# Insurance Premium Web Scraping Readme

## Overview

This Python script is designed for web scraping insurance premium information from the major NZ insurance companies’ websites. It utilizes the Selenium library to automate interactions with the website and extract insurance premium data based on provided input data.

## Instructions

### Copy this folder onto your local device

* Copy the folder onto a location in your local device (your PC or any other PC you plan to run this scraper on)

Install Python

* Download Python: If you don't have Python installed, download it from the official website: [Python Downloads](https://www.python.org/downloads/).
  + Install Python: Follow the installation instructions for your operating system.
  + NOTE: Make sure you add python to your file path (this is an option in the python installer, if you don’t do this, then uninstall and reinstall with this enabled)

### Dependencies

Before running the script, ensure you have the necessary packages installed. You can install them using the following commands in the command line (cmd):

* pip install *selenium*
* pip install *webdriver\_manager*
* pip install *pandas*
* pip install *openpyxl*

**Note: If you have not downloaded python on your device and added, pip will not work**

### WebDriver

The script uses the Chrome WebDriver. *Make sure you have Google Chrome installed on your machine*. (Google chrome (the standard webbrowser) needs to be downloaded for chromedriver to function)

### ChromeDriver

ChromeDriver is required for Selenium to automate Chrome. The script uses ChromeDriverManager to automatically download the appropriate version of ChromeDriver.

* Chromedriver should be automatically downloaded with *webdriver\_manager*

### Input Data

The test data input files (for both home and auto insurance) need to have their file paths defined. This is done at the top of the python file under the header “File path definitions”

* Notes:
  + Two backslashes (\\) are required for to define a backslash character (\) in a python string. This is because \ is an escape character in python strings (meaning it allows us to note special character, such as \n newline character). Therefore the 1st \ is just defining that the 2nd \ is actually just a character.
  + Templates for the test data input files can be found here: [S:\Library\IQS\Test data](file:///S:\Library\IQS\Test%20data).

### Running the Script

1. Turn on VPN (IPvanish) to prevent IP from being blocked by the websites
2. Either…
   1. run the script from the command prompt (cmd) - <https://www.wikihow.com/Use-Windows-Command-Prompt-to-Run-a-Python-File>
   2. Or download a python coding environment such as VScode (download VScode: <https://code.visualstudio.com/download>) and then open the file with VScode (or any other environment) and press the play button in the top right.

A screenshot of a computer screen

Description automatically generated

1. The script will open a Chrome browser, navigate to each of the insurance websites, input the data for each potential individual from the spreadsheet, and scrape the insurance premiums.

### Output

Outputs to a csv called ‘scraped\_auto\_premium.csv’.

* **NOTE: MAKE SURE this file is NOT OPEN while the code is running or the scraped premiums cannot be written to it.**

### Notes

* The script uses explicit waits to handle dynamic page loading [time.sleep(‘seconds’)]. The wait times may need adjustment based on your internet speed.
* This script is provided as-is and may require adjustments based on changes to the AMI website structure. Use it responsibly and respect the website's terms of service.

### Detailed Code Explanation