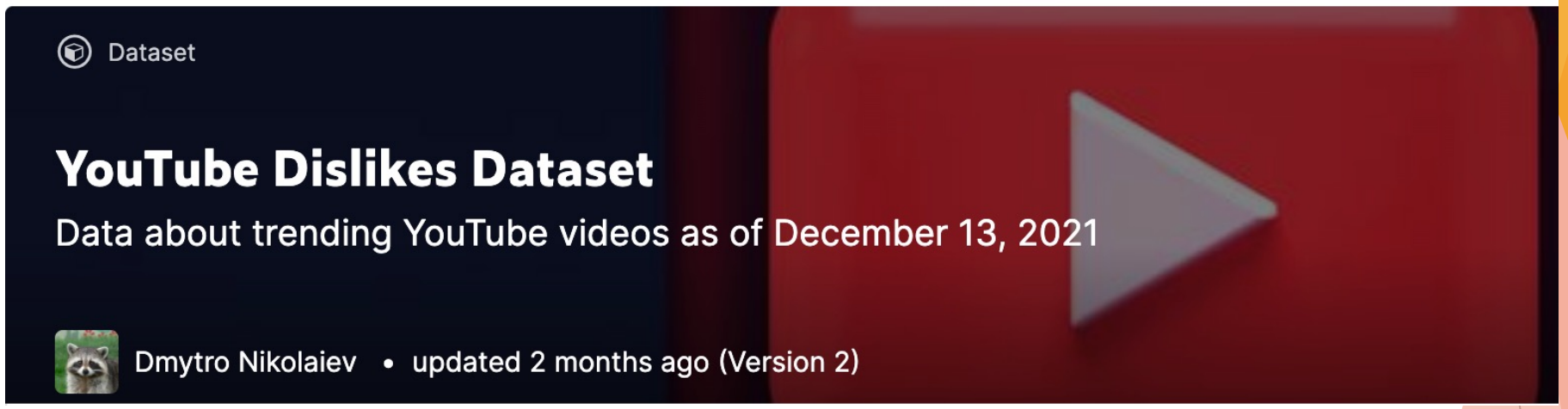
Final Project Presentation

Alison Reikher

Dataset

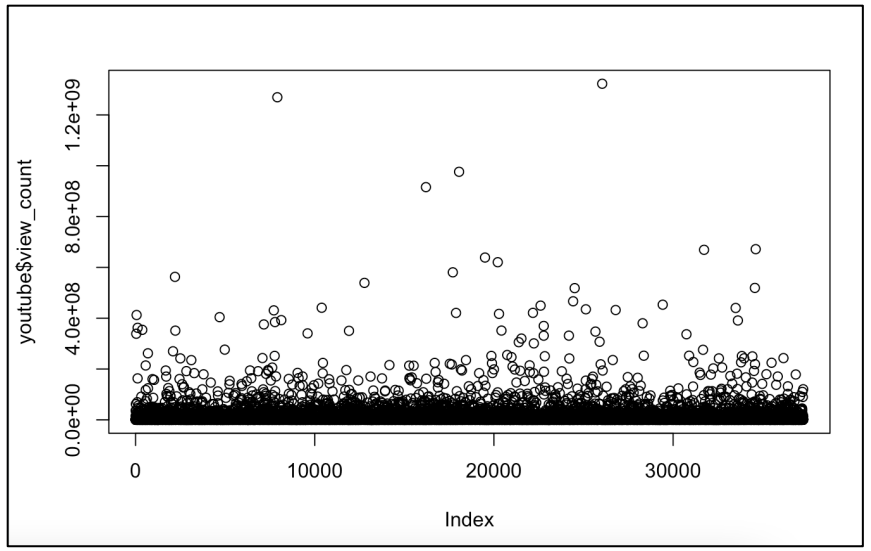


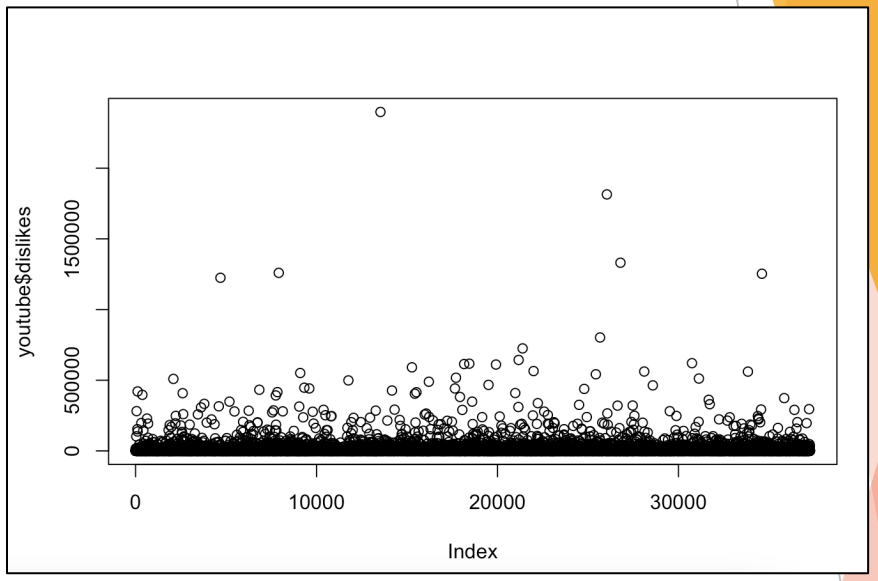
Trending YouTube videos from August 2020 to December 2021 for the USA, Canada, and Great Britain

