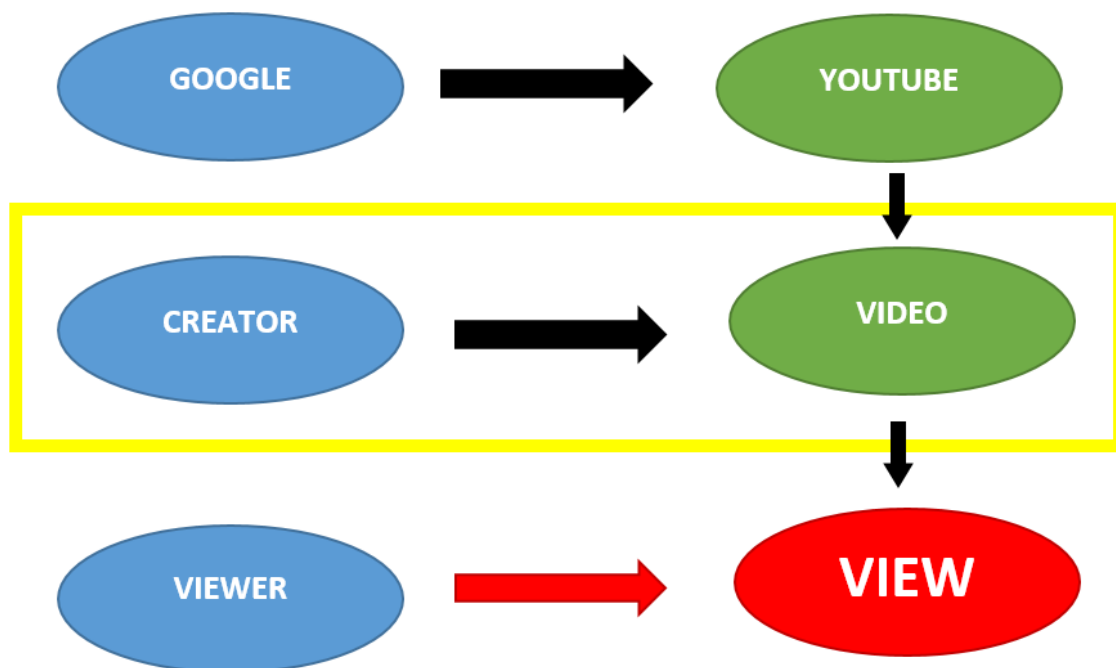


"Trends on Trends" - An Analysis of the Youtube Creator's influence on video trendability.

This project will explore how the Youtube creator affects the path from viewer to view, within the Youtube platform (excluding outside factors).

```
In [73]: from IPython.display import Image  
Image(filename='Youtube Relationship - Diagram.PNG')
```

Out[73]:



QUESTIONS TO EXPLORE

- * Are there any correlations between factors that stem from the Youtube Creator and trendability?
- * For Creators, are there any opportunities and or suggested behavior to encourage trendability?
- * For Youtube, is it worthwhile to spend additional resources on the Youtube Creator? Are there any suggested changes for existing platform?

DATA BACKGROUND & PREVIEW

What data is available?

- "Daily Record of the Top Trending Youtube Videos"
- Up to 200 trending videos per day
- Two dataframes with the same columns (19). 'df' contains one line per video with final numbers while "df_all_timestamps" contains snapshots throughout the day.

Limitations

- No data on the viewers or path to arrive to video.
- Only 1 property of channel to be attributed to Youtube creator. More information on creator could lead to further insights.

Unanswered Questions

- What defines trending?
- What is "country"?

Manipulations

DF

- Added tag_count column (which split the string in the tags column and added the individual elements)
- Replaced existing publish date column with data type "object" with new one with data type "datetime"
- Replaced existing trending date column with data type "object" with new one with data type "datetime"
- Added publish day of the week column
- Added trending day of the week column
- Added Time Live and Days Live columns

OTHERS

- Dataframe channel_count created for Multiple Trenders analysis

```
In [2]: import os
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [3]: filename1 = 'dataframe.csv'
filename2 = 'dataframe_all_timestamps.csv'

datapath1 = os.path.join '..', filename1
datapath2 = os.path.join '..', filename2
```

```
In [4]: df = pd.read_csv(datapath1)
df_all_timestamps = pd.read_csv(datapath2)
```

