
HAPPIFY INC.

Developed by-

Akshay Munde (amunde@andrew.cmu.edu)

Pranav Banthia (pbanthia@andrew.cmu.edu)

Prasun Shreshta (pshresth@andrew.cmu.edu)

Sonal Singh (sonals@andrew.cmu.edu)

Varsha Jain (varshaj@andrew.cmu.edu)

Instructions:

This file contains a step-by-step procedure to install and run the project “Happify,Inc.”

Link for the demo video:

Please refer to this youtube video containing a demonstration for running the project module along with output visuals.

<https://www.youtube.com/watch?v=Ku5LKrmLdoA&feature=youtu.be>

Installation of additional modules:

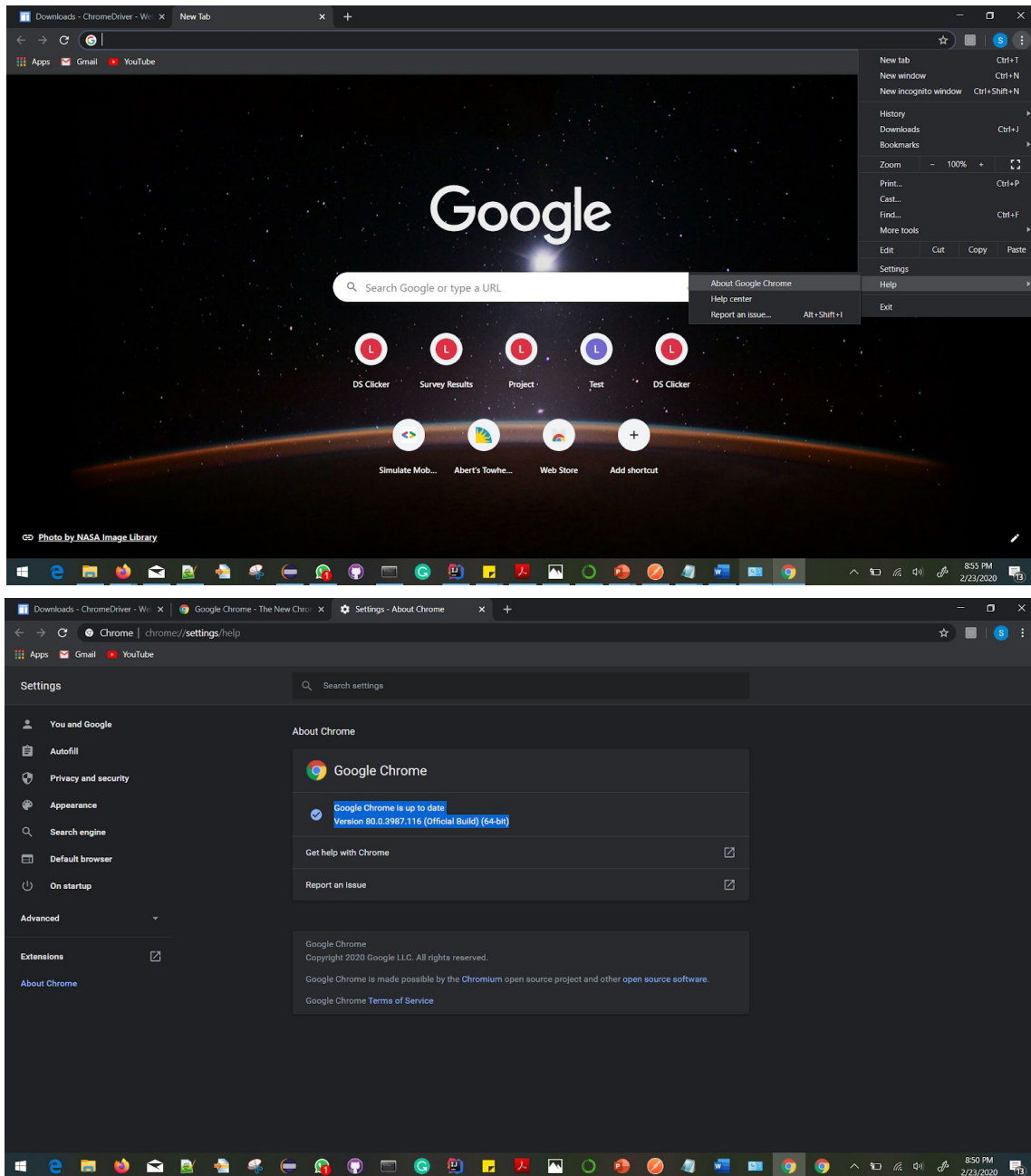
1. **Selenium Web Driver**
2. **plotly module**

