TEXT ANALYSIS – YOUTUBE DATASET

Imported libraries – **PANDAS, NUMPY , SEABORN ,MATPLOTLIB**

Read the Data by using pd.read\_csv()

Searched For Null Values

Eliminated Null Values

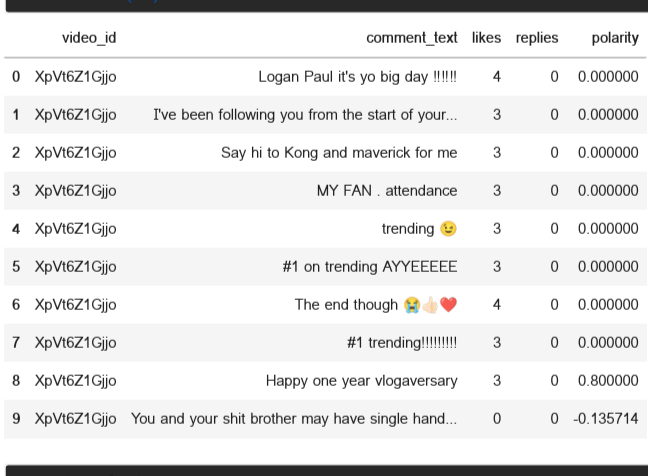
* **SENTIMENT ANALYSIS**

Imported textblob

Used **.sentiment()** this helps us to determine whether the sentiment is positive or negative which is 1 determines positive , -1 determines negative , and 0 determines neutral

Used . .Polarity() determines the polarity of the sentiment which we know is [ 1 ,0 , -1]

We created a Column polarity and added it to the main table . shown with head(10)



* **WORD CLOUD ANALYSIS**

Imported wordcloud ,STOPWORS

Filtered negative and positive comments

Removed unwanted words unessential to analysis like (a ,about ,after ,again, all , etc. ) by set(STOPWORDS)

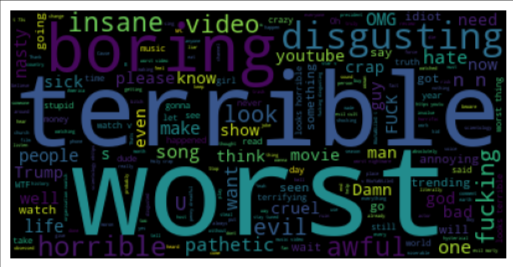
Converted the comments to string

Positive words plotting



We can see most positive words used are

1. **Awesome**
2. **Best**
3. **Perfect**



Negative word plotting

We can see most Negative words used are

1. **terrible**
2. **worst**
3. **boring**

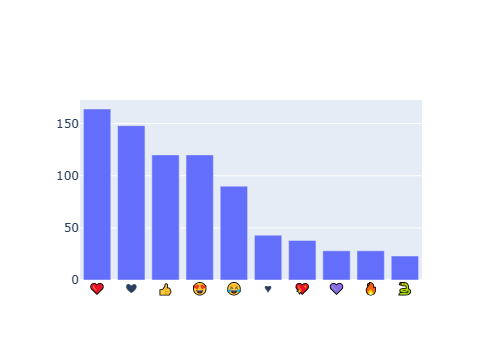
* **EMOJI’S ANALYSIS**

Installed and Imported emoji library

Identified types of emoji ‘s

Counted no of times emoji’s has been used

Represented it by bar chart



* Heart emoji was used the most
* **DATA COLLECTION**

Collected the entire data set of different countries

* **Exporting Data**

Exported all collected data to json and sqllite by using sqlalchemy

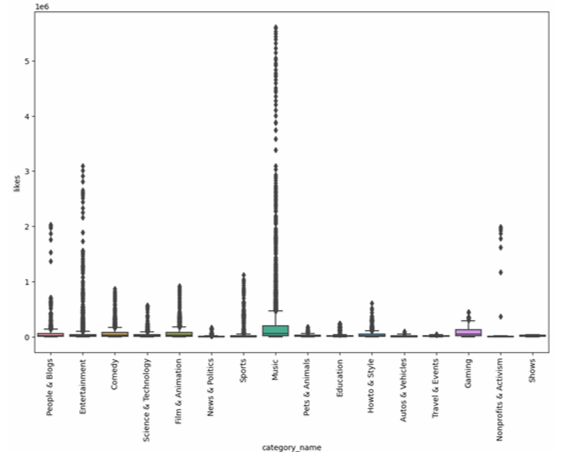
* **ANALYSING THE MOST LIKED CATEGORY**

Found the Unique Category

Read the converted json data path

Joined the category column to the table

Plotted likes vs category

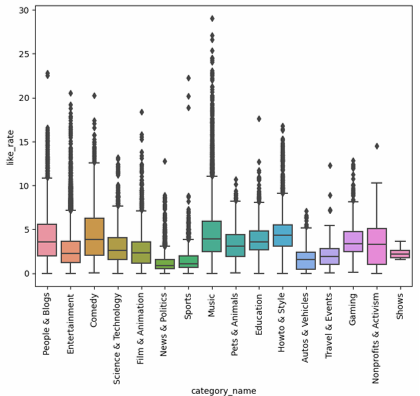


* We can clearly see that Music has the most likes
* Shows has the least likes
* **Finding Out The Audience is Engaged or Not**

Finding LIKERATE , DISLIKERATE ,COMMENT COUNT RATE

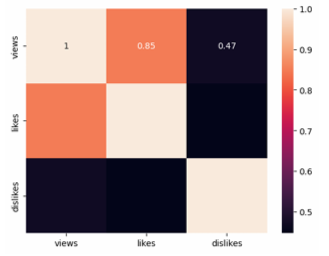
(divided by views \* 100)

Plotting likerates vs category name



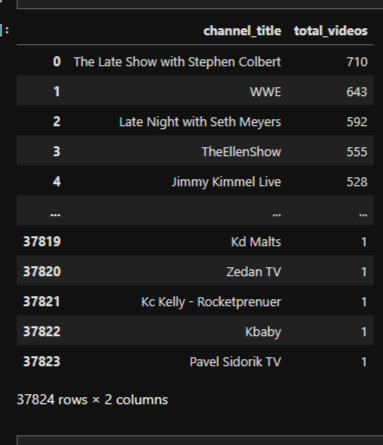
* Music Has the highest like rates
* Shows have the lowest

Plotting views vs likes vs dislikes



* **Find out which channels has the highest number of trending videos**

Used .values\_count()



Plotted the channels vs total videos

A graph of blue bars

Description automatically generated with medium confidence

* **Does Punctuations have any connections to LIKES , DISLIKES , VIEWS, COMMENTS**

Imported string lib

Used string.punctuation to detect punctuation in the texts

Counted no of punctuations

Plotted views vs count\_punctuation ,likes vs count\_punctuation ,