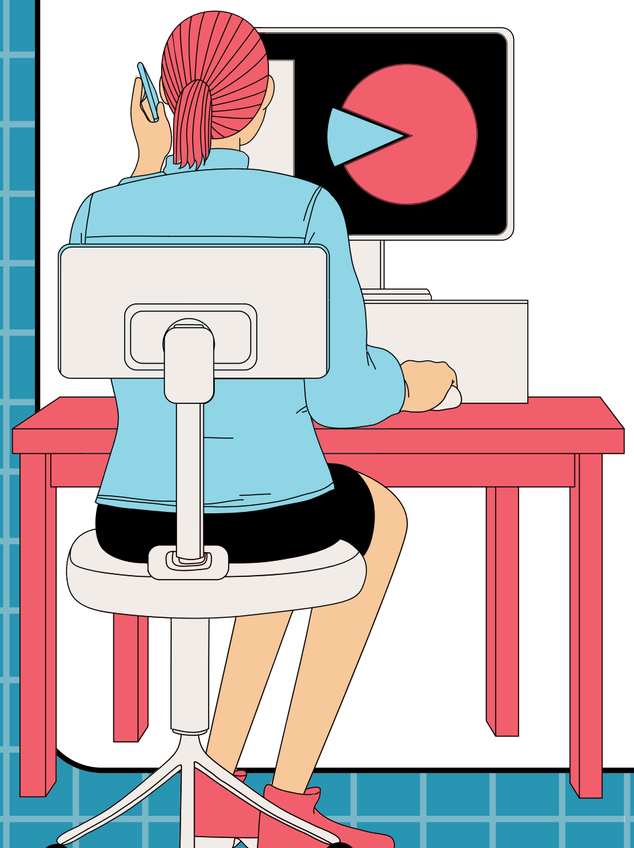
WEB SCRAPING DATA FROM CARS24.COM



by Team B

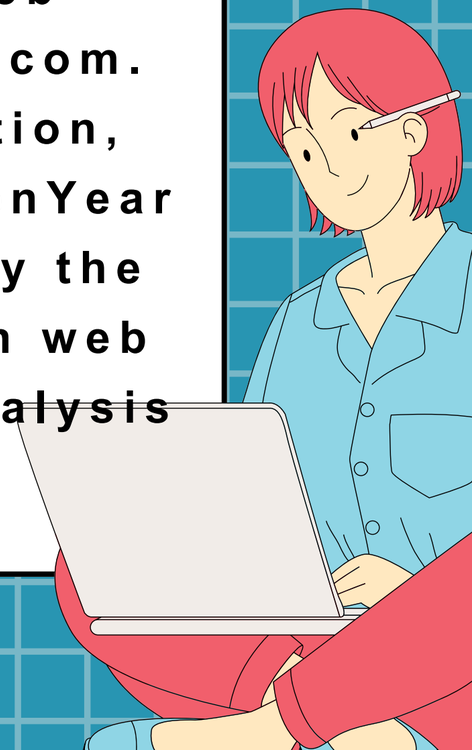
TEAM MEMBERS

- Alok Kumar
- Sajeev Kannati
- Sourabh Solanki
- Anshu Yadav
- Yeshwanth seetha
- Ranjeeth Narayanasamy
- Akshay Bharadwaj v
- Adithya Raj
- Muhamed mujawer
- Abhay Tiwari