The project requires following software to be installed in the system:

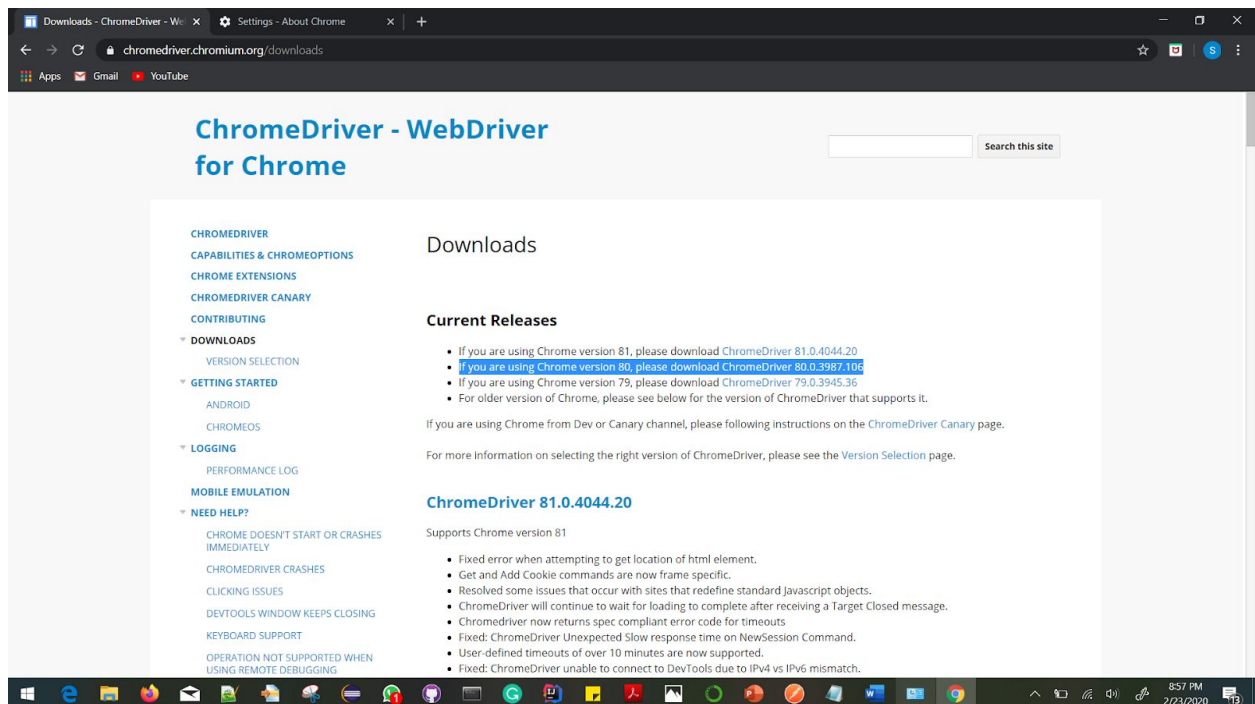
- [Google Chrome](#)
- [Selenium Web Driver for Google Chrome](#)
- [Anaconda3 2019.10](#) (Python 3.7)

Steps to install Selenium Web Driver:

