YouTube Trending Videos Analysis

**Group Name:**

Teletubbies

**Leader:**

HUANG Zefei 20449496

**Group members:**

CAI Runlin 20426550

CHEN Xiaoqi 20465106

GUO Yuju 20465769

# **Introduction**

# The media landscape has changed radically since YouTube videos were launched. We analyze YouTube videos, study what people are looking at these days, and what are the hot spots.

# **Method**

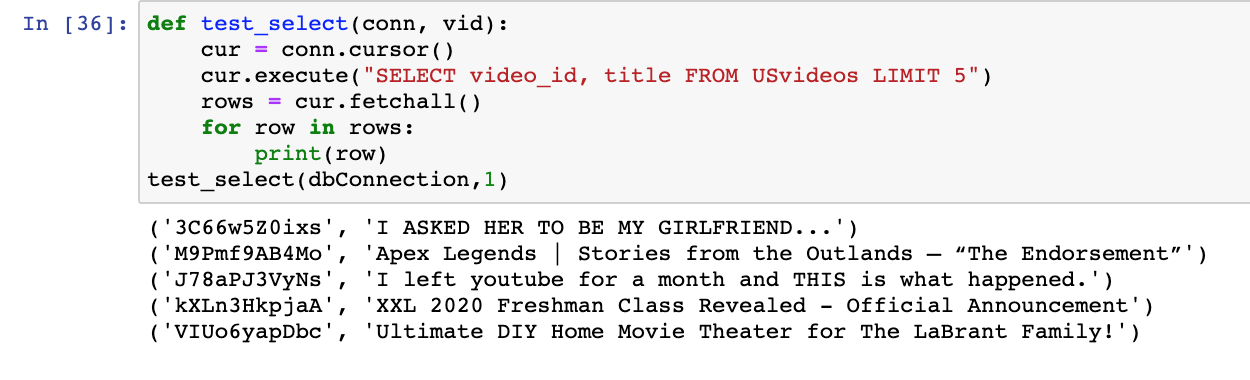
# The data we downloaded from Kaggle are two formatted files, JSON and CSV, are we read them.

Firstly, the documents both need to be cleaned, so we use the drop function to delete some useless lines, and we use panda to transform the CSV document to the data frame. After that, we parsed the JSON document to get the Id and title of the category.

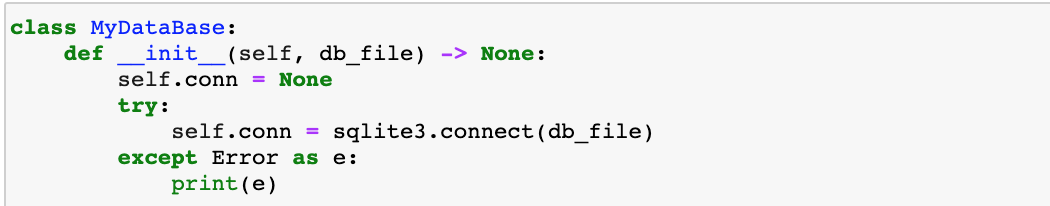
The second is the first way for big data management, we created the database and table，and we use the append（）function to Insert the data of CSV and JSON into the database.

