**Web Listening FYPJ Guide**

**By: Cheung Kai Chun Ronald**

**Hazel Phua Yi Xuan**

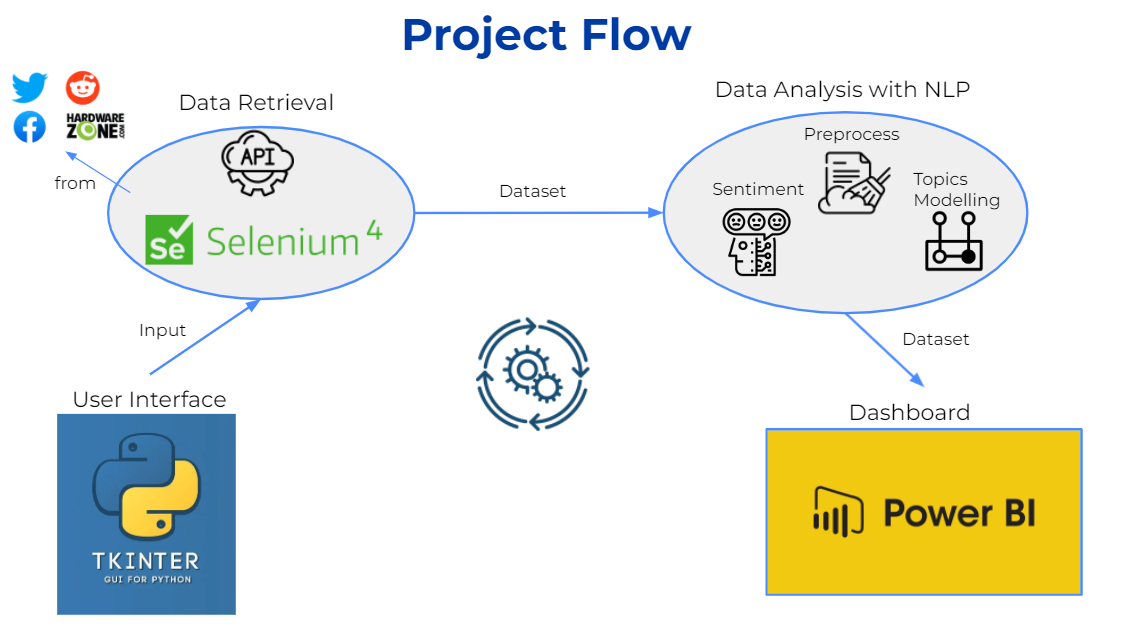
**About**

Web Listening, also known as Social listening, refers to the tracking of social media platforms for mentions and conversations related to your brand, then analyzing them for insights to discover opportunities to act. For our project, we have chosen the following social media platforms, and the use case for most of our testing is **Adidas and Nike.**

1. Twitter (using API)
2. Reddit (using API)
3. Facebook (using API)
4. HardwareZone (using Python selenium)

After retrieving relevant data from the above platforms, Natural Language Analysis (NLP) techniques will be performed on the dataset, more specifically Sentiment Analysis and Topic Modeling. After that, the dataset will be loaded into a Power Bi dashboard. The entire process starts with a user interface, where 2 keywords will be taken as input to allow the web scraping scripts to collect data related to those keywords, before moving on to the NLP analysis.

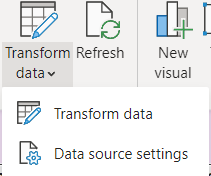
**Diagram of how the project works**

****

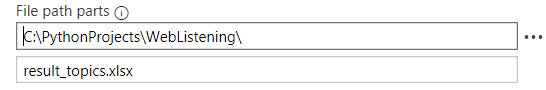
|  |  |  |
| --- | --- | --- |
| **Time Period** | **Tasks completed** | **Done by** |
| Week 1 - 4 | Research + Web scraping | Ronald, Hazel |
| Week 5 - 6 | NLP analysis | Ronald |
| Week 7 - 10 | User interface + Improve above parts | Ronald |
| Week 11 - 12 | Integration of all parts | Ronald |
| Week 5 - 12 | Power BI Dashboard | Hazel |

