APP SCRAPING PROJECT REPORT

BY: Kartikey Rastogi

TOOLS AND LIBRARIES USED:-

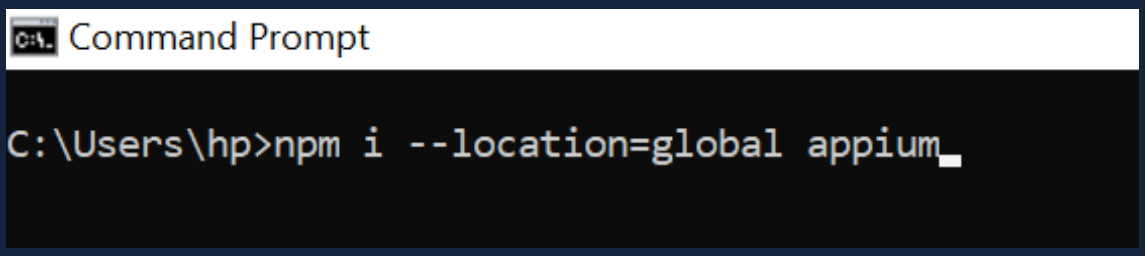
- Appium - Server GUI
- Appium Inspector
- Selenium
- Android Studio (Optional)
- Pandas
- Numpy
- Time
- Regex

IDES USED:-

- PyCharm
- VS Code(Jupyter Notebook)

SETUP REQUIREMENTS

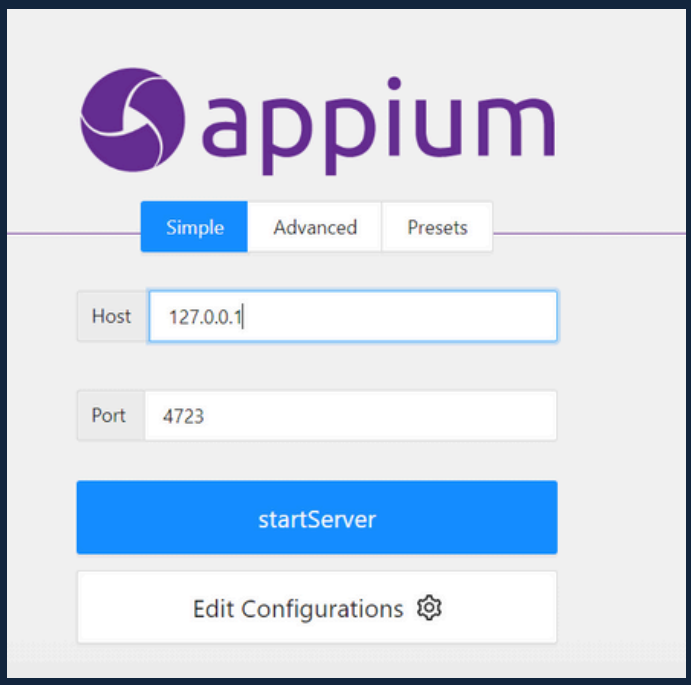
- JRE and NPM
- Appium system installation:-



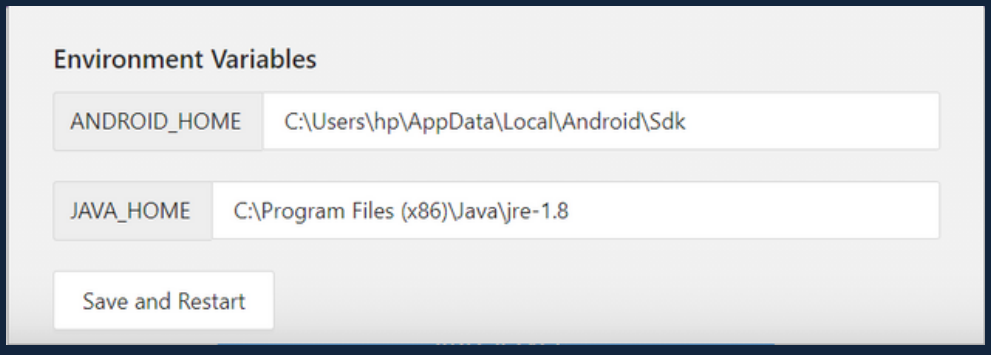
- Download and install Appium GUI:-
  - https://github.com/appium/appium-desktop/releases/tag/v1.22.3-4
- Download and install Appium Inspector:-
  - https://github.com/appium/appium-inspector/releases
- Setup system Environment Variables:-
  - Copy the file path of java and appium and paste it in the environment variables.
- Download android studio:-
  - Make sure to download adb from the SDK Manager
  - Copy the path and paste it in the environment variables
- Connect a mobile device:-
  - Create a virtual device from android studio, or
  - Connect a mobile via USB and turn on the developer options, make sure to enable USB Debugging.

