Web Scraping: Extracting TATA Cars Data's from Cars24.com

Team 'E'

Our Team Members

❑ **Web Scraping & Data Cleaning -**
1. Divay Galani (Team Lead)
2. Suman Das (Co-lead)
3. Mangaiahgari Sri Nikitha

❑ **Data Visualization –**

1. Mridula

2. Mangaiahgari Sri Nikitha

❑ **Presentation Making –**
1. Suman Das
2. Divay Galani

❑ **Project Report Making –**
1. Divay Galani

Methodology and Data Sources

Definition

Web scraping involves the automated process of extracting data from websites. It uses tools and techniques to retrieve information and convert it into a structured format for analysis.

Aim

The Moto of this project is to web scrap the car data from website 'cars24.com'. Our Team 'E' was assigned to scrap the details of 3 branded used cars such as Mahindra, Renault and Jeep in multiple locations.

The details we needed to extract are:
1. Kilometers Driven
2. Year of Manufacture
3. Fuel Type
4. Transmission
5. Price

Tools Used

Common web scraping tools we used are Python libraries like requests ,Selenium and BeautifulSoup.

These libraries provide functions for fetching web content, parsing HTML structures, and extracting data.

Project Setup and Environment

1 — Install Libraries

The project begins by setting up the development environment. Essential Python libraries, including requests, BeautifulSoup, Selenium and pandas, are installed. These libraries provide the necessary tools for web scraping, HTML parsing, and data manipulation.

2 — Website Inspection

Thorough inspection of the Cars24 website is crucial. This involves identifying HTML tags and classes associated with the targeted data. This step ensures that the scraper can accurately extract the desired information.

3 — Scraper Development

A Python script is crafted to fetch and parse the HTML content of the Cars24 website. This script leverages BeautifulSoup to extract car details based on the identified tags and classes.

4 — Data Organization and Export

The extracted data is organized into a structured format, such as a Pandas DataFrame. The data is then exported to a CSV file for further analysis and visualization.

CHALLENGES DURING THE PROJECT

- Dynamic class names for car details caused inconsistencies in data extraction.
- Incomplete data retrieval: not all listed cars were captured.
- HTML inspection revealed missing brand entries despite their presence on the webpage.
- Price tags were present in HTML but failed to yield values during scraping.
- A key part of ethical web scraping is respecting the target website's robots.txt file. The robots.txt file for cars24.com was analyzed. It contains the following lines: User-agent: * ... Disallow: /tata

Visualizations:

We extracted details of cars for 3 locations. Based on which we visualized the following data.

It is observed that the newer cars (2023–2024) hold the highest resale value, with prices nearing or crossing 10 Lakh, whereas older models before 2018 drop significantly below 5 Lakh — highlighting strong price appreciation for recent manufacturing years.

