**Project : YouTube API Project**

**Credits**

Yuteng Zhang

Dipti Sontakke

Alejandro Gutierrez

Patrick Brennan

Daniel King-Alan

### Project Description

### This project observed the YouTube ‘s Videos category, counts, likes and comments based per country and its relation to each other. This project answers the question

### Our analysis focuses on answering below questions:

### Is there a significant difference in the average like count of the top 50 videos within each designated region within YouTube?

### Ha: Average like count will be significantly different between the regions

### H0: There is no difference between the average like count by region

### Is there a significant difference in the average like count of the top 50 videos of each designated category within YouTube?

### Ha: Average like count will be significantly different between the categories

### H0: There is no difference between the average like count by category

### Additional Correlations: What is the average percent of videos liked by the number of views they have received?

### Limitation

### Could only pull 50 videos per country. Also the API which could pull the Category of videos was under maintenance by YouTube so we had to come up with list of 20 category’s (Google search) and use them in project. Also retrieved files per day changes the amount of record. So we have retrieved 5000 records for one and did the analysis.

**How this project works :**

Project is built to call the YouTube API’s dataset related to Video . Due to large size, dataset is saved in CSV file for further analysis(5000 records). Python matplotlib, bar graph, pie charts are used to analyzed the data and specify the correlation in between videos, categories, likes, views and counts. Various methods such a correlation, aggregation, t-test, comparison , ANOVA tests are used to do further analysis of data.

**Build with :**

*Python*

*YouTube API*

*GitHub*

*Jupyter Notebook*

*CSV Files*

**Some of the challenges we faced**

When setting up the pie chart for ‘Percent View Count per category’, we realize that pie chart is colliding country names label names and making impossible to read the names. We came up with solution to separate the pie chart in two sections , 1st part displays the top 5 countries while 2nd is grouped for other countries. Also, we faced issues with Heatmap.

**Features we hope to implement in the future**

Being able to implement the analysis of demographic data such as impact of having internet connectivity impacts the video’s likes, views, comments count.

**Data Visualization :**

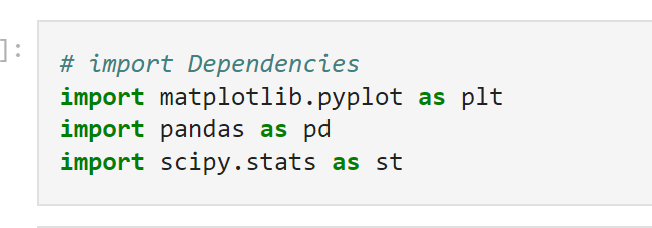
As part of Data Visualization, Bar graphs , Pie Charts, Box Plots are used to display correlation and linear regression of data. P-value and t-test , ANOVA test, linear regression tests are also calculated to show the relation between data and its hypothesis acceptance.

Graphical user interface, text, application

Description automatically generated

Table

Description automatically generated with medium confidence



A picture containing graphical user interface

Description automatically generated

Chart, pie chart

Description automatically generated

Graphical user interface, application

Description automatically generated

Chart

Description automatically generated

**Contact :**

<https://github.com/PatttyCakess/Youtube-API-project>

**References :**

YouTube API - <https://developers.google.com/youtube/v3>

Python - <https://pandas.pydata.org/>

GitHub - <https://github.com/PatttyCakess/Youtube-API-project>