Use Guide for Package fc1315

Fanglin Chen

Dec 2016

Overall description

YouTube is a video-sharing platform where members can create channels and upload videos for sharing. The program displays both text output and matplotlib popup of video statistics after a user enters a channel name and a date range. The video statistics include the number of videos, the mean, standard deviation, minimum and maximum of video durations, and the mode of video publish times (i.e. hour of the day and day of the week). The program also enables comparing the information across multiple channels. Detailed description is in the “Run the program” section.

1. Installation

(1) Fork the "final\_project" repository from the "ds-ga-1007" user on GitHub. Clone this repository onto your local system.

(2) Find the package "fc1315" in the repository, and put it in the directory where you want to run the program.

The package should contain 7 modules: \_\_init\_\_py, action.py, apidata.py, exception.py, final\_project.py, test.py and video.py.

2. Environment

You need to have the following Python tools available: Python 3.5 (Anaconda3-4.1.1), NumPy 1.11, pandas 0.18.1 and matplotlib 1.5.1.

You need to install the Google API Client Library for Python by referring to this link:

<https://developers.google.com/api-client-library/python/start/installation>

(1) Use pip to manage your installation (you might need to run sudo first):

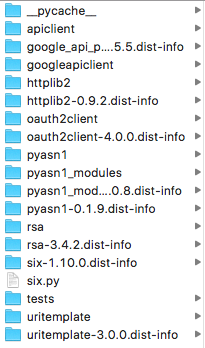
$ pip install --upgrade google-api-python-client

(2) Create a folder in the same directory where you put the package “fc1315”, such as lib/, and use pip to copy the library into the folder:

$ mkdir lib

$ pip install -t lib/ google-api-python-client

There should be a folder “lib” that contains the following libraries in the directory.



3. Create and enable an API key for YouTube

Since the dataset used in the program is directly accessed from YouTube Data API, you also need an API key for YouTube. If you do not have one, you can create an API key with the following steps:

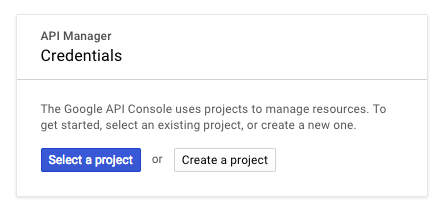
(1) Make sure you have a Google account. If not, you can create one from the link:

<https://accounts.google.com/SignUp?continue=https%3A%2F%2Faccounts.google.com%2FManageAccount>

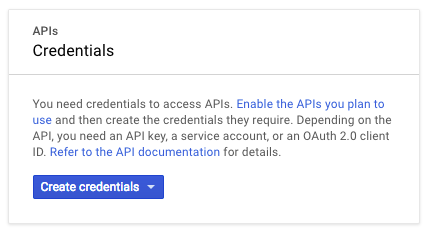
(2) Go to the “API Manager Credentials” from the link:

<https://console.developers.google.com/projectselector/apis/credentials>

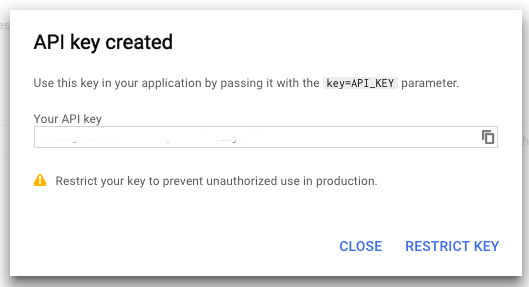
When prompted as below, choose “Create a project” and name the new project. For example, I name my project as “My Project”.



(3) After creating a project, you will be prompted as below. Click on “Create credentials” and choose “API key” in the dropdown list:

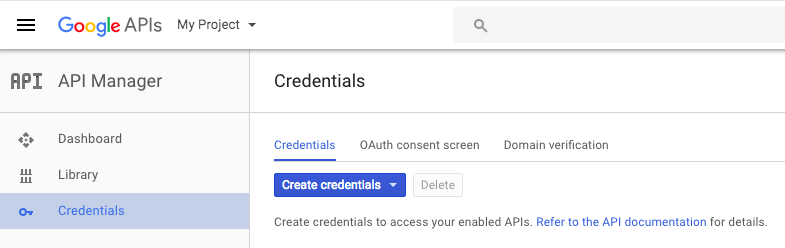


(4) When you see the following prompt, it means you have created an API key successfully. You can directly copy you API key from the prompt, or click on “CLOSE” and the key will show on the screen.