In [5]: df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 171149 entries, 0 to 171148
Data columns (total 19 columns):
video_id          171149 non-null object
trending_date     171149 non-null object
title             171149 non-null object
channel_title     171149 non-null object
category_id       171149 non-null int64
category          171149 non-null object
publish_date      171149 non-null object
publish_time      171149 non-null object
tags              171149 non-null object
views             171149 non-null int64
likes             171149 non-null int64
dislikes          171149 non-null int64
comment_count     171149 non-null int64
thumbnail_link    171149 non-null object
comments_disabled 171149 non-null bool
ratings_disabled  171149 non-null bool
video_error_or_removed 171149 non-null bool
description        171149 non-null object
country           171149 non-null object
dtypes: bool(3), int64(5), object(11)
memory usage: 21.4+ MB
```

In [6]: df_all_timestamps.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 356350 entries, 0 to 356349
Data columns (total 19 columns):
video_id          356350 non-null object
trending_date     356350 non-null object
title             356350 non-null object
channel_title     356350 non-null object
category_id       356350 non-null int64
category          356350 non-null object
publish_date      356350 non-null object
publish_time      356350 non-null object
tags              356350 non-null object
views             356350 non-null int64
likes             356350 non-null int64
dislikes          356350 non-null int64
comment_count     356350 non-null int64
thumbnail_link    356350 non-null object
comments_disabled 356350 non-null bool
ratings_disabled  356350 non-null bool
video_error_or_removed 356350 non-null bool
description        356350 non-null object
country           356350 non-null object
dtypes: bool(3), int64(5), object(11)
memory usage: 44.5+ MB
```

In [7]: df.head()

Out[7]:

	video_id	trending_date	title	channel_title	category_id	category	publish_date
0	cJon8yBKv3g	2017-11-14	Kingsman : Le Cercle d'or (Mickael J Critique)	MJ - FERMEZ LA	1	Film & Animation	2017-11-12
1	g12KeCPRB4Q	2017-11-14	Vide grenier LIVE - 12 Novembre 2017	Kirby-54	20	Gaming	2017-11-12
2	irIZRH3Jllw	2017-11-14	لقجع بثور في وجه الاتحاد الإيفواري بسبب الجماء...	أخبار اليوم	25	News & Politics	2017-11-12
3	SnHtzkQZggA	2017-11-14	عروس اسطنبول اعلان الحلقة 25 - مترجم للعربية	تركييتنا - bizim Türkiye	1	Film & Animation	2017-11-13
4	u6EoWM2KMPE	2017-11-14	Portokalli, 12 Nentor 2017 - Policat e postbll...	Top Channel Albania	23	Comedy	2017-11-13

In [8]: `df_all_timestamps.head()`

Out[8]:

	video_id	trending_date	title	channel_title	category_id	category	publish_date
0	n1WpP7iowLc	2017-11-14	Eminem - Walk On Water (Audio) ft. Beyoncé	EminemVEVO	10	Music	2017-11-10
1	cJon8yBKv3g	2017-11-14	Kingsman : Le Cercle d'or (Mickael J Critique)	MJ - FERMEZ LA	1	Film & Animation	2017-11-12
2	g12KeCPRB4Q	2017-11-14	Vide grenier LIVE - 12 Novembre 2017	Kirby-54	20	Gaming	2017-11-12
3	irlZRH3Jllw	2017-11-14	لقجع يثور في وجه الإتحاد الإفريقي بسبب الجماء...	أخبار اليوم	25	News & Politics	2017-11-12
4	SnHtzkQZggA	2017-11-14	عروس اسطنبول اعلان الحلقة مترجم للعربية	تركييتنا - bizim Türkiye	1	Film & Animation	2017-11-13

In [9]: *# converting date field*

```
def stringify(row):
    return str(row['publish_date']) + ' ' + str(row['publish_time'])

df_all_timestamps.loc[:, 'string_date_time'] = df_all_timestamps.apply(stringify, axis=1)

df_all_timestamps.loc[:, 'publish_timestamp'] = pd.to_datetime(df_all_timestamps['string_date_time'], infer_datetime_format=True, errors='coerce')

df_all_timestamps = df_all_timestamps.drop('string_date_time', axis=1)

#make a new column in df called 'publish_timestamp'
#that is a datetime object

df.loc[:, 'string_date_time'] = df.apply(stringify, axis=1)
df.loc[:, 'publish_timestamp'] = pd.to_datetime(df['string_date_time'], infer_datetime_format=True, errors='coerce')
df = df.drop('string_date_time', axis=1)
```

```
In [10]: # made new column called trending_timestamp converting trending date object to
datetime

df.loc[:, 'trending_timestamp'] = pd.to_datetime(df['trending_date'], infer_datetime_format=True, errors='coerce')
```

```
In [11]: df = df.drop('trending_date', axis=1)
```

```
In [12]: # makes new columns by taking the logarithm of the views, likes,
#dislikes and comment_count to compress the scale for easier graphing

import numpy as np
df.loc[:, 'log_views'] = np.log(df.loc[:, 'views']+1)
df.loc[:, 'log_likes'] = np.log(df.loc[:, 'likes']+1)
df.loc[:, 'log_dislikes'] = np.log(df.loc[:, 'dislikes']+1)
df.loc[:, 'log_comment_count'] = np.log(df.loc[:, 'comment_count']+1)

#show all columns
pd.set_option('display.max_columns', 100)
```

```
In [13]: df.head()
```