👉

Average Price by Year of Manufacture


Average Mileage by Fuel Type

👈
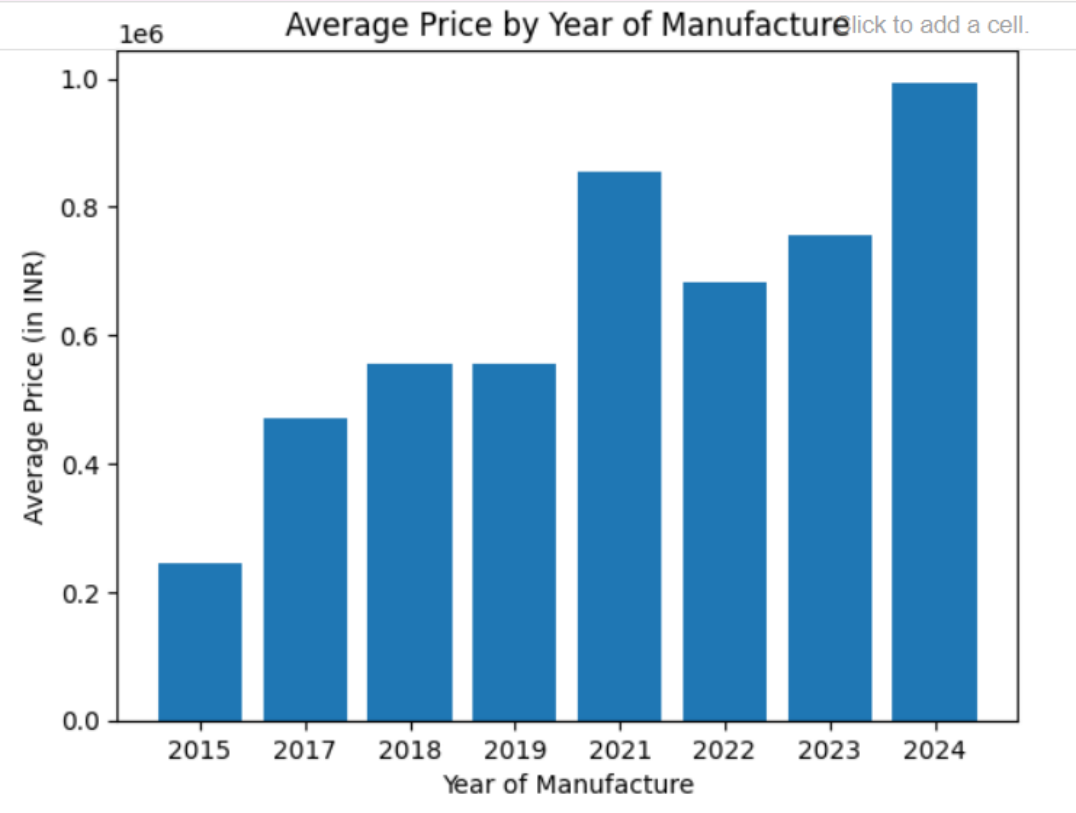
Diesel cars deliver noticeably higher mileage than petrol cars, making them more fuel-efficient for longer drives.
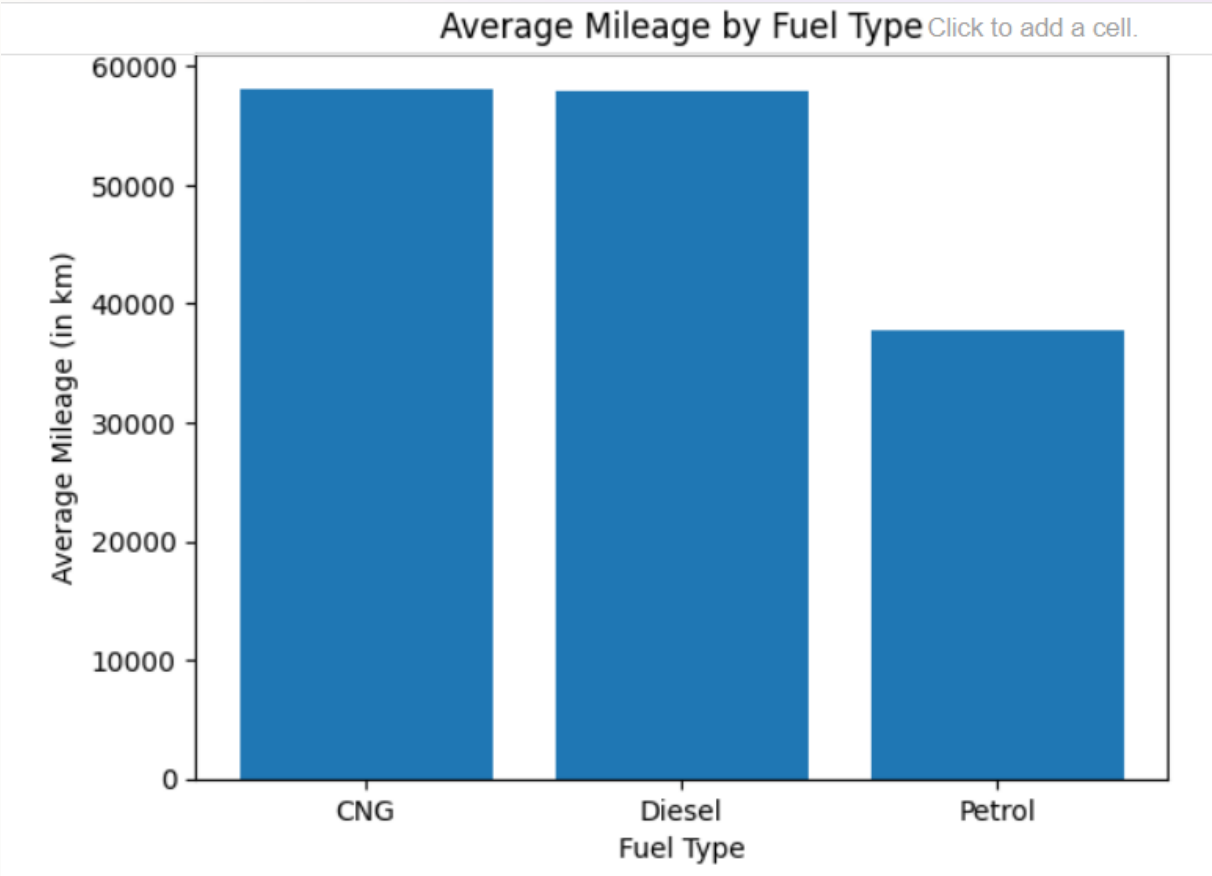
CORRELATION

