

## How to Run

Link to Dataset: <https://www.kaggle.com/datasnaek/youtube-new>

Once the dataset is downloaded from Kaggle, it needs to be unzipped and it has a folder called 'youtube-new'. Inside it are the 10 csv and 10 json files.

Alternate way: I have uploaded the unzipped 'youtube-new' folder on my drive and the link is: [https://drive.google.com/open?id=1Hpt\\_NtiZ31Uo84rDbKZOFnD8ZBAKxvso](https://drive.google.com/open?id=1Hpt_NtiZ31Uo84rDbKZOFnD8ZBAKxvso)

It can be directly downloaded from the above link to your machine.

This is the path structure that I have given in my notebook:

'C:\Users\RACHANA RAMESH\Desktop\SDSU\SDSU coursework\Big Data\project\youtube-new\USvideos.csv'

'C:\Users\RACHANA RAMESH\Desktop\SDSU\SDSU coursework\Big Data\project\youtube-new\US\_category\_id.json'

The submission consists of 2 notebooks:

1. task\_1\_usa: Consists of analysis for US data ie USvideos.csv and US\_category\_id.json files
2. task2\_all\_countries: Consists of analysis of Global trend- have combined all csv's into a single pandas dataframe

**Replacing the path at the beginning of the notebook should suffice for running provided all libraries are installed.**

- Nltk: pip install --user -U nltk
- Wordcloud: pip install wordcloud
- Textblob: pip install -U textblob
- python -m textblob.download\_corpora
- Sklearn: pip install -U scikit-learn

**Note: Detailed analysis is in the report**