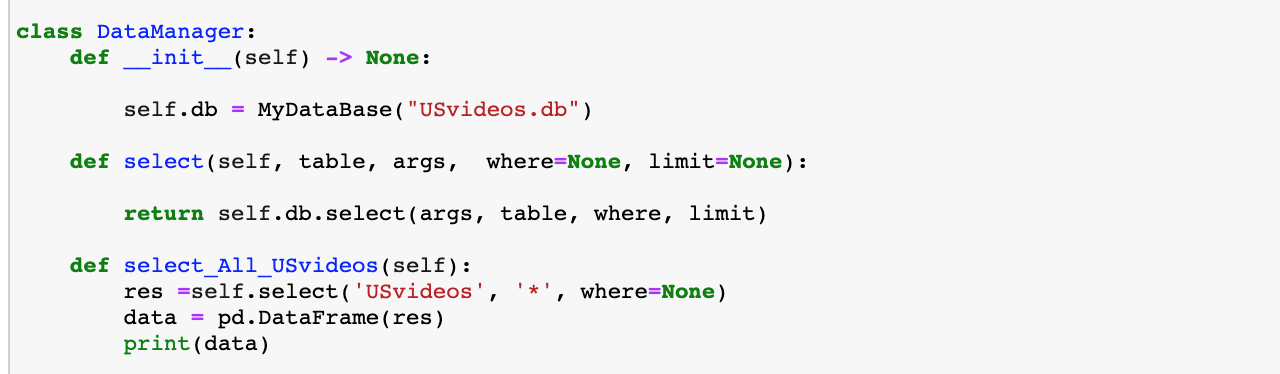


Third, for the update part of the database management, the method is as follows: after connecting to the database, we define a function to achieve UPDATE Priority, begin date, and end date of a task with the statements in SQLite. A detection feature is added to make it easy to verify that the data is updated.



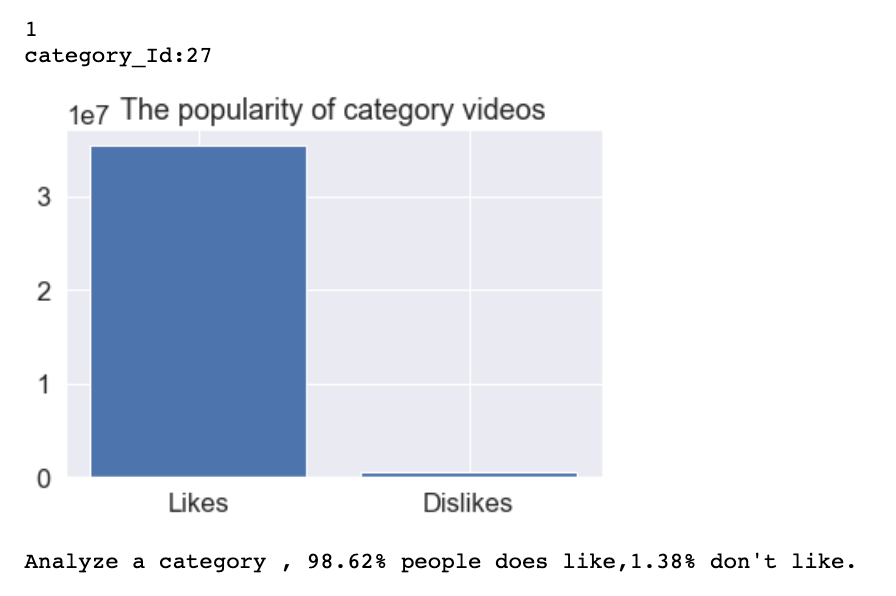
And then we create a user-view class that wraps functionality and data together because the methods of the class are different from normal functions, so they have to have an extra first argument name self .Self function is an instance of the class that we're representing and after we set up the class, we initialize and we connect to the database.





