

Mini Project Report

Web Scraping Car Details from Cars24.com

1. Title and other info

Project Title: Web Scraping Car Details from Cars24.com

Internship Program: Data Science Internship

Team: Team C – Hyundai

Team Lead: Vijayabhaskar V

Team Co-Lead: Adit Jain, Harshit Kumar

Team Members: Deep Dhar, Punit Ayare, Ravi kant kumar, Arepalli Chandra Sekhar, Abhishri Pathak, Arun Singh, Deepak Melkani, Suraj Vishwakarma.

Submission Date: 20 September 2025

2. Abstract

This project aims to extract and analyze car details from the Cars24 website using web scraping techniques. The main focus was on used-cars listed under the Hyundai brand in the Mumbai location. The process involved collecting important details such as kilometers driven, year of manufacture, fuel type, transmission, and price. Using Python libraries like Selenium, BeautifulSoup, and Pandas, the data was scraped, cleaned, and organized into a structured format. The final dataset was exported to a CSV file for further analysis. This project provided hands-on experience in real-world web scraping, data cleaning, and analysis, which are essential skills for aspiring data scientists.

3. Introduction

This mini-project, *"Web Scraping Car Details from Cars24.com"*, was developed to gain practical experience in data collection and processing from online sources. Cars24 is a popular online marketplace for buying and selling used cars. The objective of this project was to build a Python-based web scraper that extracts key car details such as year, kilometers driven, fuel type, transmission, and price for Hyundai cars listed in Mumbai. The extracted information was cleaned, structured, and saved in a CSV file for further analysis. The project demonstrates how data scientists can convert unstructured web data into valuable datasets for market analysis and business insights.

4. Objective

- To extract car details for the Hyundai brand from Cars24.com (Mumbai).
 - To identify and capture relevant data fields like year, fuel type, and price.
 - To clean, structure, and store the scraped data efficiently.
 - To analyze and visualize the dataset to gain useful insights.
 - To gain practical experience with Selenium, BeautifulSoup, and other libraries.
-

5. Tools and Technologies Used

- **Python:** Programming language used for data extraction and analysis.
 - **Selenium:** To automate browser actions and load dynamic web content.
 - **BeautifulSoup:** For parsing HTML content and extracting relevant tags.
 - **Pandas:** For data cleaning, manipulation, and exporting to CSV.
 - **Jupyter Notebook:** Used for writing, testing, and documenting the entire project.
-

6. Web Scraping Process

1. The Cars24 website was opened using **Selenium WebDriver**, and the page was scrolled gradually to load all car listings dynamically.
 2. Once the content was fully loaded, the **BeautifulSoup** library parsed the webpage's HTML source.
 3. Specific HTML tags and classes (like `<div>` for car cards, `` for names, `<p>` for details) were identified to extract required information.
 4. Extracted details such as **car name, kilometers, fuel type, transmission, and price** were stored in variables.
 5. Each car's data was stored in a dictionary and appended to a list named **all_car_data**.
 6. Finally, the list of dictionaries was converted into a **Pandas DataFrame** to form a structured table for analysis and export.
-

7. Data Export

- The final dictionary list (`all_car_data`) was converted into a Pandas DataFrame.
 - Columns included: *Car Name, Variant, Kilometers, Fuel Type, Transmission, EMI, and Price*.
 - The dataset was exported to a **CSV file** using `to_csv()` for easy access and further analysis.
 - The CSV file ensures portability and supports visualization in tools like Excel or Power BI.
-

8. Data Cleaning

- Removed duplicate and incomplete entries.
 - Cleaned unwanted symbols such as "₹", "KM", and unnecessary spaces.
 - Converted price and kilometers into numeric formats for analysis.
 - Renamed columns for readability and standardization.
 - Verified that each record contained valid and consistent information.
-

9. Data Analysis

- Checked the total number of Hyundai cars available in the dataset.
 - Performed Basic statistics for numerical columns such as Prices, Kilometers driven, & years.
 - Performed price analysis, fuel type distribution, transmission distribution, year analysis, kilometers analysis etc..
-

10. Data Visualization

- Used Matplotlib and seaborn for data visualization in python for the final dataset.
 - Used Charts like Histogram, Bar chart, Pie chart, Scatter plot for the visualization.
 - Taken Price, kilometer_driven, Year, Fuel_type distribution, transmission distribution using these charts.
-

11. Challenges and Solutions

Challenge	Solution Implemented
Dynamic content loading delayed full car data visibility	Used Selenium to scroll and load all car cards
Handling large webpage data	Implemented controlled scrolling and time delays.
Faced difficulty in managing and dividing the project tasks among 11 team members, which made coordination and progress tracking a bit challenging.	The project was divided into five parts , and two members were assigned to work on each part independently. The best version of each module was selected and integrated into the final project. Continuous communication and regular team discussions ensured smooth collaboration and timely project completion.

13. Conclusion

1. The project successfully scraped and processed car details from Cars24.com using Selenium and BeautifulSoup, and scraped key data fields such as Car_Name, Year, Fuel_Type etc...
 2. The data was cleaned and exported into a structured CSV format for analysis, The analysis provided meaningful insights into the pricing and usage trends of used Hyundai cars.
 3. This project enhanced technical proficiency in **web scraping, data preprocessing, and automation using Python.**
 4. Overall, it provided a valuable real-world experience aligned with data science and analytics practices.
-