

MINI PROJECT WEB SCRAPING CAR DETAILS

TEAM D: MARUTI SUZUKI

OVERVIEW

The objective of this mini-project is to develop skills in web scraping by extracting and analyzing car details from Cars24.com. We focused on gathering data specific to the Mumbai location, covering key details such as kilometers driven, year of manufacture, fuel type, transmission, and price.

We followed the flow of:

Research and planning	Data cleaning	Data extraction
-----------------------	---------------	-----------------

STEPS TAKEN

1. Understand structure of the website and HTML elements
2. Extraction of the required data.
3. Import Libraries – Loaded essential libraries like pandas, numpy, and re (for regex cleaning).
4. Load the Dataset – Read the dataset (Maruti_Suzuki_Cars_datai.csv) using `pd.read_csv()`.
5. Standardize Column Names
6. Remove Duplicates – Use `drop_duplicates()` to eliminate repeated rows.
7. Handle Missing Values – Replace empty strings, 'NA', or other placeholders with `np.nan` and drop rows where critical columns are missing.
8. Clean Complex Columns
9. Convert Types – Convert numeric-like text fields (e.g., “3.5 lakh”, “51.5k km”) into numerical values using functions and regex replacements.

10. Handle Outliers – Identified and remove extreme values using quantile filtering.
11. Validate and Inspect – Reviewed cleaned data with `info()`, `describe()`, and `head()`.
12. Export Cleaned Data – Save the final dataset as a new CSV file (`maruti_Suzuki_cars_data_cleaned.csv`).

INSIGHTS GAINED

Some of the insights gained from the scatter plots created from the data were:

General Price Trends

Car prices show a clear positive correlation with manufacturing year — newer models (2022–2024) typically sell between 6–13 lakh, while older ones (2010–2014) often fall below 2 lakh. Compact sedans and SUVs (like Ciaz, Ertiga, Brezza) maintain higher resale values compared to hatchbacks (like Alto, Celerio, Wagon R).

Mileage (KM Driven) Patterns

Cars driven less than 50,000 km usually have noticeably higher prices, often 25–35% more than similar models with higher usage. Extremely high mileage listings (above 1 lakh km) typically see large value drops, even among newer models.

Outlier Observations

Some listings with unusually high kilometers (1.5L–3L km) still hold moderate prices, possibly due to model durability perception in Maruti's brand appeal.