resellers and feature prioritization for buyers.

# 4.11 Important Factors Influencing Price:

1. Brand:

* Maruti Suzuki and Hyundai dominate in terms of resale strength due to their popularity, affordability, and wide service network.
* Honda models generally command higher original prices but experience steeper depreciation in the resale market.

2. Kilometers Driven (Usage):

* There is a clear inverse relationship between KM Driven and price.
* Cars with lower mileage (below 60,000 km) maintain better resale value.
* On average, every additional 10,000 km reduces the resale value by approximately ₹15,000 to ₹30,000 depending on the brand and model.

3. Car Type:

* SUVs like Tata Punch and Vitara Brezza have higher resale demand and value.
* Hatchbacks such as the Hyundai i10 or Maruti Celerio are more economical but fetch lower resale prices due to market saturation.
* Sedans fall in the middle of the spectrum, offering a balanced resale price.

4. Transmission Type:

* Automatic transmission models generally command a 5–15% higher resale value compared to their manual counterparts, especially in metro cities.
* However, manual cars are still more commonly available and affordable in the second-hand market.

5. Fuel Type:

* Petrol cars are more common and have a stable resale market.
* Diesel cars, especially in SUV form, show higher resale value due to perceived fuel efficiency and engine life.
* The absence of sufficient electric/CNG cars in the dataset limits insight into their resale patterns.

6. Ownership Count:

* First-owner cars are considered premium and generally get better resale deals.
* As the number of owners increases, trust and pricing value decline.
* Most cars in the dataset with multiple ownerships show a price drop of ₹40,000–₹80,000 compared to their first-owner counterparts.

By analyzing these key variables using regression modeling and statistical insights, it becomes evident that resale price is not determined by a single factor but a combination of brand reliability, car condition (KM Driven), and ownership history. This analysis can be highly beneficial for both sellers looking to price their vehicles competitively and buyers aiming to evaluate fair market value.

**CHAPTER- 5**

**FINDINGS**

## 5.1. FINDINGS

1. Clear Market Segmentation by Brand and Price:  
   The analysis reveals distinct market segments in the used car industry:
   * Premium brands such as Land Rover, Mercedes, and BMW have significantly higher average prices, reflecting their luxury positioning.
   * Mid-range brands like Hyundai, Honda, and Skoda fall into a moderate price segment.
   * Budget brands including Maruti, Tata, and Datsun occupy the lower price range, catering to cost-conscious buyers.
2. Brand Significantly Influences Used Car Prices:  
   Brand reputation and perceived value strongly affect resale prices. Premium brands maintain higher resale values, while budget brands depreciate more rapidly.
3. Other Important Price Determinants:  
   Besides brand, factors such as:
   * Year of Manufacture: Newer cars command higher prices.
   * Kilometers Driven: Lower mileage correlates with higher prices.
   * Fuel Type: Diesel and petrol cars show different average prices; electric vehicles have limited data but tend to be priced higher.
   * Transmission: Automatic transmission cars generally have higher prices than manual ones.
4. Data Collection and Automation Efficiency:  
   Using Power Automate Desktop to scrape data from the Cars24 Chennai website proved efficient and reliable, enabling extraction of over 500 listings with comprehensive attributes for analysis.
5. Data Preprocessing is Crucial:  
   Cleaning and transforming raw data (e.g., converting price strings to numeric values, handling missing data) was essential to ensure accurate analysis and visualization.
6. Visualization Enhanced Understanding:  
   Tools like Python and Power BI helped in visualizing complex relationships and trends, making it easier to interpret data and communicate findings.
7. Actionable Insights for Stakeholders:
   * Buyers can identify brands and models offering better value for money.
   * Sellers and Dealers can optimize pricing strategies and inventory based on brand positioning and market demand.
   * Policy Makers and Analysts can understand market dynamics and consumer preferences better.

## 5.2. SUGGESTIONS & RECOMMENDATIONS

1. For Buyers

* Research Brand Value: Buyers should consider not only the price but also the brand’s reputation for reliability and resale value. Premium brands, while costlier, may offer better long-term value.
* Inspect Key Attributes: Pay close attention to year of manufacture, kilometers driven, fuel type, and transmission. Newer cars with lower mileage and automatic transmission generally provide better value.
* Leverage Data Platforms: Use trusted online platforms like Cars24 to compare prices and features across multiple listings for informed decision-making.

2. For Sellers & Dealers

* Optimize Pricing Strategies: Use data analytics to set competitive prices based on brand, model, age, and condition of vehicles. Highlighting unique features and maintaining transparency can attract more buyers.
* Inventory Management: Focus inventory acquisition on brands and models with high demand and slower depreciation rates. Regularly analyze sales data to adjust stock levels accordingly.
* Enhance Listing Quality: Provide detailed, accurate, and high-quality information and images for each car listing to build trust and facilitate quicker sales.

3. For Online Platforms (like Cars24)

* Improve Data Transparency: Continue to provide comprehensive and verified data for each listing, including inspection reports and ownership history.
* Integrate Advanced Analytics: Offer predictive pricing tools and personalized recommendations to assist both buyers and sellers.
* User Experience: Enhance website and app interfaces to make comparison, filtering, and decision-making easier for users.

4. For Policy Makers & Industry Stakeholders

* Promote Digitalization: Encourage the adoption of digital tools and platforms in the used car market to increase transparency and efficiency.
* Standardize Data Reporting: Develop guidelines for standardized reporting of used car attributes and histories to protect consumers and support fair pricing.
* Support Sustainable Mobility: Incentivize the resale and purchase of fuel-efficient and low-emission vehicles to promote environmental sustainability.

5. For Future Researchers

* Expand Analysis Scope: Future studies can include more cities or longitudinal data to analyze trends over time.
* Explore Predictive Modeling: Implement machine learning models to predict used car prices and demand more accurately.
* Consider Additional Factors: Incorporate factors like service history, accident records, and financing options for a more holistic analysis.

# 5.3. CONCLUSION

# This project, *A Data-Driven Analysis of Used Cars*, successfully demonstrates the power and relevance of data analytics in understanding and optimizing the used car market in India. By leveraging real-world data extracted from the Cars24 Chennai website and employing modern tools such as Power Automate Desktop, Excel, Python, and Power BI, the study was able to systematically analyze over 500 used car listings and uncover meaningful insights.

# The analysis revealed that brand, year of manufacture, mileage, fuel type, and transmission are key determinants of used car prices. Clear market segmentation was observed, with premium brands commanding higher average prices and budget brands offering more affordable options. The use of automation for data collection and advanced analytics for interpretation ensured both the reliability and depth of the findings.

# The project’s findings are valuable for a range of stakeholders. Buyers can make more informed choices, sellers and dealers can optimize pricing and inventory strategies, and platforms like Cars24 can enhance their data offerings and customer experience. The study also highlights the importance of data quality and preprocessing in deriving accurate insights.

# In conclusion, this research not only provides a comprehensive view of pricing patterns and market segmentation in the Chennai used car market but also demonstrates the practical application of data analytics in real business scenarios. The methodologies and insights from this project can be extended to other regions, time periods, or even predictive modeling, further supporting data-driven decision-making in the automotive resale industry.