# Research Report: Craigslist Car Listing Scraper

## Gear Trybe Inc. - Python Developer Intern

## Muhammad Haroon

### Executive Summary

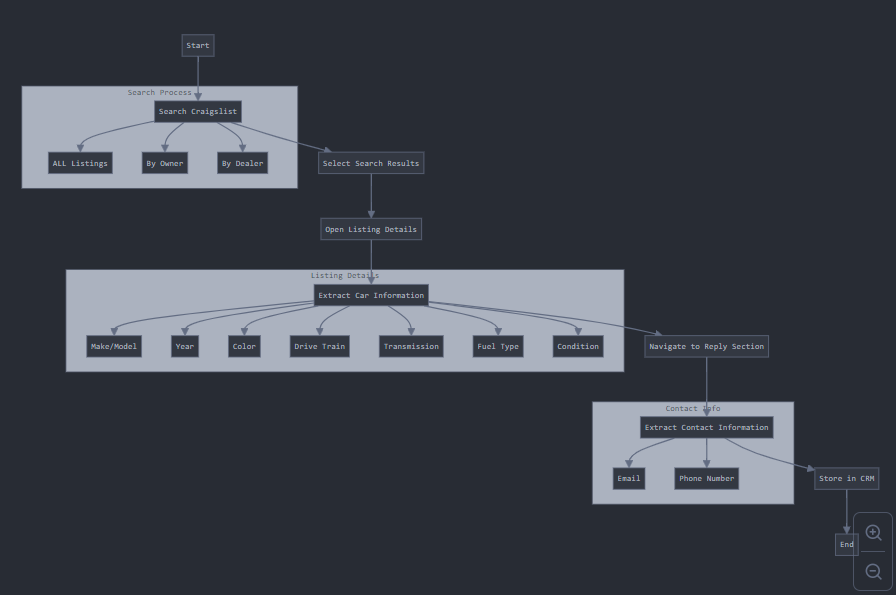
This report outlines the findings of research conducted to develop a web scraper for extracting contact information from car listings on sfbay.craigslist.org. The proposed scraper will automate the process of gathering data on vehicles listed as “needs ${autopart}” OR “need ${autopart}”, facilitating efficient lead generation for Gear Trybe Inc.

### 1. Introduction

Gear Trybe Inc. aims to extract contact information from individuals selling cars on Craigslist. This report details the research findings and proposes a methodology for developing an effective web scraper to accomplish this task.

### 2. Methodology

The research focused on identifying the key steps required to scrape relevant data from Craigslist listings. The process can be broken down into four main stages:



1. Search Query Execution

2. Result Selection

3. Detailed Listing Information Extraction

4. Contact Information Retrieval

### 3. Findings

#### 3.1. Search Query Execution

Three distinct URLs can be used on user choice for searching listings:

* ALL listings: https://sfbay.craigslist.org/search/sss?excats=5-2-13-22-26-1-26-1-1-1-3-6-11-1-5-8-1-1-1-1-1-4-1-7-1-10-2-2-2-1-1-1-1-1-1-2-3-1-1-2-2-1-1-2-1-2-1-1-1-1-1-1-3-1-1-1-1-1-4-1&query=needs%20engine
* By Owner: https://sfbay.craigslist.org/search/sss?excats=5-2-13-22-26-1-26-1-1-1-3-6-11-1-5-8-1-1-1-1-1-4-1-7-1-10-2-2-2-1-1-1-1-1-1-2-3-1-1-2-2-1-1-2-1-2-1-1-1-1-1-1-3-1-1-1-1-1-4-1&purveyor=owner&query=needs%20engine
* By Dealer: https://sfbay.craigslist.org/search/sss?excats=5-2-13-22-26-1-26-1-1-1-3-6-11-1-5-8-1-1-1-1-1-4-1-7-1-10-2-2-2-1-1-1-1-1-1-2-3-1-1-2-2-1-1-2-1-2-1-1-1-1-1-1-3-1-1-1-1-1-4-1&purveyor=dealer&query=needs%20engine

Note: The query “needs engine” is URL-encoded as “needs%20engine”.

#### 3.2. Result Selection

To select individual listings from the search results, the CSS selector can be used:

* CSS Selector: #search-results-page-1 > ol > li:nth-child(1) > div > a
* In selenium:

{ driver.find\_element\_by\_css\_selector(“ ") }

&

{ .get\_attribute('href') }

The href attribute of the selected element contains the URL for the detailed listing page.

#### 3.3. Detailed Vehicle Information Extraction

* The following potential information can be extracted from each listing page:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Make/Model | Year | Part | Color | Litre | Drive Train | Transmission | Fuel Type | Condition | VIN |

#### 3.4. Contact Information Retrieval

Contact information can be found in the reply section of each listing the CSS selector can be used:

* CSS Selector: #\34 336a589150ebebebb3b4d4fd160dbe55129de87 > div.reply-email-address > a
* In selenium:

{ driver.find\_element\_by\_css\_selector(“ ") }

&

{ .get\_attribute('href') }

The href attribute of this element contains the mailto link with the contact email address.

### 4. Technologies to Use:

* selenium==4.23.0
* fastapi==0.111.1
* uvicorn==0.30.1
* webdriver-manager==4.0.1
* chromedriver-win64

### 5. Conclusion

The proposed Craigslist car listing scraper has the potential to significantly streamline Gear Trybe Inc.’s lead generation process for vehicles needing engine work. By automating the extraction of listing details and contact information, the company can more efficiently identify and reach out to potential clients.

### 6. Next Steps

1. Project Initialization on Project Manager.
2. Begin implementation of the Python script based on the research findings.
3. Conduct thorough testing to ensure accuracy and reliability of the scraper.