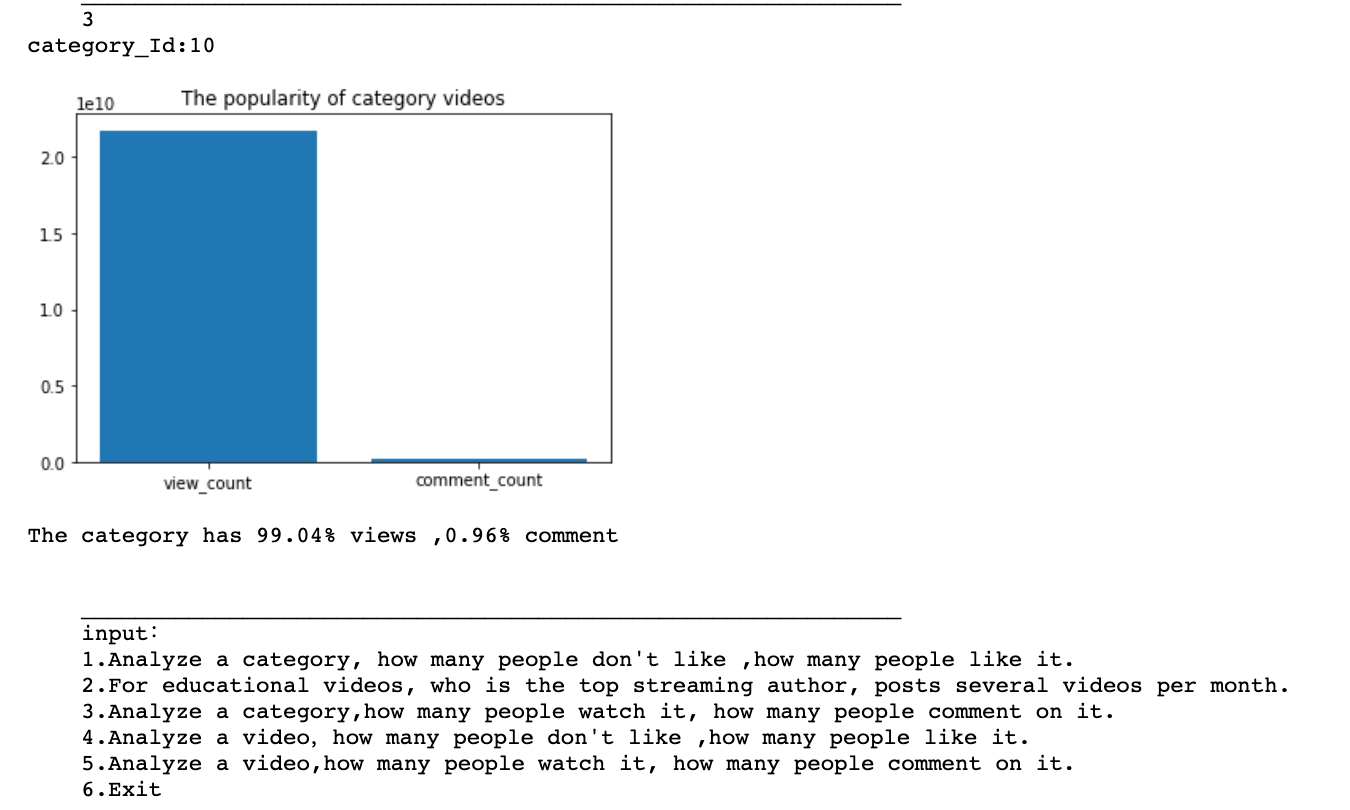
We defined several functions in the class, including querying data, querying all data, querying the popularity of different video types and the first channel played by educational class and groups uploads per month, creating new tables, extracting data from two tables to generate a data frame, and inserting the querying data into the new table.

For generating templates, we use interactive programs for user objects that allow users to select their own problems and then replace the data into the text of the settings, and visualize. 



For example, the user can select any question they want to ask, then enter the ID to generate statistical text and visualizations of the popularity of the category, and then invoke the previous insert function to insert the data into the table for the new category analysis and test whether the insert data was successful through the select function.