The correlation matrix reveals that:
- Newer cars have significantly higher prices — strong positive correlation between Year and Price (0.66).
- Higher mileage reduces resale value — Price vs Mileage shows a negative correlation (–0.24).
- Newer cars are driven less — Year vs Mileage shows moderate negative correlation (–0.34).
- Year of manufacture is the strongest price influencer among all selected factors.
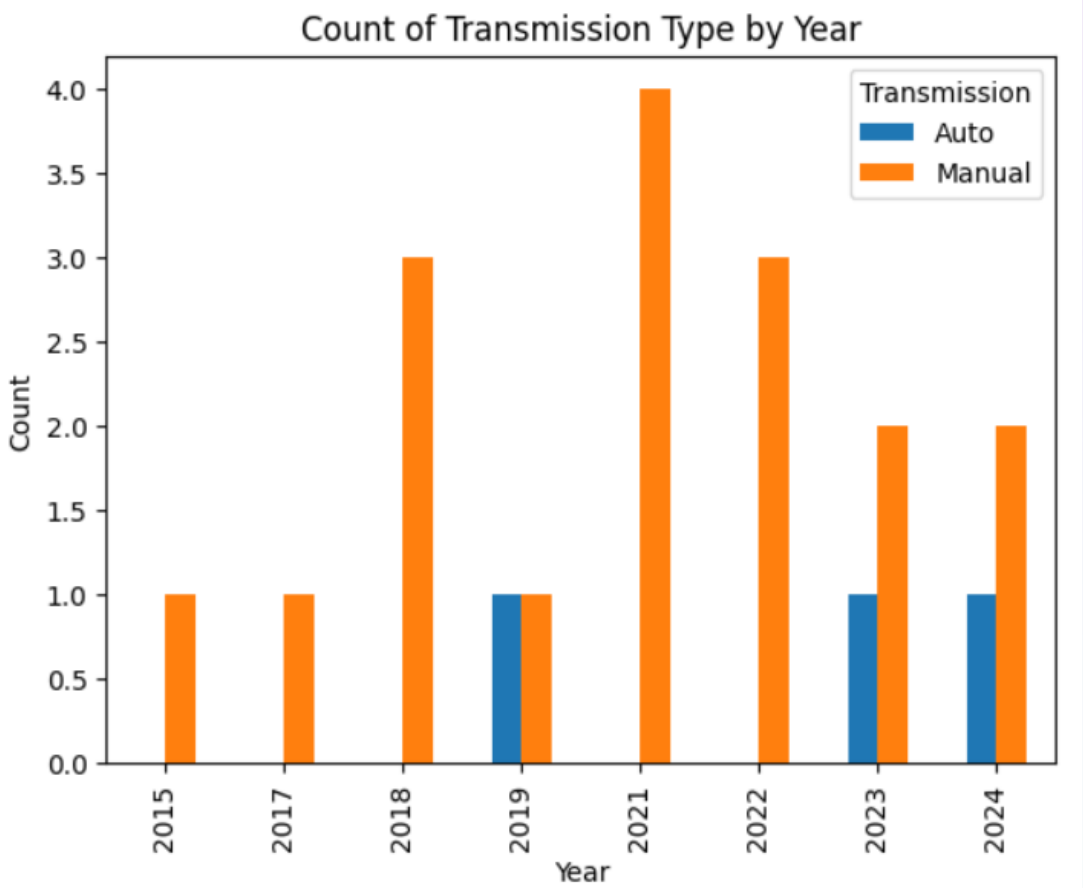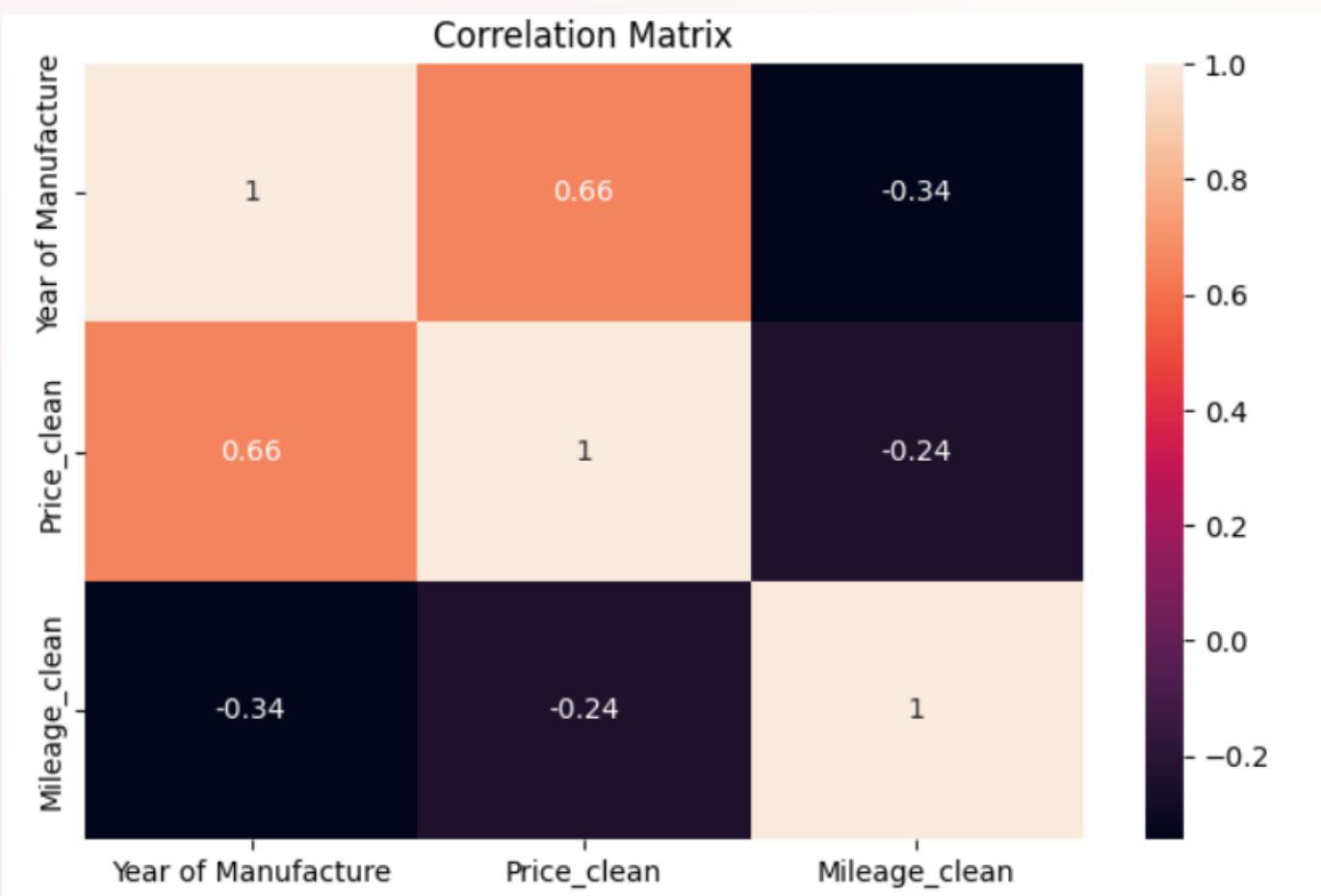
The maximum price of car was in the year 2021, and that for Automatic Transmission is far high.