INDEX

OBJECTIVES OF THE MINI PROJECT

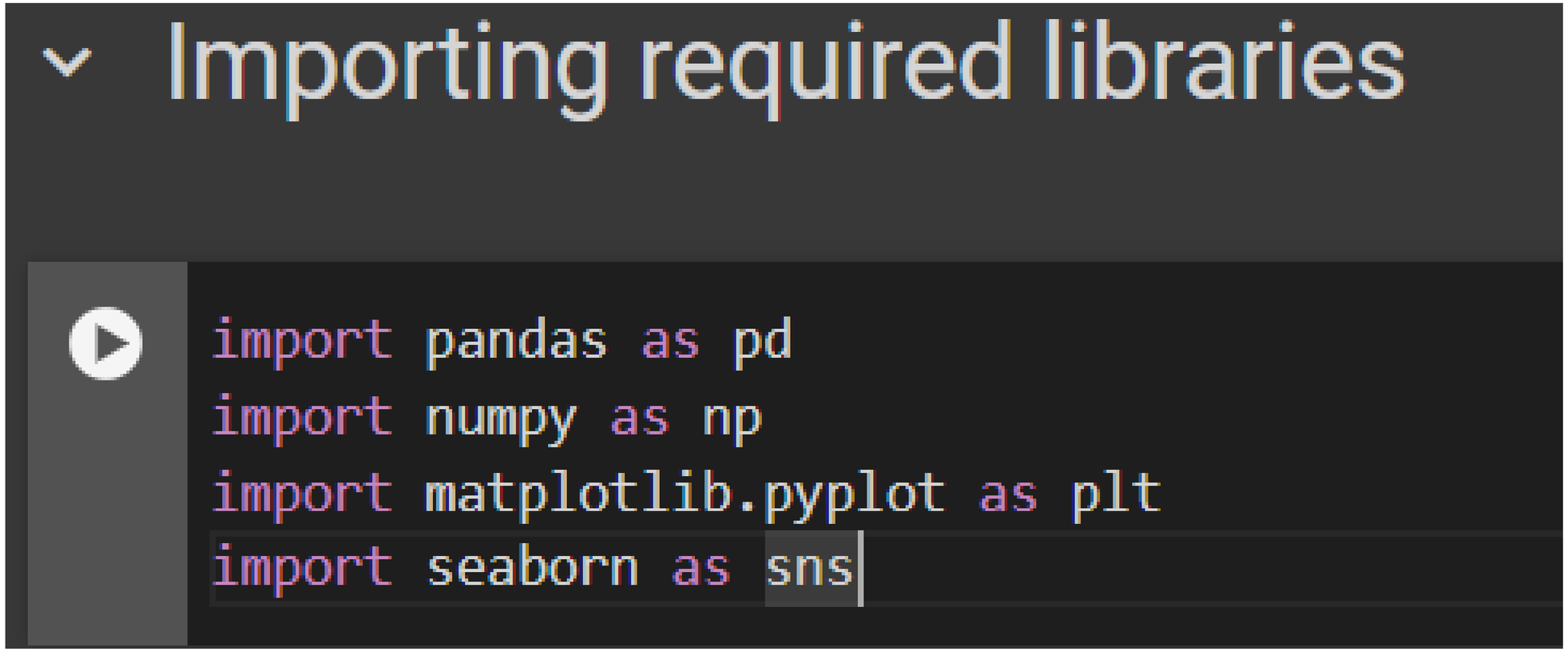
The objective of this mini-project is to develop skills in web scraping by extracting and analyzing car details from Cars24.com. All team members gathered data specific to the Mumbai location, covering key information such as Make,Model,Kilometers,DrivenYear of Manufacture,Fuel,Type,Transmission,Location and Price. By the end of this project, members will have hands-on experience in web scraping, data cleaning, and presenting structured data for analysis



REQUIREMENTS AND INSTALLTIONS

- Python: Ensure you have Python installed on your system.
- IDE: Preferably Google Colab, Jupyter Notebook or Visual Studio Code
- Libraries: The following Python libraries are required:
- Pandas
- Numpy
- bs4 (BeautifulSoup)
- re (Regular Expressions, part of the standard library)
- Matplotlib
- Seaborn

IMPORTING LIBRARIES



Your paragraph text

WORKING FLOWCHART

METHODOLOGY

- Assigned Brands: Hyundai , Kia, Nissan
- Assigned Location; Mumbai

1. Research and Planning:

- Understand the structure of https://www.cars24.com/
- Identify the HTML elements containing the required information.

2. Data Extraction:

- Write a script to scrape the necessary details for the assigned task Make, Model, Kilometers Driven ,Year of Manufacture, Fuel, Type, Transmission, Location and Price.

METHODOLOGY

3. Automate Scrolling to Load All Content:

.The script repeatedly scrolls down the page to trigger dynamic l loading of additional car listings. It waits for a short period after each scroll to ensure new content is loaded.

4. Dynamic Content Loading:

- **Many modern websites, including car listing sites, use JavaScript to load content dynamically. This means that not all the data is present in the initial HTML source but is loaded as the user scrolls or interacts with the page.**

METHODOLOGY

5. Extract Car Details: The script finds and extracts car names, details, and prices from the HTML elements. For each car, it extracts the car name and manufacturing year, kilometers driven, fuel type, and transmission type from the details list and price from the price tag.