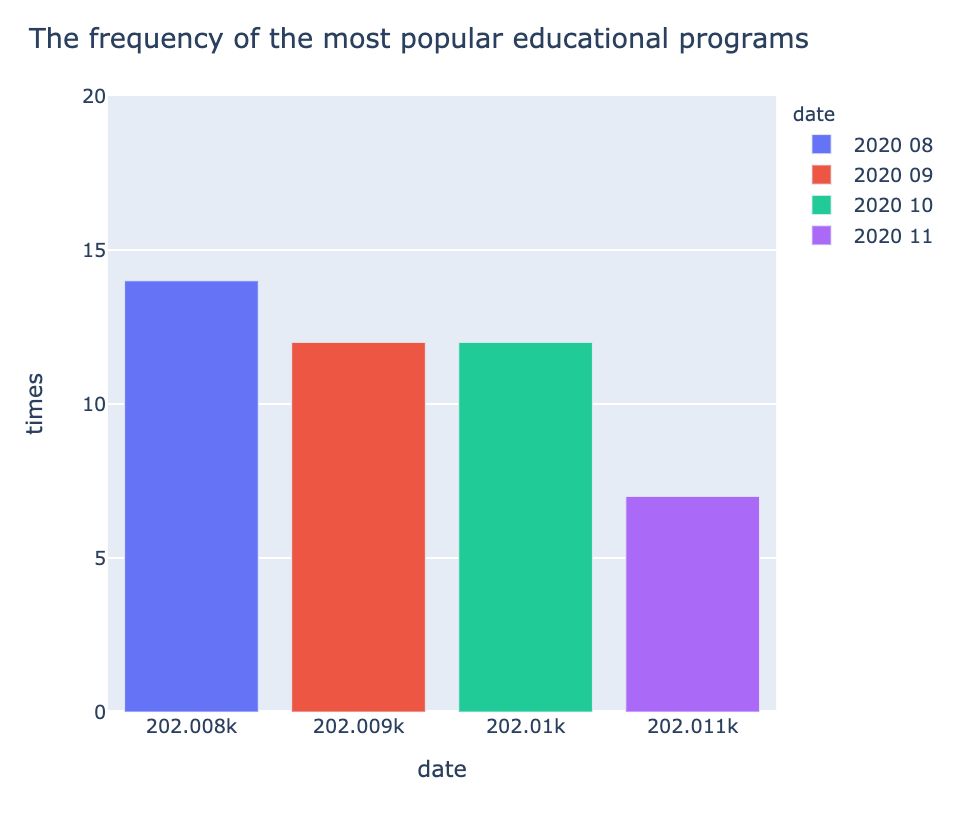




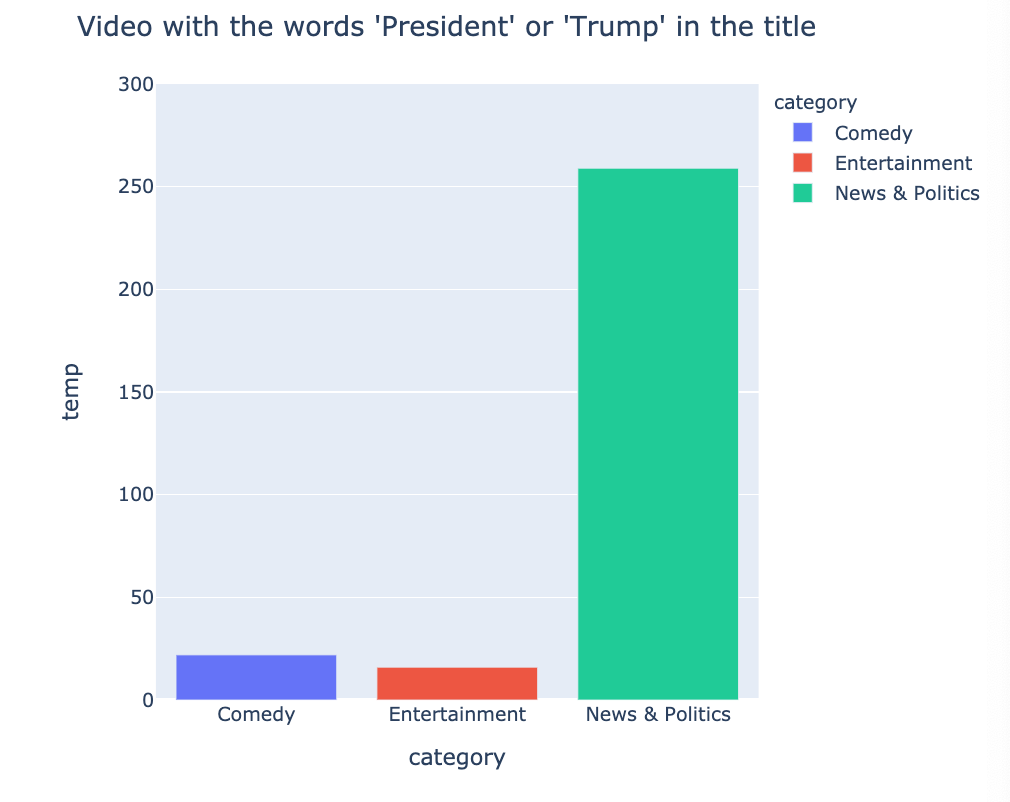
# **Result and Finding**

We have implemented the form of directly generating data frame, so we can directly use the data analysis visualization.

First, we explored the frequency of the most popular educational programs in August 2020, people visit educational programs most.



Second, we can find that videos with titles with"President" or "Trump" appear the most in News&Politics with more than 250 temps. Comedy and entertainment have a similar frequency.On YouTube, users also like to make some comedies for entertainment. And people don’t take politics so seriously and can use the president as a topic of entertainment.