Centre wise type of car(Fuel) avilibility
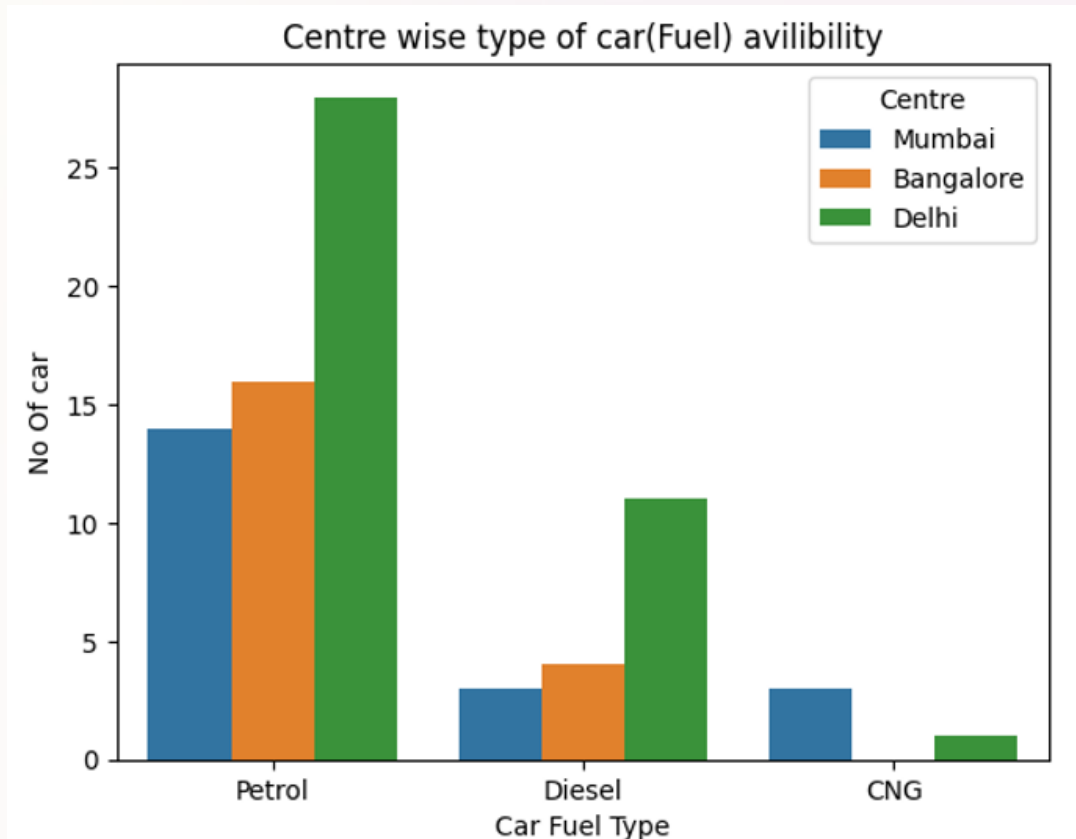
- **Petrol cars dominate** in all centres, but especially in Delhi: Delhi has ~28 petrol cars, which is substantially higher than Mumbai and Bangalore.

- **Diesel availability is modest** and again highest in Delhi (~11) compared to Bangalore (~4) and Mumbai (~3).

- **CNG availability is very low** in Bangalore (essentially none), very low in Delhi (~1) and low in Mumbai (~3).

- So: Delhi is the most diverse and abundant in terms of fuel-type availability (especially for petrol and diesel). Mumbai and Bangalore are much lower overall, and Bangalore in particular seems to have almost no CNG-cars available.

Distribution Of Price Among Various Transmission Type

Centre wise type of car(Transmission) availability

- **Manual cars are more prevalent** than automatic ones in all three centres.

- **Delhi** has the **highest overall availability** for both types.

