

Web Scraping Project using Selenium and BeautifulSoup

- By Team A - Batch 35B - EvoAstra Internship Program

Introduction

Web Scraping is an automated process through which data is extracted from websites by parsing their HTML code. This project aims to extract and analyze car details from **AckoDrive** Website through Web Scraping technique.

For this project, we covered essential steps like:

- Website Research and Planning
- Data Extraction
- Data Cleaning
- Data Presentation

Website Research

We focused on gathering the data of Maruti Suzuki car from Mumbai area, covering key details such as:

- Name
- Fuel Type
- Transmission
- Price
- Year
- Rating
- Location

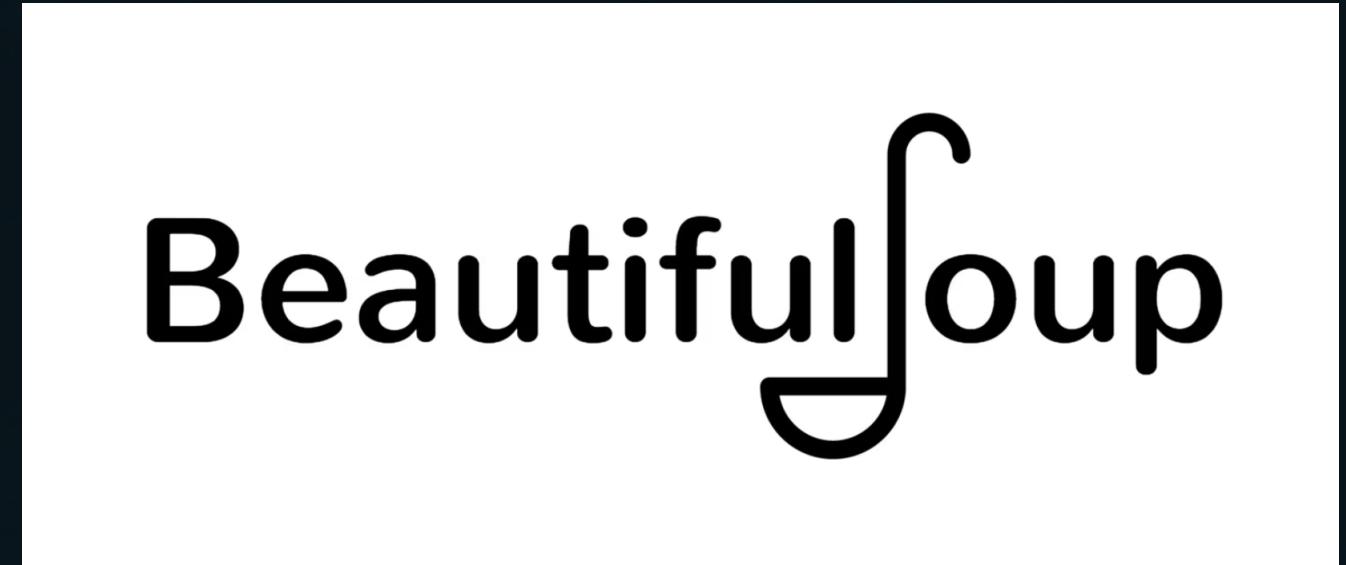
Extraction Planning

We planned to extract the specified data using the combination of Selenium and BeautifulSoup. This is because:



Selenium

AckoDrive uses dynamic JavaScript to load pages, and Selenium is preferred to open and extract real-time data.



BeautifulSoup

BeautifulSoup is fast at parsing HTML, supports CSS selectors and extracts clean text data.

Importing Modules

Imported necessary modules for data extraction like:

- Selenium
- BeautifulSoup
- Pandas
- Time
- Options

Visiting the Car Listings Page

Visited the AckoDrive website, with the location set to Mumbai and car preference set to Maruti Suzuki.

Load car cards on the website which load the content of each car.

Took and formatted the HTML content of the page to extract the data easily.

Data Extraction

1

Created an empty list to store the content of extracted car data.

2

Took a HTML tag of a car card, so that it can find all the others with the same CSS class.

3

Found and extracted the data of the necessary details in text format.

4

Created a dictionary for each car's content, which is later added to the empty list.

Data Cleaning and Presentation

1. Converted the list of dictionaries into a DataFrame
2. Saved it in the form of a CSV file
3. Removed empty columns which are no longer required
4. Replaced missing values with "NaN"

Conclusion

This project successfully demonstrated the use of Web Scraping and its tools that collect car information from the AckoDrive website. Using Selenium, the script loads all dynamically generated car listings by scrolling the page, then uses BeautifulSoup to extract key details such as name, fuel type, transmission, price, year, rating and location.

The extracted data is then converted into a CSV file and cleaned further for future use in various use cases such as Data Analysis, ML Prediction, etc. Thus, Web Scraping can be used to extract real-time data from an original website for efficient and reliable use of data.