**Web Scraping Mini Project Report: Cars24 Used Car Listings from Honda Company**

**1. Process Overview**

**a) Objective:** To Scrape Used Car listings from **cars24** for three major cities in India

**(Mumbai, Bangalore, Delhi)** and create a cleaned dataset for further analysis

1. **Tools Used:**
   * **Selenium**: To render dynamic content and handle scrolling/pagination.
   * **BeautifulSoup**: To parse HTML and extract structured data.
   * **Pandas**: To organize and save the scraped data into CSV.
2. **Steps:**
   * Open Cars24 search results using Selenium for selected cities.
   * Scroll through pages to ensure all listings load.

**2. Challenges Faced**

**a) Dynamic Content Loading:**

* Cars24 loads listings via JavaScript, so plain requests returned empty pages.

**Solution:** Used Selenium to render the page fully before parsing.

**b) Infinite Scrolling and Paging:**

* Listings did not appear all at once.

**Solution**: Implemented automated scrolling and navigation to multiple pages.

1. **Changing HTML Structure:**

* Class names like styles\_carCardWrapper\_\_sXLIp were obscure and could change.

**Solution:** Used flexible selectors (CSS + regex) and fallbacks for missing tags

1. **Data Cleaning Issues:**

* Price, KM info came in inconsistent formats (e.g., "45,000 KM", "₹ 5.5 Lakh").

**Solution:** Implemented safe parsing with regex and default values

1. **Performance & Reliability:**

* Selenium is slower than static scraping. Long sessions sometimes crashed

**Solution**: Added delays (time. sleep), limited page count, and ensured proper browser closure (driver. quit ()).

**3. Solutions Implemented**

* + **Pipeline Design:** Selenium (render pages) → BeautifulSoup (parse HTML) → Safe Parsing Functions (clean text) → Pandas (structured dataset).
  + **Error Handling:** Try-except blocks for missing values.
  + **Data Cleaning:** Regex for extracting numeric values from strings.
  + **Exporting Data:** Saved CSV with UTF-8 encoding to avoid errors in Google Colab.

**4.Insights Gained**

**From Data Itself: -** **Market/Business Insights:**

* Fuel type trend: Petrol dominates across all cities.
* Transmission: Majority are Manual, fewer Automatic.
* Location impact: 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.

**Process + Technical Insights:**

* + Car listings include structured but inconsistent data, making safe parsing essential.
  + Selenium is effective for JS-heavy websites but requires optimization for speed.
  + Data suggests that attributes like KM Driven, Year strongly influence car pricing.
  + The scraper can be extended to more cities/brands with minor modifications.