Mini project-1 Final Report -MG Cars Data Analysis(Cars24)

Project Title:

Web Scraping and Exploratory Data Analysis of MG Cars from Cars24.com

Submitted by:

Team L

Aditya Sharma(Team Leader)

Aryan Singh (Co-Lead)

Arshiya Firdous

Ojaswi B

Sahil Patil

Ritesh Kumar

Vivek

1.Problem Statement

The objective of this min-project is to:

a)Extract structured data on MG(Morris Garages) brand cars listed on Cars24 Mumbai using web scraping.

b)Perform a detailed Exploratory Data Analysis (EDA) to uncover patterns in pricing, mileage, manufacturing year,

  ownership, fuel types, and more.

c)Create a Power BI dashboard to visualize the dataset for better decision-making and market insights.

2.Tools and Technologies Used

a)Python for scripting and scraping

  Libraries: requests, BeautifulSoup, pandas, matplotlib, seaborn

b)Power BI for dashboard creation

c)Jupyter Notebook for EDA and Data Cleaning

3.Data Collection(Web Scraping)

Target Website: https://www.cars24.com

Brand Focus: MG Cars

Location: Mumbai

Fields Scraped:

Car Title (e.g., 2020 MG Hector Plus)

Variant

Kilometers Driven

Fuel Type

Transmission

Owner Count

Price

Manufacture Year (extracted from title)

Location

4. Data Cleaning and Preprocessing

Performed in Python:

Converted Kilometers Driven (e.g., "29.18k km") to integers (e.g., 29180)

Extracted Manufacture Year from car titles

Removed duplicates and cleaned whitespace

Converted Price column to numeric (₹ stripping, removing commas)

Handled missing values appropriately

5. Exploratory Data Analysis (Key Insights)

Price vs. Manufacture Year

Older models (e.g., 2019, 2020) show a predictable price depreciation

2023 models command premium prices

Fuel Type Distribution

Petrol is the most common

Diesel is present in fewer high-end variants

Transmission Type Analysis

Manual transmission cars are priced lower on average

Automatic variants (DCT/CVT) are costlier and newer

Kilometers Driven vs. Price

As expected, cars with lower mileage are listed at higher prices

However, some high-mileage cars maintain value due to variant or model year

Owner Count Insights

1st owner cars form majority

2nd/3rd owner cars show significant price drop

6. Power BI Dashboard

Created a dynamic dashboard showing:

Filters: Manufacture Year, Fuel Type, Transmission

Visuals:

Bar chart for count of cars per year

Box plot for Price distribution by Ownership

Line chart for Price trends over years

Scatter plot for Kilometers vs. Price

7.Key Findings

Most listed model: MG Hector 2020–2021

Average price: ₹12–14 Lakhs

Most common configuration: Petrol, Manual, 1st Owner

Higher resale value: 2023 Sharp variants with low kilometers

Least value retention: 2nd-owner diesel manual models from 2019

8.Future Scope

Extend scraping to other cities for comparison

Add image scraping and car feature tags (sunroof, camera, etc.)

Build a price prediction ML model

Integrate scraping → EDA → Dashboard into an automated pipeline

9.Conclusion

This project provides a full-cycle implementation from data extraction to dashboarding, with actionable insights on the MG used car market in Mumbai. It enhances hands-on skills in Python, EDA, web scraping, and visualization, and demonstrates real-world application of data analytics techniques.

The Team and There Work as well as the challenges:

We Started with a meeting and divided our work accordingly to the experience. The team was divided into 5 teams.

Team 1

Task: Role of identifying the relevant HTML tags and class names that contain the necessary information for further processing.

Members: Aryan Singh

   "To fulfill this task, I inspected the Cars24 webpage for Morris Garages (MG) cars using the browser's developer tools. Through this inspection, I located and documented all the important HTML elements (tags and classes) associated with the required data points.

```
Field                        HTML Tag        Class/Identifier/Location

-------------------------- | ----------- |
-------------------------------------------------------------------
Car Make                   | <span>       | sc-braxZu kjFjan

Car Model                  | <span>       | sc-braxZu kjFjan

Price                      | <p>          | sc-braxZu cyPhJl

Fuel Type                  | <p>          | sc-braxZu kvfdZL (2nd in list inside
`<ul class="sc-gLXSEc jkeyxT">`)
Transmission Type          | <p>          | sc-braxZu kvfdZL (3rd in list inside
`<ul>`)
Available in Color Variants | Not found   | Not present in provided HTML

Location                   | <p>          | sc-braxZu lmmumg
```