STEPS INVOLVED

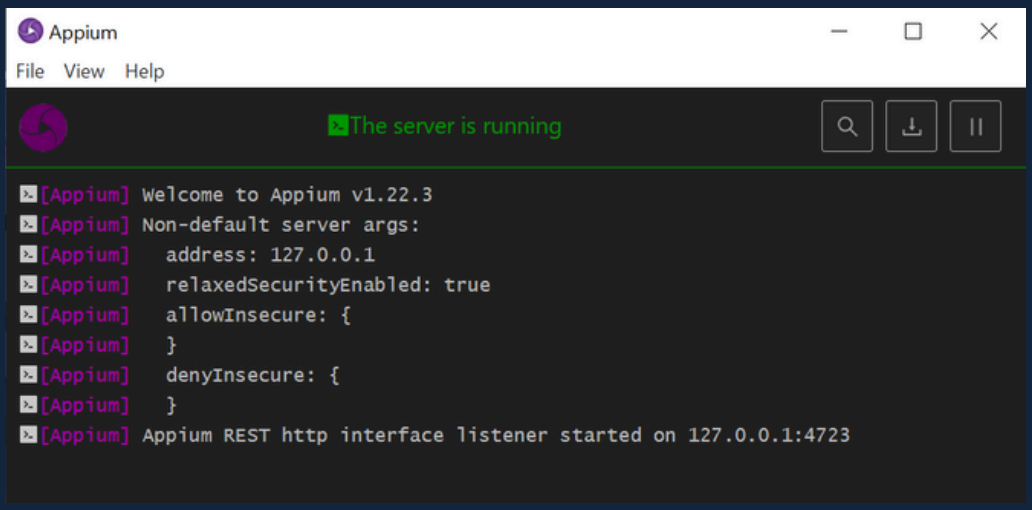
- Once the android device is setup, go to cmd and check if the device is properly connected.
- Type “adb devices” in the cmd to check the connected devices, it will give you the name and the status of the devices that are connected.
- Open the Appium GUI and start the appium server:-
  - The general host is the local host: 127.0.0.1,
  - The general port is 4723



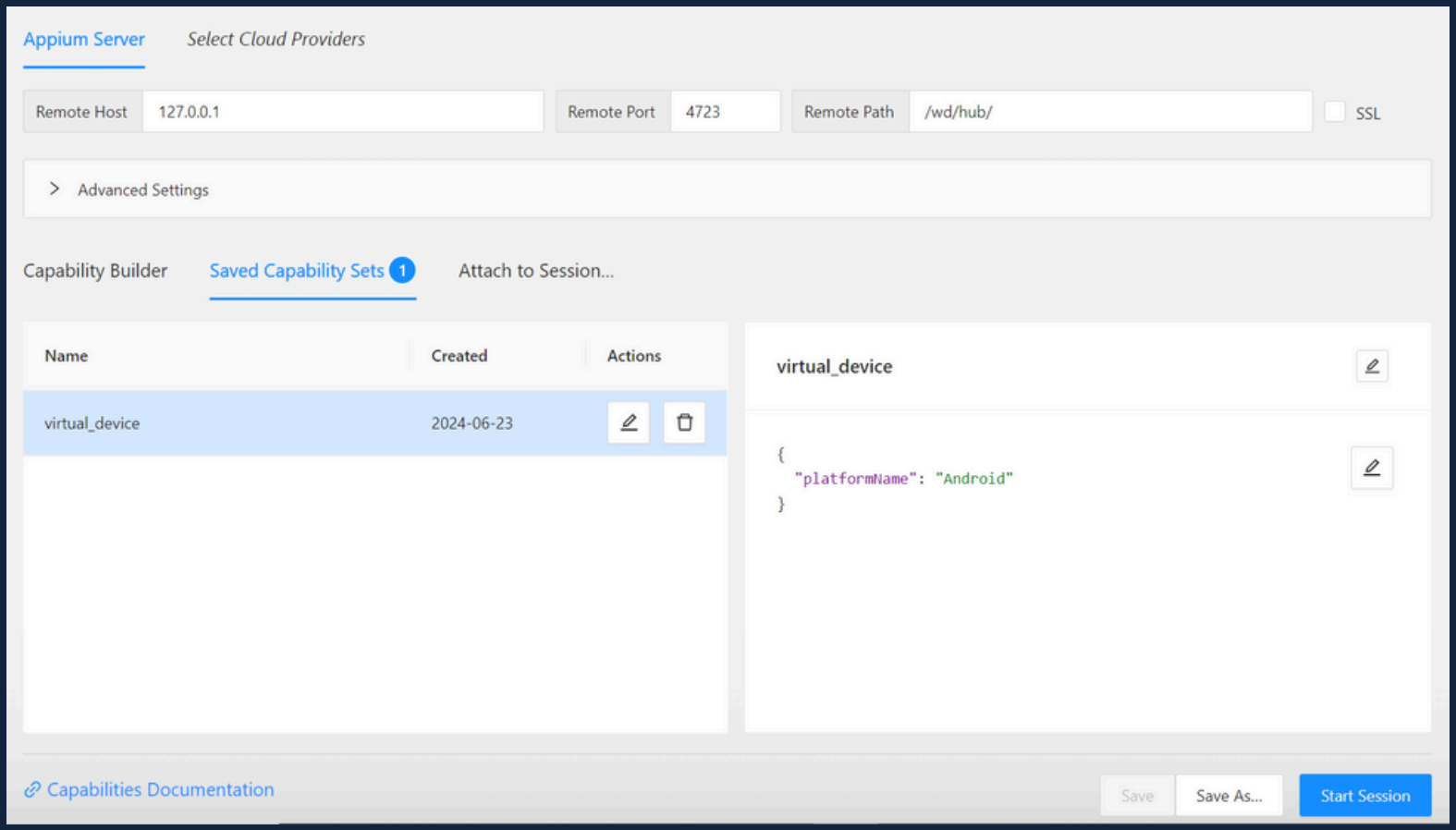
- Make sure to paste the path of your android studio SDK and Java in the Edit Configurations tab, if it is not done automatically.



