

HOW POPULAR WILL THIS VIDEO BE ON YOUTUBE ?

Team : Noura Al-Otaibi, Miatha Al-Qahtani, Alaa Al-Ghamdi





Introduction

As YouTubers earn money through the advertising and bonus of their videos, increasing the popularity of their videos is a priority. This project aims to predict the behavior of the video that is going to be uploaded to YouTube. An equation is developed to manually classify all the videos into binary groups: non-popular and popular.



01

NUMPY

04

SCIPY

02

**MATPLOTLIB AND SEABORN
FOR PLOTTING**

05

SKLEARN

03

Pandas for data manipulation

06

WORDCLOUD



Exploratory Data Analysis



Data cleaning



Modeling and compare models

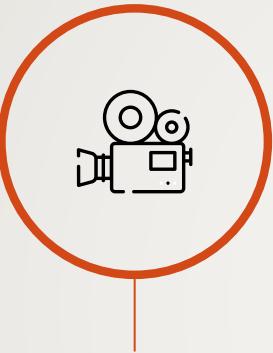


Statistical analysis



Preparing data to modeling





DATA SOURCE

Kaggle as an open-source.
we selected the trending videos of YouTube
from the United States and Canada, consisting
of 29089 unique videos.



DATASET

RECRDS: 81830
Columns: 17





	video_id	trending_date	title	channel_title	category_id	publish_time	tags	views	likes
0	2kyS6SvSYSE	17.14.11	WE WANT TO TALK ABOUT OUR MARRIAGE	CaseyNeistat	22	2017-11- 13T17:13:01.000Z	SHANtell martin	748374	57527
1	1ZAPwfrtAFY	17.14.11	The Trump Presidency: Last Week Tonight with J...	LastWeekTonight	24	2017-11- 13T07:30:00.000Z	last week tonight trump presidency last week ...	2418783	97185

