Web Scraping - Used Cars Data from Cars24

Name: Iden Abhisheik S I

College: B. S. Abdur Rahman Crescent Institute Of Science And Technology

Department: B.Tech CSE (Cyber Security)

Project Title: Web Scraping

Submission Date: 26/05/2025

1. Objective

The objective of this project is to collect relevant used car data from the Cars24 website, specifically for the Chennai region. The data includes car model, kilometers driven, year of manufacture, fuel type, transmission, and price. This dataset will be used for further analysis using Power BI to derive insights on pricing trends, usage patterns, and popular fuel types.

2. Methodology

1. Used Selenium WebDriver to automate browser interactions and extract car listings from the Cars24 Chennai page.
2. Parsed each car listing to extract details such as car model, year, kilometers driven, fuel type, transmission, and price.
3. Converted the data into structured format and stored it in a CSV file.
4. Loaded the data into Power BI and performed data cleaning (e.g., converting price and kilometers into numeric values).
5. Created visualizations like bar charts to analyze patterns.

3. Technologies Used

• Python
• Selenium WebDriver
• CSV file format
• Power BI
• Visual Studio Code
• ChromeDriver

4. Challenges Faced

• Dynamic loading of content on the Cars24 website required waits for elements to load.

• Frequent changes in class names or structure on the website caused element lookup issues.

• Unicode characters in price (₹ symbol) created formatting problems in CSV.

• Parsing 'km' values (like '1.3L km' and '24.5k km') into numeric formats for analysis.

• Handling missing or incomplete data entries without breaking the scraping process.

5. Data Analysis (Power BI)

Using Power BI, the scraped dataset was analyzed through visualizations to derive insights. The key visualizations include:

1. Average Price by Fuel Type
2. Distribution of Transmission Types
3. Bar Chart: Kilometers Driven vs Price
4. Year of Manufacture vs Price
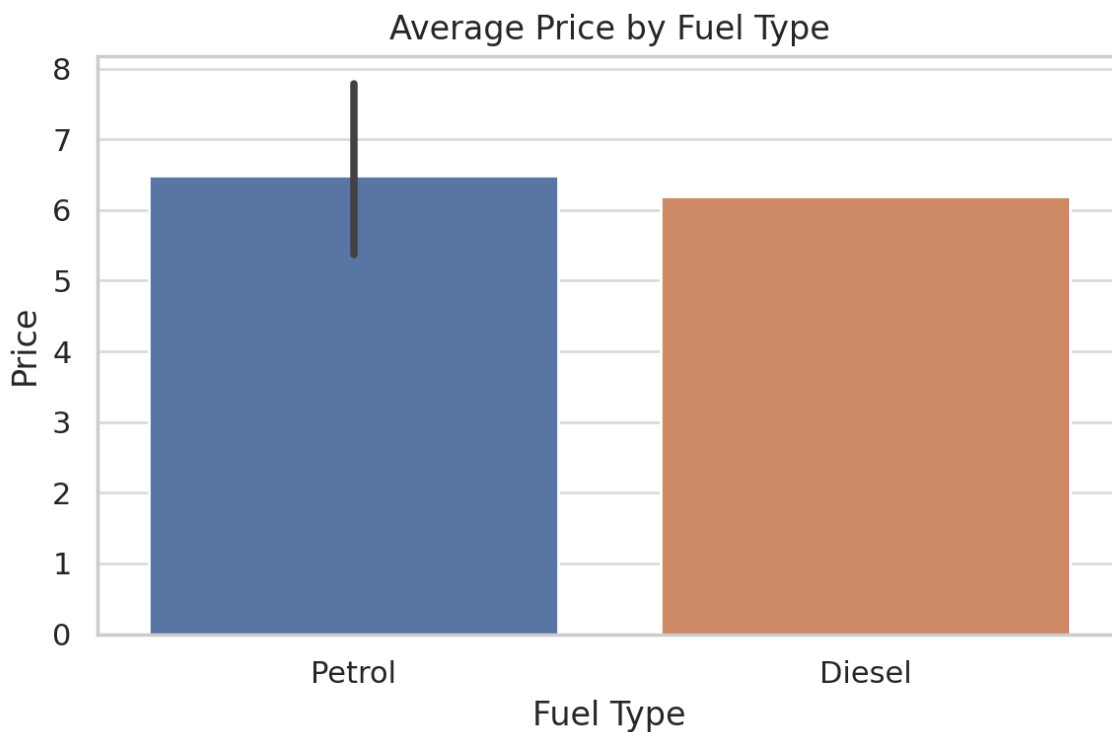5. Top 10 Car Models by Count

