Used Car Data Web Scraping Project Report

Internship Project

July 2025

1 Introduction

This project involves scraping used car data from the Cars24 website for Mumbai, covering five brands: Hyundai, Tata, Honda, Renault, and Ford. The objective is to collect 25 data points (5 models per brand) with details such as car name, model year, kilometers driven, fuel type, transmission type, price, and location. The data is processed, visualized, and saved as a CSV file using Python.

2 Methodology

2.1 Data Collection

The data was scraped from Cars24 using the following steps:

1. Defined URLs for each brand's listings in Mumbai:
   - Hyundai: https://www.cars24.com/buy-used-hyundai-cars-mumbai/?sort=bestmatch&serveWarrantyCount=true&listingSource=TabFilter&storeCityId=2378
   - Tata: https://www.cars24.com/buy-used-tata-cars-mumbai/?sort=bestmatch&serveWarrantyCount=true&listingSource=TabFilter&storeCityId=2378
   - Honda: https://www.cars24.com/buy-used-honda-cars-mumbai/?sort=bestmatch&serveWarrantyCount=true&listingSource=TabFilter&storeCityId=2378
   - Renault: https://www.cars24.com/buy-used-renault-cars-mumbai/?sort=bestmatch&serveWarrantyCount=true&listingSource=TabFilter&storeCityId=2378
   - Ford: https://www.cars24.com/buy-used-ford-cars-mumbai/?sort=bestmatch&serveWarrantyCount=true&listingSource=TabFilter&storeCityId=2378
2. Used requests to fetch HTML content for each URL.
3. Parsed HTML content using BeautifulSoup with html.parser.
4. Extracted fields: CarName, Model Year, Kilometers Driven, Fuel Type, Transmission Type, Price, and Location.

2.2 Data Cleaning

The scraped data was processed using pandas:

- Converted Kilometers_Driven to float.
- Converted Price to float.
- Saved the cleaned data as cars.csv.

2.3 Data Analysis

Visualizations were created using pandas, seaborn, and matplotlib:

- Horizontal bar plot of car counts by location using df['Location'].value$_c ounts().plot(kind =' barh').Scatterplotof Pricevs.$

3 Results

3.1 Dataset Summary

The dataset contains 25 cars with the following statistics (from df.describe()):

- Kilometers Driven:
  - Count: 25
  - Mean: 52.200800
  - Std: 26.554981
  - Min: 6.600000
  - Max: 110.000000
- Price:
  - Count: 25
  - Mean: 4.875600
  - Std: 1.435518
  - Min: 1.980000
  - Max: 7.640000

3.2 Key Findings

- The dataset includes 25 cars across five brands: Hyundai (5), Tata (5), Honda (5), Renault (5), and Ford (5).
- Locations such as Goregaon, Mumbai, and Regency Anantam, Dombivli East, have multiple car listings, as shown in the bar plot.
- The scatter plot of Price vs. Kilometers Driven shows no clear linear relationship, with prices ranging from 1.98 to 7.64 lakhs and kilometers driven from 6.6k to 110k km.

4 Conclusion

The project successfully scraped 25 used car listings from Cars24 for Mumbai, processed the data to convert numerical columns to appropriate types, and visualized location distribution and price vs. kilometers driven. The cleaned data was saved as cars.csv for further analysis.