Streaming Youtube Channel Analysis

The following is an analysis and prediction for viewCount of a live stream youtube channel. Public data from the channel is obtained through youtube V3 API with the script in the sub folder, when the program runs it will create a data folder to store the data files. The follow data shows insights on the viewership of 'Only in Japan' live stream channel for the last 300 videos. The viewership count is not the live stream viewership but after.

The benefit in obtaining data from this particular channel is the frequency that it streams at different time and also the various locations provided from the V3 API to determine the viewership.

1)Data Analysis

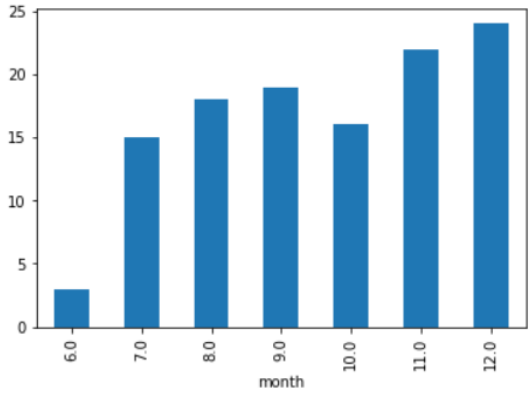
10 Top Streaming Locations

locationDescription	
Tokyo	56
Tokyo Station	14
Akihabara	6
Hibiya Park	4
Sado Island	4
Nihonbashi	3
Ginza	3
Kōyashita Station	2
Jokiin	2
Innsbruck	2

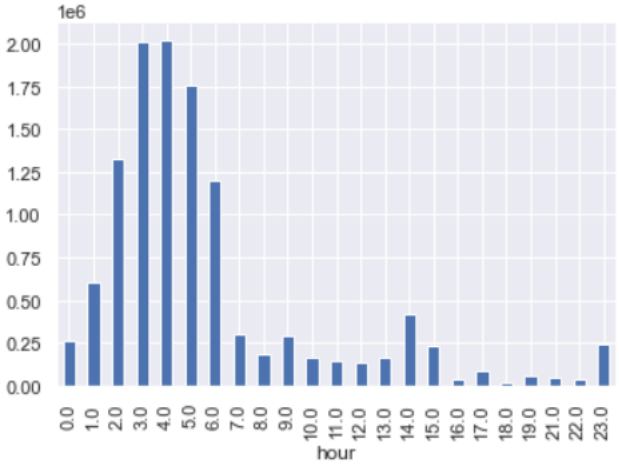
Top 10 location with most viewers

locationDescription		
Tokyo	2090067	
鷲神社(浅草酉の市御本社)		940524
Tokyo Station	720143	
Akihabara	301847	
Sado Island	289968	
Tokyo Dome	223462	
Noboribetsu Station	137922	
Shinbashi	135553	
Ginza	131222	
Sega Akihabara Building 2	121210	

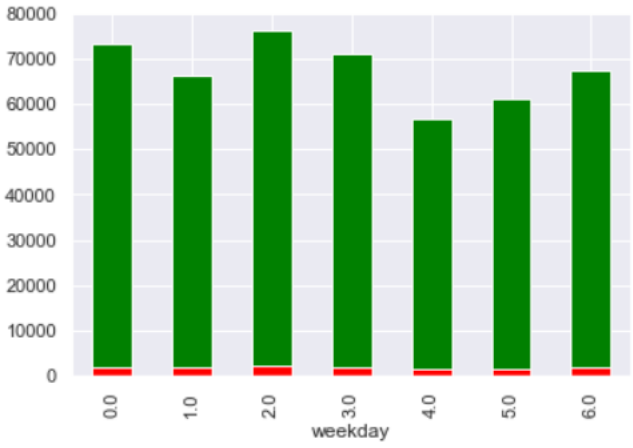
Video distribution varying by month from the last 300 vids



Most viewed hours (UTC-9)

