COVID19 Kaggle Challenge

Tour and Tasks for the next week

★ 1) Open an account, login and enroll in the challenge

Tour and Tasks for the next week

★ 2) Download the data (4GB)
   a. Contains thousands of papers (comm, non-comm, BioRxiv-MedRxiv)
   b. It also has a csv metadata file that contains information on individual research papers including their abstract
   c. All papers are in json format

Tour and Tasks for the next week

3) Learn about json format (arrays and json objects)

Tour and Tasks for the next week

4) Find out the number of files with actual research papers

Tour and Tasks for the next week

5) Download one json file and extract the main keys(columns) out of it

6) Replicate the process using a for-loop and download all the files

Tour and Tasks for the next week

7) Make a list of Research Paper IDs and compare the list with the metadata file to find out the IDs that are not available in json format

Tour and Tasks for the next week

8) Read the ten objectives on Kaggle website carefully

9) Come up with ideas on how to use NLP to extract topics

Tour and Tasks for the next week

★ 10)   Create a slack channel to share progress and brag!