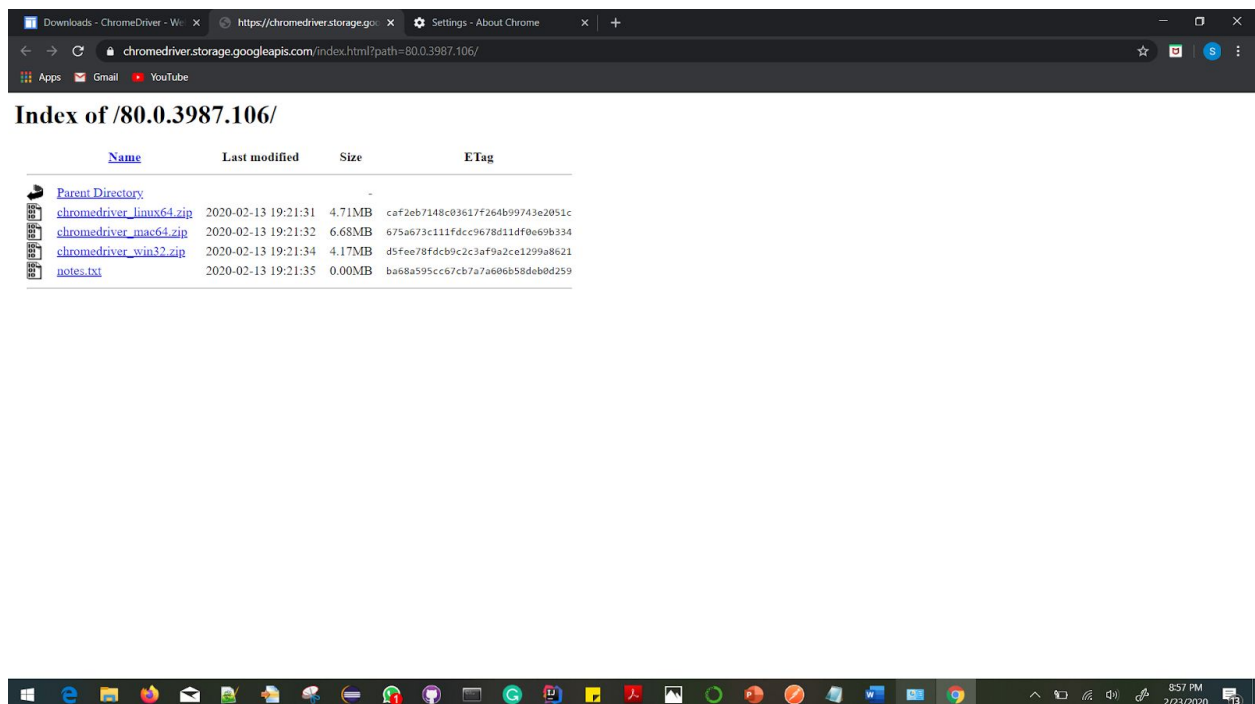
1. Open Chrome. Go to Help> About Google Chrome and check for the version for Google Chrome. Note this version while choosing the version for Web Driver



2. Open the Url: <https://chromedriver.chromium.org/downloads> for downloading the ChromeDriver and select the version from current releases

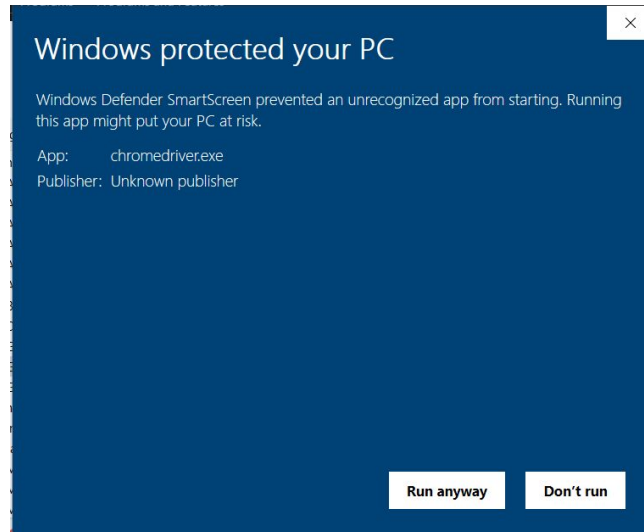


3. Select the .zip file according to your Operating system to download the driver



4. Open the zipped file (chromedriver_win32.zip in case of windows) and install the chromedriver application

- Click 'Run anyway'



- You will likely see the following message on the command prompt:

```
C:\Users\sonal\AppData\Local\Temp\Temp1_chromedriver_win32.zip\chromedriver.exe
Starting ChromeDriver 80.0.3987.106 (f68069574609230cf9b635cd784cfb1bf81bb53a-refs/branch-heads/3987@{#882}) on port 9515
Only local connections are allowed.
Please protect ports used by ChromeDriver and related test frameworks to prevent access by malicious code.
```

If you do, please save the path location of the chromedriver you had downloaded, as you will need to use the pathway to use Selenium later in the code.