Team 2

Task: Web Scraping

Members: Aditya Sharma

        Arshiya

   "As the data available on the Cars24 was huge we were having a little bit problem but as the team1 already provived as with the tags it was much helpful for us to scrap the data moreover firstly we tried it by using selenium but the errors were much so we decided it to do with BeautifulSoap and we successfully completed our task without any delay and store our dataset in a CSV file and our data was ready for EDA"

Manufacture Year	Title	Variant	Kilometers Driven	Fuel Type	Transmission	Owners	Price	Location
2023	2023 MG HECTOR PLUS	SHARP PLUS 2.0 D TURBO MT 6STR	29.18k km	Diesel	Manual	1st owner	₹16.90 lakh	Regency Anantam, Dombivli East
2019	2019 MG HECTOR	SHARP 1.5 DCT PETROL	74.19k km	Petrol	Auto	1st owner	₹10.35 lakh	Korum Mall, Thane West
2020	2020 MG HECTOR PLUS	SHARP 1.5 PETROL TURBO DCT 6-STR	32.71k km	Petrol	Auto	2nd owner	₹11.90 lakh	Mulund West, Mumbai
2022	2022 MG HECTOR	SHARP 2.0 DIESEL	22.22k km	Diesel	Manual	1st owner	₹14.40 lakh	Goregaon, Mumbai

Team 3

Task: Cleaning the data and perform EDA

Members: Ritesh Kumar

"I carried out EDA and Preprocessing on a dataset. The key challenge was that several important columns, like Kilometers Driven, Price, and Owners, contained textual data with units, symbols, or inconsistent formats. To address this, I cleaned and transformed these columns into usable numeric formats using string manipulation and regular expressions. After cleaning the data, I visualized trends using seaborn and matplotlib to analyze price distributions, manufacturing years, and fuel types"

Aditya Sharma

"Just reviewed the EDA code and just add few more insights and graphs"

Data before:

	Manufacture Year	Title	Variant	Kilometers Driven	Fuel Type	Transmission	Owners	Price	Location
0	2023	2023 MG HECTOR PLUS	SHARP PLUS 2.0 D TURBO MT 6STR	29.18k km	Diesel	Manual	1st owner	₹16.90 lakh	Regency Anantam, Dombivli East
1	2019	2019 MG HECTOR	SHARP 1.5 DCT PETROL	74.19k km	Petrol	Auto	1st owner	₹10.35 lakh	Korum Mall, Thane West
2	2020	2020 MG HECTOR PLUS	SHARP 1.5 PETROL TURBO DCT 6-STR	32.71k km	Petrol	Auto	2nd owner	₹11.90 lakh	Mulund West, Mumbai
3	2022	2022 MG HECTOR	SHARP 2.0 DIESEL	22.22k km	Diesel	Manual	1st owner	₹14.40 lakh	Goregaon, Mumbai

Data after:

Cleaned Data Sample:

	Manufacture Year	Title	Variant	Kilometers Driven	Fuel Type	Transmission	Owners	Price	Location
0	2023	2023 MG HECTOR PLUS	SHARP PLUS 2.0 D TURBO MT 6STR	2918000	Diesel	Manual	1	1689999	Regency Anantam, Dombivli East
1	2019	2019 MG HECTOR	SHARP 1.5 DCT PETROL	7419000	Petrol	Auto	1	1035000	Korum Mall, Thane West
2	2020	2020 MG HECTOR PLUS	SHARP 1.5 PETROL TURBO DCT 6-STR	3271000	Petrol	Auto	2	1190000	Mulund West, Mumbai
3	2022	2022 MG HECTOR	SHARP 2.0 DIESEL	2222000	Diesel	Manual	1	1440000	Goregaon, Mumbai

