



EVOASTRA VENTURES

EVOASTRA VENTURES

DIGITAL SOLUTIONS

A presentation by

Team M



Our Goals

Develop skills in web scraping by extracting and analyzing car details from [Cars24.com](#).



Scrape website

Successfully perform web scraping for cars24.com to scrape old car data.



Extract key details

Cover key details such as kilometers driven, year of manufacture, fuel type, transmission, and price.



Bundle data

Ensure the scraped data is clean and collect it in a structured format of a CSV file.



Analyze data

Process and organize data and perform exploratory analysis.



Methodology of the Project



Focus on Used MG Hector vehicle listings on Cars24 Website.

Remove duplicates, handle missing values, standardize formats, and transform raw HTML data into structured format.

Perform descriptive statistics, create visualizations (graphs, charts), and identify patterns and trends.

Summarize key findings, provide recommendations, and present data using visual aids.



Data Acquisition

- **Sources of Data:**

The primary source of data is the Cars24 website, specifically targeting the Buy Used Cars vehicle listings.

- **Brief Description of the Data:**

Vehicle details include vehicle name, model year, price, kilometers driven, fuel type, transmission type, and owner history.

- **Screenshot of the Raw Data:**

HTML Structure: A screenshot showing the HTML structure of a vehicle listing on the Cars24 website, highlighting key data points.

Extracted Data: A screenshot of the raw data in its initial extracted form, before cleaning and preparation.

The screenshot shows the Cars24 website interface for buying used cars. The search filters on the left indicate a search for MG vehicles (60 results) and a budget range of ₹1,00,000 to ₹30,00,000. Three vehicle listings are displayed in a grid:

- 2021 MG HECTOR PLUS SUPER 2.0 DIESEL TURBO MT 7-STR**
23,126 KM • DIESEL • MANUAL
SPL. REG. NO.
₹15.49 Lakh ₹17.40 Lakh (1.99L off)
EMIs from ₹29,484/month
Free Test Drive Today at Kakkadan, Kochi
- 2019 MG HECTOR SHARP 2.0 DIESEL**
49,559 KM • DIESEL • MANUAL
TOP MODEL
₹14.47 Lakh ₹14.91 Lakh (44k off)
EMIs from ₹27,542/month
Free Test Drive Today at Garuda Mall, Magrath...
- 2019 MG HECTOR SHARP 2.0 DIESEL**
33,306 KM • DIESEL • MANUAL
TOP MODEL
₹12.86 Lakh ₹13.50 Lakh (64k off)
EMIs from ₹27,961/month
Free Test Drive Today at Sector-18, Noida

The screenshot shows the raw HTML code extracted from the vehicle listing page. The code includes the DOCTYPE declaration, HTML head, and body content. Key parts of the body content include the vehicle's name, model year, price, and transmission type, which are highlighted in the original image.

```
<!DOCTYPE html>
<html lang="en-IN">
<head>
<link href="https://assets.cars24.com" rel="preconnect">
<link href="https://fastly-production.24c.in" rel="preconnect">
<link href="https://connect.facebook.net" rel="preconnect">
<link href="https://www.googletagmanager.com" rel="preconnect">
<link href="https://www.google-analytics.com" rel="preconnect">
<link href="https://analytics.twitter.com" rel="preconnect">
<link href="https://www.facebook.com" rel="preconnect">
<link href="https://www.google.com" rel="preconnect">
<link href="https://www.google.co.in" rel="preconnect">
<link href="https://googleads.g.doubleclick.net" rel="preconnect">
<link href="https://t.co" rel="preconnect">
<link href="https://static.ads-twitter.com" rel="preconnect">
<link href="https://tracking.cars24.com" rel="preconnect">
<link href="https://www.googleoptimize.com" rel="preconnect">
<link href="https://web-sdk.smartlook.com" rel="preconnect"/><meta charset="utf-8"/><meta content="width=device-width, initial-scale=1.0, maximum-scale=5.0, user-scalable=yes" name="viewport"><meta content="697234207046254" property="fb:app_id"/><!-- scripts from helmet --><link as="image" data-rh="true" fetchpriority="high" href="https://assets.cars24.com/production/c2b-website/240704151227/js/ad88a0b1c8ch2a335acf55a1ff27ef32.png" rel="preload" type="image/png"><link data-rh="true" href="https://www.cars24.com/buy-used-car/" rel="canonical"><script data-rh="true" type="text/javascript">(function(w,d,s,l,t){w[l]=w[l]||[];w[l].push({'gtm.start': new Date().getTime(),event:'gtm.js'});var f=d.getElementsByTagName('script')[0],j=d.createElement('script'),dl=l='dataLayer'?l+'!'+l:'';j.async=true;j.src='//www.googletagmanager.com/gtm.js?id='+idl:j.parentNode.insertBefore(j,f);})(window,document,'script','dataLayer','GTM-P3C9RJ');
```