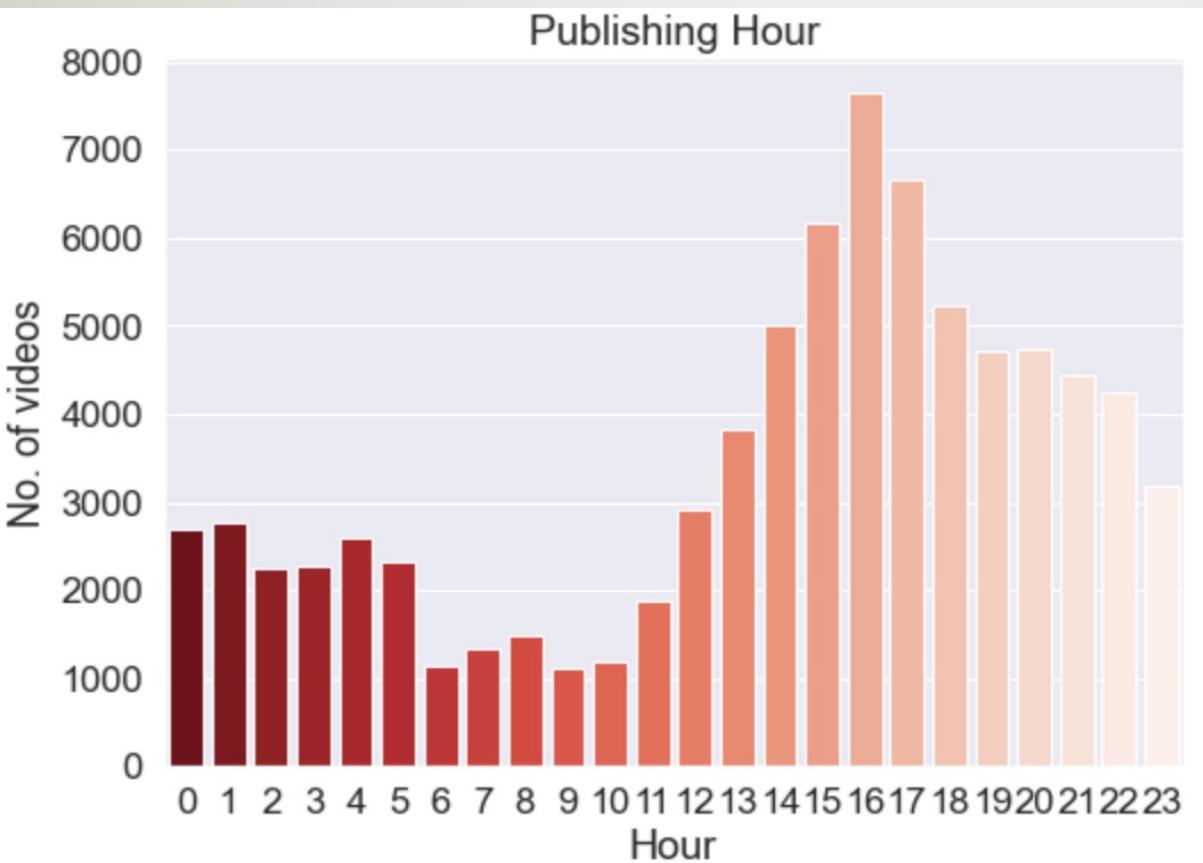
- **Formatting the datetime**

	video_id	trending_date	title	channel_title	category_id	publish_time	tags	views	likes
0	2kyS6SvSYSE	2017-11-14	WE WANT TO TALK ABOUT OUR MARRIAGE	CaseyNeistat	22	2017-11-13 17:13:01	SHANtell martin	748374	57527
1	1ZAPwfrtAFY	2017-11-14	The Trump Presidency: Last Week Tonight with J...	LastWeekTonight	24	2017-11-13 07:30:00	last week tonight trump presidency last week ...	2418783	97185

- We created a new y for prediction by eqution

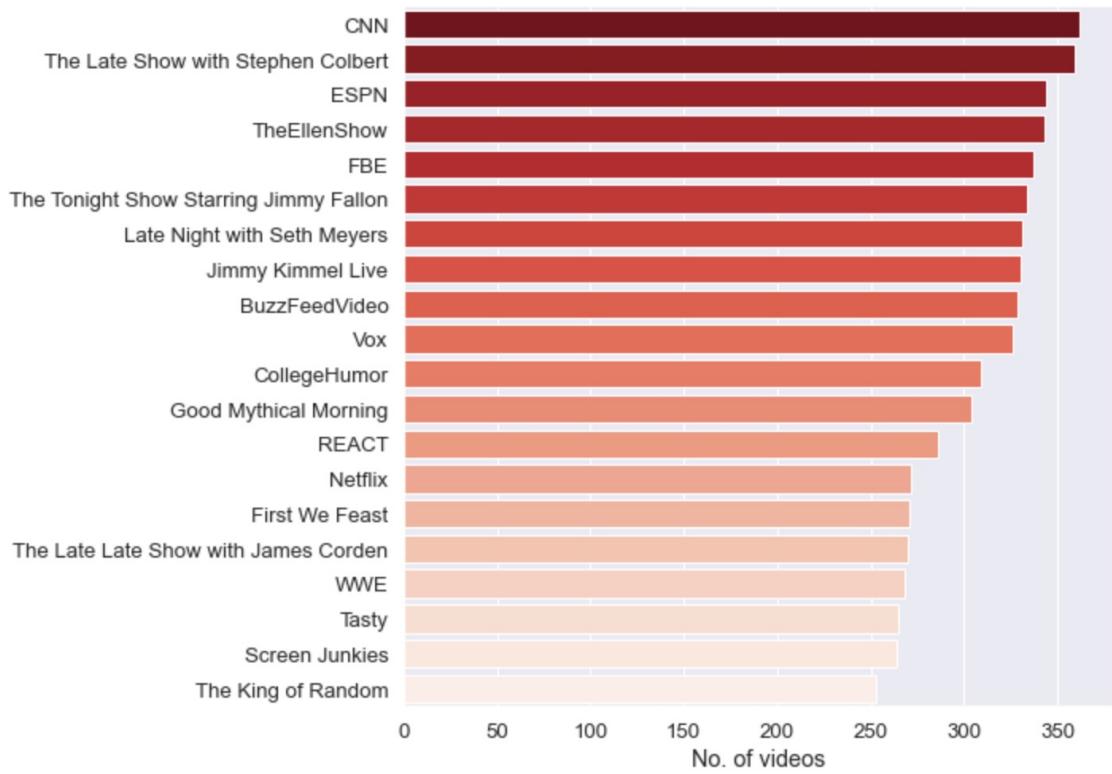
dislikes	...	trending_Year	trending_Day	trending_Month	publish_hour	publish_day	publish_year	publish_Month	score	video_type	video_type_no
2966	...	17	14	11	17	13	2017	11	1131.528370	Trending	1
6146	...	17	14	11	7	13	2017	11	461.981128	Trending	1
5339	...	17	14	11	19	12	2017	11	353.815380	Trending	1
666	...	17	14	11	11	13	2017	11	57.363326	not Trend	0