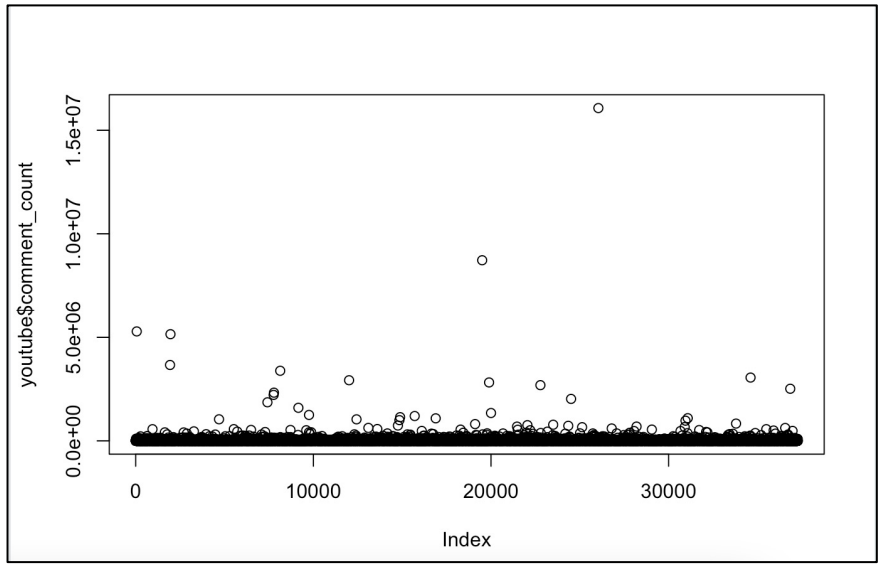
Data Summary

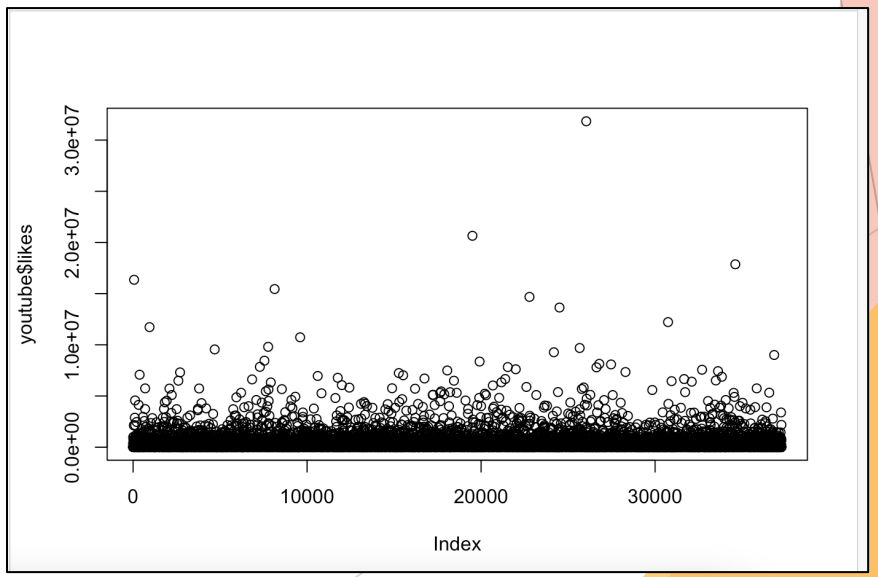
```
> summary(youtube)
    title           channel_title      published_at         view_count            likes          
 Length:37264       Length:37264       Length:37264       Min.   :2.037e+04   Min.   :      22  
 Class :character   Class :character   Class :character   1st Qu.:5.141e+05   1st Qu.:   13317  
 Mode  :character   Mode  :character   Mode  :character   Median :1.322e+06   Median :   42556  
                                                          Mean   :5.711e+06   Mean   :  167287  
                                                          3rd Qu.:3.675e+06   3rd Qu.:  130966  
                                                          Max.   :1.323e+09   Max.   :31837675  
    dislikes          comment_count          tags            description          comments        
 Min.   :      3.0   Min.   :       1   Length:37264       Length:37264       Length:37264       
 1st Qu.:    282.8   1st Qu.:     914   Class :character   Class :character   Class :character   
 Median :    798.0   Median :    2347   Mode  :character   Mode  :character   Mode  :character   
 Mean   :   4996.4   Mean   :    9967                                                            
 3rd Qu.:   2466.2   3rd Qu.:    6212                                                            
 Max.   :2397733.0   Max.   :16071029                                                            
```

