Analysis on YouTube Trending Video Statistics Dataset

Background –

YouTube is the most popular and most used video platform in the world today. YouTube has a list of trending videos that is updated constantly. Here I will use Python with some packages like Pandas and Matplotlib to analyze a dataset and represent observation via different graph.

The dataset contains data more than 40,000 trending videos. I will analyze this data to get insights into YouTube trending videos, to see what is common between these videos. Those insights might also be used by YouTube creators who want to increase popularity of their videos on YouTube.

Data Source-

I will be using "Trending YouTube Video Statistics" dataset (link –https://www.kaggle.com/datasnaek/youtube-new)

This dataset includes several months of data on daily trending YouTube videos. Data is included for the USA, Great Britain, Germany, Canada, and France etc.., with up to 200 listed trending videos per day.

Each region's data is in a separate file. Data includes the video title, channel title, publish time, tags, views, likes and dislikes, description, and comment count. The data also includes a "category_id" field, which varies between regions which can be retrieved by respective JSON files.

Metadata of dataset-

Columns -

- video_id
- trending_date
- title
- channel_title
- category_id
- publish_time
- tags
- views
- likes
- dislikes
- comment_count
- thumbnail_link
- comments_disabled
- ratings_disabled
- video_error_or_removed
- description

CSV files (each having 16 columns and around 40000 rows)

- CAvideos.csv
- DEvideos.csv
- FRvideos.csv
- GBvideos.csv
- GBvideos.csv
- JPvideos.csv
- KRvideos.csv
- MXvideos.csv
- RUvideos.csv
- USvideos.csv

JSON files-

- CA_category_id.json
- DE_category_id.json
- FR_category_id.json
- GB_category_id.json
- IN_category_id.json
- JP_category_id.json
- KR_category_id.json
- MX_category_id.json
- RU_category_id.json
- US_category_id.json

Goal of analysis –

Here, I am proposing of using YouTube trending video statistics dataset and analyzing in different ways such as

- How long usually a video can trend in different countries?
- How many likes, dislikes, views and comments get by different countries?
- How are views, likes, dislikes, comment count, title length, and other attributes correlate with (relate to) each other? How are they connected?
- Correlation of trending video in between countries.
- Videos from which category has longer trend?
- Users like videos from which category the most?
- What is the ratio of Likes-Dislikes and Views-Comments in different categories?
- What's the tags in the most negative and most positive category? What's the most discuss words for Science & Technology?

etc.

Outcome-

Following outcomes can be observed-

- Top category of all countries.
- We will get to know about correlation between the number of views and the number of likes of trending videos.
- Correlation between the number of likes and the number of comments, and between the number of dislikes and the number of comments.

Above mentioned observations are helpful to find out how YouTube creator's channel is doing.

The YouTube Trend data provide lots of information of the user, which also reveals the insights of video trend and how visitor behaviors.

This provides a straightforward visualization of the data and can be interacted with the user, which helps the YouTube decision maker to better understand customer behaviors and make strategies.