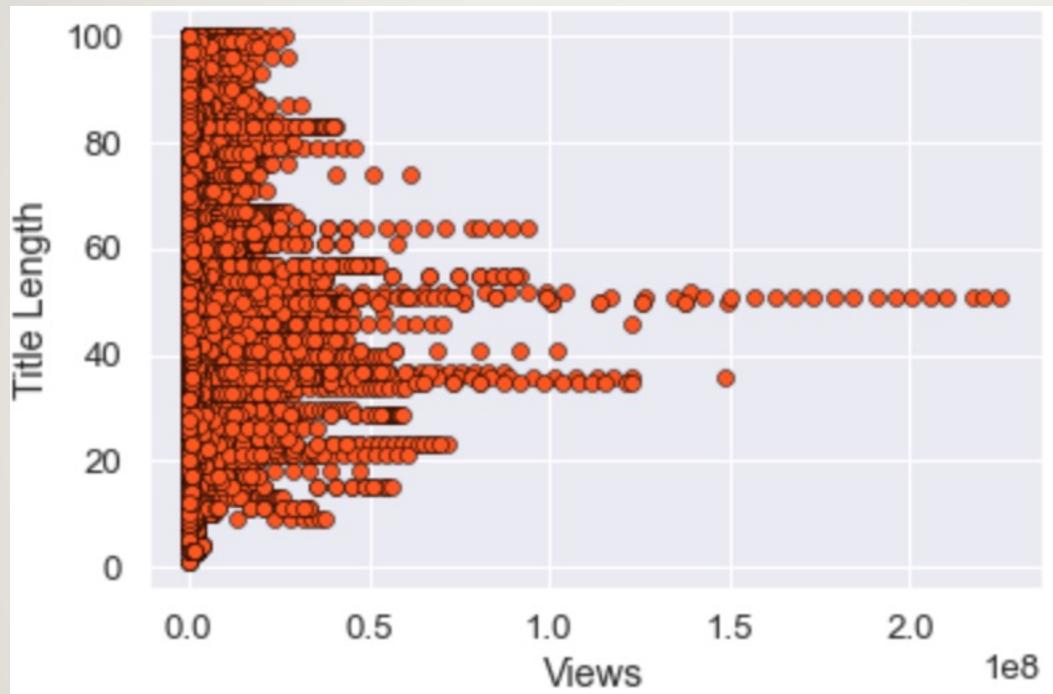
observation that number of videos published at 4 pm, 3 pm, 5 pm and 6 pm trend more than the number of videos uploaded 8 am, 7 am, 6 am.