Out[13]:

	video_id	trending_date	title	channel_title	category_id	category	publish_date
0	cJon8yBKv3g	2017-11-14	Kingsman : Le Cercle d'or (Mickael J Critique)	MJ - FERMEZ LA	1	Film & Animation	2017-11-12
1	g12KeCPRB4Q	2017-11-14	Vide grenier LIVE - 12 Novembre 2017	Kirby-54	20	Gaming	2017-11-12
2	iriZRH3Jllw	2017-11-14	لقجع يثور في وجه الإتحاد الإيفواري ببسيب الجماء...	أخبار اليوم	25	News & Politics	2017-11-12
3	SnHtzkQZggA	2017-11-14	عروس اسطنبول اعلان الحلقة 25 - مترجم للعربية	تركييتنا - bizim Türkiye	1	Film & Animation	2017-11-13
4	u6EoWM2KMPE	2017-11-14	Portokalli, 12 Nentor 2017 - Policat e postbll...	Top Channel Albania	23	Comedy	2017-11-13

```
In [14]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 171149 entries, 0 to 171148
Data columns (total 25 columns):
video_id                171149 non-null object
trending_date           171149 non-null object
title                   171149 non-null object
channel_title           171149 non-null object
category_id             171149 non-null int64
category                171149 non-null object
publish_date            171149 non-null object
publish_time            171149 non-null object
tags                    171149 non-null object
views                   171149 non-null int64
likes                   171149 non-null int64
dislikes                171149 non-null int64
comment_count           171149 non-null int64
thumbnail_link          171149 non-null object
comments_disabled       171149 non-null bool
ratings_disabled        171149 non-null bool
video_error_or_removed  171149 non-null bool
description              171149 non-null object
country                 171149 non-null object
publish_timestamp       171149 non-null datetime64[ns]
trending_timestamp      171149 non-null datetime64[ns]
log_views               171149 non-null float64
log_likes               171149 non-null float64
log_dislikes            171149 non-null float64
log_comment_count       171149 non-null float64
dtypes: bool(3), datetime64[ns](2), float64(4), int64(5), object(11)
memory usage: 29.2+ MB
```

```
In [15]: df_all_timestamps.head()
```

```
Out[15]:
```

	video_id	trending_date	title	channel_title	category_id	category	publish_date
0	n1WpP7iowLc	2017-11-14	Eminem - Walk On Water (Audio) ft. Beyoncé	EminemVEVO	10	Music	2017-11-10
1	cJon8yBKv3g	2017-11-14	Kingsman : Le Cercle d'or (Mickael J Critique)	MJ - FERMEZ LA	1	Film & Animation	2017-11-12
2	g12KeCPRB4Q	2017-11-14	Vide grenier LIVE - 12 Novembre 2017	Kirby-54	20	Gaming	2017-11-12
3	irIZRH3Jllw	2017-11-14	لقجع يثور في وجه الإتحاد الإفريقي بسبب الجماء...	أخبار اليوم	25	News & Politics	2017-11-12
4	SnHtzkQZggA	2017-11-14	عروس اسطنبول اعلان الحلقة - مترجم للعربية	تركييتنا - bizim Türkiye	1	Film & Animation	2017-11-13

EFFECT OF CONTENT

The Youtube Creator has a lot of control over the properties of the channel and video. The following properties are manually entered by the creator: Title, Description, Thumbnail, Tags, Category, Comments Disabled, and Ratings Disabled.

This section will explore if the content properties from the creator have any correlation with trendability.

Summary of Findings

- **CATEGORY:** Subjective selection. Although Entertainment is the biggest category, a qualitative review shows that it is used too broadly. How can channels utilize this? Should Youtube do something to prevent this?
- **DESCRIPTION:** Every video has a description.
- **TAGS:** 50% of the videos had 7-25 tags, with the minimum being 1 and the maximum of 152.
- **OPENNESS TO FEEDBACK:** Strong correlation. 98% of videos are open to comments and 98% of videos are open to feedback. (ONLY 0.67% are closed to both).

... CATEGORY ...

- Category is subjectively chosen by the creator.
- Although 'Entertainment' is the most popular category, a closer look shows that it is used broadly instead of better matching categories
- With this in mind, is a better solution to make categories more descriptive for Youtube OR should the creator play around this?
- If 'Entertainment' videos trend better and assuming category is used in suggestion algorithms, the Youtube creator should consider if he/she would prefer to use a wider category that may reach more people but might not fit what the viewer is looking for OR if they prefer to select the category that better matches their video, targets less viewers, is less competitive, and perhaps it is more relevant to what the viewer is looking for.

