# Craigslist Scraper Research Report

## Gear Trybe Inc. - Python Developer Intern Project Muhammad Haroon

### Overview

This report summarizes the initial research and development of a web scraper designed to extract contact information from car listings on sfbay.craigslist.org. The project aims to automate the process of gathering seller information for potential business opportunities.

### Current Implementation

The current implementation uses Selenium WebDriver to navigate and interact with the Craigslist website. Key features include:

1. **Search Functionality**: The scraper can perform searches based on a given query and seller type (all, owner, or dealer).
2. **Pagination**: It navigates through multiple pages of search results.
3. **Page Navigation**: It opens each listing in a new tab, allowing for potential data extraction.
4. **Error Handling**: Implements retry mechanisms for handling stale elements and timeouts.
5. **Performance Optimization**: Disables image loading to improve speed.

### Technical Details

* **Language**: Python
* **Main Library**: Selenium WebDriver
* **Browser**: Chrome (with ChromeDriver)
* **Key Classes Used**:
  + webdriver.Chrome
  + WebDriverWait
  + expected\_conditions

### Considerations

1. The current implementation does not actually extract contact information.
2. There’s no data storage mechanism implemented yet.
3. The scraper may be detected as automated behavior by Craigslist.
4. Execution time may be significant due to the sleep intervals.