

## **SI Final Project Report - AJ Estes**

### **Overview**

My goals for this project were to use the YouTube API in order to demonstrate the relationship between a YouTube channel's video duration and number of comments. I would demonstrate this by using MatLab in order to create a bar graph of the relationship and output the bar graph as an image file. The file 'final-project.py' contains the code in order to properly achieve these goals, and is able to be run without issues.

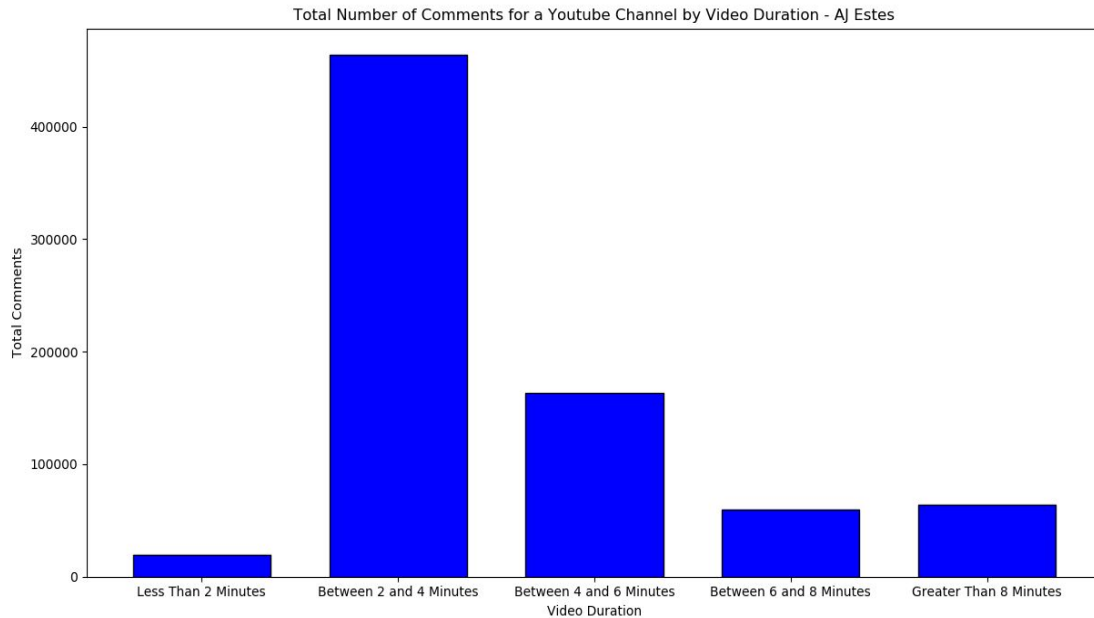
### **Problems**

Initially, I was planning on showing the relationship between the number of views that a channel's videos had and the day of the week that it was posted. I was unable to use the number of views that a channel had because YouTube is currently resolving an error in which some of their keys in the video statistics json outputs, including video views, where they are unable to be called. I then planned of switching to the video's number of comments instead, as there were no issues with this metric. I also faced the issue where YouTube was revoking public access to certain aspects of video information, which included the metric of when the video was published at. I then switched to duration of the video, as it was still able to be accessed. From here, I was then able to properly use YouTube's API in order to reach my goal of demonstrating the relationship between a YouTube channel's video duration and number of comments.

### **Report**

After running the program, I was able to see a clear relationship between the total number of comments that a YouTube channel has according to the duration of their videos. In this example I used Ariana Grande's YouTube channel to show this relationship. As can be seen in Figure 1, I categorized the duration of the channel's videos into 5 segments: Less Than 2 Minutes, Between 2 and 4 Minutes, Between 4 and 6 Minutes, Between 6 and 8 Minutes, and Greater Than 8 Minutes. For Ariana Grande's channel, videos between 2 and 4 minutes were the most popular with a total of 464,070 comments. From there, Ariana Grande's videos between 4 and 6 minutes had 163,235 comments, videos greater than 8 minutes had 64,159 comments, videos between 6 and 8 minutes had 59,372 comments, and videos less than 2 minutes had 19,327 comments. These metrics can be found in the 'YoutubeDict.csv' file.

