

YouTube Videos - Likes and Views Prediction

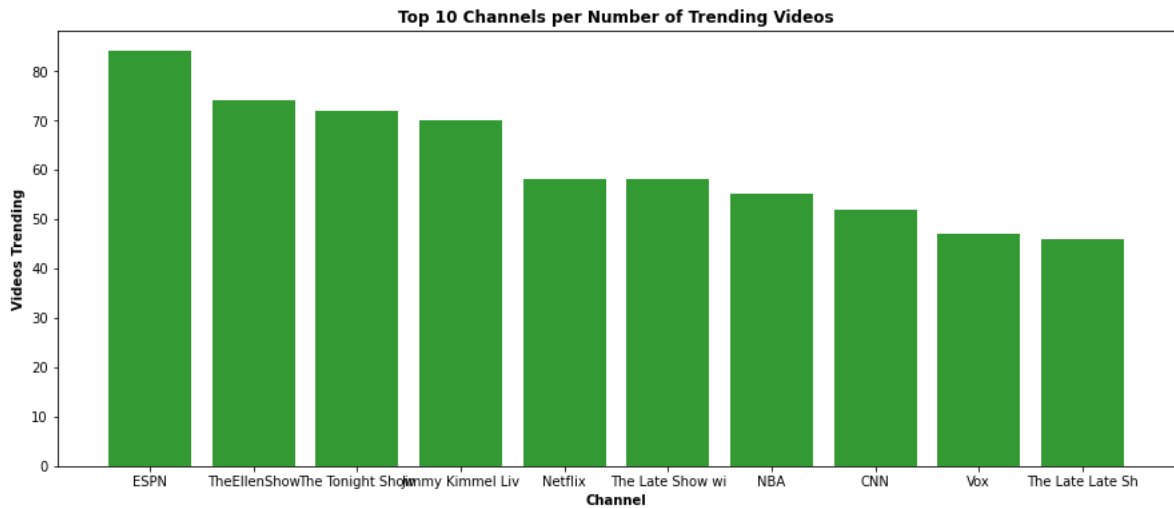
Vikas Ranjan

DSC680, Summer 2021

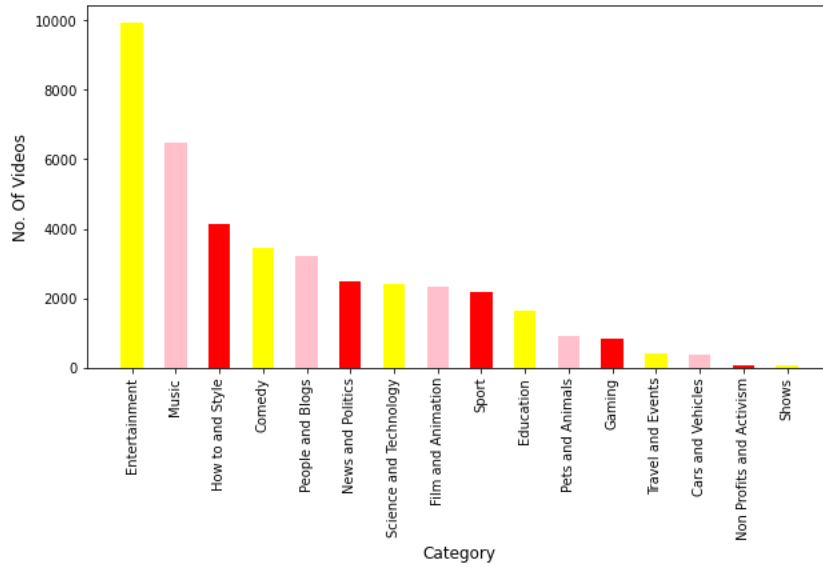
Bellevue University, NE

Objective:

1. Which are top 10 channels based on the number of trending videos?
 - As we can see below, top 10 channels which have greatest number of trending videos on YouTube. ESPN and late night comedy shows leading the chart with most number of views.

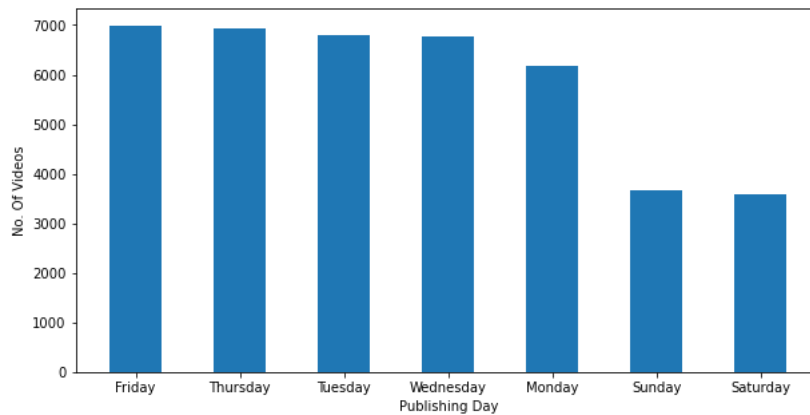


2. Analyze and determine which categories has the greatest number of videos?
 - Above plot shows that the Entertainment category contains the largest number of trending videos with around 10,000 videos, second is Music category with around 6,200 videos, followed by How to & Style category with around 4,100 videos.