**Prerequisites to run the project**

1. Install **Power BI desktop**
2. Install **Python** (3.8 and above) and set as project interpreter
3. Install the missing libraries in all the python files after opening the project
4. Ensure your **chrome version** is **102**, if not, [download](https://chromedriver.chromium.org/downloads) a new chromedriver that matches your chrome version and replace the current chromedriver (chromedriver.exe) in the project
5. Change the file path of Power BI dashboard
   1. Open **Dashboard.pbix**
   2. Click on Transform Data > Data source settings



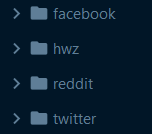
* 1. Right click on the current path and select Change Source
  2. Change the **top path** to **your** project path (data file to be in .xlsx)



* 1. Save your changes and close the file

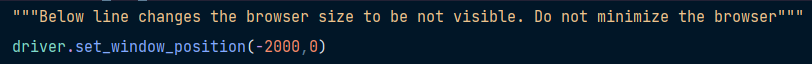
**Running the project**

1. Let’s see what will be happening in the background first.
   1. Web scraping codes can be found here. Their respective csv files will be found inside them under the dataset folder after the scraping process is done.



**Note 1**: You will need to login to Facebook and copy valid cookies using an extension, for example EditThisCookie. Paste them in cookies.json under the Facebook folder. Take note that the web scraping codes for Facebook will only work if the page for the inputted keywords exists in Facebook, for example, both Adidas and Nike have official pages in Facebook.

**Note 2**: The web scraping codes for HardwareZone will not always work due to the existence of Google ReCAPTCHA, preventing selenium from using the browser to scrape data. If you would like to see what’s happening when HardwareZone is being scraped. Comment the line below in both hwz1.py and hwz2.py.



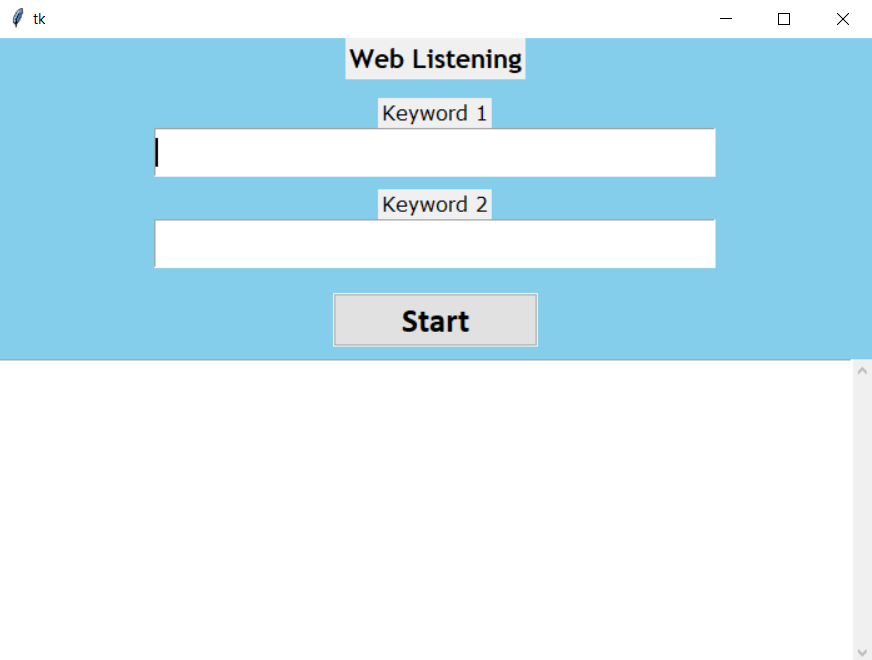
* 1. NLP Analysis codes can be found here