5. Open Anaconda Prompt Powershell and run the command:

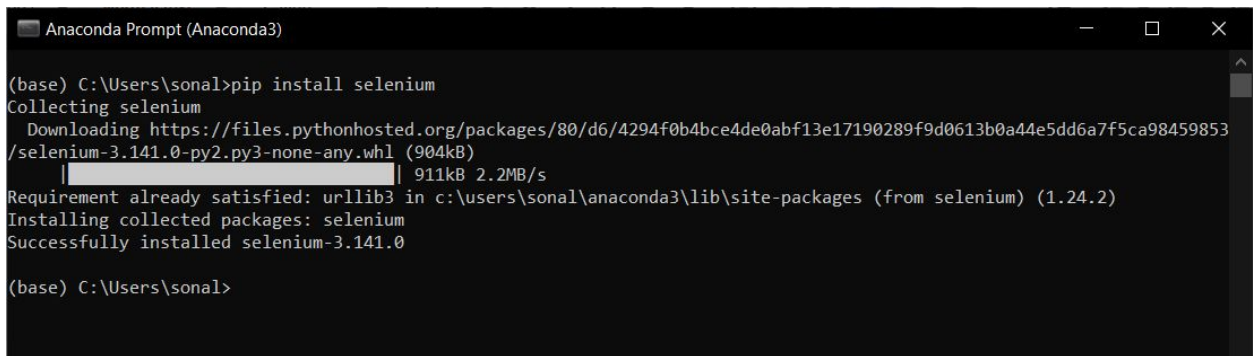
-> **pip install selenium**



```
Anaconda Prompt (Anaconda3)

(base) C:\Users\sonal>pip install selenium
```

6. This will install the latest version on selenium:



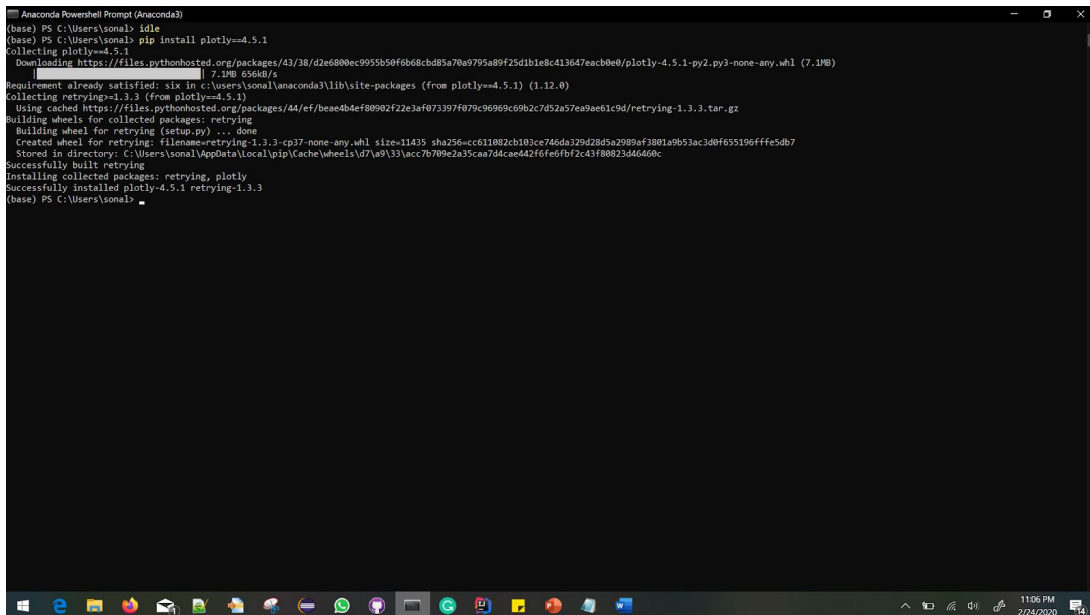
```
Anaconda Prompt (Anaconda3)

(base) C:\Users\sonal>pip install selenium
Collecting selenium
  Downloading https://files.pythonhosted.org/packages/80/d6/4294f0b4bce4de0abf13e17190289f9d0613b0a44e5dd6a7f5ca98459853/selenium-3.141.0-py2.py3-none-any.whl (904kB)
    | 911kB 2.2MB/s
Requirement already satisfied: urllib3 in c:\users\sonal\anaconda3\lib\site-packages (from selenium) (1.24.2)
Installing collected packages: selenium
Successfully installed selenium-3.141.0

(base) C:\Users\sonal>
```

