5/30/20

* Emmanuel led the discussion today. He had a nice agenda typed up for us to follow. It is posted on our Slack channel.
* We spent a few minutes figuring out the YouTube API permissions
* Polina set up the git repository and invited us to join it. We made sure everyone had it cloned on their computer. This took a while unfortunately. Everyone created their own folders to put their work in to help keep things organized.
* We discussed what our main objective is…what is the why?
  + We explored this topic to answer the following question…
  + A person wants to be able to create a viral video.
  + What should a person look at when trying to create a YouTube video that will get them the most views in the United States?
* We decided to break it down into different topics to explore outside of class:
  + Emmanuel-
  + Katie-
  + Michelle-
  + Polina-
* We are using the following CSV files:
* We decided to break it down into different topics to explore outside of class:
  + <https://www.kaggle.com/datasnaek/youtube-new?select=USvideos.csv>
* Katie recommended combining the CSV’s we are using into one master csv file for us to pull our information from.
* We need to do an API call for the extra data we need and then combine everything into one master CSV.
* After spending a little bit of time spinning our wheels we decided to take the time to:
  + How to combine the files?- We decided to only use the United States
  + Do we need to do the API call for information?
  + Is the file we have providing enough information to answer the questions we have?
  + We need to match up the Category ID’s
  + We need to group by video title because some videos were trending for several days…how many unique videos are there actually in the list
  + We need to sort thru the data and each come up with three things we might want to do an analysis on. Ideas so far:
    - Number of days it is trending affect how many views there are?
    - Do some channels have more trending videos
    - Is there a correlation between likes/dislikes and the total number of views
    - Is there a correlation between the Category ID and the total number of views
    - Is there a correlation between the Publish time and the total number of views
    - Polina suggested a line chart of different videos and when they were trending.
  + Katie created a Jupyter notebook and we sorted it different ways to start learning about our data.
* Is our data sample big enough?- We discovered that there are over 6000 individual videos which spanned 180 days which we decided was a plenty big sample.
* **We are all going to take some time to explore the data. Have your three ideas on slack by Monday night so by Tuesday night we can all have begun sorting our data and creating our Jupyter notebooks.**