- **Mumbai** has the **lowest number of automatic cars**, indicating a possible preference or lower demand/supply.
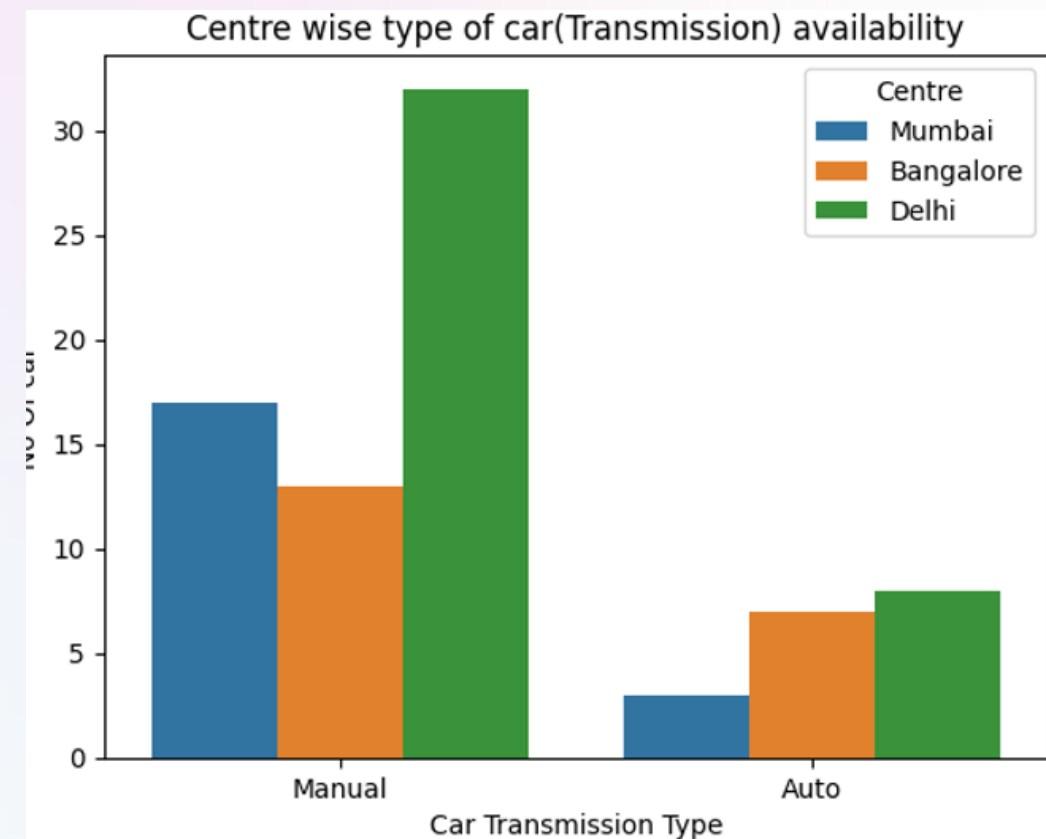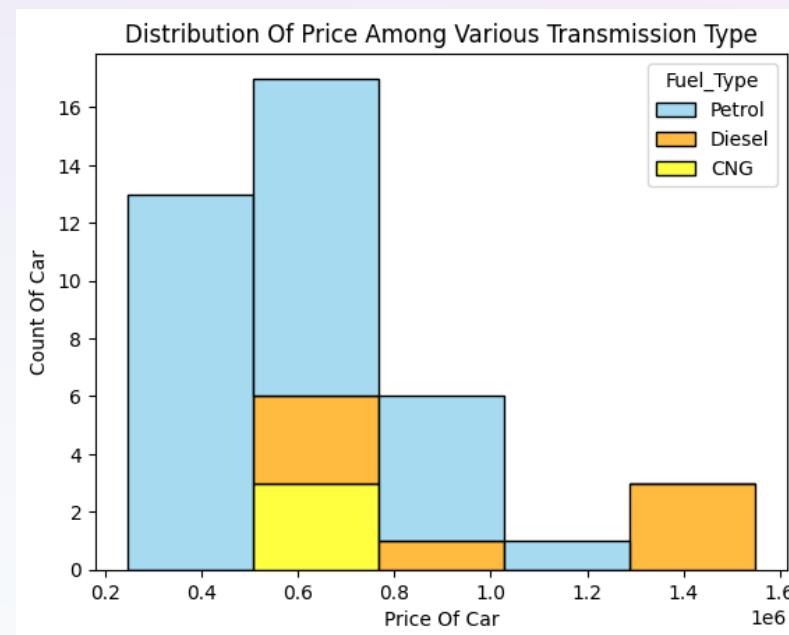
Output:- .CSV file without Data Cleaning.

	Name	of Manufa	Mileage	Fuel	ransmissio	Price
1	**Name**	**of Manufa**	**Mileage**	**Fuel**	**ransmissio**	**Price**
2	2015 Tata	2015	40.97k km	Petrol	Manual	₹2.45 lakh
3	2024 Tata	2024	8.33k km	Petrol	Auto	₹9.86L
4	2023 Tata	2023	42.18k km	Petrol	Manual	₹6.27L
5	2019 Tata	2019	39.65k km	Petrol	Manual	₹4.24L
6	2024 Tata	2024	2.93k km	Petrol	Manual	₹8.67L
7	2021 Tata	2021	54.67k km	Diesel	Manual	₹14.03L
8	2017 Tata	2017	19.89k km	Petrol	Manual	₹4.70 lakh
9	2018 Tata	2018	71.61k km	Petrol	Manual	₹5.75L
10	2023 Tata	2023	90.53k km	CNG	Manual	₹7.66L
11	2024 Tata	2024	3.09k km	Petrol	Manual	₹11.29L
12	2023 Tata	2023	42.47k km	Petrol	Auto	₹8.79L
13	2021 Tata	2021	63.27k km	Petrol	Manual	₹7.51L
14	2021 Tata	2021	63.58k km	Diesel	Manual	₹7.99L
15	2018 Tata	2018	55.42k km	Diesel	Manual	₹6.32L
16	2022 Tata	2022	32.44k km	CNG	Manual	₹6.74L
17	2021 Tata	2021	50.42k km	Petrol	Manual	₹4.66L
18	2018 Tata	2018	98.18k km	Petrol	Manual	₹4.63L
19	2022 Tata	2022	5.83k km	Petrol	Manual	₹7.29L
20	2022 Tata	2022	51.35k km	CNG	Manual	₹6.46L
21	2019 Tata	2019	39.75k km	Petrol	Auto	₹6.87L

Output:- .CSV file with Data Cleaning.