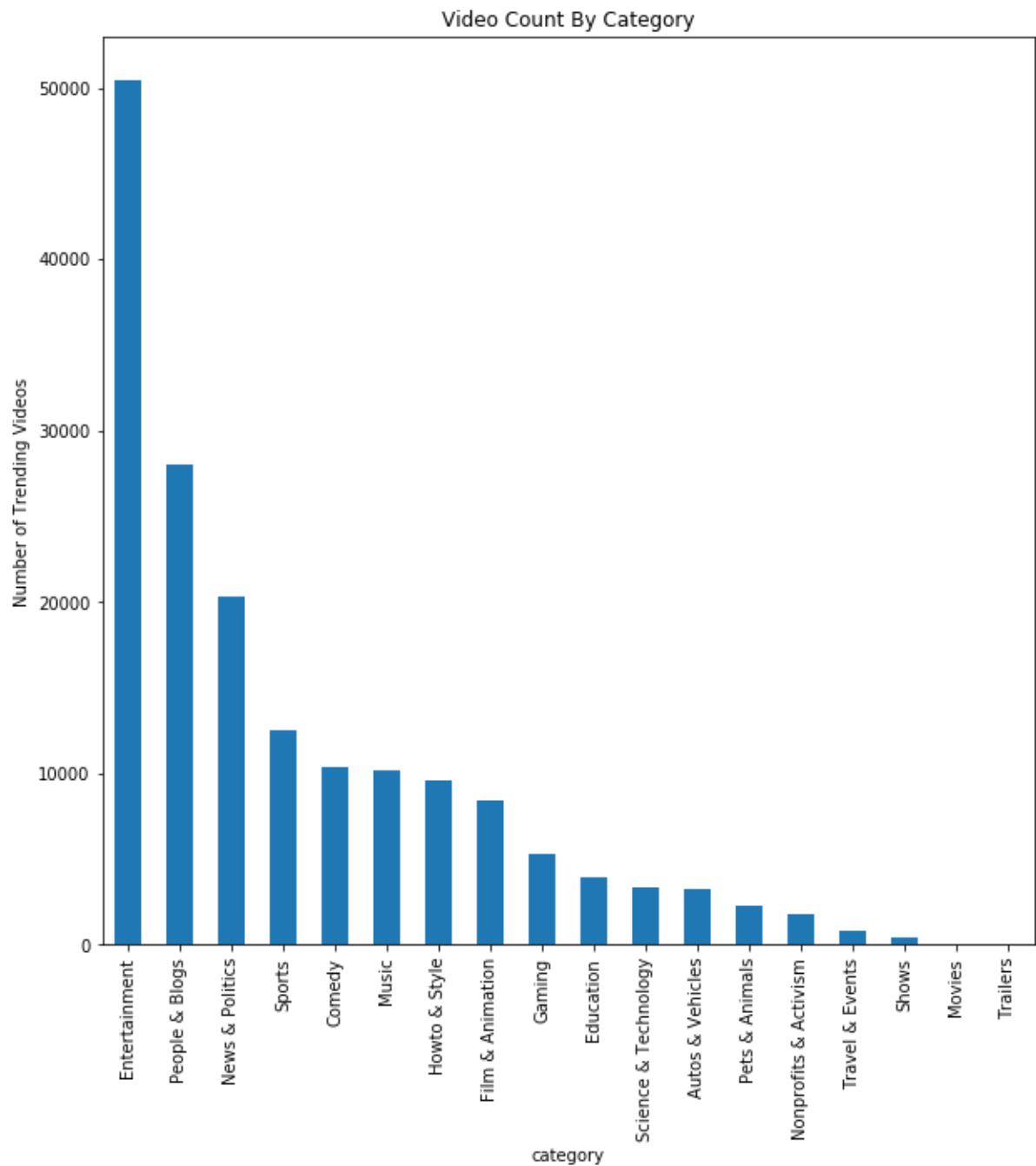
```
In [19]: category_count = df['category'].value_counts()
print(category_count)
```

```
Entertainment      50468
People & Blogs      28045
News & Politics     20327
Sports             12561
Comedy             10358
Music              10145
Howto & Style       9580
Film & Animation    8460
Gaming             5318
Education          3897
Science & Technology 3404
Autos & Vehicles    3274
Pets & Animals      2271
Nonprofits & Activism 1779
Travel & Events     812
Shows              435
Movies              12
Trailers            3
Name: category, dtype: int64
```

```
In [20]: # Bar plot per category

fig = plt.figure(figsize=(10,10))
ax = fig.gca()
category_count.plot(kind = 'bar', ax = ax)
ax.set_title('Video Count By Category') # Give the plot a main title
ax.set_xlabel('category') # Set text for the x axis
ax.set_ylabel('Number of Trending Videos') # Set text for y axis
```

Out[20]: Text(0, 0.5, 'Number of Trending Videos')