Team 4

Task: Creating The Dashboard

Members: Sahil Patil

"I visualized the data by creating a clean, KPI-focused dashboard highlighting total listings, average price, and key trends like fuel type, transmission, and price variations. The biggest challenge was the small dataset, which could distort visuals. To overcome this, I adjusted scales, used clear labels, and focused on delivering meaningful insights visually."
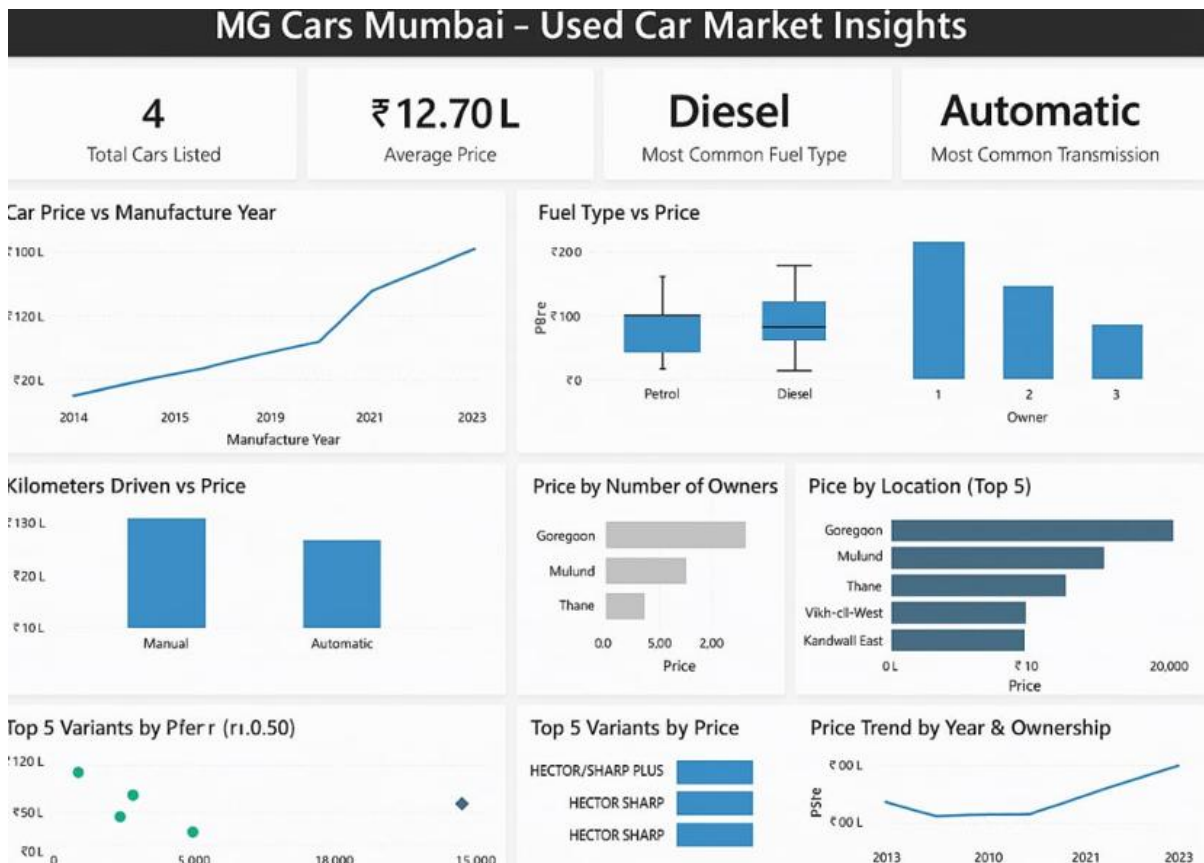
Team 5

Task: Creating the presentation

Members: Ojaswi B

" worked on putting together the final presentation for our data science project. I focused on understanding the dataset, organizing our insights, and explaining things like pricing trends, fuel type impact, and ownership patterns. I structured the PPT to cover our task, team roles, process, challenges, and key takeaways. The main challenge was making sure all the data points made sense to someone seeing them for the first time. So I kept the layout clean and focused on clarity."

MG Cars Resale Analysis – Cars24 Mumbai

- By Team L
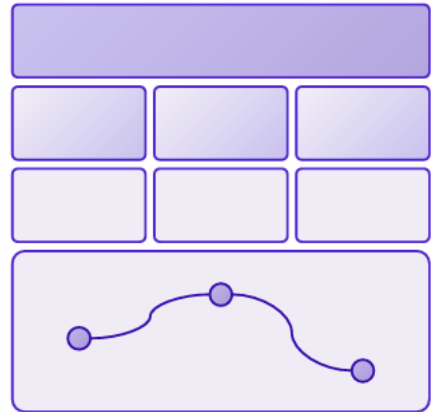
Report: All Shared their experience and collectively help in creating the report.