INFO 5002 Final Project Summary

Youtube Trending Video Analysis

Team: Anything (Anqi Guan, Mengyue Liu, Minji Kim, Run Hou)

Project Introduction

During our project, our team decided to focus on the Youtube Trending Video Analysis within United States. Social media has taken a vital role in people's lives today. Among all the social media platforms, Youtube is one of the biggest platforms that most people use. We wanted to find out what videos people are most interested in and how these trending videos are categorized.

User Story

Through our project, we can see different graphs and tables that identify the largest number of trending videos, the hours of the published videos, the categories of the videos, the number of videos during the past years, the title lengths, and the likes and views of the videos. Since graphs and tables are clearly shown, users can easily compare the trending videos and how these videos are chosen through different circumstances.

Programming Language

Our project was done in python through the Jupyter notebook. We imported the US Youtube trending data and used matplotlib to visualize the information we imported. The scatter plot was used to show if the title length affects the views of the videos. Plt.subplot was used to create the grids and ax.scatter was used to insert the information needed for the scatter plot visualization. Sns.barplot was used to visualize the bar plot for the number of videos collected during the specific year. This sets the x label and the y label along with the values. Sns.displot method is for the distribution plots. We would set the x and y values and have the collected data distributed, which shows the views of the collected data videos. For the data without non-numerical columns, we grouped the data we collected with the groupby variable or used the correlation variable (corr) to illustrate the data in tables.

Challenges

There were few challenges that we went through during our project. Inputing the information within the graph methods was a challenge. When one input results to an error, we had to go back and check again in order to create the graph we want. For example, for the publishing times and trending videos graph, it was confusing to extract the information in the publishedAt column.

Further Development

In the future, if we have a better knowledge in python coding, I believe the project can be done better by having different comparisons from various countries. Since each countries have different trending videos, we could compare them and find out what videos are commonly trending in the countries and what videos differ from each other.