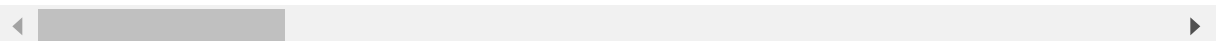
```
In [21]: # Is Entertainment category more popular OR is it used very leniently, based on US?
# Qualitative Assessment
# Checking US top 5 and bottom 5, you can see a movie trailer miscategorized as entertainment (675)
# It also seems to be a catch all of hard-to-categorize videos such as #666 (Shows, People and Blogs, News)

entertainment = df['category'] == 'Entertainment'
US = df['country'] == 'US'
df[entertainment & US]
```

Out[21]:

	video_id	trending_date	title	channel_title	category_id	category
322	zYWt2mnalP8	2017-11-14	How To Do Thanksgiving Makeup That Has Nothing...	Sailor J	24	Entertainment
666	cOc3tsFWoRs	2017-11-14	Jason Momoa & Lisa Bonet: Love at First Sight	The Late Late Show with James Corden	24	Entertainment
672	F-j_6luaYfw	2017-11-14	I Miss the Old Taylor	Wong Fu Productions	24	Entertainment
674	Y6eKxjMA9ek	2017-11-14	Dropping And CATCHING A Paper Airplane At 2,00...	Tucker Gott	24	Entertainment
675	zy0b9e40tK8	2017-11-14	Dark Official Trailer [HD] Netflix	Netflix	24	Entertainment
...
169876	vRWH9GQ2J90	2018-06-14	100 LAYERS OF DEEP FRIED PIZZA CHALLENGE!!!	HelthyJunkFood	24	Entertainment
169878	n-DTjpde9-0	2018-06-14	Spinning a Record to Pieces at 12,500fps - The...	The Slow Mo Guys	24	Entertainment
169879	UTVpWVO4vZU	2018-06-14	SIDEMEN FC VS YOUTUBE ALLSTARS 2018 (Goals & H...	Sidemen	24	Entertainment
169885	t_bnyOOFDNM	2018-06-14	COLLEGE KIDS REACT TO ROSEANNE CANCELED?! (Twi...	FBE	24	Entertainment
169886	a30K69hUJyo	2018-06-14	Jurassic World Meets Parkour in Real Life	devinsupertramp	24	Entertainment

1319 rows × 25 columns



... DESCRIPTION ...

No insight as all videos have a description

```
In [22]: no_description = df['description'] == None
df[no_description]
```

Out[22]:

video_id	trending_date	title	channel_title	category_id	category	publish_date	publish_time
							

... TAGS ...

- Tag count does not directly correlate with views but we can see patterns on how many tags 'trending' views tend to have.
- Assuming tags factor into Youtube search algorithms, we can make recommendations on tag_count.
- The average tag count of trending videos is 17 and it should be kept within 7 and 25 (the range followed by 50% of trending videos). Every trending video has at least 1 so it should be not left blank.

```
In [23]: # every video has tags

no_tags = df['tags'] == None
df[no_tags]
```

Out[23]:

video_id	trending_date	title	channel_title	category_id	category	publish_date	publish_time
							

```
In [24]: # make new column called tag_count

df['tag_count'] = None
```

```
In [25]: # put data into the column

df['tag_count'] = df['tags'].str.split(r'|')
df.loc[:, 'tag_count'] = df['tag_count'].apply(lambda x: len(x))
```

```
In [26]: df.head()
```