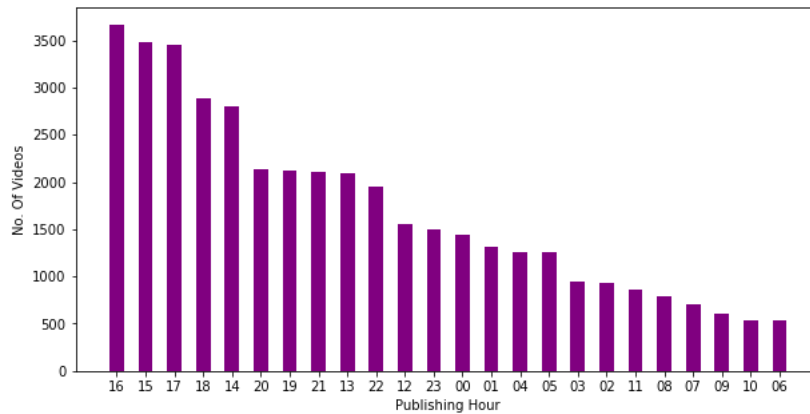
3. Analyze and determine the distribution of uploaded videos based on the publishing day of the month.

- Here we can see the days of the week which had the largest numbers of trending videos. There is a trend that on weekends, there are lesser videos being uploaded.



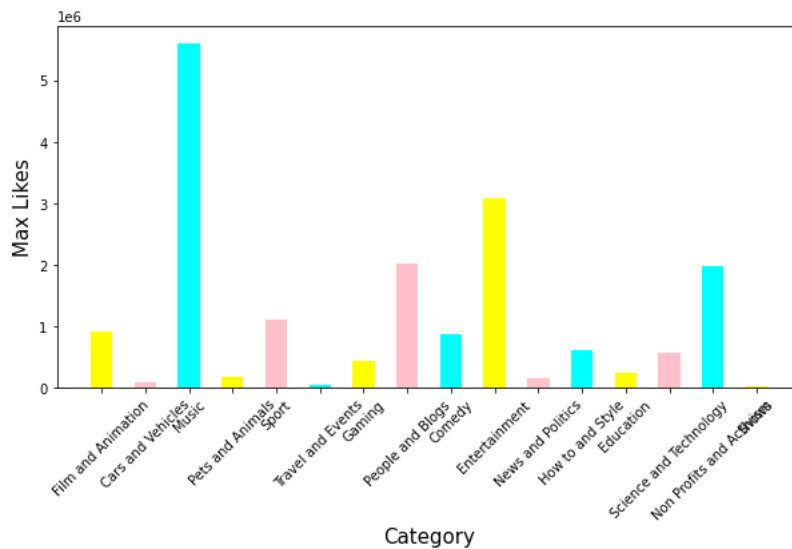
4. Determine the times at which most of the videos are being published? And analysis publishing time distribution.

- We can see the trend that most of the videos are being uploaded between 14:00 and 16:00, highest number of videos being uploaded are around 16:00.



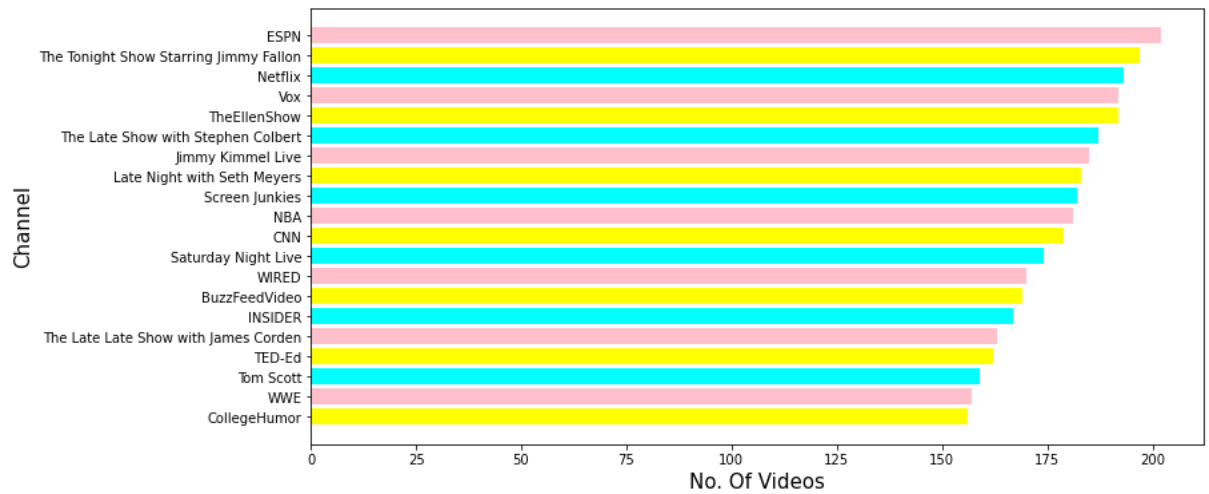
5. Determine top 3 categories in which videos have gathered the greatest number of likes?