Centre	car_name	Year_of_Manufacture	Kilometers_Driven	Fuel_Type	Transmission	Price
Mumbai	2015 Tata Zest	2015	40.97k km	Petrol	Manual	2.45 lakh
Mumbai	2024 Tata NEXON	2024	8.33k km	Petrol	Auto	9.86L
Mumbai	2023 Tata PUNCH	2023	42.18k km	Petrol	Manual	6.27L
Mumbai	2019 Tata TIGOR	2019	39.65k km	Petrol	Manual	4.24L
Mumbai	2024 Tata ALTROZ	2024	2.93k km	Petrol	Manual	8.67L
Mumbai	2021 Tata Harrier	2021	54.67k km	Diesel	Manual	14.03L
Mumbai	2017 Tata NEXON	2017	19.89k km	Petrol	Manual	4.70 lakh
Mumbai	2018 Tata NEXON	2018	71.61k km	Petrol	Manual	5.75L
Mumbai	2023 Tata PUNCH	2023	90.53k km	CNG	Manual	7.66L
Mumbai	2024 Tata NEXON	2024	3.09k km	Petrol	Manual	11.29L
Mumbai	2023 Tata ALTROZ	2023	42.47k km	Petrol	Auto	8.79L
Mumbai	2021 Tata NEXON	2021	63.27k km	Petrol	Manual	7.51L
Mumbai	2021 Tata NEXON	2021	63.58k km	Diesel	Manual	7.99L
Mumbai	2018 Tata NEXON	2018	55.42k km	Diesel	Manual	6.32L
Mumbai	2022 Tata TIGOR	2022	32.44k km	CNG	Manual	6.74L
Mumbai	2021 Tata TIGOR	2021	50.42k km	Petrol	Manual	4.66L
Mumbai	2018 Tata NEXON	2018	98.18k km	Petrol	Manual	4.63L
Mumbai	2022 Tata ALTROZ	2022	5.83k km	Petrol	Manual	7.29L
Mumbai	2022 Tata TIGOR	2022	51.35k km	CNG	Manual	6.46L
Mumbai	2019 Tata NEXON	2019	39.75k km	Petrol	Auto	6.87L
Bangalore	2019 Tata Tiago	2019	46.15k km	Petrol	Manual	4.76 lakh
Bangalore	2018 Tata Tiago	2018	48.85k km	Petrol	Manual	3.75 lakh
Bangalore	2018 Tata NEXON	2018	91.95k km	Petrol	Manual	5.51 lakh
Bangalore	2021 Tata NEXON	2021	29.21k km	Petrol	Manual	7.79 lakh
Bangalore	2021 Tata Safari	2021	64.89k km	Diesel	Manual	15.50 lakh
Bangalore	2021 Tata NEXON	2021	38.85k km	Petrol	Manual	7.99 lakh
Bangalore	2020 Tata Harrier	2020	38.20k km	Diesel	Auto	13.86 lakh
Bangalore	2018 Tata NEXON	2018	40.44k km	Petrol	Manual	5.04 lakh
Bangalore	2020 Tata Tiago	2020	65.54k km	Petrol	Auto	4.70 lakh
Bangalore	2021 Tata TIGOR	2021	45.81k km	Petrol	Auto	5.39 lakh
Bangalore	2018 Tata NEXON	2018	44.22k km	Petrol	Manual	5.70 lakh
Bangalore	2022 Tata TIGOR	2022	38.15k km	Petrol	Manual	5.69 lakh
Bangalore	2019 Tata NEXON	2019	93.55k km	Petrol	Manual	5.35 lakh
Bangalore	2020 Tata NEXON	2020	64.41k km	Diesel	Manual	7.13 lakh
Bangalore	2019 Tata NEXON	2019	73.33k km	Petrol	Auto	6.19 lakh
Bangalore	2018 Tata Tiago	2018	54.70k km	Petrol	Auto	4.70 lakh
Bangalore	2017 Tata Tiago	2017	31.96k km	Petrol	Auto	4.50 lakh
Bangalore	2018 Tata Tiago	2018	57.34k km	Petrol	Auto	4.20 lakh
Bangalore	2019 Tata NEXON	2019	1.1L km	Diesel	Manual	6.30 lakh
Bangalore	2021 Tata Tiago	2021	79.23k km	Petrol	Manual	4.95 lakh
Delhi	2017 Tata Tiago	2017	85.64k km	Petrol	Manual	
Delhi	2024 Tata Tiago	2024	26.27k km	Petrol	Manual	
Delhi	2020 Tata NEXON	2020	76.27k km	Petrol	Auto	
Delhi	2022 Tata NEXON	2022	65.15k km	Petrol	Manual	
Delhi	2021 Tata ALTROZ	2021	43.03k km	Diesel	Manual	
Delhi	2020 Tata NEXON	2020	35.01k km	Petrol	Auto	
Delhi	2022 Tata ALTROZ	2022	28.00k km	Petrol	Auto	
Delhi	2022 Tata Tiago	2022	22.19k km	Petrol	Auto	
Delhi	2021 Tata Tiago	2021	45.57k km	Petrol	Manual	
Delhi	2023 Tata NEXON	2023	24.88k km	Petrol	Manual	
Delhi	2022 Tata Tiago	2022	27.38k km	CNG	Manual	
Delhi	2017 Tata Tiago	2017	59.93k km	Diesel	Manual	
Delhi	2023 Tata NEXON	2023	9.49k km	Petrol	Manual	
Delhi	2024 Tata NEXON	2024	5.20k km	Petrol	Manual	
Delhi	2023 Tata PUNCH	2023	12.82k km	Petrol	Manual	
Delhi	2024 Tata NEXON	2024	3.98k km	Petrol	Manual	
Delhi	2023 Tata NEXON	2023	32.15k km	Petrol	Manual	
Delhi	2019 Tata NEXON	2019	82.99k km	Petrol	Auto	
Delhi	2021 Tata Harrier	2021	1.1L km	Diesel	Auto	

Output:- .CSV file with Data Cleaning (continued).

