**MODULE 3.5 SUBMISSION**

**What did I learn:**

* using following APIs
  + twitter api
  + reddit api
  + quandl api
  + yfinance api
  + googlenews api
  + snscrape api
  + pandas usage
  + dataframes usage
  + saving a csv file
* git lfs for large file storage
* creating environments and kernels in Anaconda
* installing packages to specific environment
* how to search datasets on google dataset search
* how to search for datasets on Kaggle
* use Textblob for finding sentiment analysis

**Preparation steps:**

1. Installed Anaconda on my mac machine
2. Created a ucsd-1 environment
3. Added necessary packages to the environment
4. Created a kernel for the environment
5. Started Jupyter-lab and selected the above kernel
6. Created the various notebooks for writing code

**References**:

[Link your Virtual Environment to Jupyter Using Kernels | by Adam Sabra | Towards Data Science](https://towardsdatascience.com/link-your-virtual-environment-to-jupyter-with-kernels-a69bc61728df)

[python - Using Pip to install packages to Anaconda Environment - Stack Overflow](https://stackoverflow.com/questions/41060382/using-pip-to-install-packages-to-anaconda-environment/44066694)

**my git repository link:**

<https://github.com/rbyakod/mec-mini-projects.git>

Read a lot of articles on google and **learned**:

* Using Twitter API
* Reddit API
* Googlenews API
* Kaggle datasets
* FinViz API usage
* Yahoo finance API usage

Created **API code for each of them** (jupyter notebooks) to retrieve data from the data sources. Learnt how to create stream listeners for Twitter as well.

Complied and ran the code to make sure they work well.

Downloaded some sample data from each of the sites mentioned above and saved as CSV files.

Searched Kaggle for data on terms “finance”, “stock”, “stock market”. Identified a few good data sets – both small and large and up to date datasets. Downloaded a few.

**Uploaded large files using “git lfs” to my github repository**

Pushed all my changes to the git repository.

**Note**: Even with Git LFS, there is still a file size limit of 2GB, which is a restriction placed by Github. Anything bigger, it is probably time to look into cloud storage.

**I stored a large (> 2GB) file in my gDrive and am sharing it here as well:**

[Large Dataset ( > 2GB)](https://drive.google.com/file/d/1vGh1WGx0JGKoqESEbnIzQF03rxfXcyFV/view?usp=sharing)