7. Install plotly module from Anaconda Powershell Prompt (Anaconda 3) using the following command:

pip install plotly==4.5.1

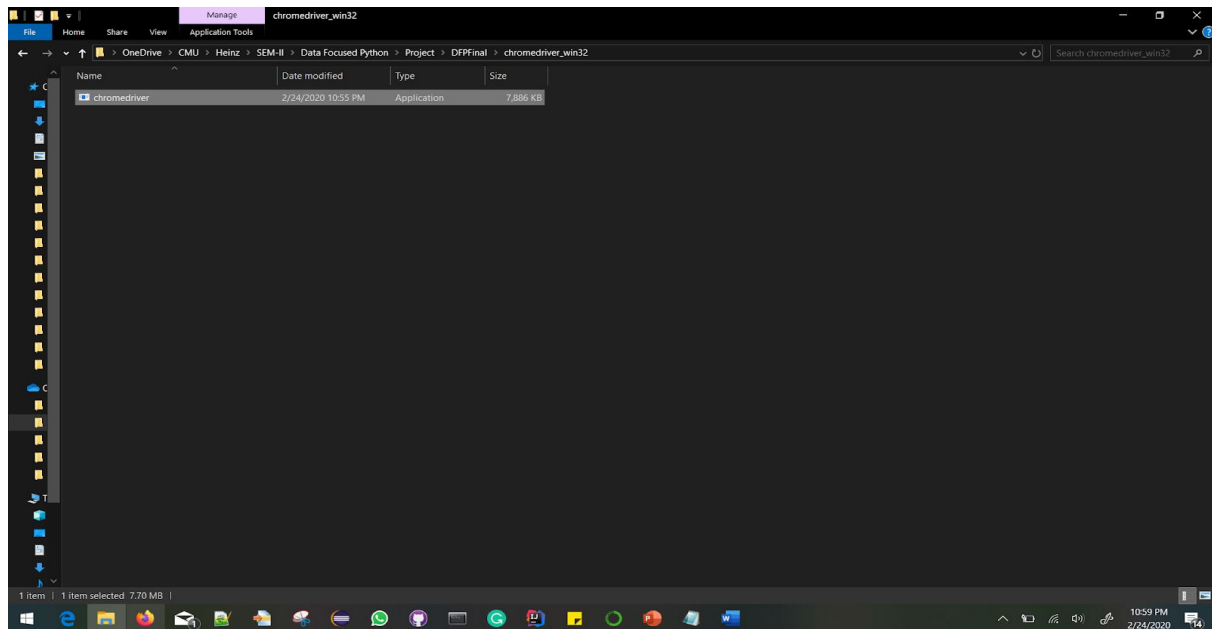


```
Anaconda Powershell Prompt (Anaconda3)

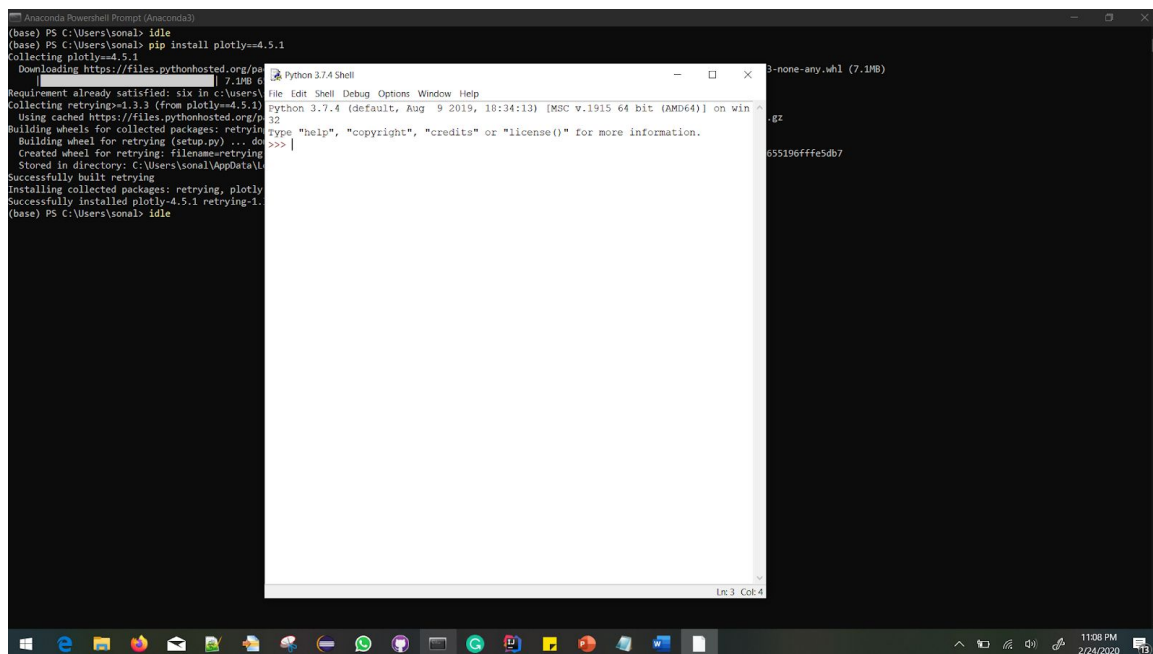
(base) PS C:\Users\sonal> idl
(base) PS C:\Users\sonal> pip install plotly==4.5.1
Collecting plotly==4.5.1
  Downloading https://files.pythonhosted.org/packages/41/38/d2e6800ec9955b50f6b68cd85a70a9795a80f25d1b1e8c413647each0e/plotly-4.5.1-py2.py3-none-any.whl (7.1MB)
    | 7.1MB 656kB/s
Requirement already satisfied: six in c:\users\sonal\anaconda3\lib\site-packages (from plotly==4.5.1) (1.12.0)
Collecting retrying==1.3.3 (from plotly==4.5.1)
  Using cached https://files.pythonhosted.org/packages/44/ef/beae4bdef80902f22e3af073397f079c9696c69b2c7d52a57ea9ae61c9d/retrying-1.3.3.tar.gz
Building wheels for collected packages: retrying
  Building wheel for retrying (setup.py) ... done
  Created wheel for retrying: filename=retrying-1.3.3-cp37-none-any.whl size=11435 sha256=cc611802cb103ce746da329d28d5a2989af3801a9b51ac3d0f655196fffe5db7
  Stored in directory: C:\Users\sonal\AppData\Local\pip\Cache\wheels\d7\aa\33\acc7b709e2a35caa7d4cae442f6f6b72c43f8882d46460c
Successfully built retrying
Installing collected packages: retrying, plotly
Successfully installed plotly-4.5.1 retrying-1.3.3
(base) PS C:\Users\sonal>
```

Running the project:

1. After installing the above software, download the project.



2. Open Idle

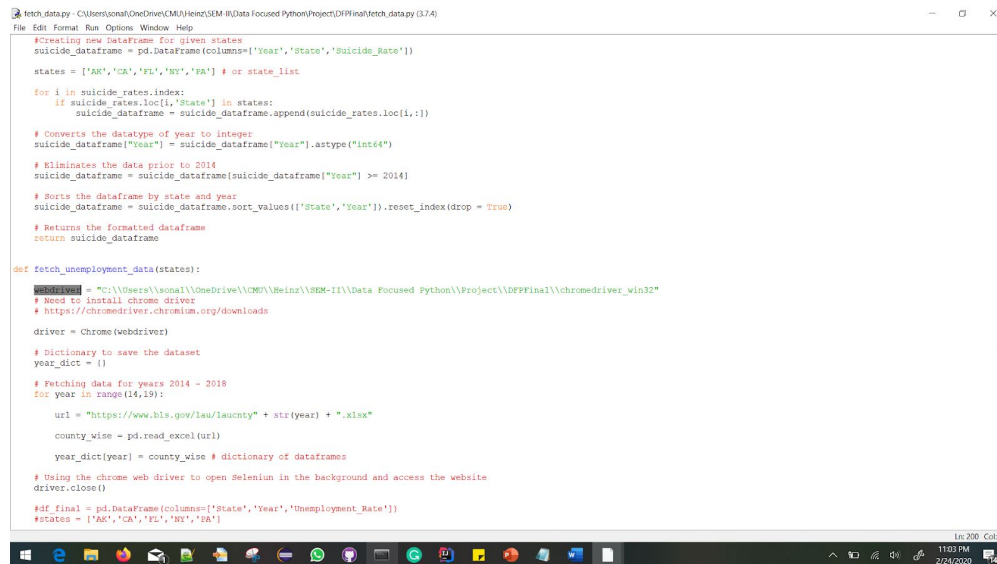


3. Update the path of the chrome driver in the fetch_data.py file

Path in local: **C:\Users\sonal\OneDrive\CMU\Heinz\SEM-II\Data Focused Python\Project\DFPFinal\chromedriver_win32**