Plots

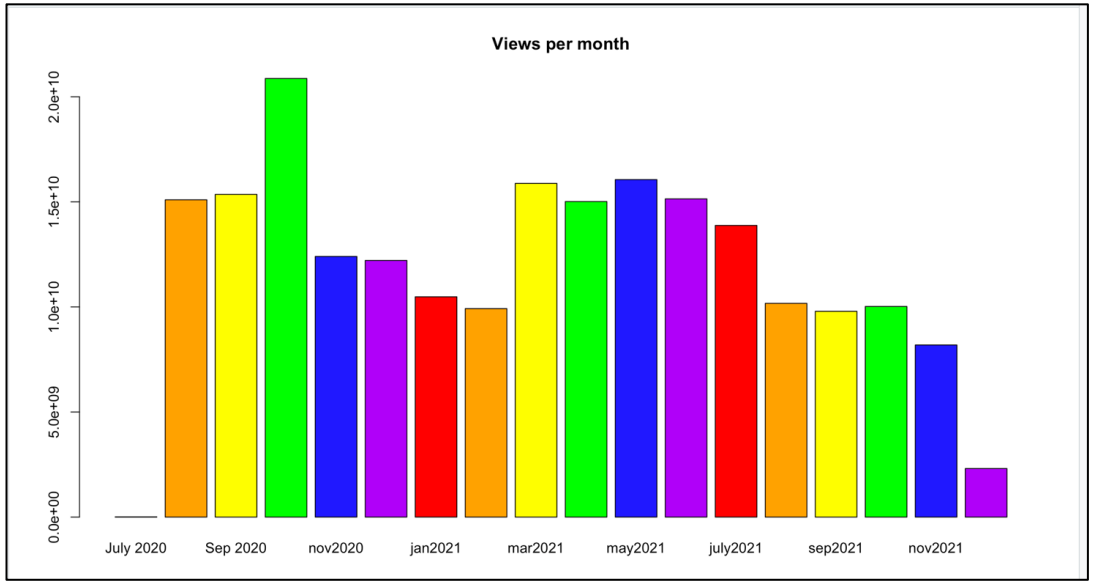

youtube$view_count
0.0e+00
4.0e+08
8.0e+08
1.2e+09
0
10000
20000
30000
Index

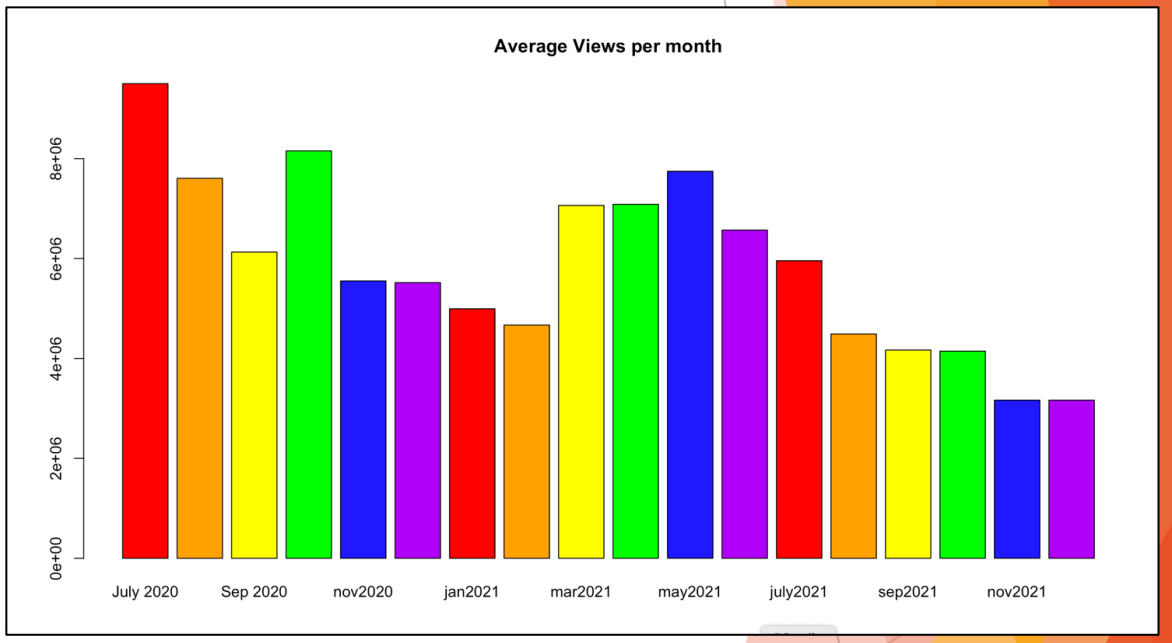

youtube$dislikes
0
500000
1500000
0
10000
20000
30000
Index


youtube$comment_count
0.0e+00
5.0e+06
1.0e+07
1.5e+07
0
10000
20000
30000
Index


youtube$likes
0.0e+00
1.0e+07
2.0e+07
3.0e+07
0
10000
20000
30000
Index

Views Distribution





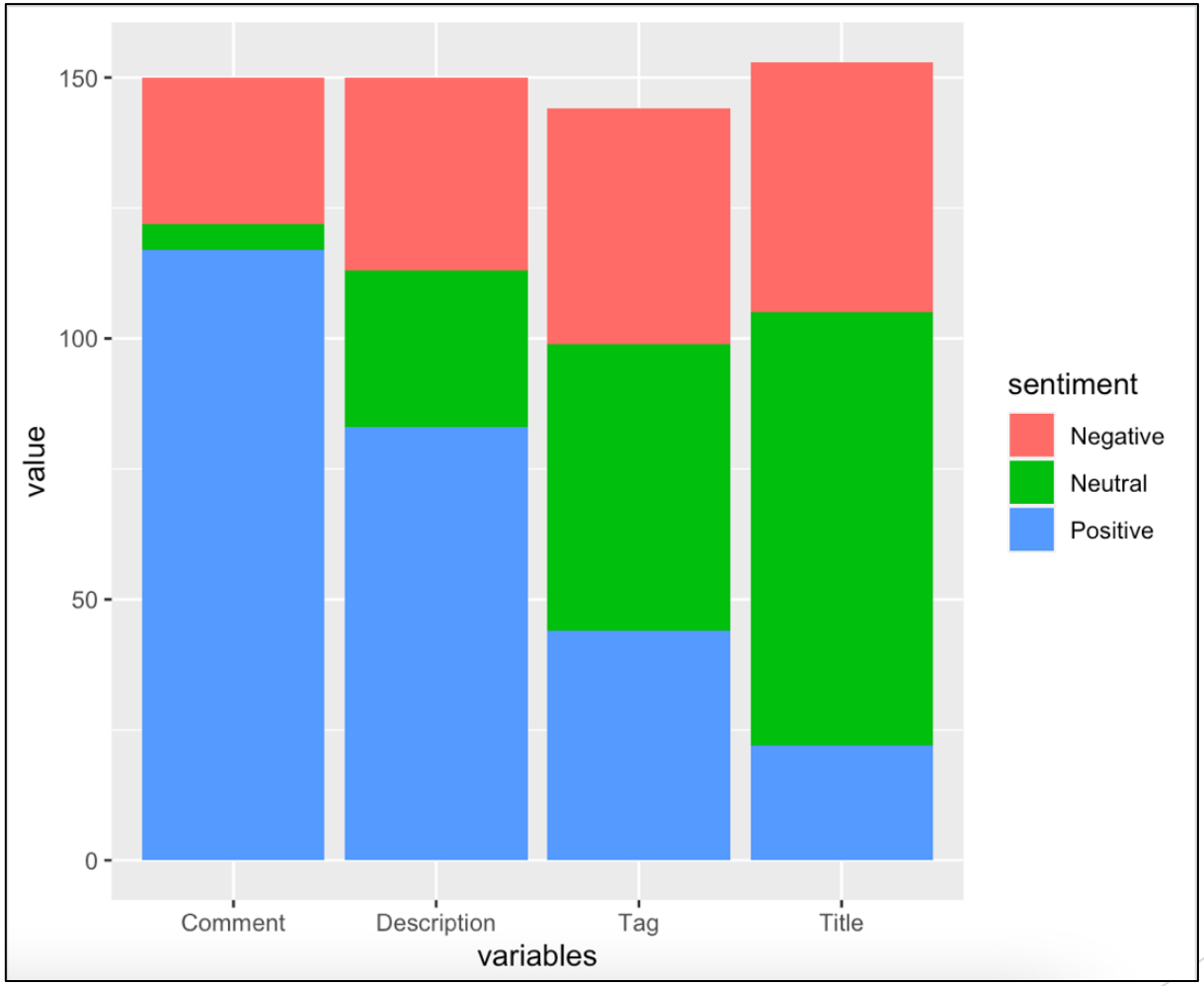
Methods for Analysis

- Associative Rule Mining
- Sentiment Analysis
