## Google Trends Web Scraping

Prerequisites:

1. Install Anaconda for python 3
2. Open Anaconda prompt from Start menu and run as administrator.
3. Install Selenium: conda install -c conda-forge selenium
4. Install fake-useragent: pip install fake-useragent
5. Upgrade chrome browser to the latest version.
6. Download chromedriver.exe from <http://chromedriver.chromium.org/downloads>  
     
   The chromedriver version should be same as your chrome browser version. If you are using windows it is version 75.

Directory Setup:

1. Put chromedriver in C:/driver/chromedriver.exe
2. Put folder for data collection "Robin Google Trend II" in some directory

* all\_data: All data will be downloaded in this folder. Do not manually modify its contents.
* reference: Contains url\_machine\_6.xlsx
* extraction\_machine\_6.ipynb: Contains code that has to be run

1. Update the path to "Robin Google Trend II" within code

Launching jupyter notebook:

1. Open Anaconda prompt
2. Goto drive inside prompt where you have created "Robin Google Trend II"
3. Launch jupyter lab by entering command: jupyter lab
4. Goto directory "Robin Google Trend II" and open extraction\_machine\_6.ipynb

Code editing and running instructions are provided within jupyter file as comments.

Output:

* The output file is the all\_data folder. It must contain data\_400000.csv to data\_499999.csv
* Caution must be observed while updating numbering within code. It should be accurate, complete and no duplicates. Regarding this, instructions are given within jupyter file.
* Zip the all\_data folder. Rename the zip as ‘Machine\_6\_data’ and send to us.
* If any further doubts regarding numbering, any specific error/exception or other you can ask.

Note:

* If code does something that seems weird, let us know. There might be bugs or issues that might not have been handled.
* If you see any numbering discrepancies or files that are not getting renamed, let us know.