Figure 1: **Average Price by Fuel Type**
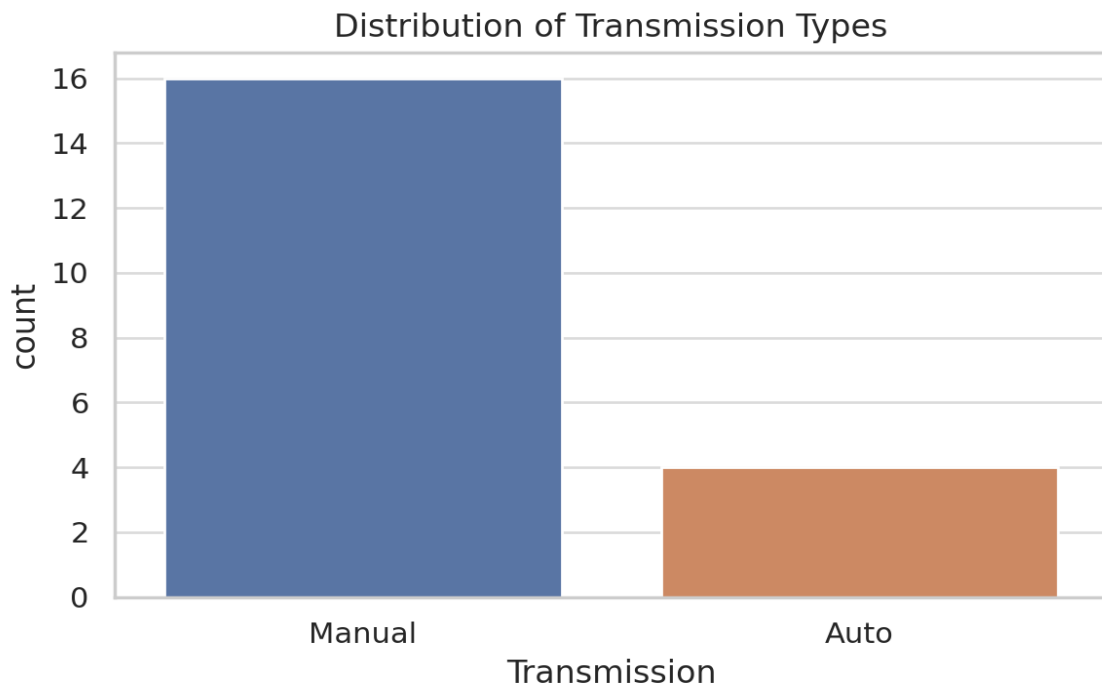
Figure 2: **Distribution of Transmission Types**:

