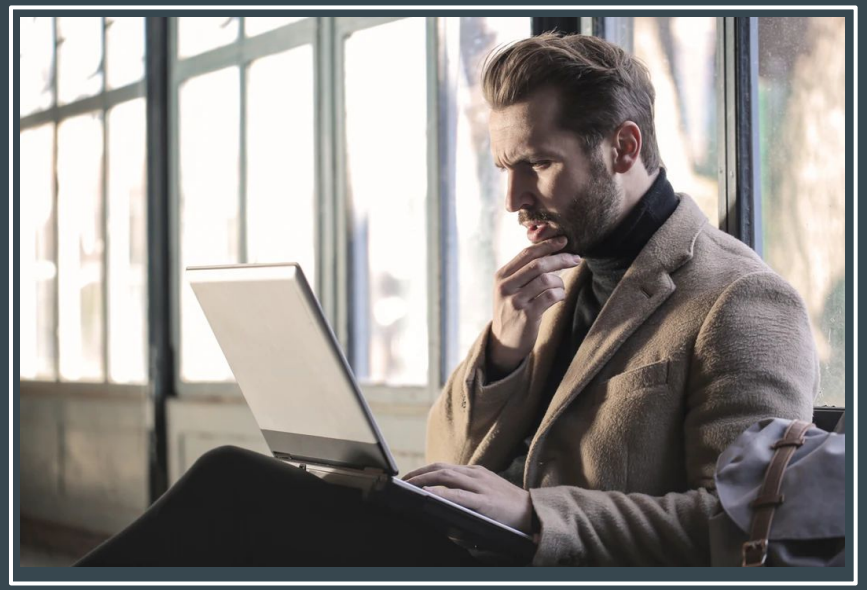
Predicting YouTube Views?

...

Analysis by Oliver Hernandez

What is the goal?

- My goal was to provide valuable insight for YouTubers
- Question: Can we predict views, based on video metrics?



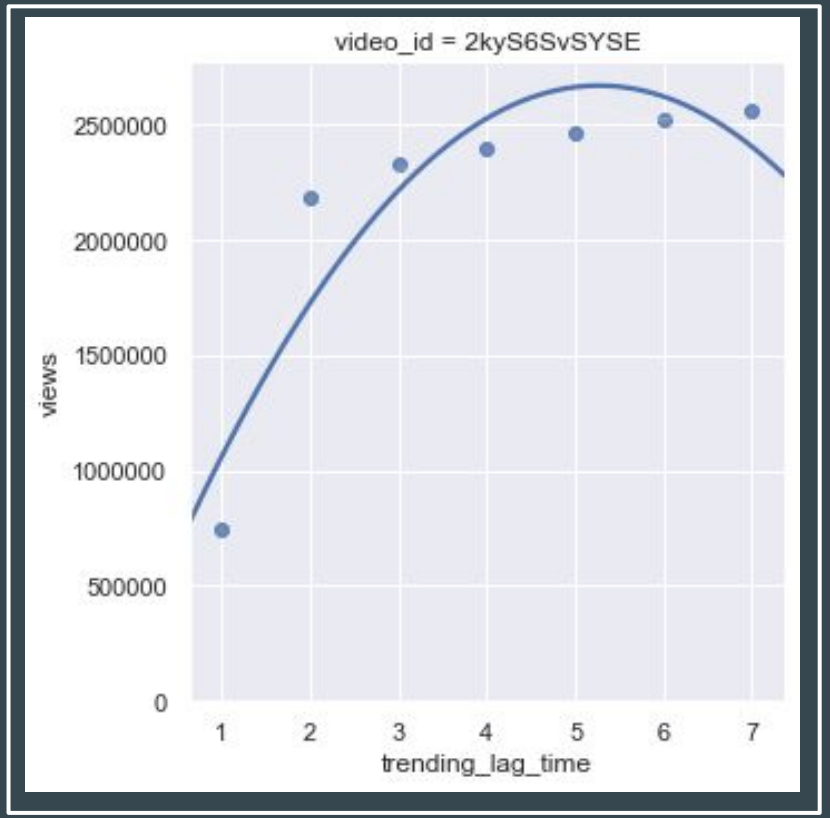
Feature Engineering

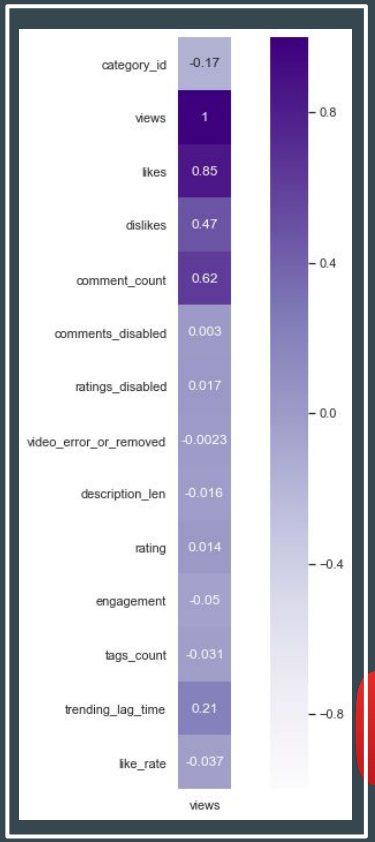
- Engagement rate (number of comment divided by the current amount of views)
- Rating (The amount of likes over the total likes and dislikes)
- Trending lag time (The difference in days a video went from publish to viral)
- Tags count (number of tags for video)
- Like rate (likes per view)