Data Analysis

Summary statistics:

- Techniques Used for Analysis:
 - Pandas for data manipulation.
 - NumPy for numerical operations.
- Average Distance Travelled: **35535 KM**
- Average Price: **14.65 LACK**

• **Minimum Value_count**

Cars MG HECTOR SMART HYBRID 1.5 PETROL
Kilometers 3306
Year 2022
Fuel Type Diesel
Transmission Manual
Price(Lakh) 12.86
dtype: object

• **Maximum Value_count**

Cars MG HECTOR SHARP 1.5 DCT PETROL
Kilometers 127904
Year 2019
Fuel Type Petrol
Transmission Automatic
Price(Lakh) 18.08
dtype: object



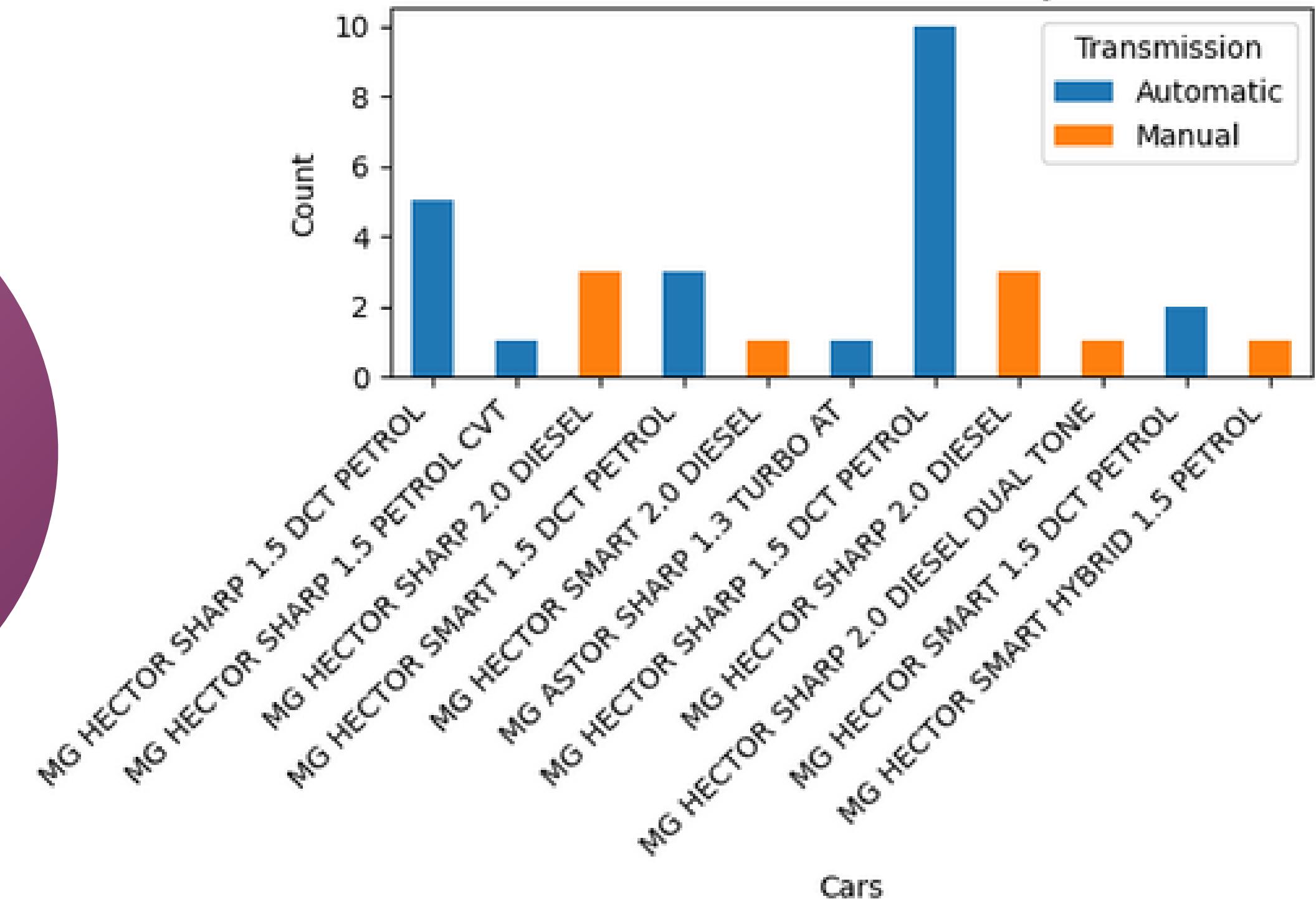
Data Analysis

Key Findings and Insights

Back to Agenda

TRANSMISSION DISTRIBUTION BY CAR:

Transmission Distribution by Car



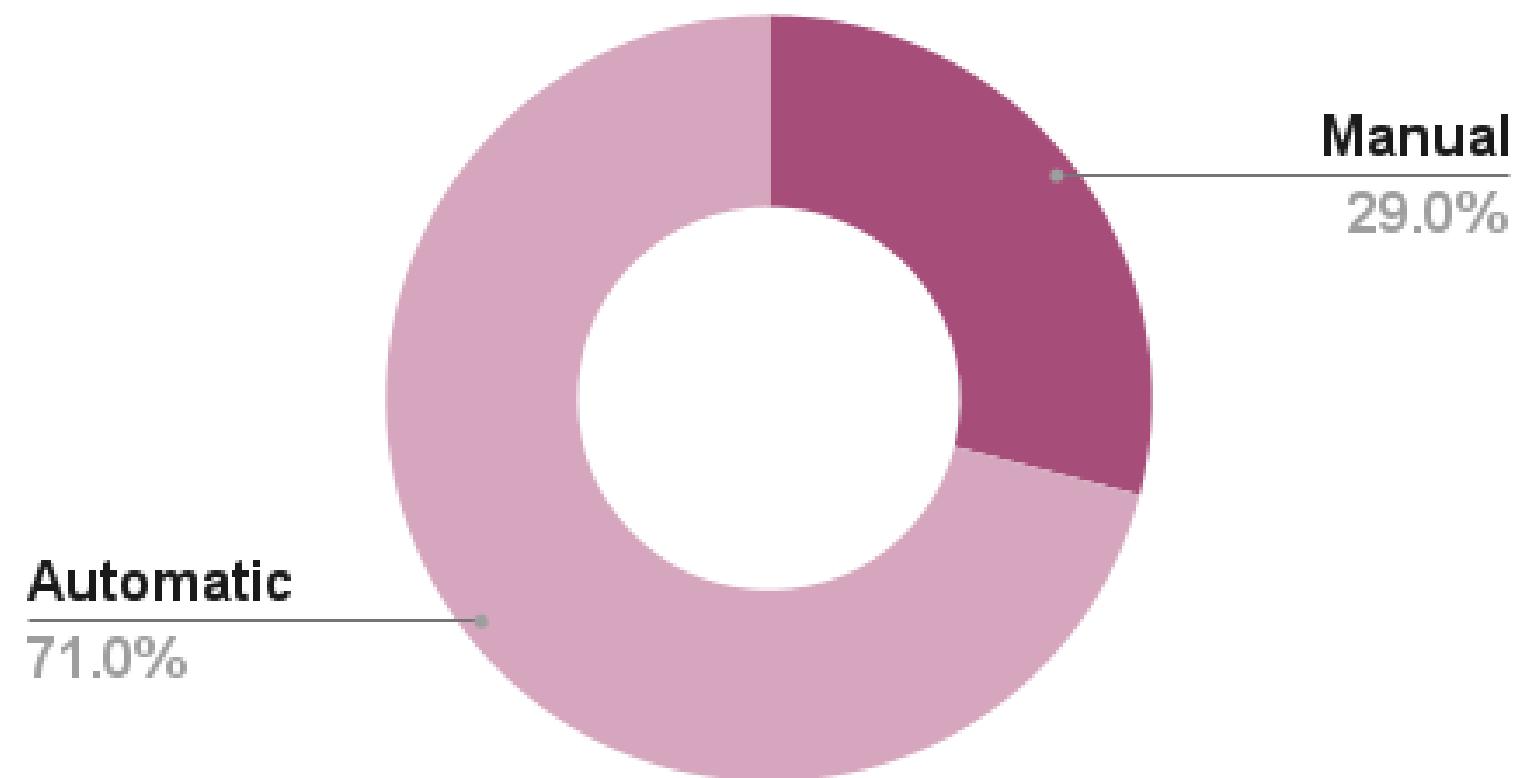
- MG HECTOR SHARP 1.5 PETROL HAS MAXIMUM NO OF AUTOMATIC CARS (10 CARS)