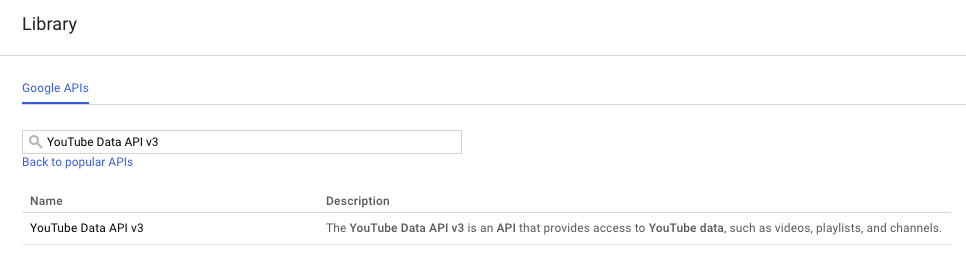


You need to enable the YouTube Data API for the new project with the following steps:

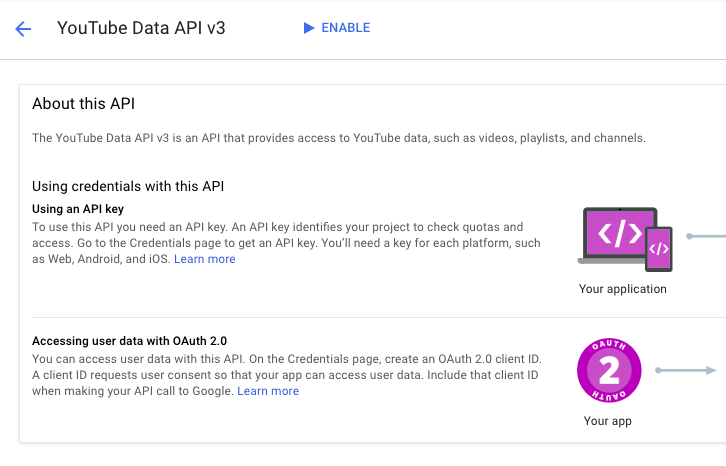
(1) Select the project that you just created (“My project” in my example), and click on “Library” on the left side.



(2) Search for “YouTube Data API v3” in the upper left side, and click on the name “YouTube Data API v3”:



(3) Click on “ENABLE” on the top, and you have enabled your API key for YouTube Data API.



4. Run the program

You can run the program directly from the command line:

$ python final\_project.py

The program will ask you to enter one YouTube channel name, or multiple YouTube channel names separated by commas. Make sure the channel name(s) are valid, and there is only comma (no space) between any two channel names. For example:

Google

Google,PewDiePie

Google,PewDiePie,CaseyNeistat

The program will ask you to enter a start date in the form of YYYY-MM-DD, such as 2010-09-28, and an end date in the form of YYYY-MM-DD, such as 2010-10-04. Make sure the dates are valid, for example, 2015-02-31 (no such date) or 2017-08-02 (date after today) or 1968-01-01 (before YouTube was created) is not allowed. Also, The end date should be later than the start date.

The program will ask you to enter your key for YouTube Data API. Please enter the API key that you already have or just created. Note that if you enter an invalid API key, the API would give an output of zero videos rather than raise an error.

The program will ask you to choose one of the following actions: number, duration, or published. Note that you may need to wait for several minutes to get the output, depending on the number of channel names and the length between dates that you input, so please be patient.

(1) If you enter “number”, the program will show the output in the terminal, which is a dictionary of the channel names and the number of videos published by these channels between the start date and the end date.

(2) If you enter “duration”, the program will show the output in the terminal, which contains the count, mean, standard deviation, minimum and maximum of video durations. The durations are expressed in seconds rather than minutes. The summary statistics are calculated with respect to videos published by each channel between the start date and the end date. It will also show histograms of video durations in the matplotlib pop-up.

(3) If you enter “published”, the program will show the output in the terminal, which contains the mode of video publish time, namely hour of the day (0 to 23) and day of the week (Monday to Sunday) when videos were published. Similarly, it is calculated with respect to videos published by each channel between the start date and the end date. It will also show two barplots (weekday and hour) of publish time of these videos in the matplotlib pop-up.

To quit the program, enter “quit” when prompted for channel names, dates or API key, and enter “quit” again when prompted for action.