61	Delhi	2019 Tata NEXON	2019	67.71k km	Diesel	Manual
62	Delhi	2023 Tata NEXON	2023	18.11k km	Petrol	Manual
63	Delhi	2024 Tata NEXON	2024	19.26k km	Petrol	Manual
64	Delhi	2019 Tata Harrier	2019	65.57k km	Diesel	Manual
65	Delhi	2018 Tata Tiago	2018	56.83k km	Petrol	Manual
66	Delhi	2020 Tata Harrier	2020	84.47k km	Diesel	Manual
67	Delhi	2021 Tata Safari	2021	59.36k km	Diesel	Manual
68	Delhi	2020 Tata ALTROZ	2020	34.30k km	Petrol	Manual
69	Delhi	2022 Tata PUNCH	2022	51.70k km	Petrol	Manual
70	Delhi	2024 Tata ALTROZ	2024	16.92k km	Petrol	Manual
71	Delhi	2020 Tata Harrier	2020	95.79k km	Diesel	Manual
72	Delhi	2020 Tata NEXON	2020	81.67k km	Petrol	Auto
73	Delhi	2024 Tata Harrier	2024	7.83k km	Diesel	Manual
74	Delhi	2022 Tata TIGOR	2022	26.64k km	Petrol	Manual
75	Delhi	2023 Tata NEXON	2023	9.09k km	Petrol	Manual
76	Delhi	2021 Tata ALTROZ	2021	73.82k km	Diesel	Manual
77	Delhi	2022 Tata PUNCH	2022	36.84k km	Petrol	Manual
78	Delhi	2022 Tata Harrier	2022	89.66k km	Diesel	Manual
79	Delhi	2019 Tata Tiago	2019	25.73k km	Petrol	Auto
80	Delhi	2020 Tata ALTROZ	2020	71.50k km	Petrol	Manual
81	Delhi	2023 Tata Tiago	2023	46.11k km	Petrol	Manual

Data Analysis: Unveiling Market Trends

Data Cleaning

1

Before analysis, the extracted data is cleaned and processed. This involves converting data types, removing inconsistencies, and handling missing values. This ensures accurate and reliable results.

Data Visualization

2

Visualizing the data provides a clear understanding of trends and patterns. Histograms, scatter plots, and count plots are used to analyze the distribution of kilometers driven, year of manufacture, fuel type, transmission, and price.
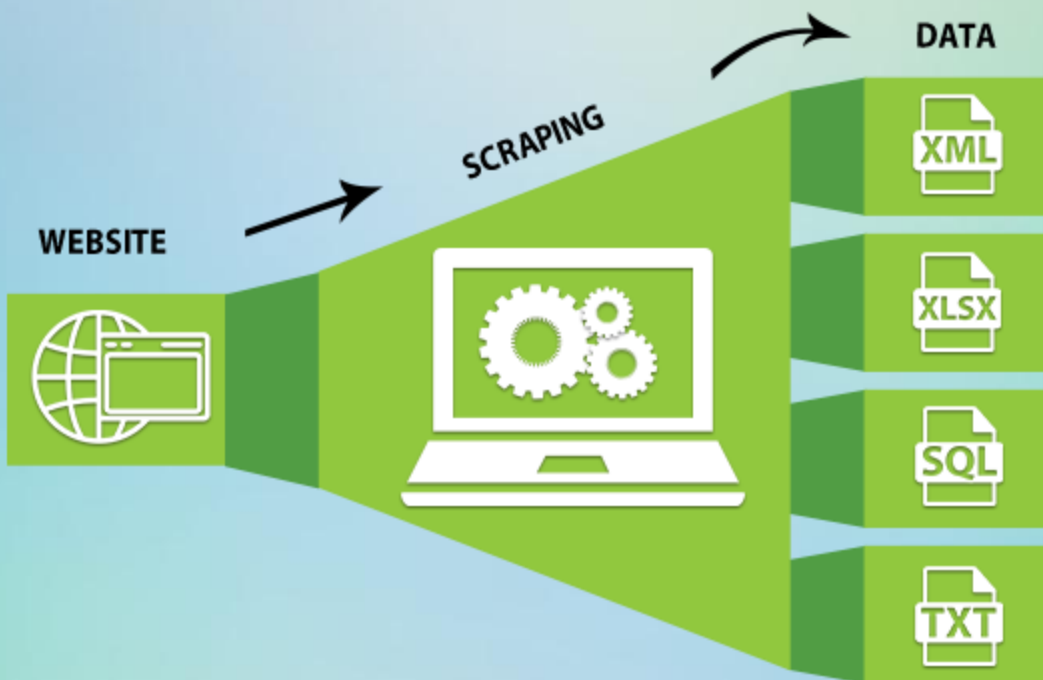
Insights and Observations