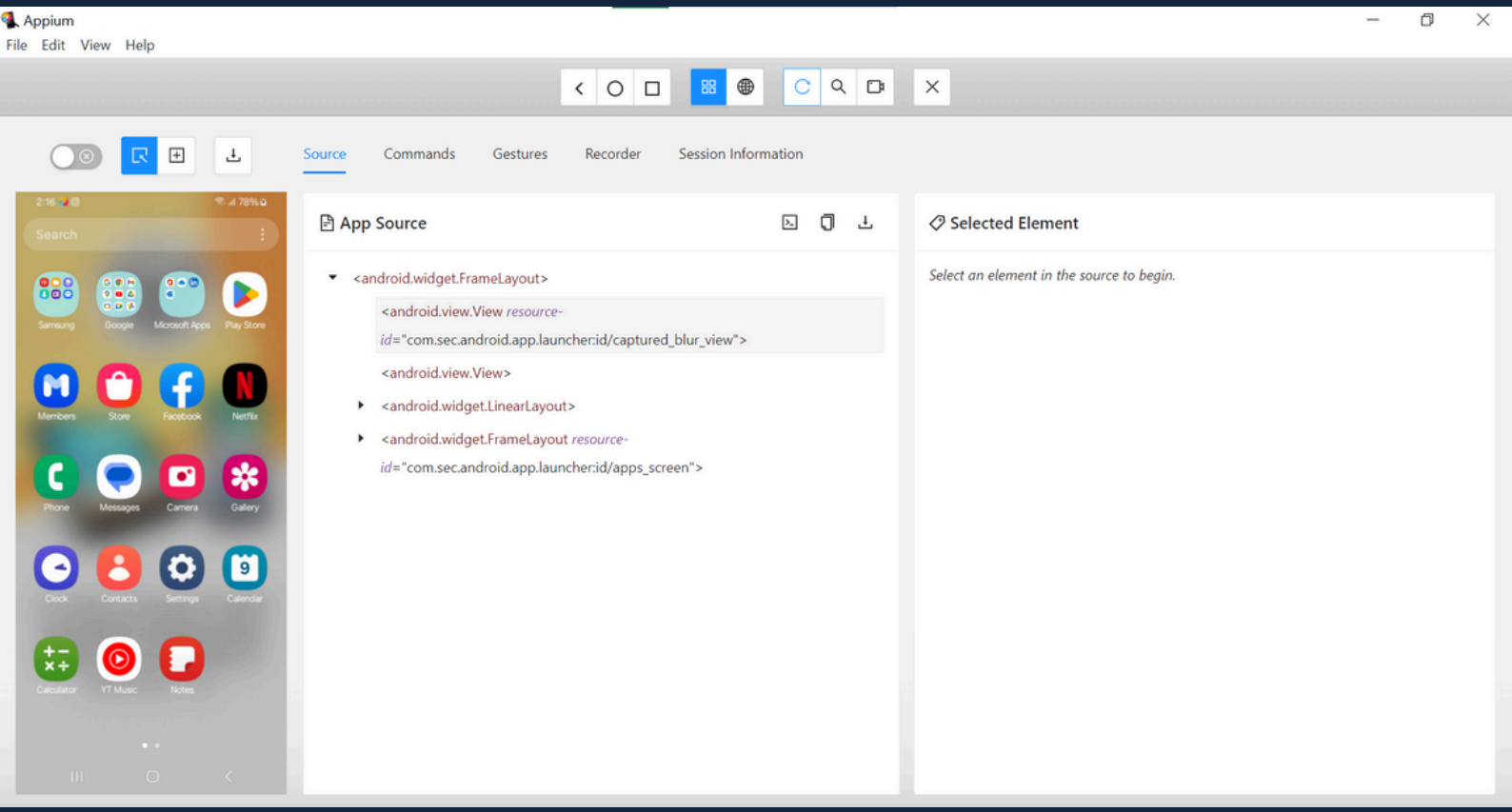
- Start the server



- Open Appium Inspector:-
  - Remote Host: 127.0.0.1
  - Remote Port: 4723
  - Remote Path: /wd/hub/
  - Capability Builder: {"platformName":"Android"}
  - Start Session



- Once the device is connected it will look like this:-



- Open your desired IDE, install these packages:-
  - Appium-Python-Client
  - Selenium

- Import these libraries:-
  - from appium import webdriver
  - from appium.options.android import UiAutomator2Options
  - from appium.webdriver.common.appiumby import AppiumBy
  - from selenium.webdriver.support.ui import WebDriverWait
  - from selenium.webdriver.common.by import By
  - from selenium.webdriver.support import expected_conditions as EC
  - import time
  - import pandas as pd
  - import numpy as np
  - import re

- We need to set these desired capabilities in order to connect and interact with our device, you may make changes as per requirements.

```
desired_caps = {
    'platformName': 'Android',
    'deviceName': 'R9ZWA07928P',   # Replace with your device name(adb devices --> Name)
    'udid': 'R9ZWA07928P',         # Replace with your device UDID(adb devices --> Name)
    'platformVersion': '14',     # Replace with your Android version (e.g., 11)
    'appPackage': 'com.instagram.android',
    'appActivity': '.activity.MainTabActivity',
    'automationName': 'UiAutomator2',    # Use UiAutomator2 for Android
    'noReset': True,                     # Keeps the app data between sessions
    'newCommandTimeout': 6000,           # Timeout for new commands to the server
    'adbExecTimeout': 20000,             # Timeout for ADB commands (adjust as needed)
    'autoGrantPermissions': True,        # Grant necessary permissions automatically
    'disableWindowAnimation': True,      # Disable window animations for faster test execution
    'unicodeKeyboard': True,             # Enable Unicode input (if needed)
    'resetKeyboard': True                # Reset keyboard after test (if needed)
}
options = UiAutomator2Options().load_capabilities(caps=desired_caps)