Name	Nissan MAGNITE XV MT	TURBO CVT	PREMIUM	Nissan MicraXV PETROL	Nissan MAGNITE XV MT	TURBO CVT
Details	Alloy wheels33,160 kmPetrol1st owner	Reg. serviced15,037 kmPetrol1st owner	Top Model47,127 kmPetrol1st owner	Top Model40,385 kmPetrol2nd owner	Alloy wheels33,160 kmPetrol1st owner	Reg. serviced15,037 kmPetrol1st owner
Price	₹6.58L	₹7.74L	₹6.60L	₹1.91L	₹6.58L	₹7.74L
Location	Free Test DriveTomorrowatGoregaon, Mumbai	Free Test DriveTomorrowatGoregaon, Mumbai	Free Test DriveTomorrowatMulund West, Mumbai	Free Test DriveTomorrowatSeawood	Free Test DriveTomorrowatGoregaon, Mumbai	Free Test DriveTomorrowatGoregaon, Mumbai
Year	2022	2022	2021	2012	2022	2022
Kilometers Driven	33,160	15,037	47,127	40,385	33,160	15,037
Fuel Type	Petrol	Petrol	Petrol	Petrol	Petrol	Petrol
Owner	1st owner	1st owner	1st owner	2nd owner	1st owner	1st owner
Company	Nissan	Nissan	Nissan	Nissan	Nissan	Nissan
Location	Goregaon, Mumbai	Goregaon, Mumbai	Mulund West, Mumbai	Seawood	Goregaon, Mumbai	Goregaon, Mumbai

METHODOLOGY

6. Data Cleaning: Ensure the scraped data is clean and organized for analysis. The extracted data is cleaned (e.g., removing the make year from car names) and structured into a list.

	Name	Details	Price	Location	Year	Kilometers Driven	Fuel Type	Owner
0	Nissan MAGNITEXV MT	Alloy wheels33,160 kmPetrol1st owner	₹6.58L	Free Test DriveTomorrowatGoregaon, Mumbai	2022	33,160	Petrol	1st owner
1	Nissan MAGNITEXL TURBO CVT	Reg. serviced15,037 kmPetrol1st owner	₹7.74L	Free Test DriveTomorrowatGoregaon, Mumbai	2022	15,037	Petrol	1st owner
2	Nissan MAGNITEXV PREMIUM	Top Model47,127 kmPetrol1st owner	₹6.60L	Free Test DriveTomorrowatMulund West, Mumbai	2021	47,127	Petrol	1st owner
3	Nissan MicraXV PETROL	Top Model40,385 kmPetrol2nd owner	₹1.91L	Free Test DriveTomorrowatSeawood	2012	40,385	Petrol	2nd owner
4	Nissan MAGNITEXV MT	Alloy wheels33,160 kmPetrol1st owner	₹6.58L	Free Test DriveTomorrowatGoregaon, Mumbai	2022	33,160	Petrol	1st owner
...	...	...	...	...	...	...	...	...
165	Hyundai Elite i20SPORTZ 1.2	Safety specs.19,288 kmPetrol1st owner	₹5.34L	Free Test DriveTomorrowatGoregaon, Mumbai	2017	19,288	Petrol	1st owner
166	Hyundai i20 Active1.4 SX	Top Model77,631 kmDiesel2nd owner	₹5.01L	Free Test DriveTomorrowatGoregaon, Mumbai	2015	77,631	Diesel	2nd owner

METHODOLOGY

7. Saving Data into file:The structured data is converted into a pandas DataFrame and saved to a CSV file/ipynb

```
# Save the final DataFrame to a CSV file
final_df.to_csv('Nissan_cars_data.csv', index=False)

print("Data processing complete and saved to 'Nissan_cars_data.csv'.")
```