3

Analyzing the visualized data reveals valuable insights into market trends, pricing patterns, and consumer preferences. This information can inform business decisions related to inventory management, marketing strategies, and pricing strategies.
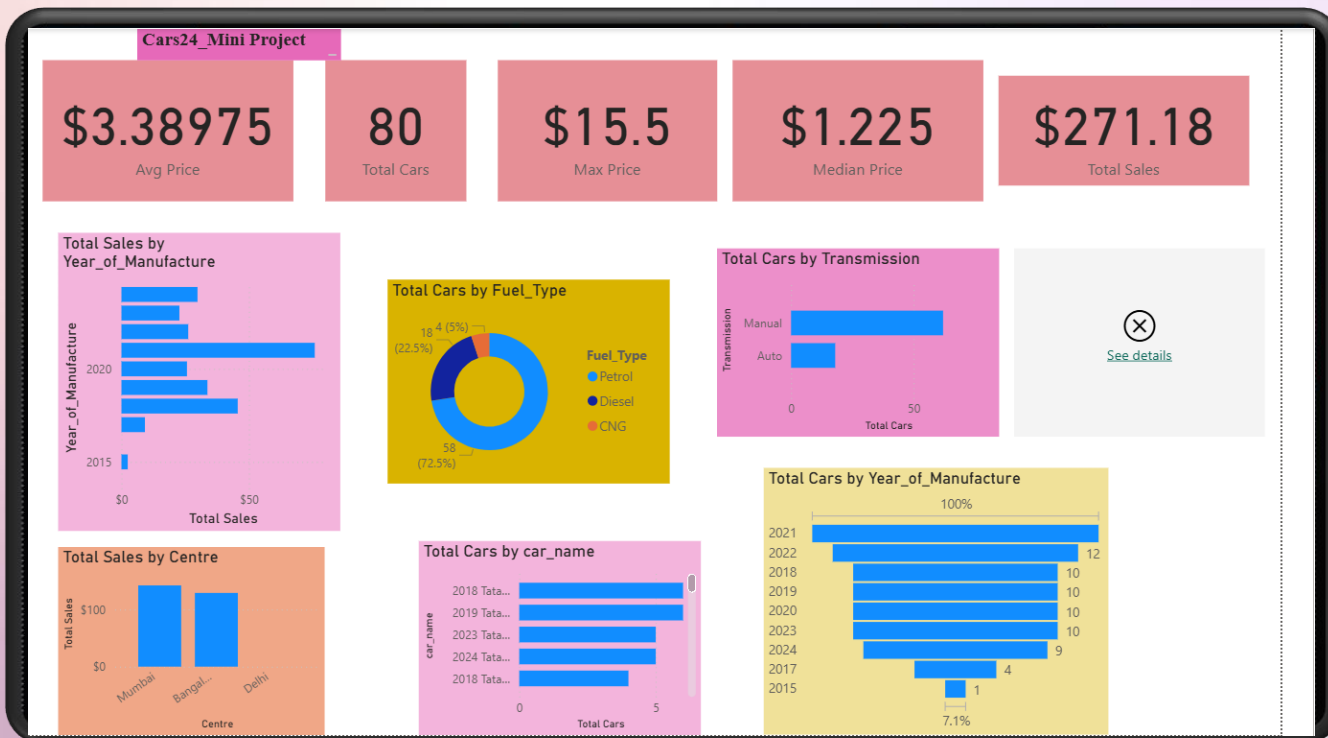
Outputs and Findings:

Efficient Scripting

Optimized scripts are developed to reduce scraping time and improve data extraction efficiency. This involves utilizing efficient algorithms, minimizing redundant code, and optimizing data processing techniques. Through this we were able to scrap the data's of 3 Locations , such as

1. Mumbai
2. Delhi
3. Banglore

The prime factors responsible for the price differences are Ownership, Kilometers driven, Fuel Type and Transmission type.

Key Takeaways: Insights from the Data

Petrol is the dominant fuel type, but the preference is stronger in Mumbai.
In Mumbai, Petrol cars (approx. 160) are more than double the number of Diesel cars (approx. 75). While Petrol is also the leader in Delhi (approx. 125), the gap with Diesel (approx. 80) is much smaller, suggesting a more balanced fuel-type market in Delhi compared to Mumbai.

Manual transmission is heavily favored in Mumbai, while Delhi shows a higher adoption of automatics. Manual transmission is the clear winner in both markets. However, the preference is far more skewed in Mumbai, with more than twice as many Manual cars (approx. 180) as Automatic (approx. 80). In Delhi, the ratio is much closer (approx. 140 Manual to 85 Automatic), indicating a significantly stronger relative demand for automatic transmission cars in the Delhi market.

CNG remains a niche fuel option in both locations. Despite being major metropolitan areas, CNG-powered vehicles represent the smallest fraction of the listings in both datasets. With only around 20-25 listings in each city, CNG is clearly a minor player compared to the popularity of Petrol and Diesel.

Challenges in Web Scraping

1   Dynamic Content

Websites often utilize JavaScript and other technologies to dynamically load content. Handling dynamic content poses a challenge, as traditional scraping methods may not capture the complete information.

2   Request Limits

Web scraping should be conducted responsibly to avoid overwhelming websites with excessive requests. Websites often implement rate limits to protect their resources and prevent abuse.

3   Legal and Ethical Guidelines

Web scraping must adhere to legal and ethical guidelines. It's essential to respect website terms of service, robots.txt files, and privacy policies. Avoid scraping sensitive data or engaging in activities that could harm websites or their users.

4   Data Cleaning and Validation

The scraped data may contain inconsistencies, errors, or missing values. Data cleaning and validation processes are crucial to ensure the accuracy and reliability of the extracted information.

:(

Your PC ran into a problem and needs to restart. We're just collecting some error info, and then we'll restart for you.

20% complete

For more information about this issue and possible fixes, visit https://www.windows.com/stopcode

If you call a support person, give them this info:
Stop code: CRITICAL_PROCESS_DIED

Conclusion: Harnessing Data for Informed Decisions

This web scraping project demonstrates the power of data extraction and analysis. By leveraging Python libraries and techniques, valuable insights into the automotive market are extracted from Cars24. These insights empower businesses to make informed decisions related to pricing, inventory management, and marketing strategies, ultimately leading to enhanced profitability and customer satisfaction.

Applications:

Web scraping has diverse applications, including data mining, market research, price monitoring, and competitive analysis. It enables businesses to gain valuable insights from publicly available online data.

Thank You