User Guide

Technical requirements : in order to be able to run this program, you need to install the seaborn package (`pip install seaborn` or `conda install seaborn` if using conda)

This program enables the user to perform many actions based on an initial data set containing meta-data (title, description, number of likes, duration,...)on more than 240.000 YouYube videos

Here is one row of the data we have on the videos

	video_category_id	title	description	published_at	viewCount	likeCount	dislikeCount	favoriteCount	commentCount	definition	caption	licensedContent	topicIds	relevantTopicIds
0	19	Disneyland Railroad at night and first look at...	Disneyland Railroad at night and first look at...	2014-11-15T17:00:06.000Z	3826	154	0	0	26	hd	False	True	/m/02fzs; /m/0550hw; /m/028k8w	/m/010jjr; /m/01zx_2; /m/09b3v; /m/01jddz

The actions the user can perform thanks to this program are the following :

1) Learn about the data set (which features are available, ...) and visualize the raw data set
2) Perform various analysis / plots on the data set :
   - General analysis
   - Analysis per video category
   - Analysis per feature
3) Test the model we have built on this data set to predict the category of a video based on the input of its title and description

For the two first options, just let you guide by the instructions displayed by the program. Note that to see the plots generated, you can find them in the file YoutubeData.

For the third option, some further precisions are necessary.
The program enables the user to test the model we have built. In order to do that, the user should go on YouTube and pick a video of his/her choice.
Just copy the title and the description of the video (separately) and paste it when the program asks you to.
Please be aware to **copy only the description** and not the text above nor below (as shown on the following illustration) since it contains the Video Category the program has to guess !

To check whereas the prediction is accurate, just compare the prediction of the model to the true video category (that you can find right below the description of the video, as shown below)

Please be also aware that a wrong prediction could have 2 reasons :

1) Some videos are misclassified by the people who uploaded them
2) We only implemented a "light" version of our prediction model in order for the user to be able to run it in a reasonable amount of time. The accuracy is therefore lower than our full model (around 65%)

Illustration showing which information to copy-paste in the program