- Correlation Matrices

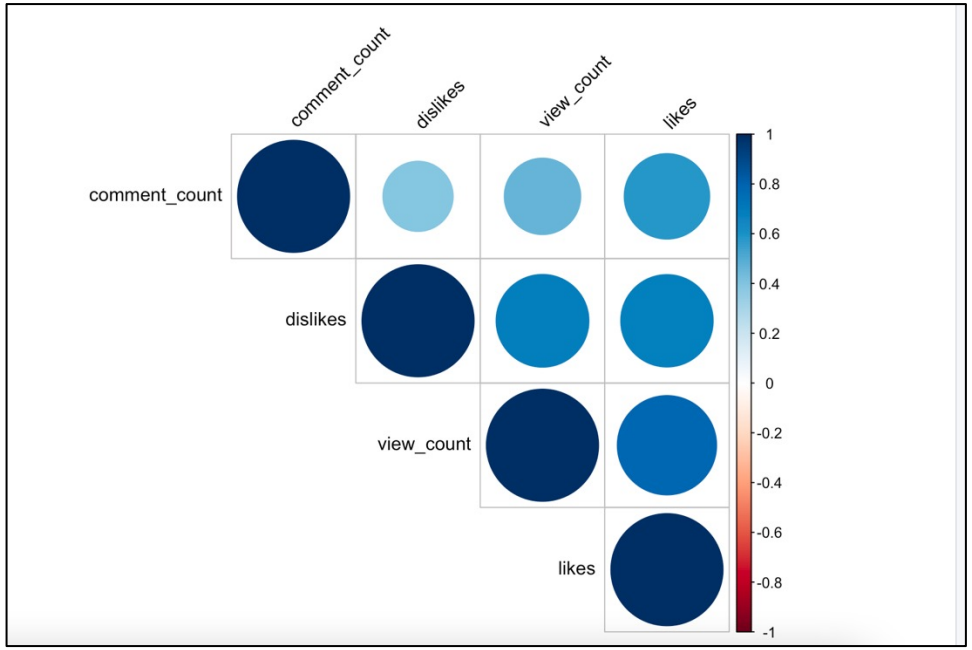
Associative Rule Mining

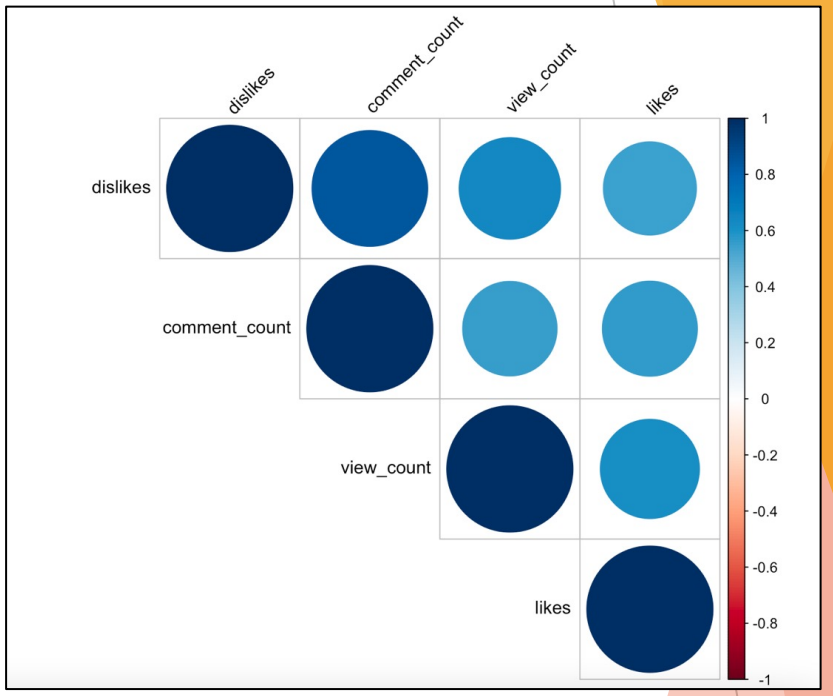
- Rule with the highest confidence, lift, and support:
  - {view_count=[2.55e+06,1.32e+09], dislikes=[1.63e+03,2.4e+06], comment_count=[4.38e+03,1.61e+07]} => {likes=[8.78e+04,3.18e+07]}
    - Support: 0.1917400, Confidence: 0.9379102, Lift: 2.813580
- Most Interesting Rules:
  - {channel_title=Sky Sports Football} => {likes=[22,2e+04)}