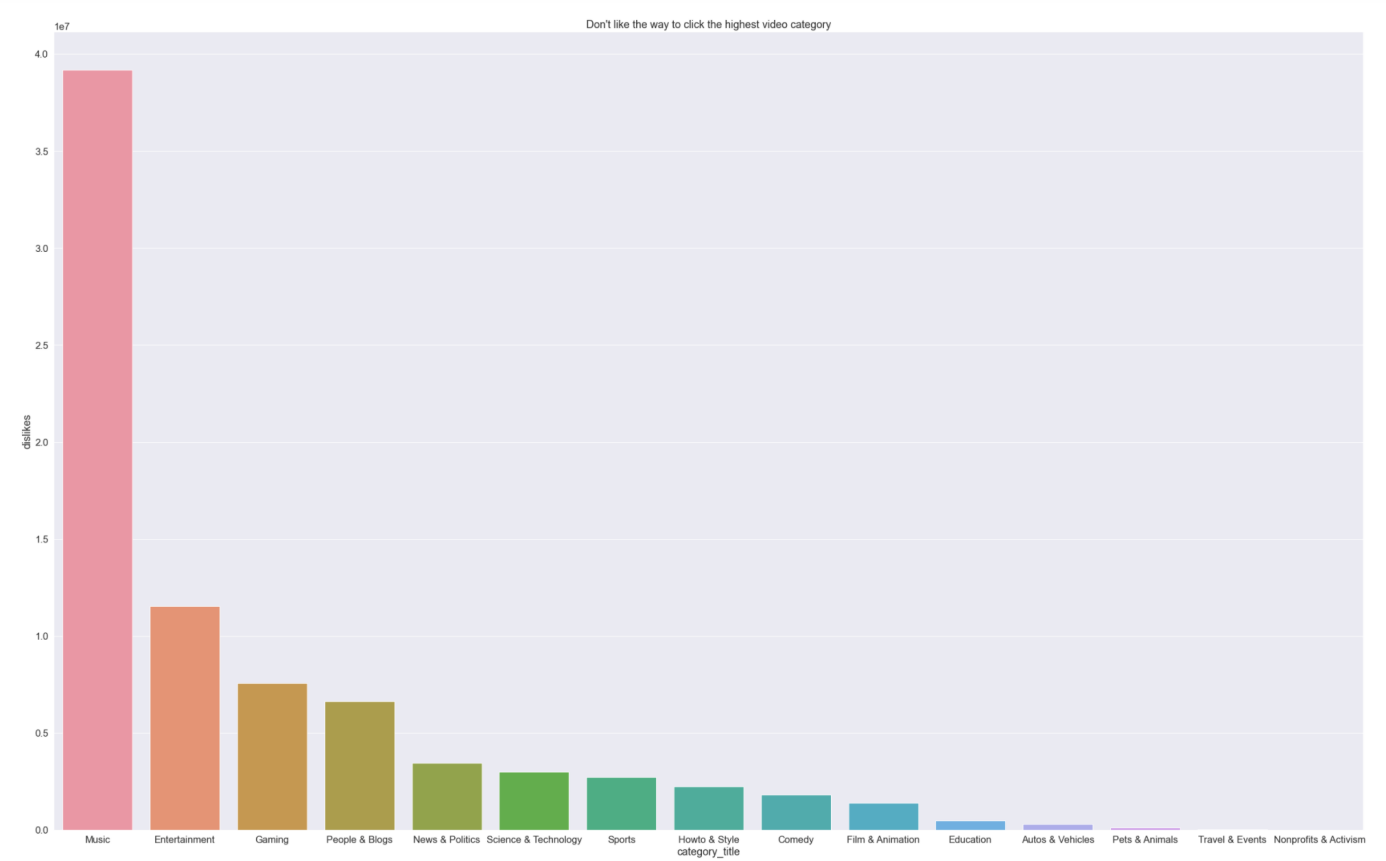


Third,we use the previously defined select\_all\_table () to query all data for the visualization part.Query the titles of all popular videos and perform word cloud analysis.The results show that people are most interested in art, music, and video.

