

YOUTUBERS' GUIDE

SAP SVNT TAP DAY PROJECT – Ruxin Shen DS

WHO WANTS TO BE POPULAR !?

What are some of the common characteristics of a popular video?

- Tags
- Descriptions
- Categories
- Channel_titles

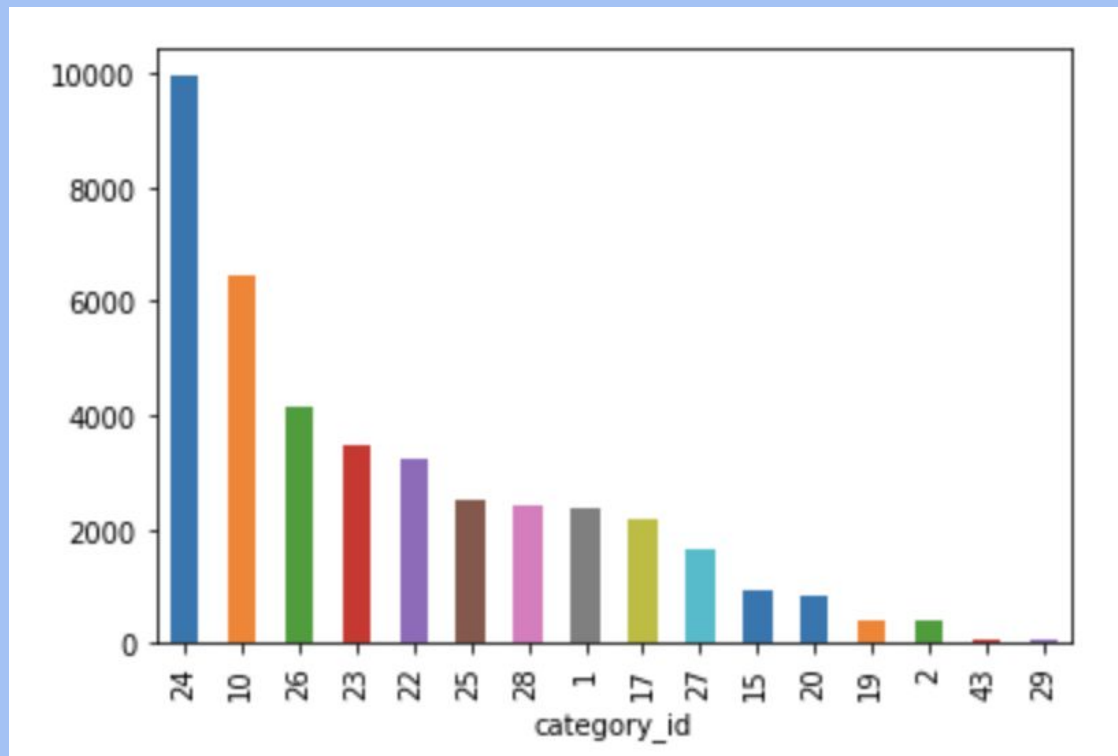
BROAD CHOICES - USER'S DECISION

Popularity measures:

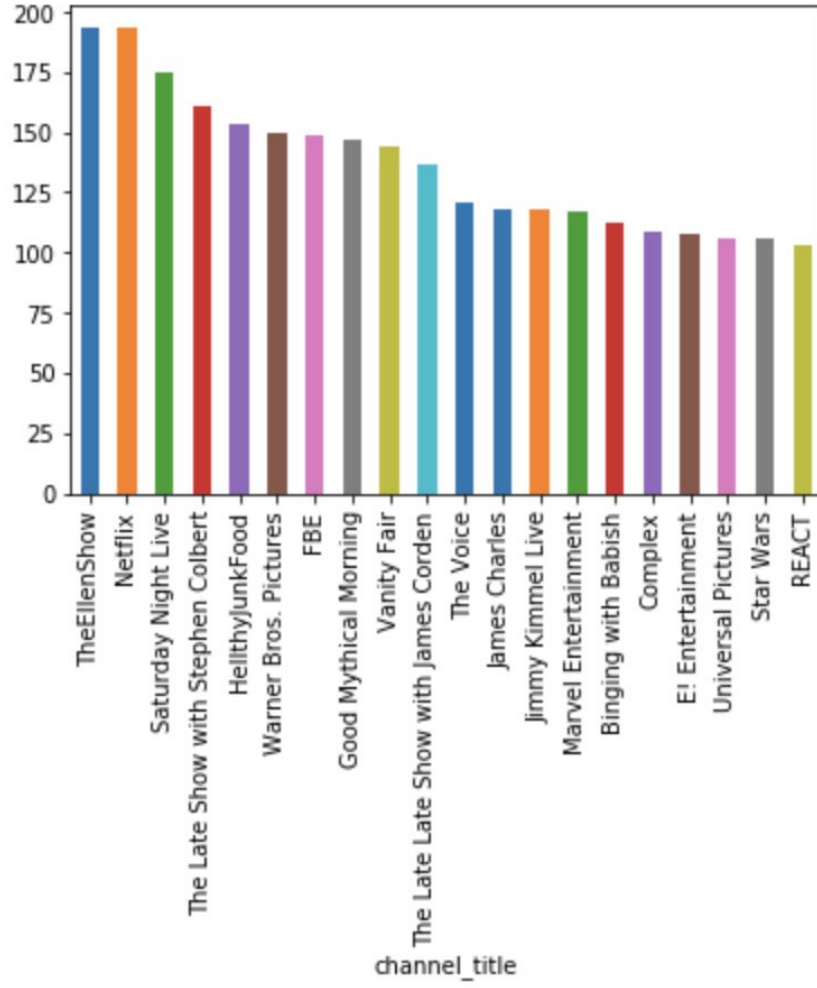
- By category
- By tag
- By trending_date
- By channel

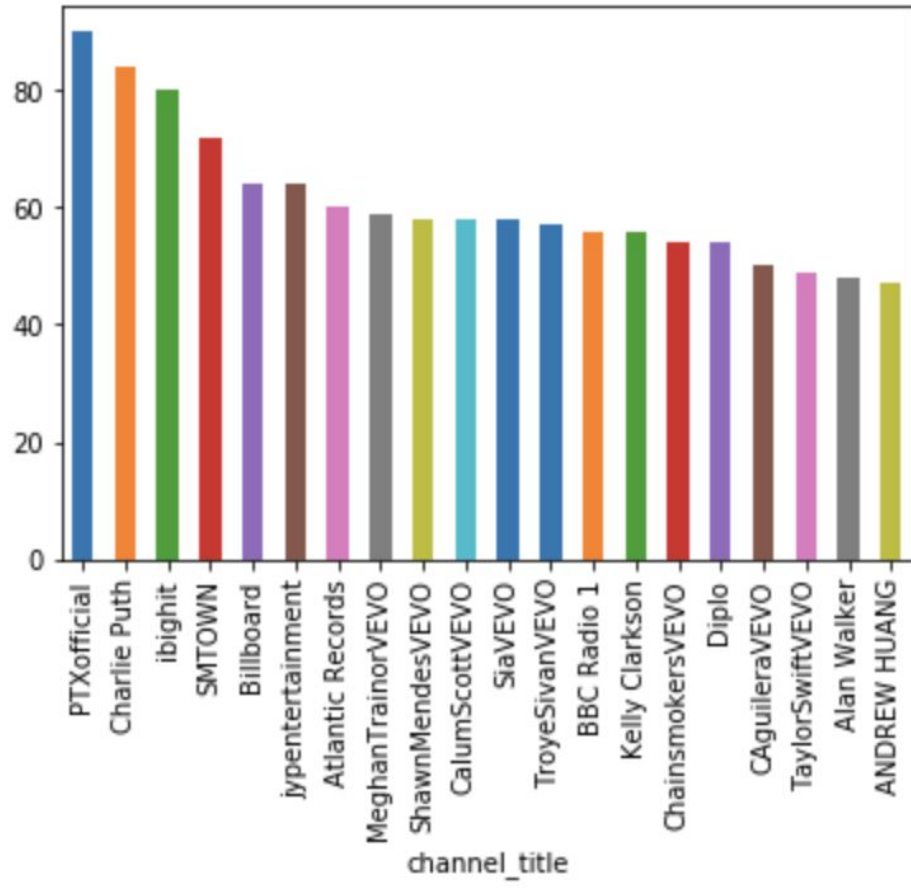
USER'S PERSONAL PREFERENCE SHOULD BE
CONSIDERED!