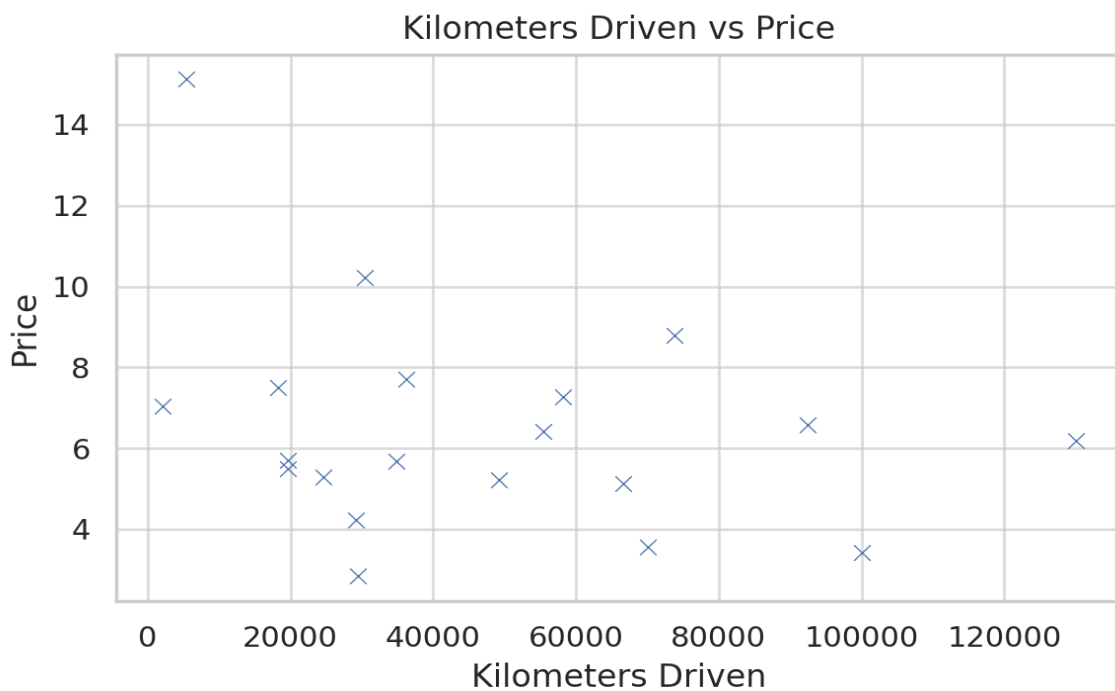Figure 3: **Kilometers Driven vs Price**
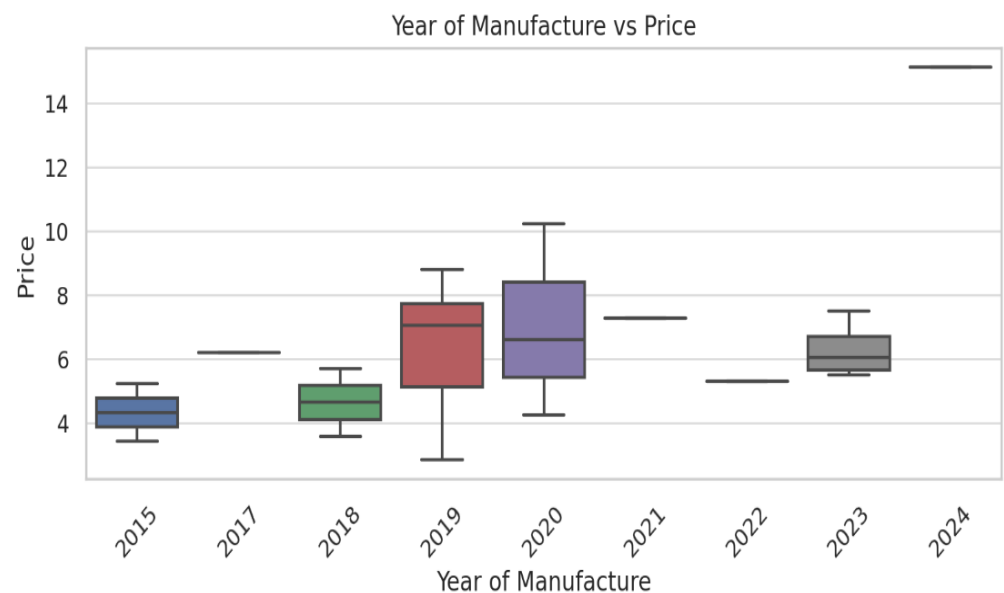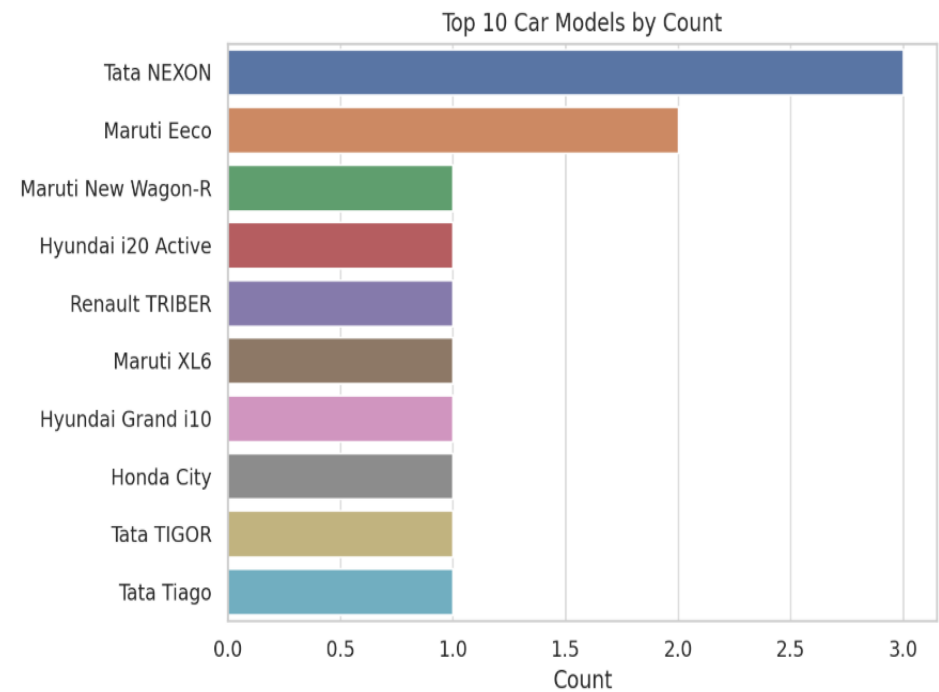
Figure 4: **Year of Manufacture vs Price**

