

Project Title	Twitter Scraping
Skills take away From This Project	Python scripting, Data Collection, MongoDB, Streamlit
Domain	Social Media

Problem Statement:

Today, data is scattered everywhere in the world. Especially in social media, there may be a big quantity of data on Facebook, Instagram, Youtube, Twitter, etc. This consists of pictures and films on Youtube and Instagram as compared to Facebook and Twitter. To get the real facts on Twitter, you want to scrape the data from Twitter. You Need to Scrape the data like (date, id, url, tweet content, user,reply count, retweet count,language, source, like count etc) from twitter.

Approach:

- By using the “**snsrape**” Library, Scrape the twitter data from Twitter [Reference](#)
- **Create a dataframe with** date, id, url, tweet content, user,reply count, retweet count,language, source, like count.
- **Store each collection of data** into a document into Mongoddb along with the hashtag or key word we use to Scrape from twitter. eg:({“LIC corporation+current Timestamp”: [{1000 Scraped data from past 100 days }])
- **Create a GUI using [streamlit](#)** that should contain the feature to enter the keyword or Hashtag to be searched, select the date range and limit the tweet count need to be scraped. After scraping, the data needs to be displayed in the page and need a button to upload the data into Database and download the data into csv and json format.

Results: You have to build a solution that should be able to scrape the twitter data and store that in the database and allow the user to download the data with multiple data formats.

Project Evaluation metrics:

- You are supposed to write a code in a modular fashion (**in functional blocks**)
- Maintainable: It can be maintained, even as your codebase grows.
- Portable: It works the same in every environment (operating system)
- You have to maintain your code on **GitHub**. (Mandatory)
- You have to keep your **GitHub** repo public so that anyone can check your code. (Mandatory)
- Proper readme file you have to maintain for any project development (Mandatory)
- You should include basic workflow and execution of the entire project in the readme file on **GitHub**
- Follow the coding standards: <https://www.python.org/dev/peps/pep-0008/>
- You need to Create a Demo video of your working model and post in **LinkedIn** (Mandatory)

