

DS250: Assignment 1 (10 marks)

Task: Automatic Playlist creation

Motivation: Your friend/cousin/neighbor approaches you to recommend the best set of online video lectures (a customized playlist) that he/she can watch to learn 3 different subjects/topic. These can be engineering, math, science, history or any other field of education. You are welcome to choose different values for the duration of individual playlists in the range [5,15] hours.

Methodology

For this assignment, you have to use the *Google developer's* account to gather information (published date, duration etc.) and statistics about different lectures available on *youtube*. To make this more interesting, you must also monitor the recent trends on these videos over the course of few days and make your final recommendations for each playlist.

Directions for submission

a. Please submit the Python notebook you use for gathering data, parsing, analysis, plotting etc. with good explanations, comments, notes etc. The name of your notebook should be <INSTITUTEID>_Solution1.ipynb

b. The cleaned data must be saved to a <INSTITUTEID>_data.csv file. There must be at least 3000 rows in the file. The minimum columns/fields must be in the specified order:

Topic, TimeQueried, Video_Title, Video_ID, Published_At, duration, viewCount, likeCount, dislikeCount, favoriteCount, Score, ...

Where Topic is the topic/subject you queried using the API

TimeQueried is the time when your query was made

Score is a value you compute which you may use to make your recommendations

Other fields are as defined in the Google API docs.

c. If you chose additional information/fields to make your decisions/recommendations, clearly document them and append them to the right of the specified fields in the file <INSTITUTEID>_data.csv

d. Please submit a 1-page document titled <INSTITUTEID>_Approach.pdf explaining your approach to solve the problem. You are welcome to use more sophisticated recommendation techniques than what have been taught in the class so far. Provide clear documentation and references if you do so.

e. Finally, place all your files in a folder called DS250_<INSTITUTEID>_Solution1, zip it and create a file <INSTITUTEID>_Solution1.zip

This assignment is strictly on individual effort. Please don't consult each other or post hints/solutions on Piazza. You can send private question to instructors (including Tas) so that one of us can reply. If you used web resources, clearly document/reference the websites/code in your notebook and also in the 1-page document.

f. Use the following code to sign up for the Google classroom for DS250: 5gicx7c

Please do this as soon as possible. All submissions have to be made on Google Classroom. You will be able to see this assignment there soon. The last submission time for this assignment is 8/21 (4:00 p.m.). Any delay will result in a penalty. We will not grade any solution submitted after 8/24 (4:00 p.m.).

NOTE: Google classroom will only be used for Assignments, Exams and Quizzes.

Helpful Resources:

<https://developers.google.com/youtube/v3/getting-started>

<https://developers.google.com/youtube/v3/docs>

<https://www.json.org/json-en.html>

<https://docs.python.org/3/library/json.html>

Hints:

a. Install the google api client for python with the following command:

`pip install google-api-python-client`

b. Use a Python notebook “YoutubeSearchAPI_Starter” on Piazza which can get you started

c. Read the docs to understand how to use filters in your API calls to make them more focused

d. You may need to use multiple APIs which provide different levels of details

e. Learn how to parse the JSON objects returned by the API calls