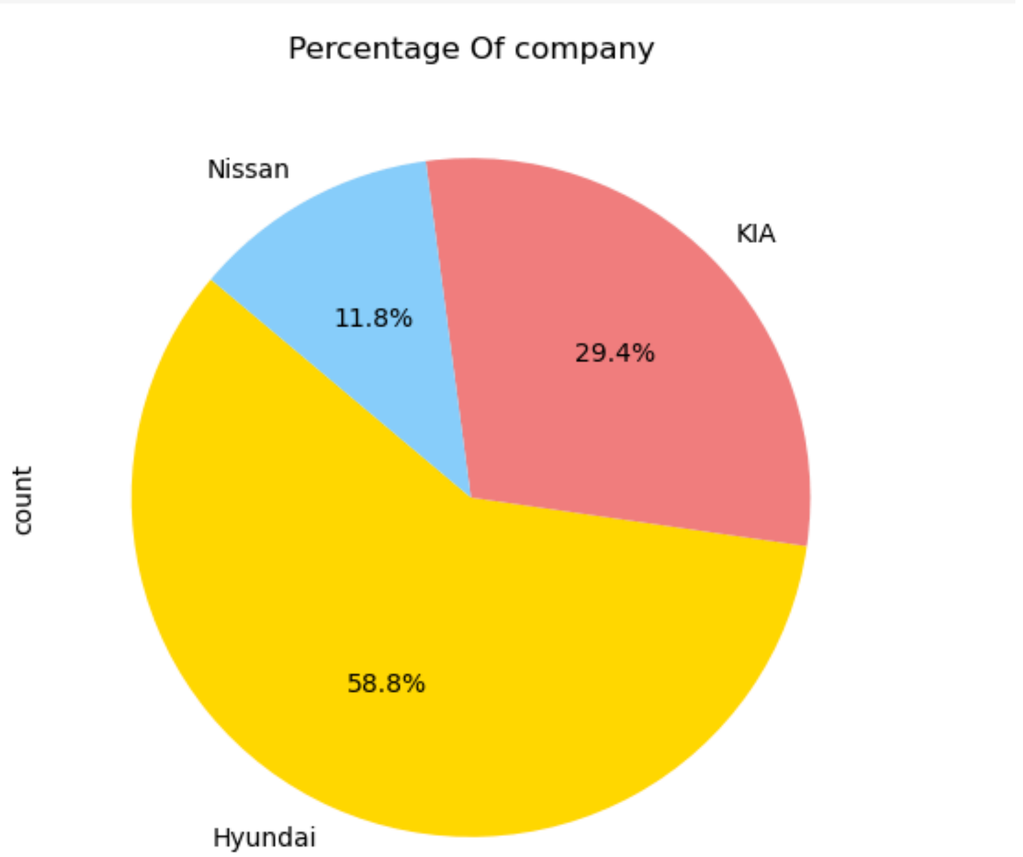
METHODOLOGY

8. Cars kms driven fuel type

	Name	Details	Price	Location	Year	Kilometers Driven	Fuel Type	O
0	Nissan MAGNITEXV MT	Alloy wheels33,160 kmPetrol1st owner	₹6.58L	Free Test DriveTomorrowatGoregaon, Mumbai	2022	33,160	Petrol	1st o
1	Nissan MAGNITEXL TURBO CVT	Reg. serviced15,037 kmPetrol1st owner	₹7.74L	Free Test DriveTomorrowatGoregaon, Mumbai	2022	15,037	Petrol	1st o
2	Nissan MAGNITEXV PREMIUM	Top Model47,127 kmPetrol1st owner	₹6.60L	Free Test DriveTomorrowatMulund West, Mumbai	2021	47,127	Petrol	1st o
3	Nissan MicraXV PETROL	Top Model40,385 kmPetrol2nd owner	₹1.91L	Free Test DriveTomorrowatSeawood	2012	40,385	Petrol	o
4	Nissan MAGNITEXV MT	Alloy wheels33,160 kmPetrol1st owner	₹6.58L	Free Test DriveTomorrowatGoregaon, Mumbai	2022	33,160	Petrol	1st o
...	...	...	...	...	...	...	...	
165	Hyundai Elite i20SPORTZ 1.2	Safety specs.19,288 kmPetrol1st owner	₹5.34L	Free Test DriveTomorrowatGoregaon, Mumbai	2017	19,288	Petrol	1st o
166	Hyundai i20 Active1.4 SX	Top Model77,631 kmDiesel2nd owner	₹5.01L	Free Test DriveTomorrowatGoregaon, Mumbai	2015	77,631	Diesel	

