**Group members**:

Kexuan Chen, kexuanch@andrew.cmu.edu

Yanijing Huang, [yanjingh@andrew.cmu.edu](mailto:yanjingh@andrew.cmu.edu)

Swapnil Panwala, spanwala@andrew.cmu.edu

Hetong Wang, hetongw@andrew.cmu.edu

**Manual installation command for modules and packages:**

Install the following modules and packages below with your shell or terminal if you have not installed them before

Selenium: pip install selenium

os: pip install os

Pandas: pip install pandas

BeautifulSoup: pip install beautifulsoup4

Scikit-learn: pip install scikit-learn

numpy: pip install numpy

matplotlib: pip install matplotlib

praw: pip install praw

**Data Source 1: Web Scraping at** [**Twitter Trend**](https://trends24.in/united-states/)  (Time to extract data: 10-15 Seconds)

Note: For twitter scraping, when the website automatically jumps out, it means the application is getting real-time data from the website, just waiting for the window to close automatically (approximately 10-15 seconds). Don’t interrupt during scrapping.

Chromedriver for Twitter scraping:

Since there is dynamic scraping like button presses and getting data from websites, it needs to download Chrome driver, and also import some packages in python code.

**For Windows:**

(i) **Download Chrome driver**:

This is the download links, highly recommended download the **win64** resources

| Platform | URL |
| --- | --- |
| **linux64** | https://storage.googleapis.com/chrome-for-testing-public/129.0.6668.91/linux64/chromedriver-linux64.zip |
| **mac-arm64** | https://storage.googleapis.com/chrome-for-testing-public/129.0.6668.91/mac-arm64/chromedriver-mac-arm64.zip |
| **mac-x64** | https://storage.googleapis.com/chrome-for-testing-public/129.0.6668.91/mac-x64/chromedriver-mac-x64.zip |
| **win32** | https://storage.googleapis.com/chrome-for-testing-public/129.0.6668.91/win32/chromedriver-win32.zip |
| **win64** | https://storage.googleapis.com/chrome-for-testing-public/129.0.6668.91/win64/chromedriver-win64.zip |

**(ii) Set path for chrome driver:**

# This is in the relative path that recommend to do:

service = Service('./chromedriver-win64/chromedriver.exe')

# If you want to put it in anywhere you like, make sure the path is correct, for example:

service = Service ('C:/Users/14550/Downloads/chromedriver-win64/chromedriver.exe')