**Figure 1**



### Instructions To Run

In order to run the file 'final-project.py', you must first have the file 'client\_secret.json' in the directory. This file contains your credentials for the YouTube API, and can be created by following the instructions through [Quick-Starting](#) the API and understanding your [client secret information](#). Once this is set up, you must run the code in a program that has a built-in terminal, such as [Microsoft Visual Studio](#). This is necessary because the YouTube API generates an authentication key each time you run the code, which must be inputted in the terminal window in order for it to run. Once this is complete, you are able to run the code. In the function 'get\_all\_video\_in\_channel', you must assign the variable 'api' to a string that is your unique Youtube API key. This key can be found by following the instructions [here](#).

As of now, Ariana Grande's channel ID is being run through the program to create the bar graph image file seen above. This can be seen in line 320 of the code, where her Channel ID is being saved as a variable name, and called to run into the program on line 321. One is able to run the program with any YouTube channel ID by creating a variable name that is assigned to a string that is their channel ID, and is passed into a print statement of the function runProject. More information can be found in the docstrings of the file 'final-project.py'.

## Functions

Below are descriptions of the functions that are used in the project with the exception of the YouTube API provided functions, which can be viewed [here](#). More information can be found in the docstrings of the file 'final-project.py'.

def get\_all\_video\_in\_channel(channel\_id): Function get\_all\_video\_in\_channel that takes a Youtube Channel ID as an input and outputs a list of all of the the channel's Video IDs.

def get\_stats(video\_list, cacheDict, fname): Function get\_stats takes in an input of a list of Youtube Video IDs, a dictionary cacheDict to cache video information, and a file fname to output the dictionary to a json file. The function outputs a dictionary, where the keys are the Video Ids, and the values are a list of dictionaries of the statistics for the video.

def setUpYoutubeTable(youtubeDictionary, conn, cur): Function setUpYoutubeTable takes a dictionary where the keys are the Video Ids, and the values are a list of dictionaries of the statistics for the video, a sqlite3 connection, and a conn.cursor(). This function then creates a sqlite table where the columns are the Youtube Video Id, the duration of that video, and the number of comments.

def getYoutubeDict(cur): Function getYoutubeDict takes in a sqlite connection as an input and selects the duration and number of comments for each video. This function then puts this data into a dictionary where the keys are the length of the video, which are as follows: Less Than 2 Minutes, Between 2 and 4 Minutes, Between 4 and 6 Minutes, Between 6 and 8 Minutes, and Greater Than 8 Minutes. The values for these keys are the total number of comments received for the videos. This is calculated by iterating through the sqlite data and using regular expressions to decode the given duration information from YouTube's API and adding each video's comment count to the total number of comments for each key. This function returns a dictionary of the categories of video duration with their values being the total number of comments for each duration category.

def drawBarChart(youtubeDict): Function drawBarChart takes in a dictionary of duration categories with their values being the total number of comments for each category. This function then outputs an image of a bar graph that displays the total number of comments for a youtube channel by video duration.

def createCSV(finalDict): Function createCSV takes in the dictionary of the 5 duration categories and their total amount of comments as an input and creates a CSV file that displays them by duration and total amount of comments.

def runProject(YoutubeID): Function runProject takes in the a Youtube Channel's ID and passes it through all of the above functions in order to reach the end goal, which is to create a bar graph image file that displays the total number of comments for a youtube channel by video duration. This function then outputs a message confirming that the function has run.

if \_\_name\_\_ == '\_\_main\_\_': The following line of code runs the entire project. The variable ArianaGrandeChannelId is the YouTube Channel ID for pop-sensation Ariana Grande and is used in the function runProject in order to return a bar graph image file that displays the total number of comments for Ariana Grande's channel by video duration.

## Documentation

12/3/2018 | Setting up Youtube API |

<https://developers.google.com/youtube/v3/quickstart/python> | Able to Quickstart Youtube API

12/3/2018 | Getting Client Secret |

[https://developers.google.com/api-client-library/python/guide/aaa\\_client\\_secrets](https://developers.google.com/api-client-library/python/guide/aaa_client_secrets) | Able to get my client secret information in json format

12/3/2018 | Getting API Key | <https://developers.google.com/youtube/v3/getting-started> | Retrieve unique API Key

12/3/2018 | Getting YouTube Video API Keys |

<https://developers.google.com/youtube/v3/docs/videos/list> | Get the keys for the retrieved youtube information, such as 'commentCount' and 'duration'