Figure 5: **Top 10 Car Models by Count**

From the above analysis:

 1. Average Price by Fuel Type: Petrol and Diesel cars have a similar average price, with Petrol slightly higher.

2. Distribution of Transmission Types: Manual cars dominate the dataset, while Auto transmissions are less common.

 3. Kilometers Driven vs Price: No strong linear correlation; some low-mileage cars are very expensive.

4. Year of Manufacture vs Price: Newer cars tend to be more expensive, with 2024 models having the highest prices.

5. Top 10 Car Models by Count: Tata NEXON and Maruti Eeco appear most frequently.

6.Conclusion

The objective of this project was to automate the extraction of relevant car details—such as kilometers driven, year of manufacture, fuel type, transmission type, and price—from the Cars24 website using Python and Selenium. The project successfully demonstrated how web scraping can be used to gather large amounts of structured data from dynamic web pages.

Through this process, we were able to collect real-world data for over 20+ used cars listed in Chennai. This data was then cleaned, formatted, and stored in a CSV file, making it ready for visualization and analysis using Power BI.

The project not only enhanced our technical proficiency in web scraping but also improved our understanding of handling challenges such as dynamic content, missing data, inconsistent formatting (like km and price values), and encoding issues. Furthermore, the data visualization part of the project enabled meaningful insights, like identifying the most common fuel types, transmission preferences, and how mileage affects car pricing.

Overall, this project serves as a practical example of how data scraping and analysis can be used to support decision-making in real-world scenarios like used vehicle markets.