    - Support: 0.01199549, Confidence: 0.8386492, Lift: 2.516218
  - {channel_title=Sky Sports Football} => {dislikes=[3,404)}
    - Support: 0.01148562, Confidence: 0.8030019, Lift: 2.412956
- Most informative Rules:
  - {view_count=[2.55e+06,1.32e+09]} => {dislikes=[1.63e+03,2.4e+06]}
    - Support: 0.2671211, Confidence: 0.8013202, Lift: 2.403251
  - {view_count=[2.55e+06,1.32e+09]} => {likes=[8.78e+04,3.18e+07]}
    - Support: 0.2668527, Confidence: 0.8005152, Lift: 2.401417

Sentiment Analysis



Correlation Analysis





Youtube Dataframe

	view_count	likes	dislikes	comment_count
view_count	1.0000000	0.7841987	0.6845758	0.4644415
likes	0.7841987	1.0000000	0.6796676	0.5811171
dislikes	0.6845758	0.6796676	1.0000000	0.3919985
comment_count	0.4644415	0.5811171	0.3919985	1.0000000

Dislikes Dataframe

	view_count	likes	dislikes	comment_count
view_count	1.0000000	0.6172731	0.6465151	0.5599469
likes	0.6172731	1.0000000	0.5437888	0.5667356
dislikes	0.6465151	0.5437888	1.0000000	0.8411579
comment_count	0.5599469	0.5667356	0.8411579	1.0000000