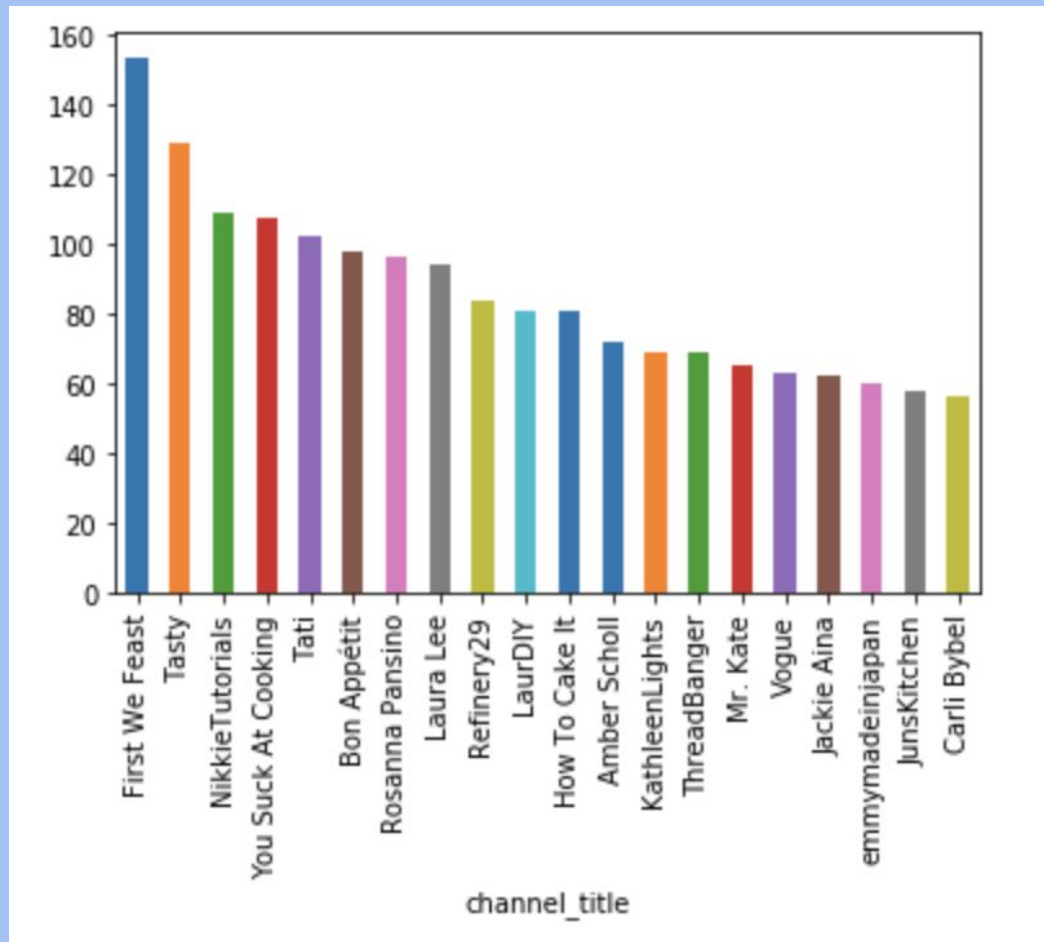
POPULARITY BY CATEGORY



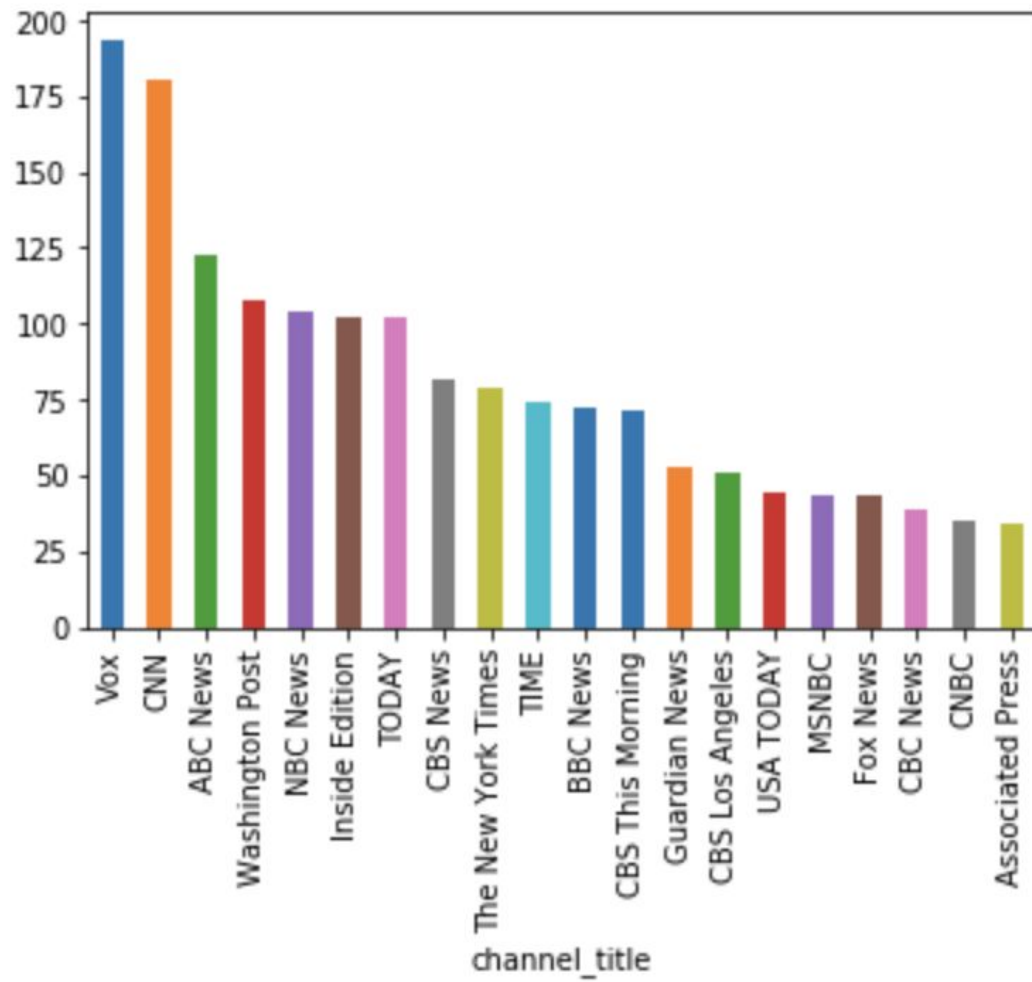
24







25



TAGS

```
((['none'],), 1535),
```

```
((('ABC', 'americanidol', 'idol', 'american idol', 'ryan',  
  'seacrest', 'ryan seacrest', 'katy', 'perry', 'katy perry', 'luke', 'bryan',  
  'luke bryan', 'lionel', 'richie', 'lionel richie', 'season 16', 'american idol XVI',  
  'television', 'ad', 'spring', '2018', 'music', 'reality', 'competition',  
  'song', 'sing', 'audition', 'auditions', 'performance', 'live',  
  'fox', 'AI', 'hollywood', 'contestant', 'official', 'american',  
  'official american idol', 'hollywood week', 'hometown audition'), 87),
```

```
((('Jacksfilms', 'Jack Douglass', 'YGS', 'YGS 100', 'YGS 50',  
  'The Best of Your Grammar Sucks', 'Your Grammar Sucks',  
  'YIAY', 'Yesterday I Asked You', 'Fidget Spinners', 'Emoji Movie',  
  'Kermit Sings', 'JackAsk', 'Jack Ask', 'Dubstep Solves Everything',  
  'Frozen 2', 'iPhone Parody', 'Apple Parody'), 80),
```

```
[('funny', 3603),  
 ('comedy', 2931),  
 ('how to', 1604),  
 (['none'], 1535),  
 ('music', 1302),  
 ('Pop', 1280),  
 ('2018', 1275),  
 ('humor', 1185),  
 ('food', 1159),  
 ('science', 1111),  
 ('review', 1005),  
 ('makeup', 990),  
 ('news', 988),  
 ('celebrity', 930),  
 ('vlog', 928),  
 ('video', 890),  
 ('tutorial', 864),  
 ('live', 862),  
 ('comedian', 861),
```

TAGS' EFFECTS

With 56693 unique tags used 808183 times, I used doc2vec to transform tags to vectors and compared cosine similarity.

Previously I tried to compare jaccard distances of integer indexed tags, but it runs forever;

The user can look for similar videos based on tags, and check out videos of similar tag set but of different popularity and try to figure out why ?