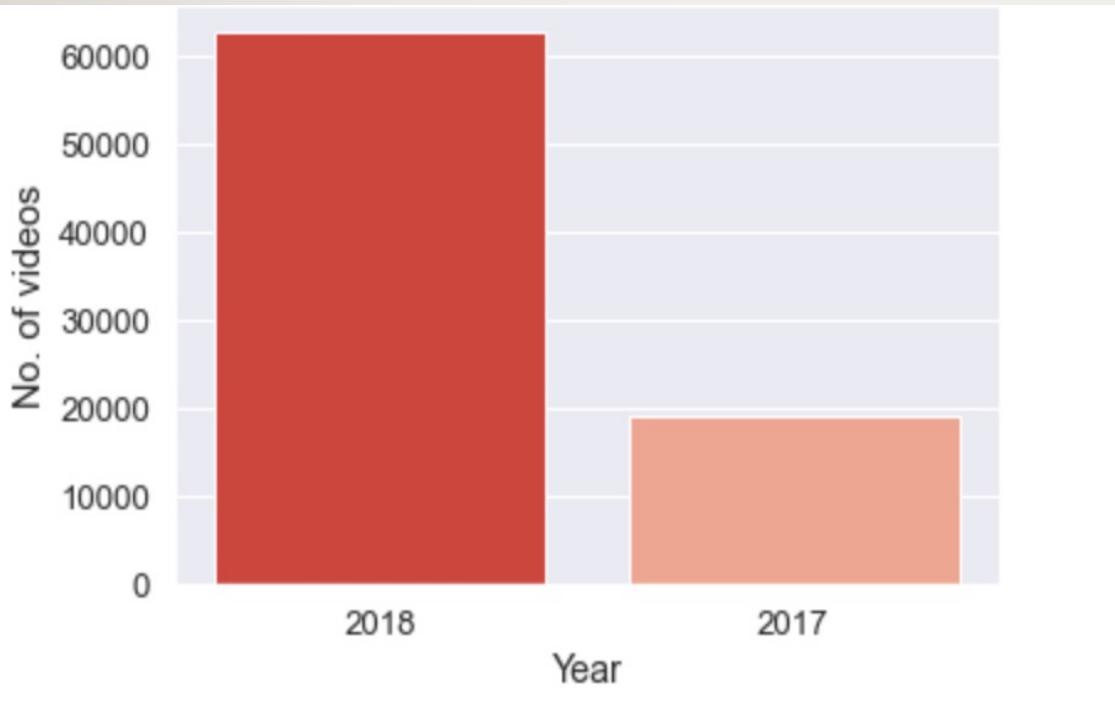
As we see the highest channels watched CNN, ESPN , The late show with Stephen.





plot shows videos that have 100,000,000 views and more have Title Length between 35 and 60 characters approximately.





We can see that the dataset was collected in 2017 and 2018 with 77% of it in 2018 and 23% in 2017



Episode Trailer ft

Live Punjabi Show Week Night Review Ball Royal Love War Day Secret Explained Make Starcast Girl NBA March Smith Dog Theory Teaser Breaking Song Lyric Award Play Now Old Inside Food Tutorial HD Street Bad Fan Logan Made Wild Question America Back Never Black Michael Minute

Offici1al

Full Video New Challenge

Time Part TV ESPN Got School Talk Audio Lyric S Music Kardashian Moment X Final Life Home Cardi

MAKEUP President Tour People Question Logan Made Wild Street Bad Fan Logan Made Wild Question America Back Never Black Michael Minute