- As we can see in below plot, Music and entertainment videos have most likes.



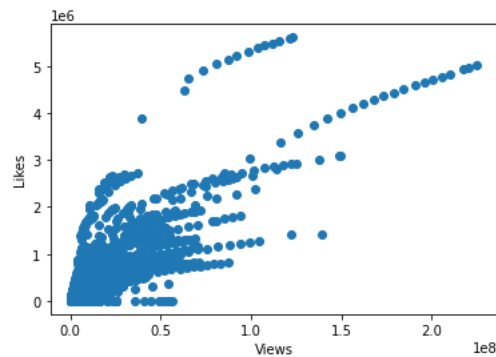
6. Determine and plot the top 20 channels with most number of video uploads?

- Below chart shows the top 20 channels with most number of videos.



7. Determine the correlation between number of likes vs number of views on a video?

- Based on the below scatterplot, there seems to be positive correlation between the no of views and the no of likes of a video.



	views	likes	dislikes	comment_count
views	1.00	0.85	0.47	0.62
likes	0.85	1.00	0.45	0.80
dislikes	0.47	0.45	1.00	0.70
comment_count	0.62	0.80	0.70	1.00

8. Determine the correlation between number of dis-likes vs number of views on a video?

- As we can see there is not strong correlation between dislikes and number of views.

	views	likes	dislikes	comment_count
views	1.00	0.85	0.47	0.62
likes	0.85	1.00	0.45	0.80
dislikes	0.47	0.45	1.00	0.70
comment_count	0.62	0.80	0.70	1.00

9. Perform text analysis and build word cloud to identify most used words in the description?

- Below word cloud shows the most used words in the description.

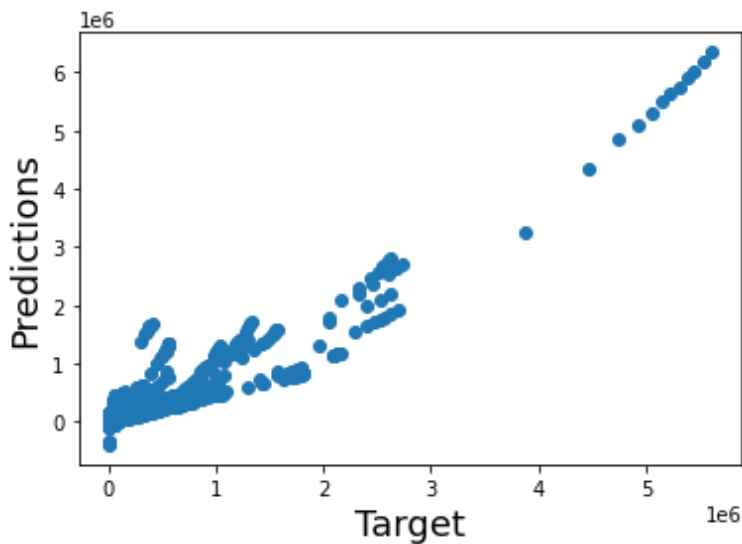


10. Apply, evaluate, and determine which models will be most suitable for the prediction of number of likes and number of views for a video.

- Linear Regression is a good fit when modeling to predict number of likes.

Mean Absolute Error on Training Set: 30058.4653286869
Mean Absolute Error on Testing Set: 32831.88943833925
R-Squared Score on Training Set: 0.8793238222682671
R-Squared Score on Testing Set: 0.887092843540879

Graph below shows the evaluation of the linear regression model.



- To predict the number of views, Random Forest Regressor has better score than linear regression.

Mean Absolute Error on Training Set: 979261.5911413658
Mean Absolute Error on Testing Set: 1017251.9907693673
R-Squared Score on Training Set: 0.7632860415673052
R-Squared Score on Testing Set: 0.7626145637025763

Linear regression model to predict number of views had a R-Squared score of .76 on training dataset. Since linear regression is not accurate enough, I applied Random Forest Regressor.

Graph below shows the evaluation of the Random Forest Regressor model.