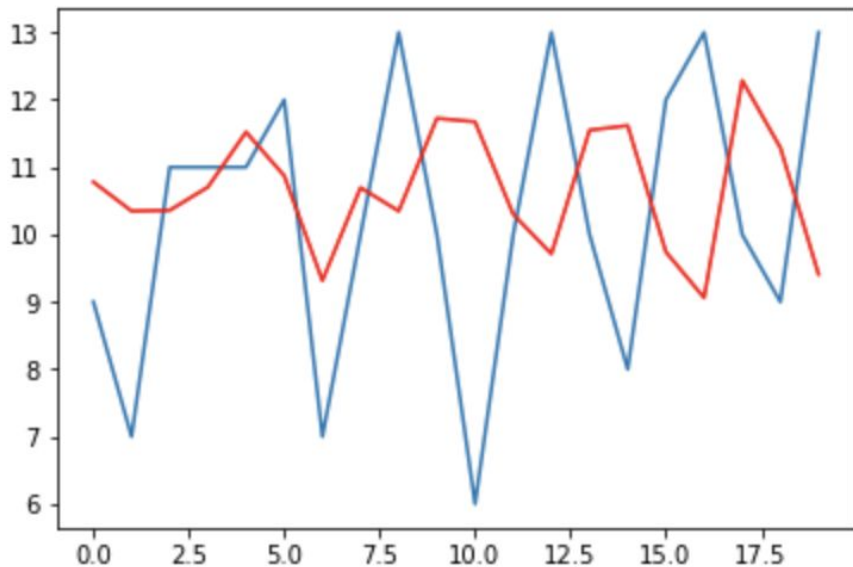
So the user can learn a lesson and know how to make a good video!

TAGS' EFFECT

Furthermore, if we find a proper encoding of tags, when the user wants to put out a new video, he can check out different combination of tags and see the predicted popularity/view.

TIME SERIES FORECAST ON THE POPULARITY OF CATEGORY

Prediction result of category 1



```
predicted=10.779973, expected=9.000000
predicted=10.349011, expected=7.000000
predicted=10.359837, expected=11.000000
predicted=10.704273, expected=11.000000
predicted=11.522250, expected=11.000000
predicted=10.871884, expected=12.000000
predicted=9.316713, expected=7.000000
predicted=10.695818, expected=10.000000
predicted=10.348496, expected=13.000000
predicted=11.725470, expected=10.000000
predicted=11.673552, expected=6.000000
predicted=10.301193, expected=10.000000
predicted=9.712612, expected=13.000000
predicted=11.549576, expected=10.000000
predicted=11.616909, expected=8.000000
predicted=9.742402, expected=12.000000
predicted=9.061501, expected=13.000000
predicted=12.286700, expected=10.000000
predicted=11.287969, expected=9.000000
predicted=9.409167, expected=13.000000
Test MSE: 6.741
```

OUCH! FAILED APPROACH...

Inter-category measure

What categories have
shared users?

user		recom
0	22	[1667, 2121, 610, 1598, 235, 822, 1754, 983, 1...
1	24	[1898, 2121, 651, 1754, 983, 1686, 856, 224, 1...
2	23	[1898, 1667, 610, 1598, 651, 235, 822, 1754, 9...
3	28	[1898, 1667, 2121, 610, 1598, 235, 1754, 983, ...
4	1	[1898, 2121, 610, 1598, 235, 983, 1686, 2075, ...
5	25	[1898, 1667, 2121, 610, 1598, 651, 235, 822, 1...
6	17	[1898, 1667, 2121, 610, 1598, 651, 235, 822, 1...
7	10	[1898, 1667, 610, 1598, 651, 235, 822, 1754, 9...
8	15	[1898, 1667, 2121, 610, 1598, 651, 235, 822, 1...
9	27	[1898, 1667, 2121, 610, 1598, 651, 235, 822, 1...
10	26	[1898, 1667, 2121, 610, 1598, 235, 822, 1754, ...
11	2	[1898, 2121, 610, 822, 1754, 983, 1686, 856, 2...
12	19	[1898, 1667, 2121, 610, 1598, 235, 822, 1754, ...
13	20	[1898, 1667, 2121, 610, 1598, 651, 235, 822, 1...
14	29	[1898, 1667, 2121, 610, 1598, 651, 235, 822, 1...
15	43	[1898, 1667, 2121, 610, 1598, 651, 235, 822, 1...

FUTURE WORK

If I have the data on users, it would be interesting to do a collaborative filtering model so we can target our audience!

What kind of users would your videos attract?

And who would you like to attract?