**Title**

Used Car Listings Scraping and Analysis from Cars24

**1. Objective**

The goal of this project was to extract, clean, and analyze used car listings for a specific brand across multiple cities from the Cars24 platform.  
We aimed to:

* Build a robust **web scraping pipeline** (Selenium + BeautifulSoup).
* Collect listings for a target brand across selected cities.
* Structure raw listing text into meaningful attributes: **year, brand/model, kilometers driven, fuel type, transmission, ownership, price, and location**.
* Generate exploratory insights into used car availability and pricing trends.

**2. Data Collection**

* **Source:** Cars24 (brand-specific + city-specific URLs).
* **Tools Used:**
  + Python 3.12
  + Selenium for automated browsing and dynamic page loading.
  + BeautifulSoup for HTML parsing.
  + pandas for data wrangling.
* **Scope:** One brand scraped across **3 cities** (Mumbai, Pune, Thane).
* **Size:** Final dataset contains **N = {len(df)} rows** (number from your file).

**3. Data Schema**

The cleaned dataset contains the following columns:

| **Column** | **Description** |
| --- | --- |
| Year | Manufacturing year of the car |
| Make | Brand variant (e.g., City) |
| Model | Engine details |
| Price | Listed price (in INR) |
| KM Driven | Kilometers driven |
| Fuel Type | Fuel type (Petrol/Diesel/CNG/Electric) |
| Transmission | Manual / Automatic |
| Location\_code | Pincode |
| Location | Area of the city |
| Link | Hyperlink of the listing |

**4. Methodology**

1. **Web Scraping:**
   * Dynamic rendering handled by Selenium with headless Chrome.
   * Multiple city-wise URLs looped with pagination support.
2. **Cleaning & Extraction:**
   * Regex rules to detect year, kms, owners.
   * Token matching for fuel and transmission.
   * Price normalized to integer values.
3. **Storage:**
   * All listings combined into one CSV.
   * Preserved raw\_text for reproducibility.

**5. Key Insights**

* **Average Price:** 5.6 lakhs INR approx..
* **Most Common Model:** Honda City
* **Average Age:** 7.5 years (based on year vs current year)
* **Fuel type trend: Petrol dominates across all cities.**
  + **Fuel Split:** 88.75% Petrol, 6.25% Diesel, 5% CNG.
* **Transmission: Majority are Manual, fewer Automatic.**
* **City-level Insights:**
  + Delhi has the largest supply but Bangalore has higher average prices (6.4 lakhs).
  + Delhi has relatively newer cars.
  + Mumbai shows more budget-friendly listings.

**6. Challenges**

* Dynamic content loading required **delays and retry logic**.
* Parsing inconsistencies (data format required cleaning).
* Handling multiple city formats in URLs.

**7. Conclusion**

* A working **end-to-end scraping + analysis pipeline** has been built.
* Dataset provides actionable insights into used car trends.
* Can be extended to multiple brands/cities with minimal modification.

**8. Future Work**

* Automate data refresh (daily/weekly scrape).
* Expand to multiple brands.
* Integrate ML models for **price prediction** based on car attributes.
* Build dashboards for visualization (Streamlit/PowerBI).