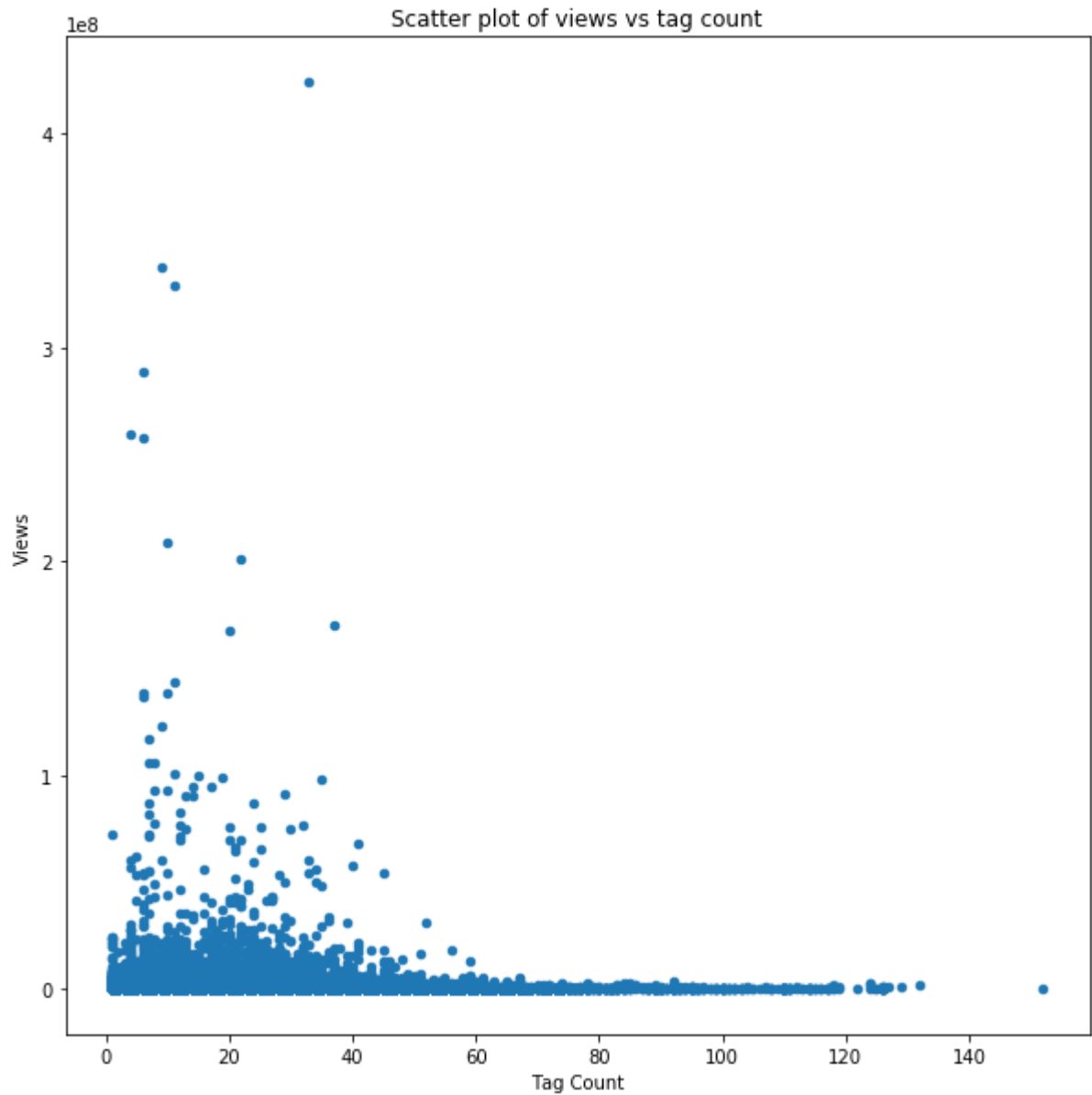
Out[26]:

	video_id	trending_date	title	channel_title	category_id	category	publish_date
0	cJon8yBKv3g	2017-11-14	Kingsman : Le Cercle d'or (Mickael J Critique)	MJ - FERMEZ LA	1	Film & Animation	2017-11-12
1	g12KeCPRB4Q	2017-11-14	Vide grenier LIVE - 12 Novembre 2017	Kirby-54	20	Gaming	2017-11-12
2	irIZRH3Jllw	2017-11-14	لقجع بثور في وجه الاتحاد الإفواري بسبب الجماء...	أخبار اليوم	25	News & Politics	2017-11-12
3	SnHtzkQZggA	2017-11-14	عروس اسطنبول اعلان الحلقة 25 مترجم للعربية	تركييتنا - bizim Türkiye	1	Film & Animation	2017-11-13
4	u6EoWM2KMPE	2017-11-14	Portokalli, 12 Nentor 2017 - Policat e postbll...	Top Channel Albania	23	Comedy	2017-11-13

In [27]: *# Tag count vs views*

```
fig = plt.figure(figsize=(10, 10))
ax = fig.gca()
df.plot(kind = 'scatter', x = 'tag_count', y = 'views', ax = ax)
ax.set_title('Scatter plot of views vs tag count')
ax.set_xlabel('Tag Count')
ax.set_ylabel('Views')
```

Out[27]: Text(0, 0.5, 'Views')



```
In [28]: df['tag_count'].describe()
```

```
Out[28]: count      171149.000000
mean         17.148222
std          13.284787
min           1.000000
25%           7.000000
50%          15.000000
75%          25.000000
max          152.000000
Name: tag_count, dtype: float64
```

... OPENNESS TO FEEDBACK (COMMENTS & RATINGS) ...

- The grand majority of videos are open to feedback with 98% open to comments and 98% open to ratings
- Less than 1 percent (0.67%) are closed to both ratings and comments.
- For Youtube creators: it is important to note that trending videos overwhelming tend to be open to feedback and if trendability/popularity is the goat, their videos should follow this pattern.