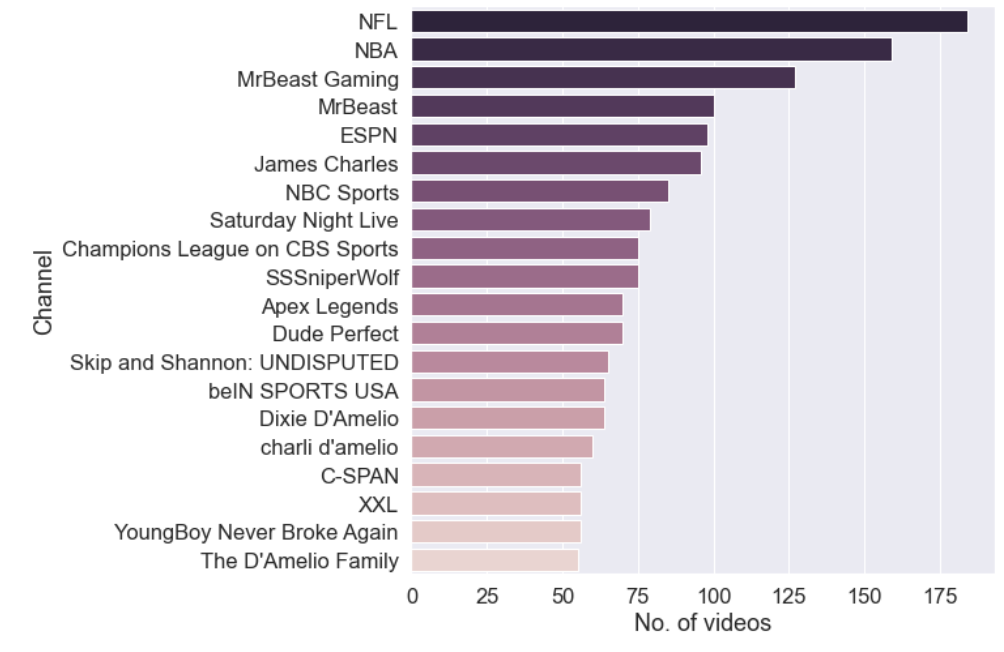
On the other hand, people prefer to watch videos and music. People will also pay attention to some interests.



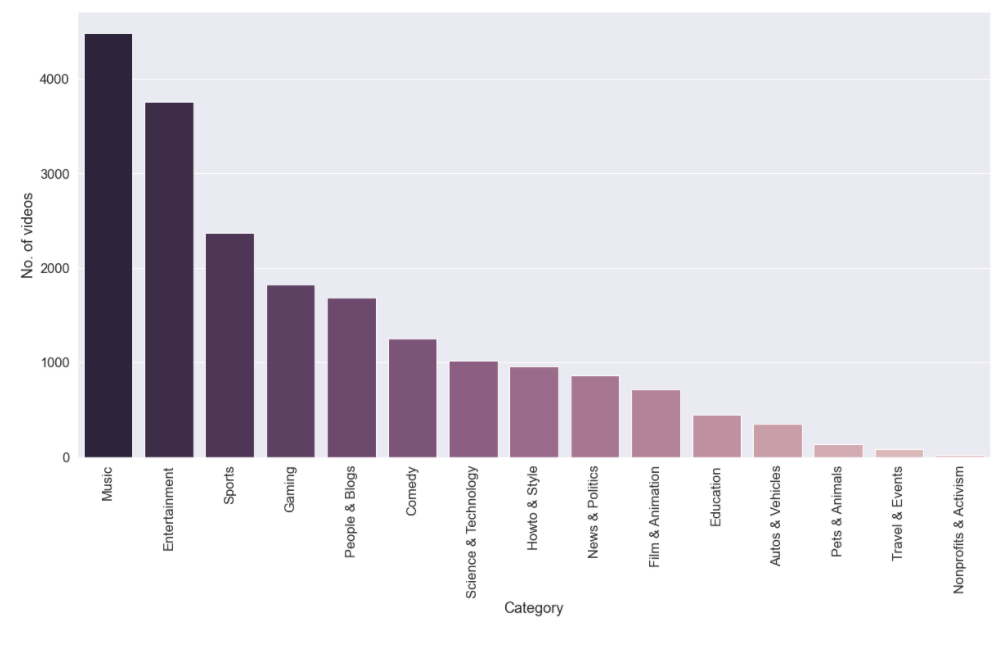
Fourth,we counted the types of videos with the highest number of clicks on "Dislike". We found that the music category is the highest. People have multiple experiences and comments on different types of music or music video production styles.