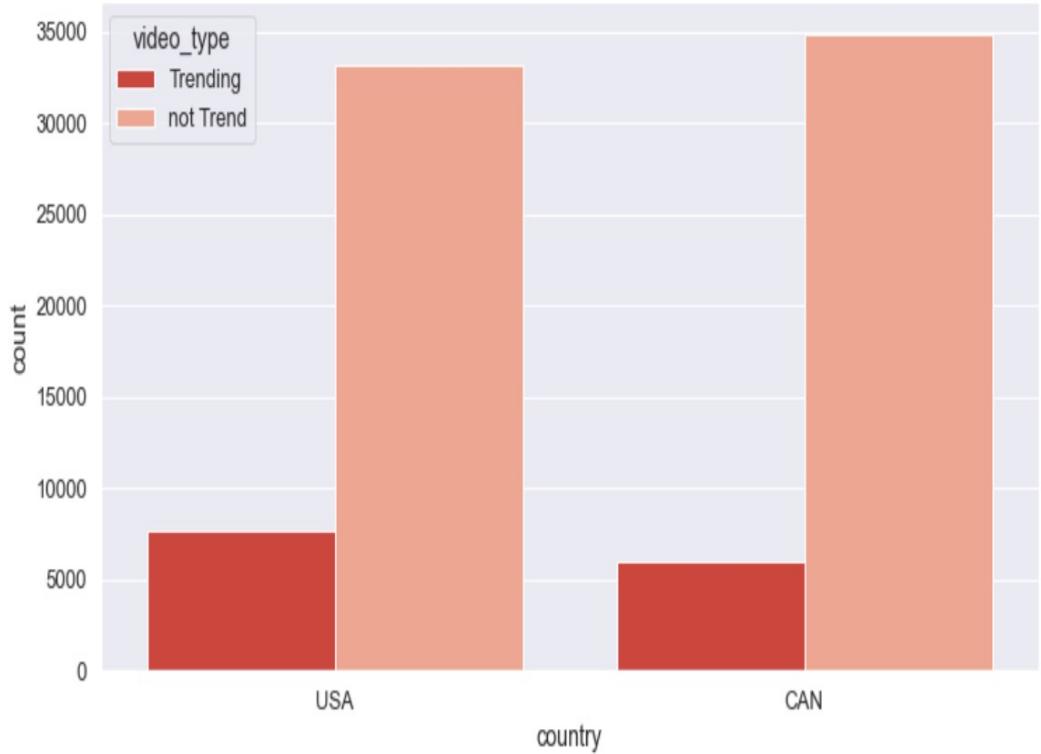
April Try Voice Open U See Every January BTS YouTube Everything Highlight Comment Commercial Meghan World La Movie Back Never Black Michael Minute

Battle Netflix Gets Interview Infinity Hit View Joe James Honest One Bowl Chris Go Mom Friend Car Win Big Car Win American Wrong Box REACT Wedding Call I Trump Tried feat Year Golden Man Drama Ep Face Prince John Kevin First

Christmas Super DIY iPhone Best King Avenger

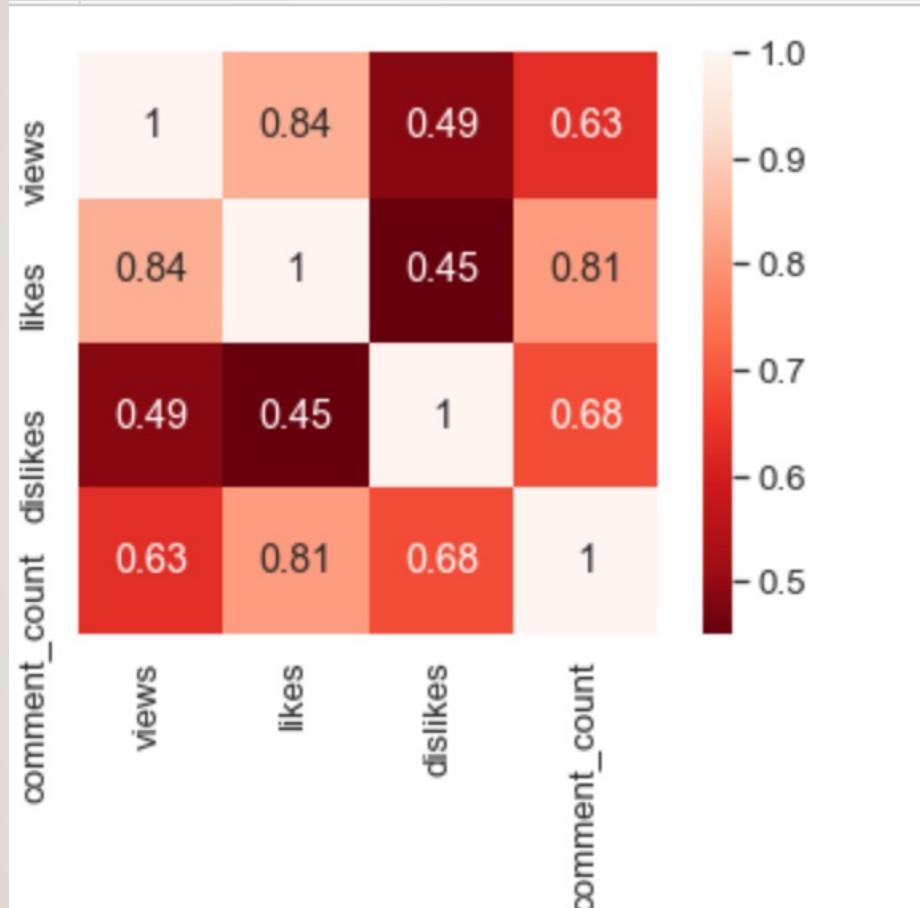
As we can see Trending videos contain words like Official, Trailer, Live, New, Challenge, HD etc. which helps attracting the viewer's attention.