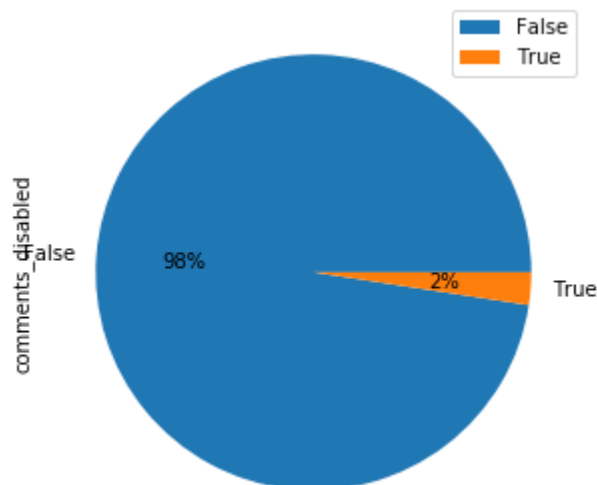
```
In [29]: # open to comments? - count
```

```
comments_disabled_count = df['comments_disabled'].value_counts()
print(comments_disabled_count)
```

```
False    167027
True       4122
Name: comments_disabled, dtype: int64
```

```
In [30]: # open to comments? - proportion
```

```
df2 = pd.DataFrame(comments_disabled_count)
plot = df2.plot.pie(y='comments_disabled', figsize=(5, 5), autopct='%1.0f%%')
```



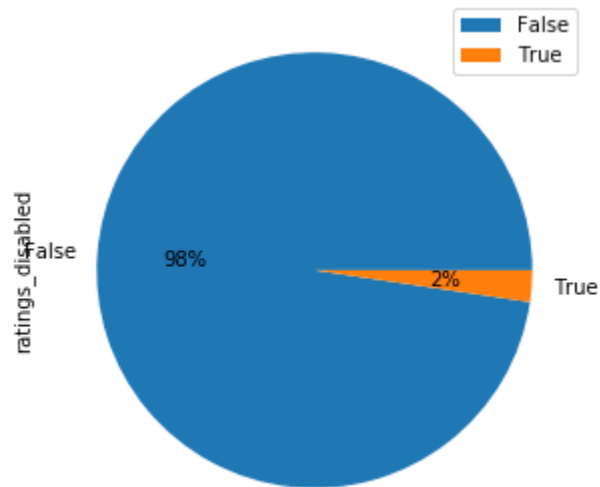

```
In [31]: # open to ratings? - count

ratings_disabled_count = df['ratings_disabled'].value_counts()
print(ratings_disabled_count)
```

```
False    167159
True       3990
Name: ratings_disabled, dtype: int64
```

```
In [32]: # open to ratings? - proportion

df3 = pd.DataFrame(ratings_disabled_count)
plot = df3.plot.pie(y='ratings_disabled', figsize=(5, 5), autopct='%1.0f%%')
```



```
In [33]: # What is the overlap? Closed to both comments are ratings.

df4 = df[(df.ratings_disabled == True) & (df.comments_disabled == True)]

len(df4)
```

Out[33]: 1153

```
In [34]: # closed videos / total videos
# as a percentage

(len(df4)/len(df))*100
```

Out[34]: 0.6736819963891113

EFFECT OF TIME

Some video properties stem from the Youtube Creator's actions: particularly publish time and frequency of uploads. This section will explore possibly opportunities related to the timeline of video uploads.

Summary of Findings

PUBLISH DATE:

- Creator's slightly prefer to publish on weekdays over weekends with Friday being the most popular day.
- The overwhelming majority of trending videos achieve trending within the day (56%), with 85% within 3 days.
- Considering the strong correlation, we can deduce that videos that will trend will most likely to do within days if not hours from the date published.
- In this perhaps an indication of viewer's attention span?

MULTIPLE TRENDERS:

- Aggregating the number of videos by channel, we can see that almost 50% (48%) of the channels had more than 1 video trending at a time.
- This either provides insight on existing algorithm or shows that a video trending builds natural moment for the channel.
- It also provides an opportunity for the creator to achieve trendability with more videos

... PUBLISH DATE ...

- Publish date is generated when the Creator uploads the video and hence in his control.
- The day of the week is pretty evenly split so the difference between days is not drastic. However... when placed in video count order, there seems to be a preference and/or strategic decision to publish on weekdays (Saturday and Sunday had the least amount)
- Friday is the most popular day to publish, perhaps getting ahead of the weekend?

```
In [35]: df['publish_day'] = None
```

```
In [36]: df['publish_day'] = df['publish_timestamp'].dt.dayofweek
```