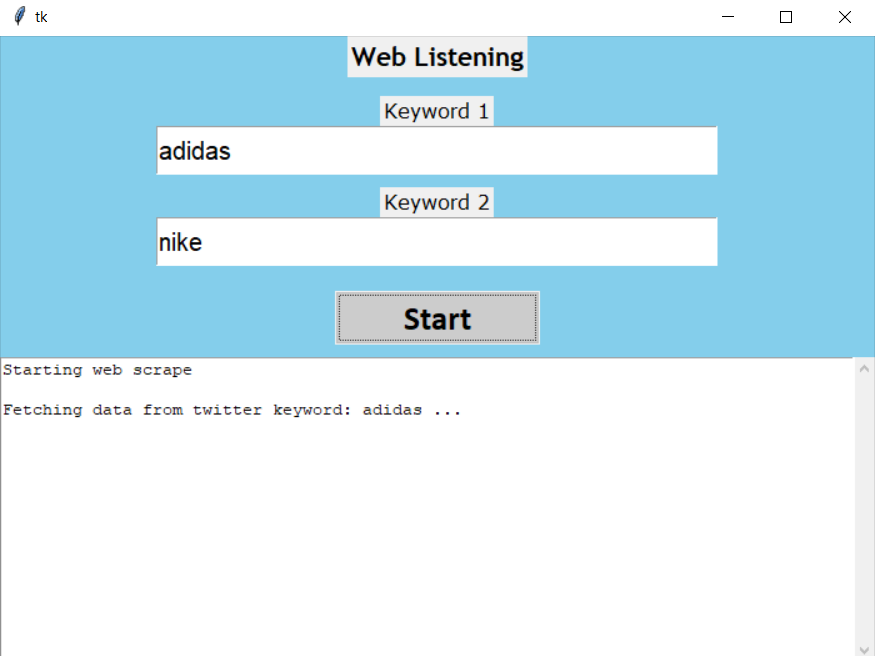
**Note**: If you want to study the codes, use the jupyter notebooks found in the **analysis\_notebooks** folder. The code in the analysis\_in\_use folder is taken from analysis\_notebooks and converted to python files so they can be run easily. To view the jupyter notebooks you will need to install jupyter on your python interpreter.

* 1. **Scraper.py** is the integration of everything, such that the run\_all() function in scraper.py can be executed in **app.py** to start the whole process. The combine\_csv() function in Scraper.py takes all the csv files generated from the web scraping codes and combines them into result.csv.

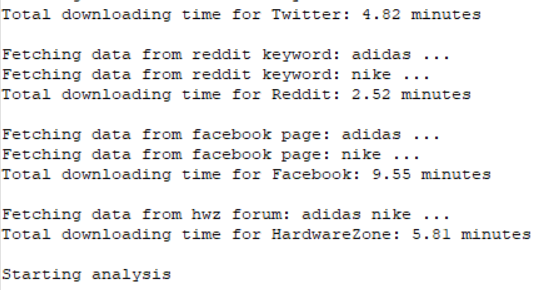
1. Run **app.py** and you should see this window

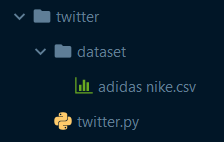


1. Input any two related brands (note: final dashboard summaries are done in regards to Adidas and Nikebut feel free to use other keywords) and click Start, Click yes to the prompt



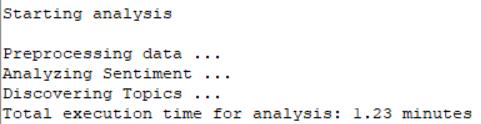
1. The entire process will take about 20-30 minutes. You will see updates on the interface regarding what is currently being done.
2. When the process is done up to this point. Csv files for each social media will be found or updated under the dataset folder in their individual folders. Result.csv should be generated or updated in the project as well.

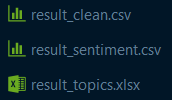






1. When the process is done up to this point. Result\_clean.csv, result\_sentiment.csv and result\_topics.xlsx should be generated in the project after each analysis step.





1. Power BI should be launched right after this and all you have to do is to click **Refresh**, and your new dataset will be loaded into the dashboard



**References**

Web Scraping

<https://github.com/kevinzg/facebook-scraper> (Facebook)

<https://github.com/JustAnotherArchivist/snscrape> (Twitter)

<https://melaniewalsh.github.io/Intro-Cultural-Analytics/04-Data-Collection/14-Reddit-Data.html> (Reddit)