You can test the program directly from the command line:

$ python test.py

5. Examples

$ python final\_project.py

Please enter one YouTube channel name, or multiple YouTube channel names separated by commas:

> Google,PewDiePie,CaseyNeistat

Please enter a start date in the form of YYYY-MM-DD:

> 2010-09-28

Please enter an end date in the form of YYYY-MM-DD:

> 2010-10-04

Please enter your key for YouTube Data API:

> AIzaSyCx264wDZhofuRTgiocIClsa3wy8hiVw-E

Please choose one of the following actions: number, duration, or published

> number

{'CaseyNeistat': 0, 'Google': 2, 'PewDiePie': 1}

$ python final\_project.py

Please enter one YouTube channel name, or multiple YouTube channel names separated by commas:

> Google

Please enter a start date in the form of YYYY-MM-DD:

> 2015-02-28

Please enter an end date in the form of YYYY-MM-DD:

> 2015-04-01

Please enter your key for YouTube Data API:

> AIzaSyCx264wDZhofuRTgiocIClsa3wy8hiVw-E

Please choose one of the following actions: number, duration, or published

> duration

Google

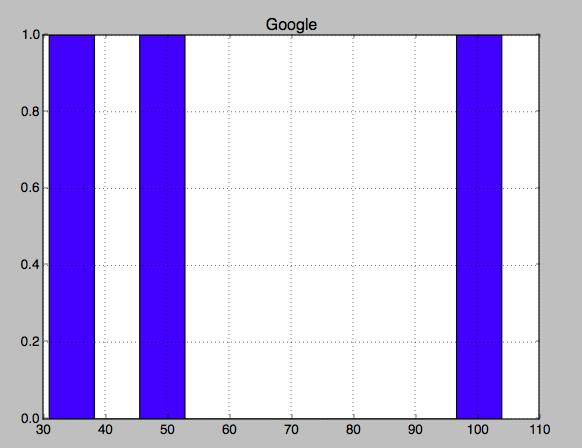
count 3.0

mean 61.0

std 38.2

min 31.0

max 104.0



$ python final\_project.py

Please enter one YouTube channel name, or multiple YouTube channel names separated by commas:

> Google,CaseyNeistat

Please enter a start date in the form of YYYY-MM-DD:

> 2016-01-19

Please enter an end date in the form of YYYY-MM-DD:

> 2016-02-20

Please enter your key for YouTube Data API:

> AIzaSyCx264wDZhofuRTgiocIClsa3wy8hiVw-E

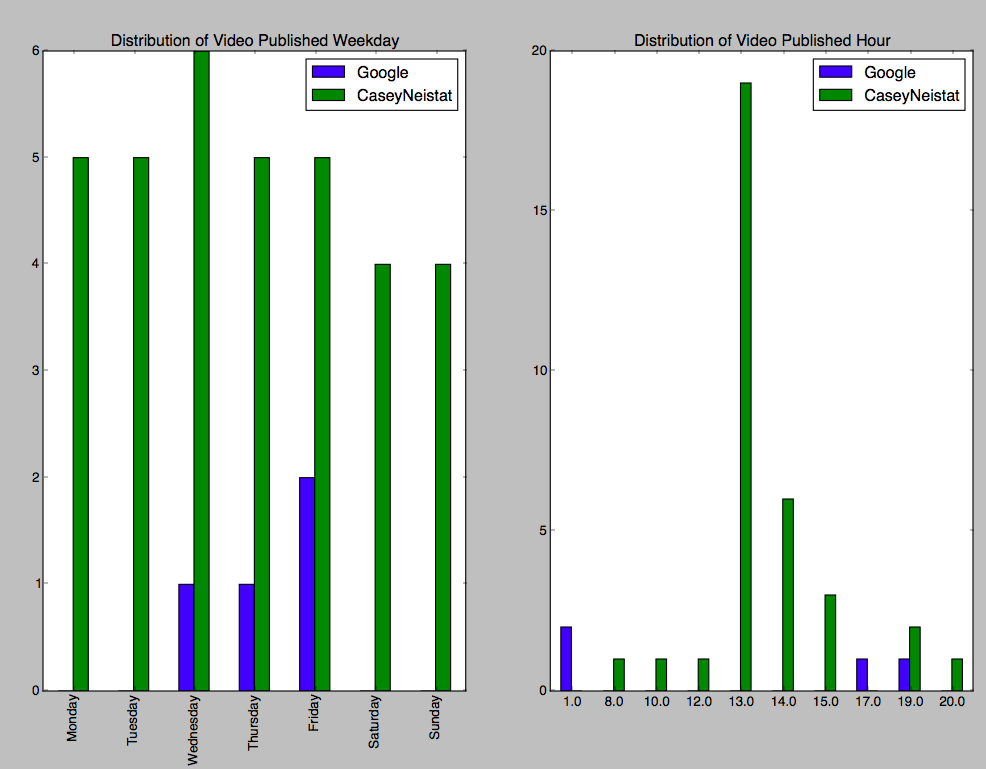
Please choose one of the following actions: number, duration, or published

> published

Google CaseyNeistat

0 Friday Wednesday

1 1 13



$ python final\_project.py

Please enter one YouTube channel name, or multiple YouTube channel names separated by commas:

> PewDiePie

Please enter a start date in the form of YYYY-MM-DD:

> quit

Please choose one of the following actions: number, duration, or published

> quit

$ python test.py

Ran 13 tests in 34.577s

OK