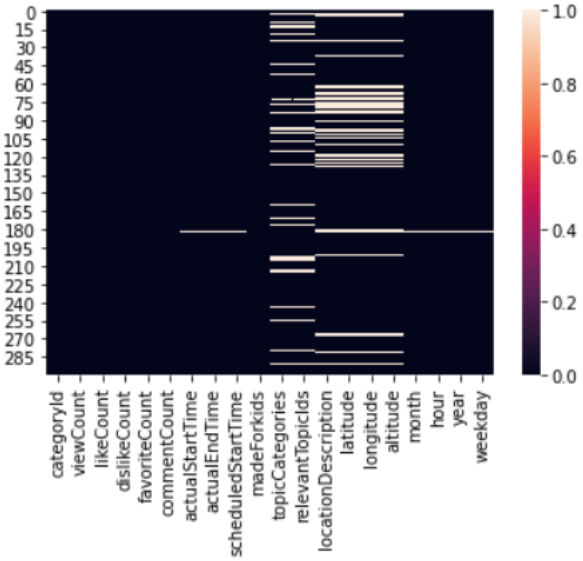


Like/Dislike

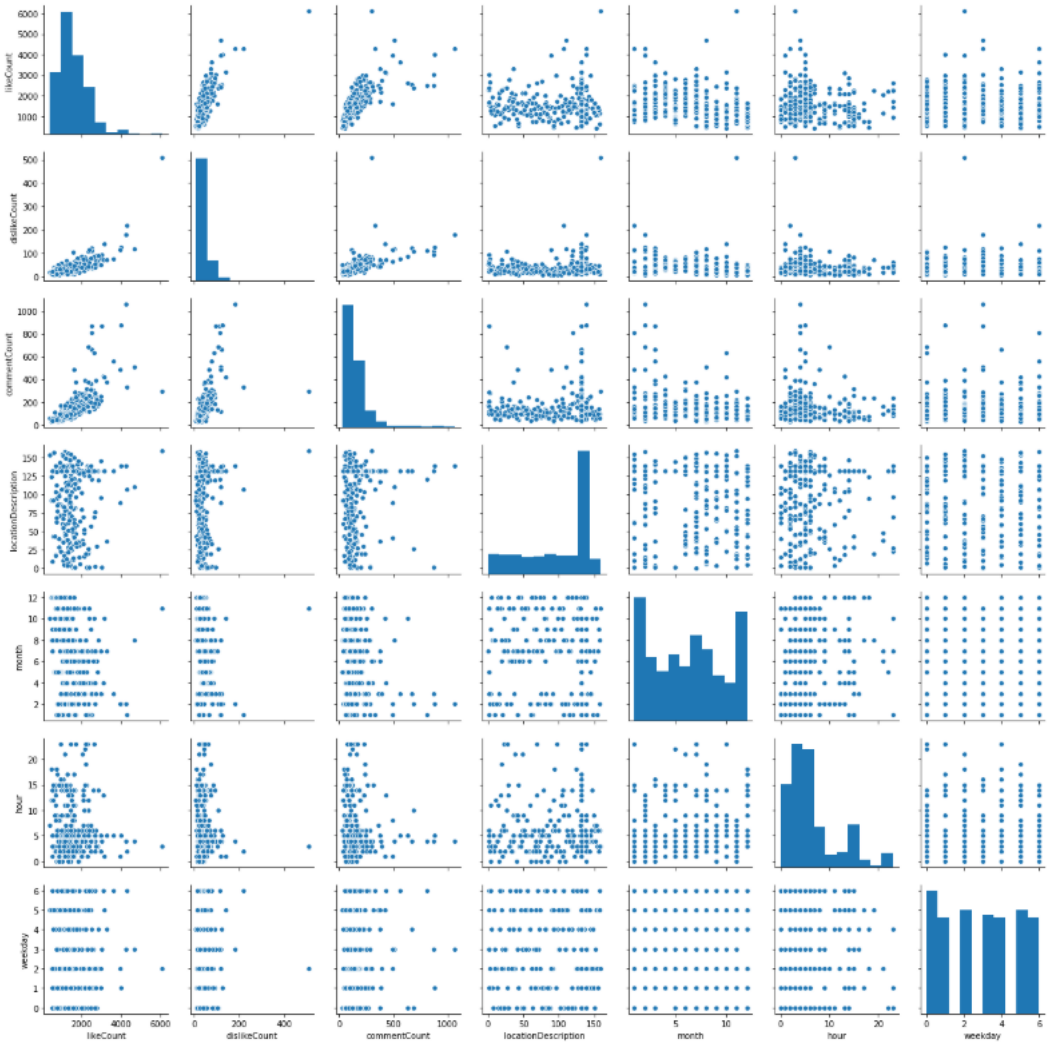


2) Viewer Count Prediction

Minimal Cleaning for the data



Correlation plot



The correlation relations between the numeric features are shown above. From the scatter plots there appears to be high linearity between like, dislikes, and the comment counts. Besides that other feature correlation are not as apparent. The only other meaningful correlation could be seen in the hour feature with like, dislike, and comments. Based on these observations additional features are generated.

Model prediction

Linear regression was used to predict viewer counts. Standard procedure with splitting training and test set, fitting the train set to the model and predicting the results against the test labels. The overall result was decent achieving R2 of 0.908. Lastly the plot between the predicted result and testing label forms a linear form demonstrating a good initial impression.

```
MAE 7941.796437654413
MSE 132890843.09785065
R2 0.9077712747806662
```