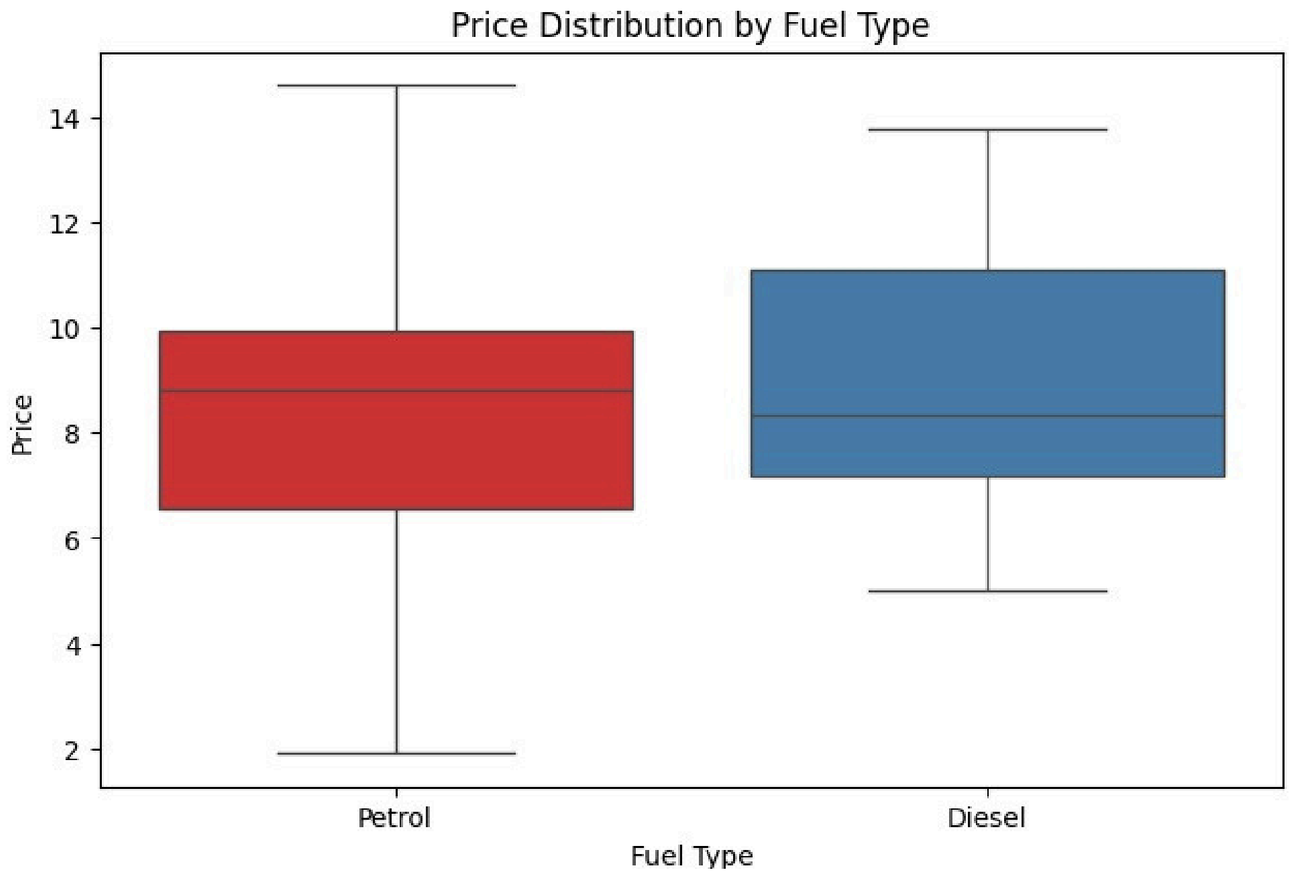
METHODOLOGY

9. Pie Chart



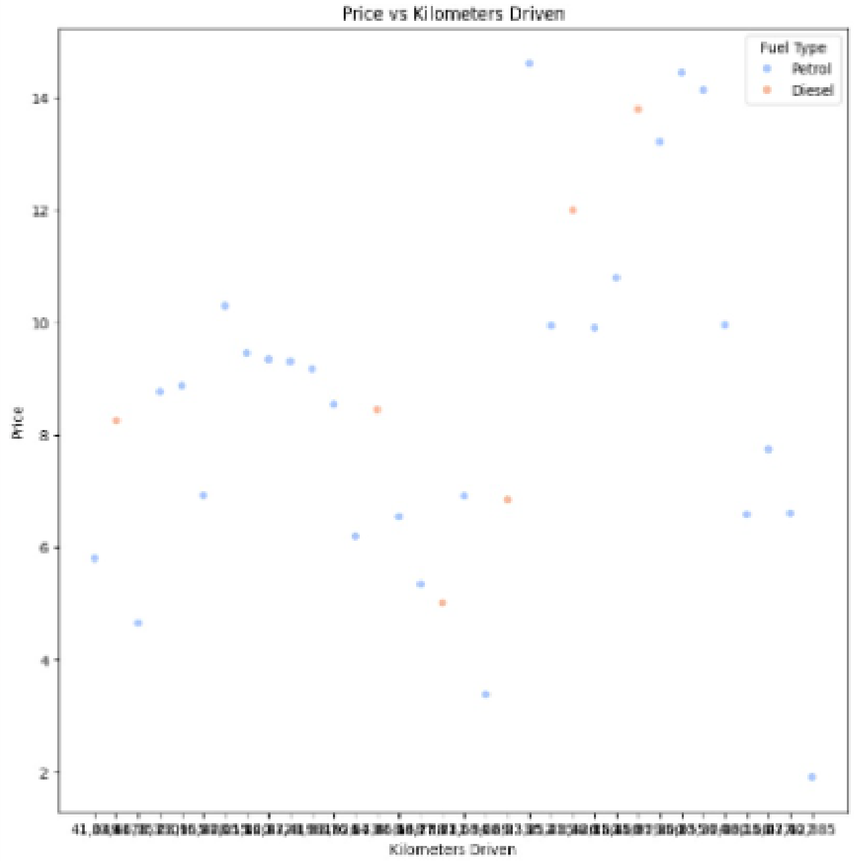
METHODOLOGY

10. Box Plot



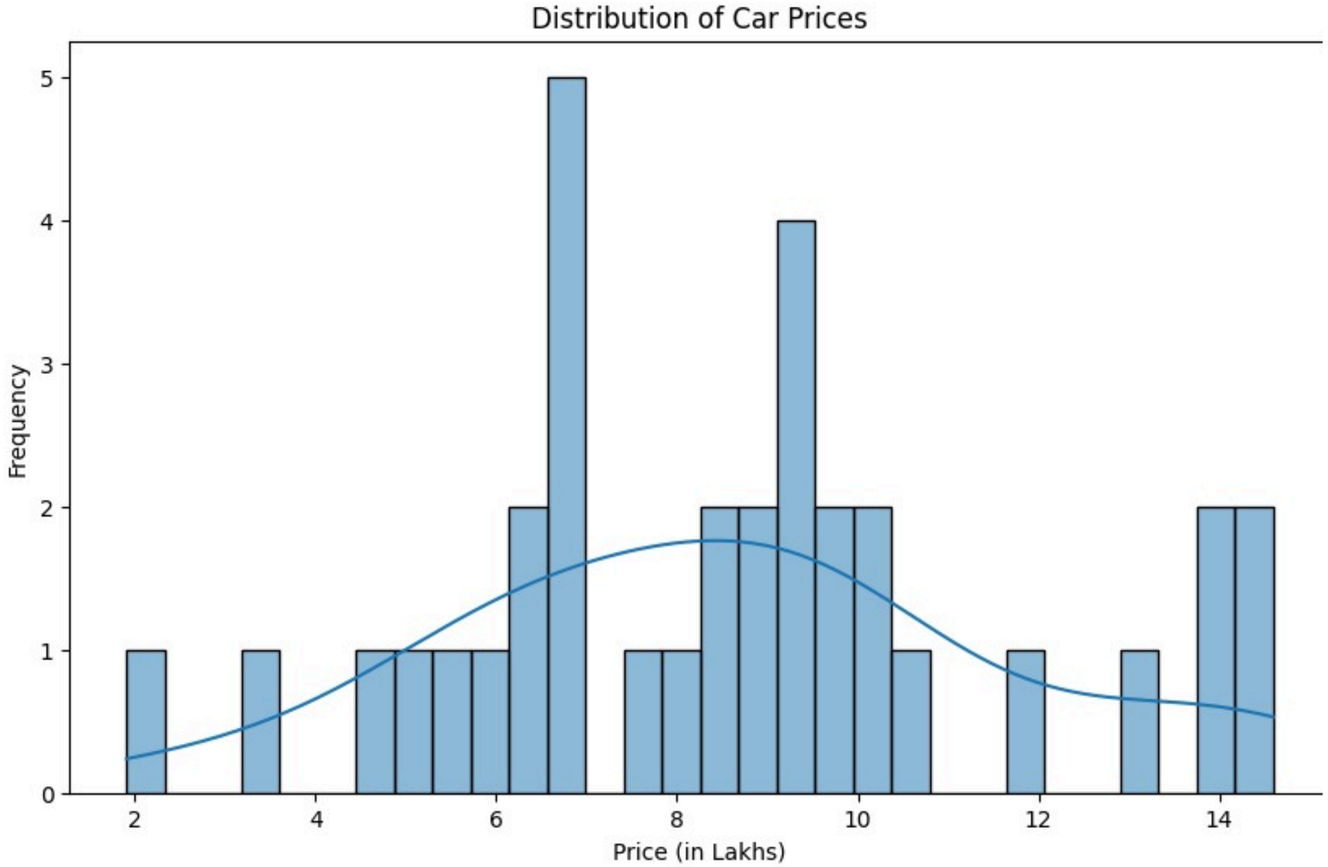
METHODOLOGY

11. Scatter plot



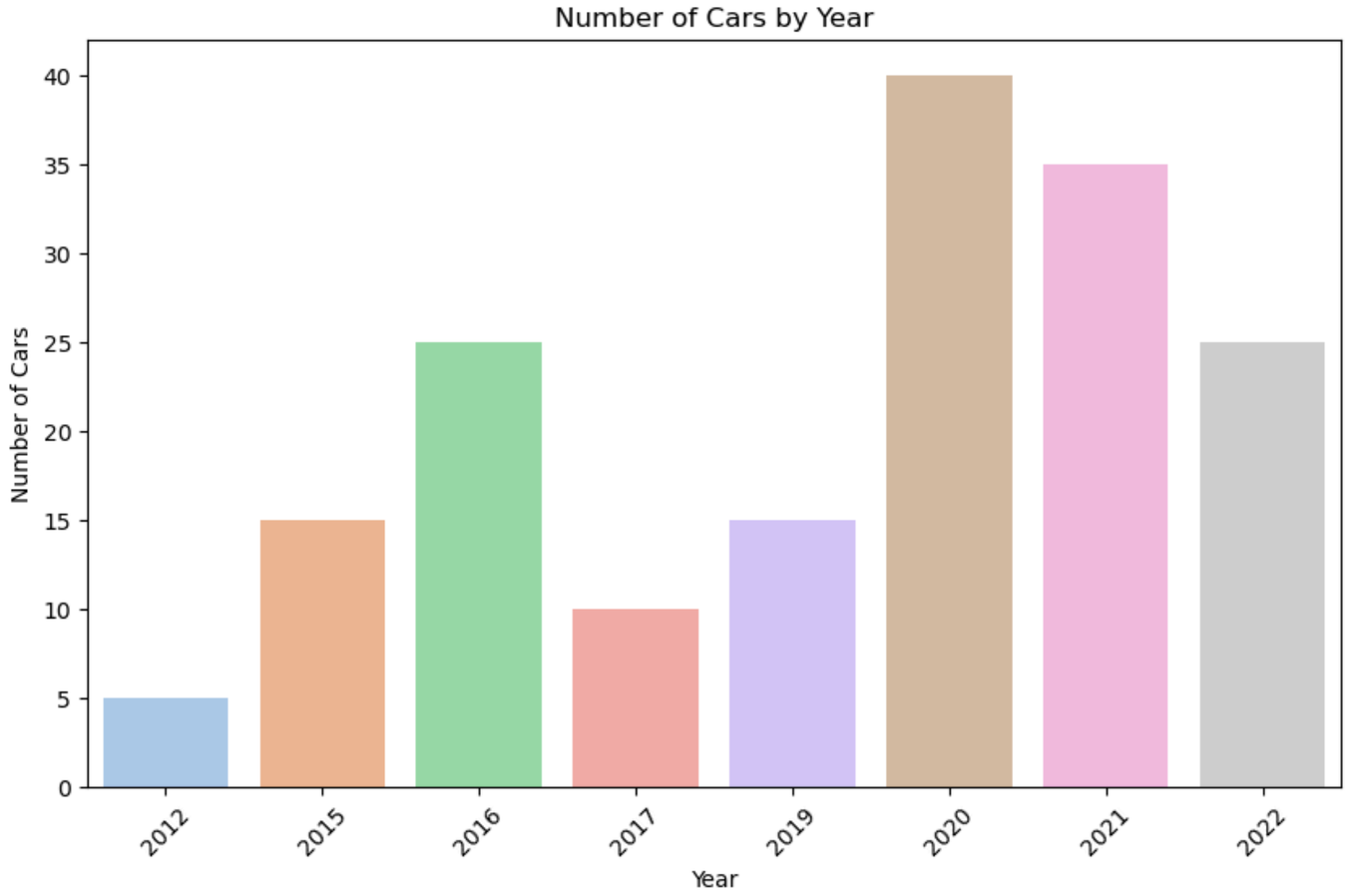
METHODOLOGY

12. Histogram



METHODOLOGY

12. Bar graph

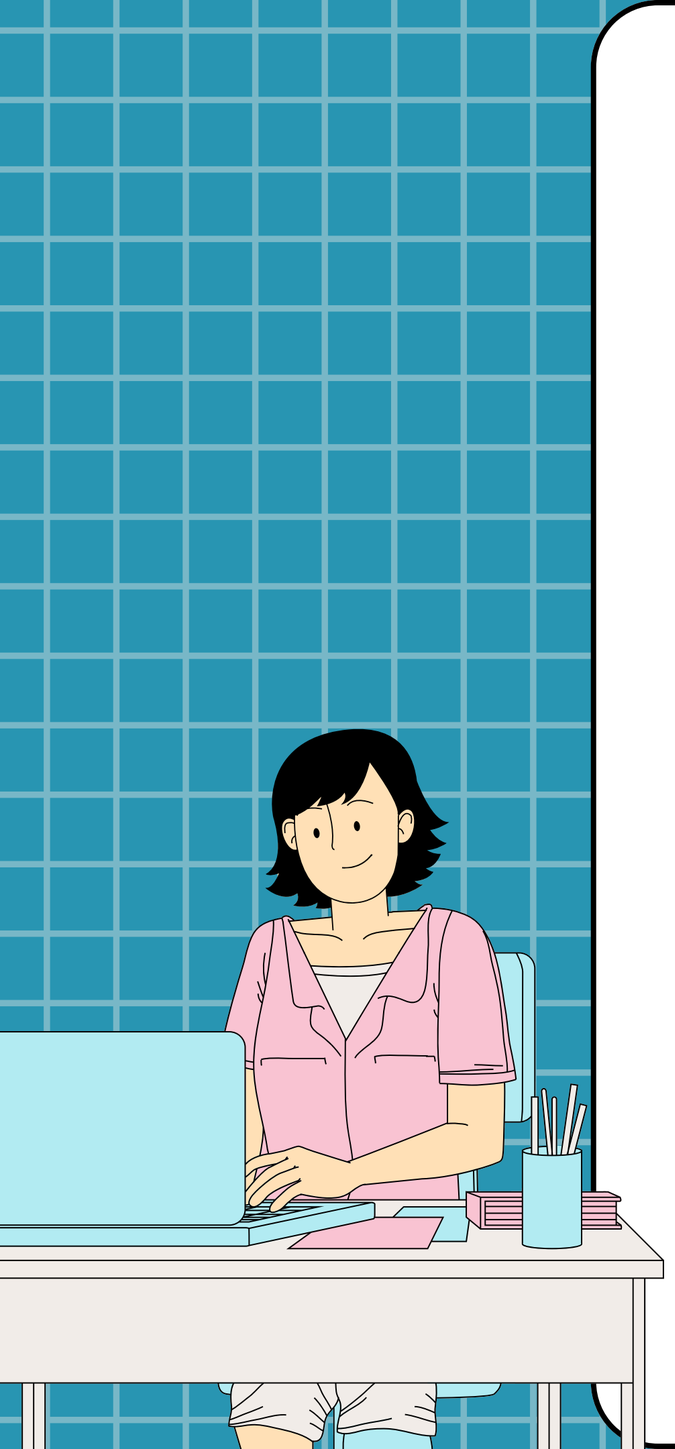


CHALLENGES AND OVERCOMING THEM