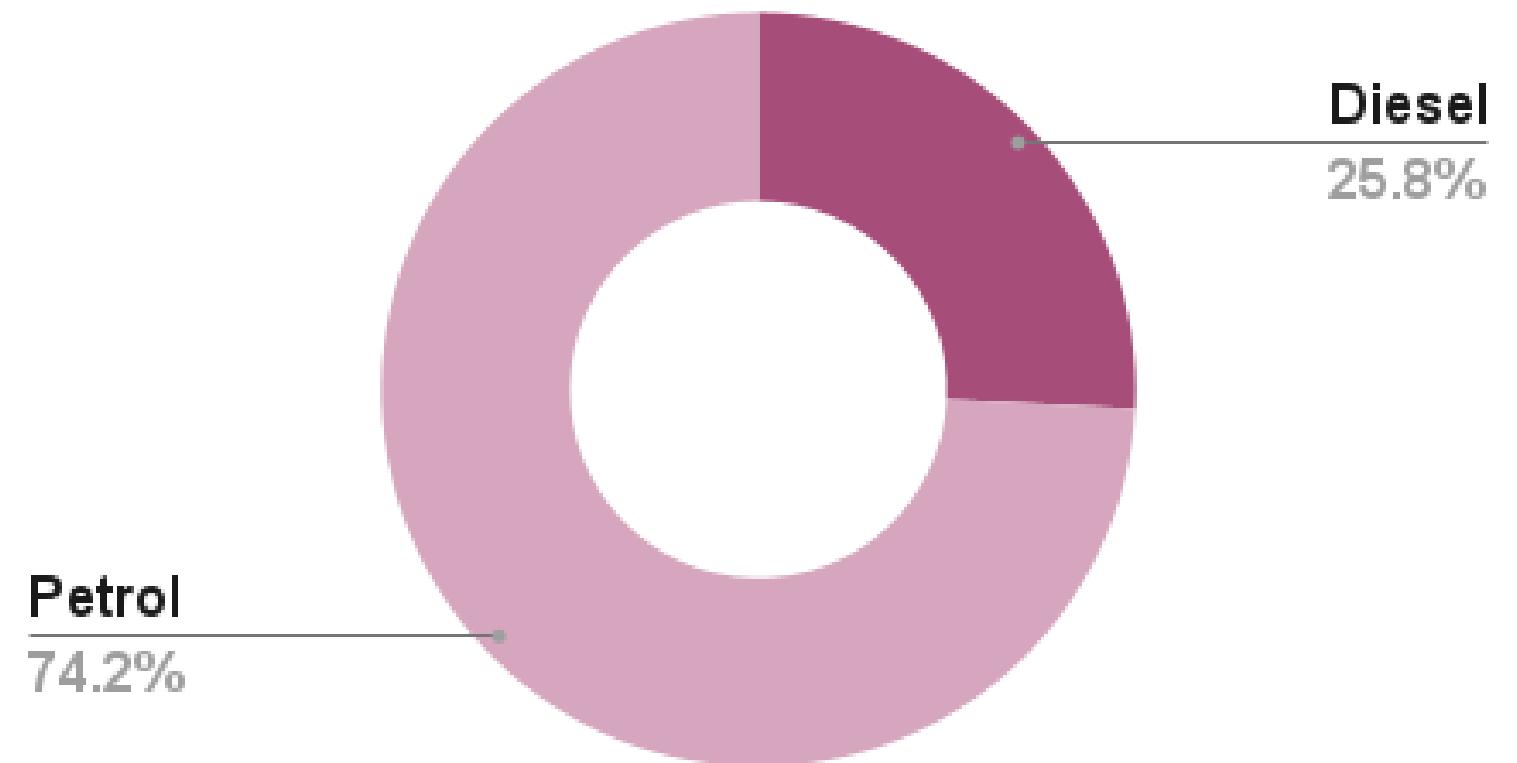
Data Analysis

Analysis Results

Transmission Ratio of MG Hector cars



Fuel Type



- OUT OF ALL MG HECTOR CARS THERE ARE 71% AUTOMATIC CARS AND 29% MANUAL CARS

- OUT OF ALL MG HECTOR CARS 74.2% CARS USE FUEL TYPE PETROL WHILE 25.8% USE DIESEL.



Our Experience and Learnings



Experience

- **Experience:** Improved skills in web scraping techniques using libraries like requests and BeautifulSoup.
- **Data Handling:** Gained experience in handling and parsing HTML data for specific information extraction.



Learnings

- **Libraries:** How to use requests for making HTTP requests and BeautifulSoup for parsing HTML.
- **Web Scraping:** Techniques for navigating and extracting data from web pages.
- **Data Cleaning:** Basic concepts in cleaning and structuring the scraped data for analysis.



Challenges Faced

- ◆ Handling Complex HTML Structures
- ◆ Dealing with rate limiting and potential IP blocking.
- ◆ Ensuring the accuracy and completeness of the scraped data.
- ◆ Combining Datasets from Multiple Locations into One Dataset

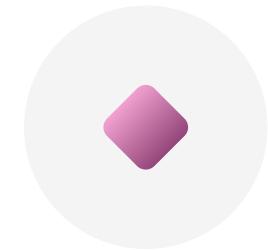
[Back to Agenda](#)



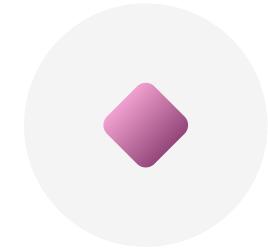
Conclusion



Successfully scraped and analyzed data from Cars24.



Derived meaningful insights about MG Hector vehicle listings.



Identified key challenges and learned valuable techniques for future projects.



THANK YOU

*Huge Thanks to Aniket Manwatkar Sir,
Team M is Greatful to work with
EvoAstra Ventures.*