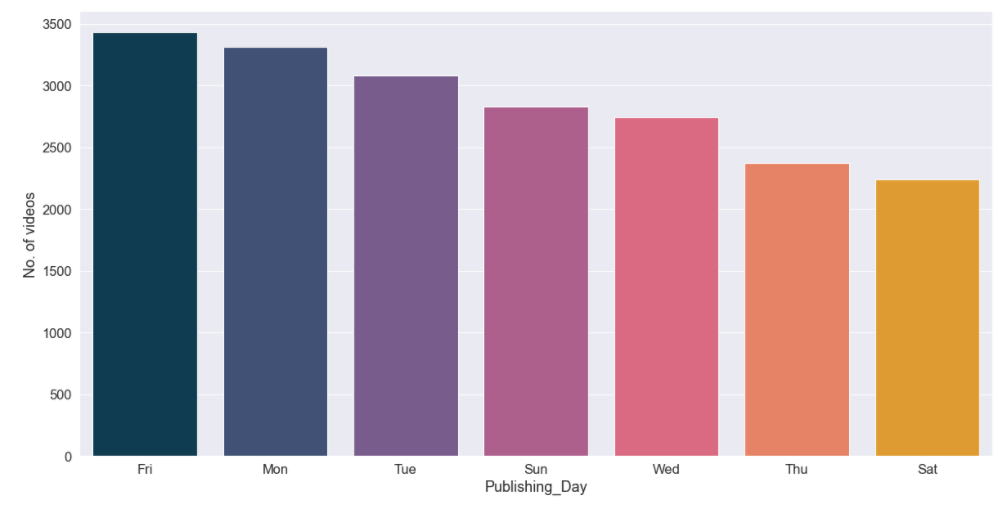
Fifth,the most popular channel is NFL and next one is NBA and MrBeast Gaming.



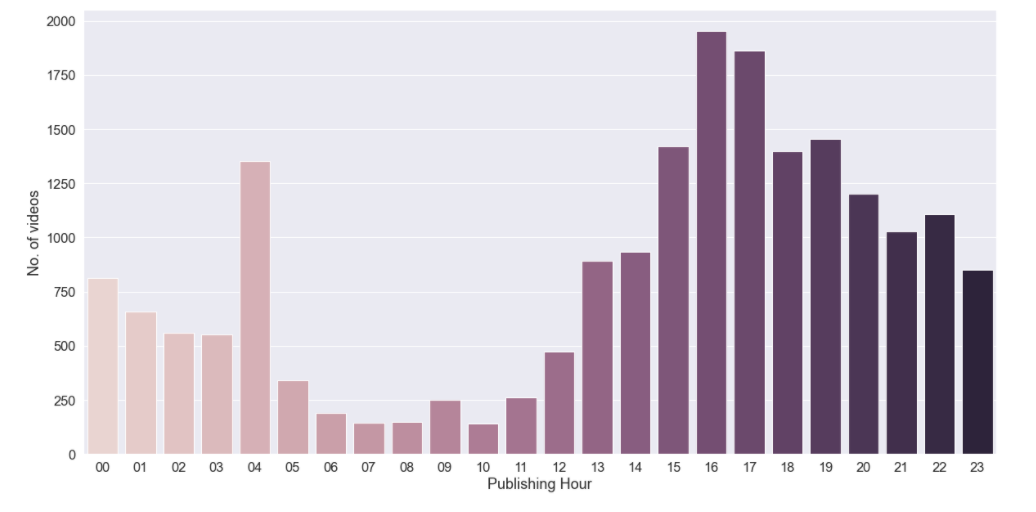
Sixth,the picture shows that the most famous category is Music and entertainment. The third is sports, because of some official sports accounts(NFL and NBA). Gaming is also hot.



Seventh,we use the publishing\_hour column to see which publishing period has the most trending videos. The number of trending videos published on Sunday and Saturday are noticeably less than the number of trending videos published on other days of the week.



Finally,we use publishing\_hour column to see which publishing hours had the largest number of trending videos,and people watch YouTube videos post from 16:00-17:00 most,because people finish work or class and they are free to watch videos.



# **Conclusion**

People want to get official news and the latest news on YouTube, and they also watch some videos about their hobbies or listen to some music.

Music is the hottest part of YouTube, but the 'dislike' of the music is the highest. People have multiple experiences and comments on different types of music or music video production.