As we see here the number of videos that become trending between the United States and Canada, the United States is higher in the emergence of trend

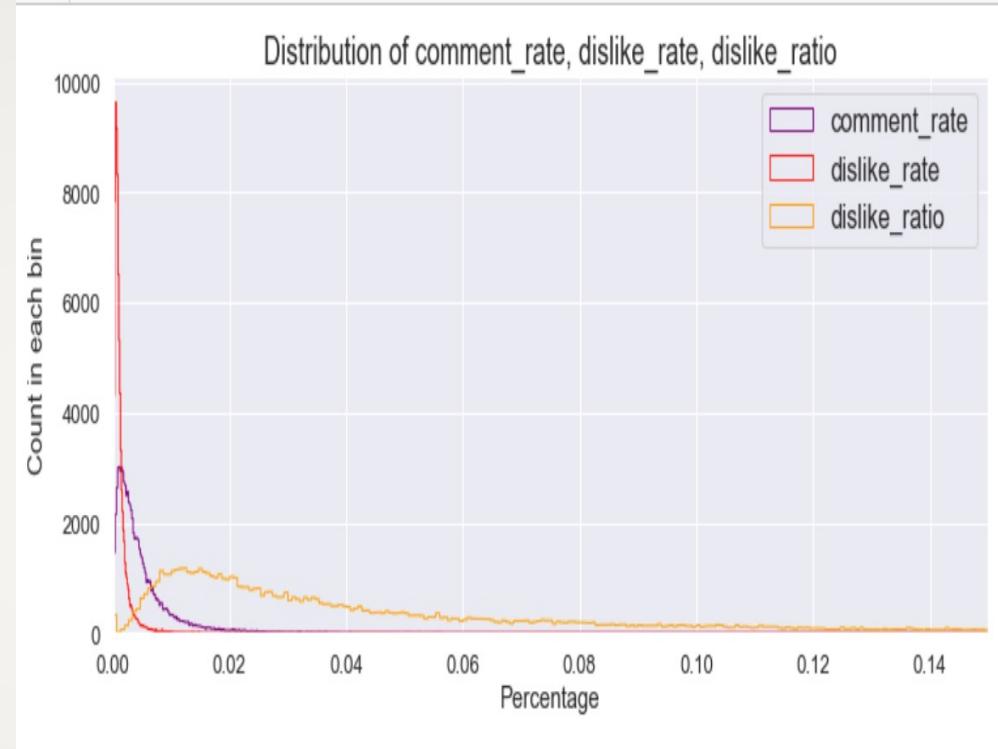




As we can see it is high correlation between Views and likes and likes and comment_count



- ▶ we observe that the majority of trending videos have both low dislike rates and low comment rates. And we can see that the shape of these two distributions are similar (both right skewed).



MODELS RESULTS

Recall_ score	Model name
0.993336	Random_forest
0.991258	Voting
0.989913	Bagging
0.935563	knn
0.851134	Decision_Tree
0.332625	logistic





- **We can help a YouTuber to make their video as a trending one by telling them what day what time and what category of videos they should upload**
- **We can say if a video becomes a trending one when most of the people like that video**

It's tell about audiences satisfaction





THANKS!
DO YOU HAVE ANY QUESTIONS?