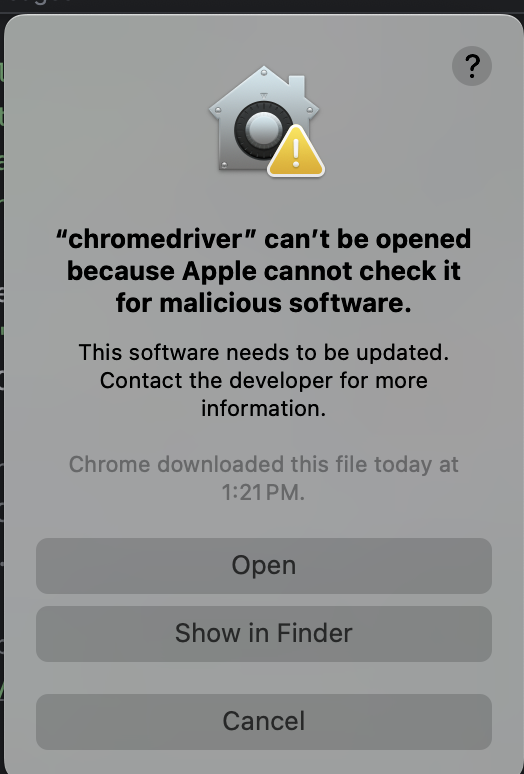
**Make sure that:**

**1. path end at the file: ‘/chromedriver.exe’ or ‘/chromedriver’**

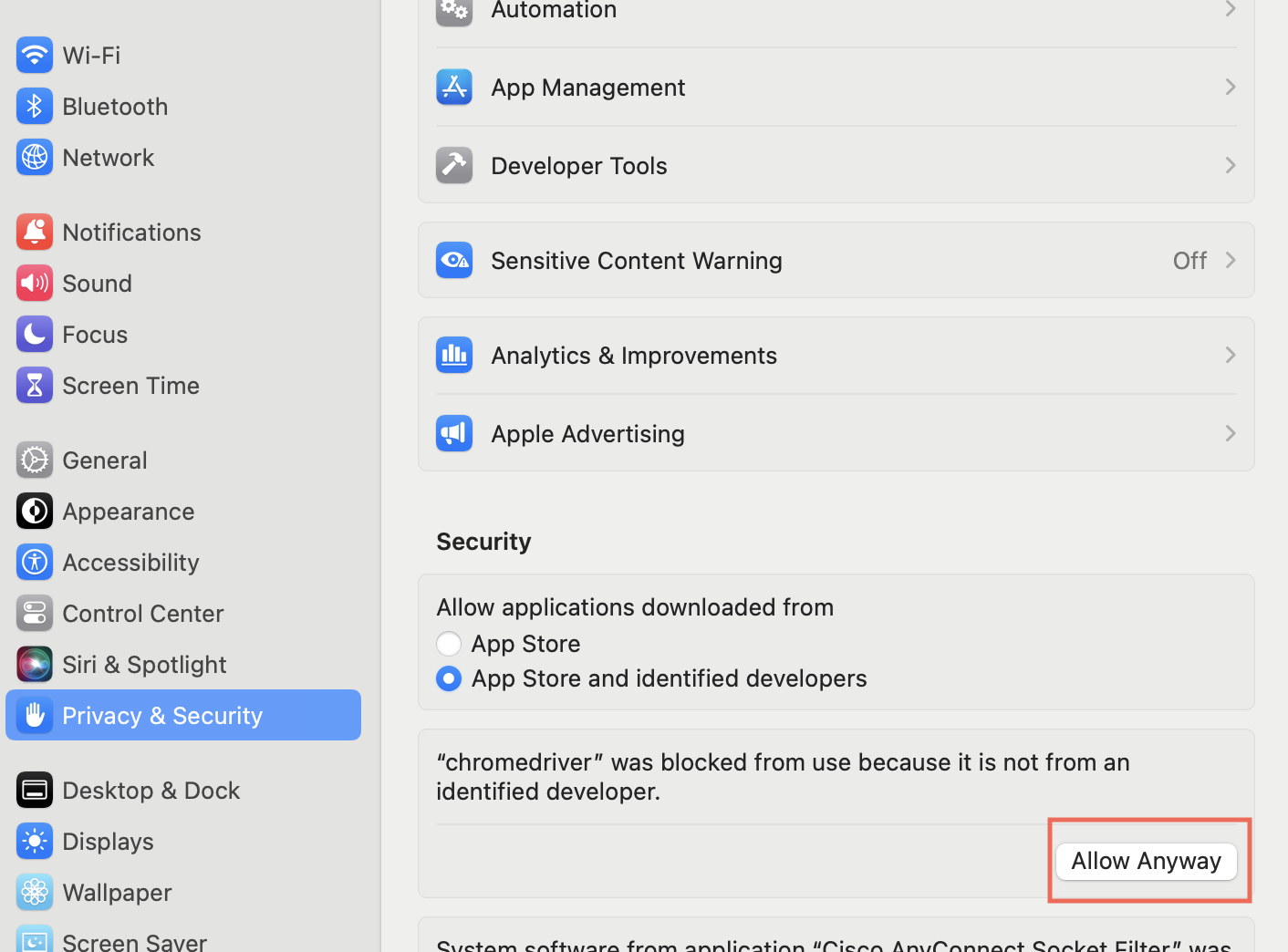
**2. ‘.exe’ file is extracted out of zip file. (path with “/chromedriver-win64.zip” is not working)**

**For Mac:**

You may be blocked by the Mac system when running and using the chrome driver for the first time. This is the pop up window you may see:

****

To relove this, go to **System Preferences > Security & Privacy > General**, look for a message about chromedriver being blocked, and click "Open Anyway" to allow it

****

**Data Source 2:** [**Instagram API**](https://rapidapi.com/social-api1-instagram/api/instagram-scraper-api2) **access from** [**Rapid API**](https://rapidapi.com/hub)(Time to extract data: 30 Seconds)

API Access request is already within the python file. No additional installation required.

**Data Source 3: Reddit API** (Time to extract 1000 data points: ~10 minutes)

API Access request is already within the python file. praw installation required.

**Trello Board:** <https://trello.com/b/z9t7pmua/trendwave>

**Demo Video:** <https://youtu.be/GQiltUAjpw4>