Model Results

	Linear Regression	KNeighborsRegressor	DecisionTree Regressor	BaggingRegressor	RandomForestRegressor	AdaBoostRegressor	SVR	Keras Regressor
R2 Train	0.778436	0.975715	1.000000e+00	0.998399	0.998752	0.071224	0.659420	0.980821
R2 Test	0.786858	0.958741	9.565481e-01	0.992966	0.985983	0.238648	0.572746	0.978653
RMSE Train	0.029869	0.000478	5.460960e-16	0.000002	0.000001	0.432877	0.034909	0.000262
RMSE Test	0.039104	0.001881	3.477376e-03	0.000071	0.000286	0.433385	0.066025	0.000483

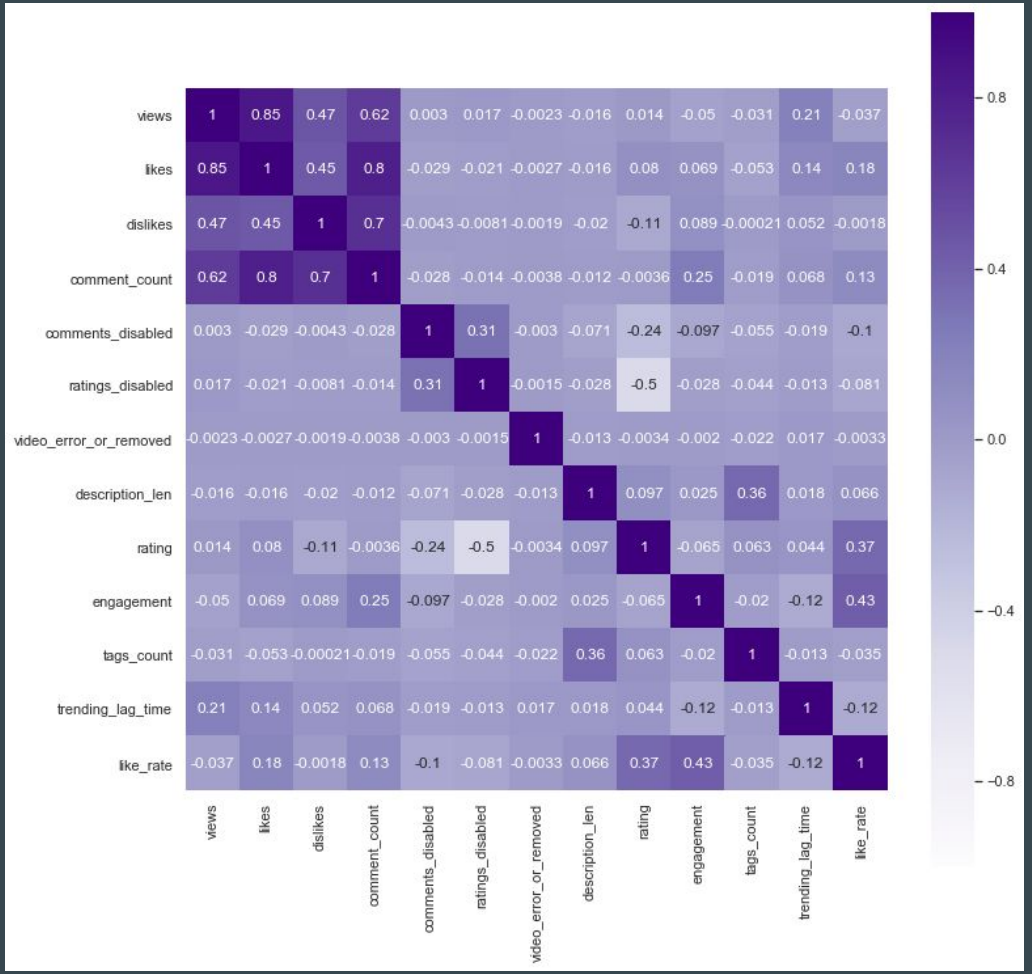
What did we learn?





Correlation

- Likes are correlated with the views
- The views are correlated with the comment
- Comments to dislikes
- Vice Versa



OLS Assumptions

- Linearity
- Independence of error (No endogeneity)
- Normality and homoscedasticity (heteroscedasticity)
- No Autocorrelation
- No Multicollinearity

Violation of Key Assumptions

- Multicollinearity
-
- Indirect Autocorrelation
-
- All the variable a highly correlated with the views
- The variables a rely on views, which is a fundamental flaw and not practical.

Next Steps

- Utilize NLP to identify trends via social media (twitter & reddit)
- Scrape YouTube for the same timeline (look at trending topic videos)
- Create recommendation system to recommend video topics for YouTubers based on the things that are trending.