Replace `'\'` by `\\`

webdriver = "C:\\Users\\sonal\\OneDrive\\CMU\\Heinz\\SEM-II\\Data Focused Python\\Project\\DFPFinal\\chromedriver_win32"



```
fetch_data.py - C:\Users\sonal\OneDrive\CMU\Heinz\SEM-II\Data Focused Python\Project\DFPFinal\fetch_data.py (3.7.4)
File Edit Format Run Options Window Help

#Creating new DataFrame for given states
suicide_dataframe = pd.DataFrame(columns=['Year', 'State', 'Suicide_Rate'])

states = ['AK', 'CA', 'FL', 'NY', 'PA'] # or state_list

for i in suicide_rates.index:
    if suicide_rates.loc[i, 'State'] in states:
        suicide_dataframe = suicide_dataframe.append(suicide_rates.loc[i, :])

# Converts the datatype of year to integer
suicide_dataframe['Year'] = suicide_dataframe['Year'].astype("int64")

# Eliminates the data prior to 2014
suicide_dataframe = suicide_dataframe[suicide_dataframe['Year'] >= 2014]

# Sorts the dataframe by state and year
suicide_dataframe = suicide_dataframe.sort_values(['State', 'Year']).reset_index(drop = True)

# Returns the formatted dataframe
return suicide_dataframe

def fetch_unemployment_data(states):
    webdriver = "C:\\Users\\sonal\\OneDrive\\CMU\\Heinz\\SEM-II\\Data Focused Python\\Project\\DFPFinal\\chromedriver_win32"
    # Need to install chrome driver
    # https://chromedriver.chromium.org/downloads
    driver = Chrome(webdriver)

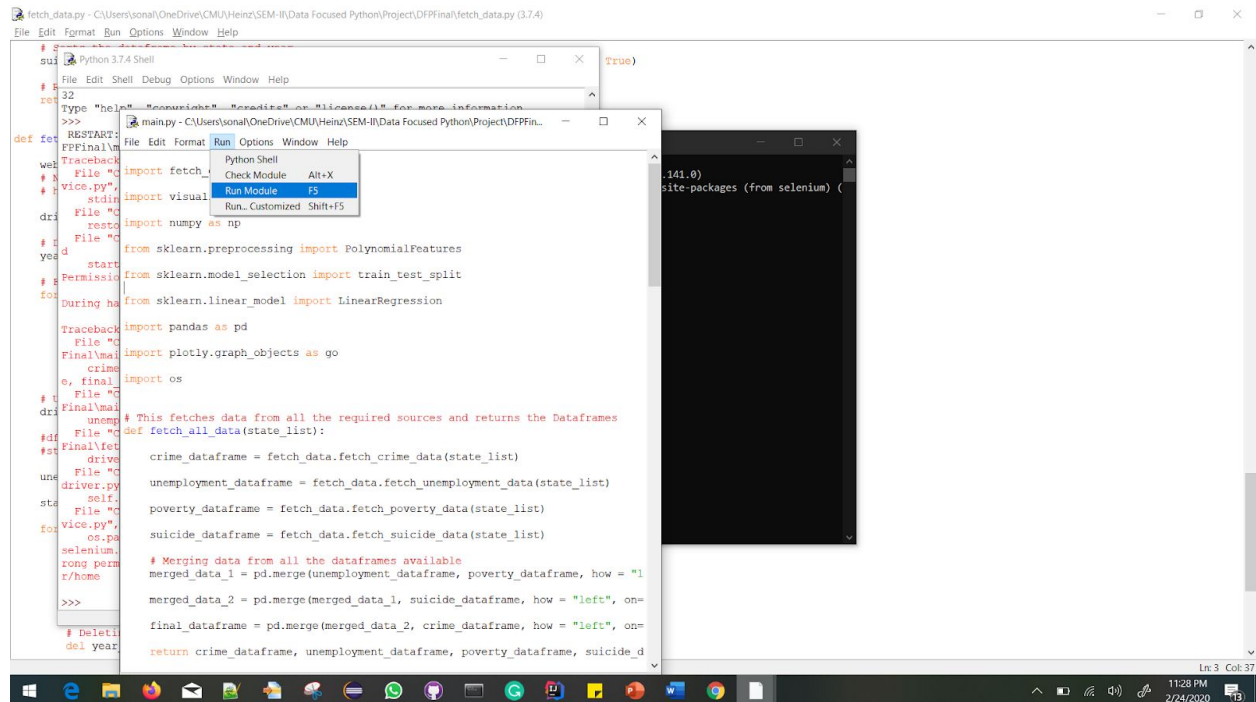
    # Dictionary to save the dataset
    year_dict = {}

    # Fetching data for years 2014 - 2018
    for year in range(2014, 2019):
        url = "https://www.bls.gov/lau/laucnty" + str(year) + ".xlsx"
        county_wise = pd.read_excel(url)
        year_dict[year] = county_wise # dictionary of dataframes

    # Using the chrome web driver to open Selenium in the background and access the website
    driver.close()

    #df_final = pd.DataFrame(columns=['State', 'Year', 'Unemployment_Rate'])
    #states = ['AK', 'CA', 'FL', 'NY', 'PA']
```

4. Run the module, main.py in idle:



```
main.py - C:\Users\sonal\OneDrive\CMU\Heinz\SEM-II\Data Focused Python\Project\DFPFinal\main.py (3.7.4)
File Edit Format Run Options Window Help

# Importing modules
import sys
import os
import pandas as pd
import numpy as np
from sklearn.preprocessing import PolynomialFeatures
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
import plotly.graph_objects as go
import os

# This fetches data from all the required sources and returns the Dataframes
def fetch_all_data(state_list):
    crime_dataframe = fetch_data.fetch_crime_data(state_list)
    unemployment_dataframe = fetch_data.fetch_unemployment_data(state_list)
    poverty_dataframe = fetch_data.fetch_poverty_data(state_list)
    suicide_dataframe = fetch_data.fetch_suicide_data(state_list)

    # Merging data from all the dataframes available
    merged_data_1 = pd.merge(unemployment_dataframe, poverty_dataframe, how = "left", on = "State")
    merged_data_2 = pd.merge(merged_data_1, suicide_dataframe, how = "left", on = "State")
    final_dataframe = pd.merge(merged_data_2, crime_dataframe, how = "left", on = "State")

    return crime_dataframe, unemployment_dataframe, poverty_dataframe, suicide_dataframe
```