Capstone Project # 1
Predicting Popularity of YouTube Videos

Problem Statement

The purpose of this project is two folds:

1. Find the most viewed genre of YouTube videos in an economically challenged country like Mexico and an economically stable country like the USA? By focusing on this data YouTube can advertise to the audience of these countries with the videos that can potentially get them more followers/subscribers and in turn more ad revenue
2. Predict popularity class of YouTube videos

Client

The stakeholder of this project is YouTube. The company can use the findings and promote the videos that belong to the popular genres among USA and Mexico viewers using their “You may also like” suggestion area. The probability of getting clicks could increase with directing the appropriate content into appropriate domains. Advertisements can be targeted accordingly to get better revenue. By predicting popularity of a video YouTube can make sure to push the most popular videos to their audiences and attract more views.

DataSet

Kaggle has an enormous amount of data on YouTube’s Video Statistics. It can be found at [Trending YouTube Video Statistics]

The specific dataset would be for the USA and Mexico. The record of interests from these two datasets will be number of views, number of likes, Category_id and publish_time.
As the project will take its shape in next few weeks it is possible that a change in these records of interests is warranted and therefore the end result might look a little different.

Deliverables

1. GitHib repository with the project code https://github.com/fariha23/YouTube_Data_Analysis_Video_Categories
2. Code format will be either Jupyter Notebooks or python script
3. A report in PDF format explaining the methodology and results will be published in GitHub repository

    a. The results will include several clear data visualizations that goes with the data story of the project
    b. Any assumptions that were made for the resolution of the problem will be clearly stated
    c. The recommendations to implement the solution, if there are any, will be included too
4. A slide deck for presentation will also be included and saved in the GitHub repository