Dynamic Data Loading: BeautifulSoup alone was inadequate due to dynamic content loading on the website.

.

Data Cleaning: The extracted data required significant cleaning, especially in standardizing car names and separating the manufacture year from the car name for better readability and conversion to correct data types.

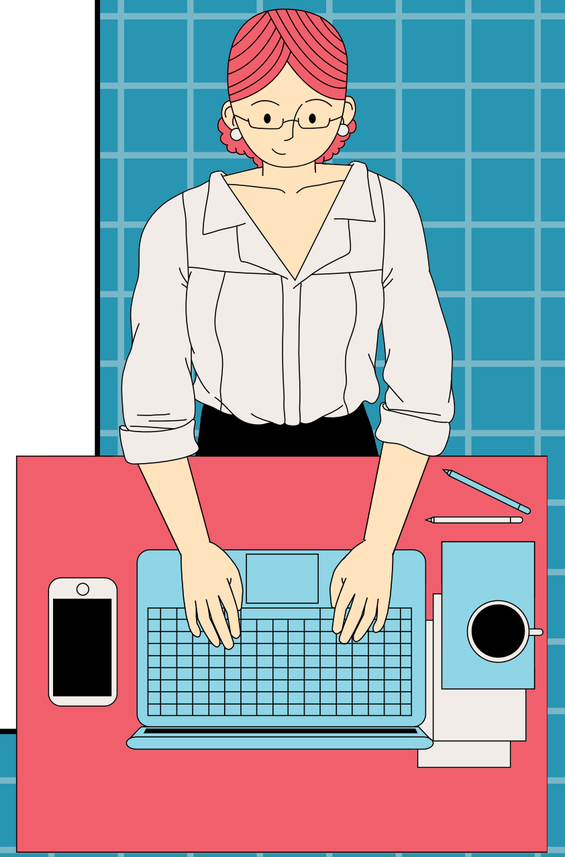


CHALLENGES AND OVERCOMING THEM

During an experiment, data is typically collected and organized using data tables with the independent variable on the left side and the dependent variable on the right side -- with units included!

INSIGHTS

1. *Purpose*: Extract data from websites for analysis.
2. *Methods*: Use static (HTML parsing) or dynamic (Selenium) scraping.
3. *Tools*: Beautiful Soup, Scrapy,
4. *Challenges*: Handle anti-scraping measures, respect legal/ethical guidelines.
5. *Best Practices*: Use headers, respect rate limits, handle errors, and automate tasks.



CONCLUSION

Web scraping extracts data from websites for analysis, involving HTML navigation and handling anti-scraping measures and legal guidelines. When done ethically, it supports applications like market research and data aggregation, making it valuable in data science.