```

- Once the options are created, create a driver to do the tasks:-
  - driver = webdriver.Remote('http://127.0.0.1:4723/wd/hub', *options*=options)
- Now we are ready to start our Scrapping!

PROJECT SUMMARY

In this project I have created a complete automation process to scrape the data from the instagram accounts. As many small business owners are self employeed professionals always need leads for the outreach of their business, however to manully collect the leads is a very hectic tast, hence using this program they will be able to automate this process.

This programe will do the following tasks:-

- User only need to input their device details such as:-
  - Adb device name
  - Android Version
  - Number of Leads they want
  - Search hashtag
- The program will start from the home screen.
- Then it will open instagram.
- It will move to the search screen.
- It will enter the search tag in the search bar.
- Then once the results are available
- It will move to the “Posts” section and will start visiting the account of each post.
- From the account it will gather these infos:-
  - Username
  - No. of Followers
  - No. of Following
  - Email &Phone No. if the contact button is there.

In this version, the program will only pick the unique accounts, otherwise it will move to the next post, hence only collecting the data of unique accounts. In this version the program will start from the posts section only, however in the upcoming versions we can get the input from the user weather they want to start scraping the data from the posts or account section.

The program will collect the data in respected variables and once the scrapping is completed, it will give you all the data into an Excel file.

You will be able to see the complete video on the working of the program, on my linkedIn page, and the link for the video is given on the cover page.

I am sure there can be many improvements in this program that can be done, if you feel like contributing anything feel free to reach out to the email given below:-

Email: kartik.keyrast@gmail.com