In [37]: `df.head(50)`

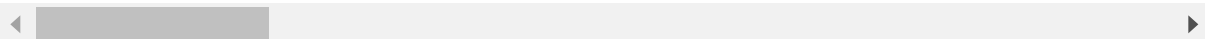
Out[37]:

	video_id	trending_date	title	channel_title	category_id	category	rank
0	cJon8yBKv3g	2017-11-14	Kingsman : Le Cercle d'or (Mickael J Critique)	MJ - FERMEZ LA	1	Film & Animation	
1	g12KeCPRB4Q	2017-11-14	Vide grenier LIVE - 12 Novembre 2017	Kirby-54	20	Gaming	
2	irIZRH3Jllw	2017-11-14	لقّع بثور في وجه الإتحاد الإفريقي بسبب الجماء...	أخبار اليوم	25	News & Politics	
3	SnHtzkQZggA	2017-11-14	عروس اسطنبول اعلان الحلقة 25 - مترجم للعربية	تركييتنا - bizim Türkiye	1	Film & Animation	
4	u6EoWM2KMPE	2017-11-14	Portokalli, 12 Nentor 2017 - Policat e postbll...	Top Channel Albania	23	Comedy	
5	k0OruM7tneA	2017-11-14	الإعلام المصري : شكون ... فالمنتخابات العربية لي	AR-EN-FR LANGUAGES	17	Sports	
6	z4PWAu9HfxM	2017-11-14	No hablo (Vidéoclip)	Thomas Gauthier	23	Comedy	
7	4HeOSihd32A	2017-11-14	LES GAMERS [Version intégrale] - Les Reportage...	ERTV Officiel	25	News & Politics	
8	_wChN4yabJM	2017-11-14	DALS S08 - Lenni-Kim danse un Tango avec Denit...	Danse avec les stars	24	Entertainment	
9	3KvSVqN43t0	2017-11-14	VOL, BAGARRE ET CRACHAT - Draw My Life	Jeel	20	Gaming	
10	hWtTJ3Q37rY	2017-11-14	MAGIE - IL FAIT APPARAÎTRE UN AVION DEVANT 230...	Cyril Ayrau	24	Entertainment	
11	OMgw9gPX3xk	2017-11-14	MA PLAYLIST ! Dansez avec nous 2 - Pembe Cherole	Pembe Cherole	22	People & Blogs	
12	PqDTqJMMSrY	2017-11-14	Samsung Galaxy Teaser	Concept Creator	28	Science & Technology	
13	RKgk0DXZCvQ	2017-11-14	Réalité 3 Fatou Barcelon Herman Belvie Coquett...	esepelisa	1	Film & Animation	
14	O7JJllvepts	2017-11-14	Michel Sardou : Les révélations inédit de Loua...	koi de neuf? Buzz	24	Entertainment	
15	J_I00wRkYFY	2017-11-14	VLOG MAROC 2-0 COTE D'IVOIRE (RUSSIA 2018) - بف...	Taha Essou	22	People & Blogs	

	video_id	trending_date	title	channel_title	category_id	category	
16	1hNAMEGnDSg	2017-11-14	MOTO VLOG#40 PRÉSENTATION DU S1000R	BIBIX	2	Autos & Vehicles	
17	mAbw28ePp5c	2017-11-14	LES 5 MEILLEURS MAGICIENS DU MONDE PASSÉ SUR I...	HEROZ	24	Entertainment	
18	xBwL10BMXeI	2017-11-14	Kurt Volker et la stratégie américaine en UKra...	Stratpol	25	News & Politics	
19	4VVSNyAGYeg	2017-11-14	maroc vs cote d'ivoire 2-0 Résumé du match cot...	kora goal top	17	Sports	
20	O90_g03eo2I	2017-11-14	L'Antidote - Dopage au quotidien : une béquill...	canal9valais	25	News & Politics	
21	8bSyJzEC1vo	2017-11-14	Débriefing F1 - GP du Brésil 2017	Motorsport tv France	2	Autos & Vehicles	
22	ondy6jke_SQ	2017-11-14	LES DILEMMES EN LIVE ! (REDIFFUSION)	Les dilemmes	24	Entertainment	
23	HjEYwXJh-Og	2017-11-14	Dans la Tête d'un Pro : Gaëlle Baumann aux WSO...	Winamax	17	Sports	
24	7CdZE2zKQnk	2017-11-14	Carla Bruni- Sarkozy en larmes face à Delahouss...	Ça Zap !	24	Entertainment	
25	icKOIAc4rAA	2017-11-14	Le client est roi... vraiment ? - Tout Compte Fa...	Tout Compte Fait	24	Entertainment	
26	VG1bq1mD5bk	2017-11-14	Iphone X : une addiction comme une autre ?	Le Fil d'Actu - Officiel	24	Entertainment	
27	k_eAklnTFn8	2017-11-14	FAIRE UNE MUSIQUE EN 10 MINUTES ! Feat Medusa	BIRDYY	24	Entertainment	
28	NKxUBmxLdRw	2017-11-14	ჩემი ცოლის დაქალები 12 სეზონი 24 სერიი / Chemi...	My PAGE	1	Film & Animation	
29	O9g-hqEnfUo	2017-11-14	FUERTE TEMBLOR EN IRAK HOY 12 DE NOVIEMBRE 21...	Noticias Mundiales 2017	22	People & Blogs	

	video_id	trending_date	title	channel_title	category_id	category	
30	0pNCxK0HdI8	2017-11-14	DECORACIONES PARA NAVIDAD - IDEAS PARA DECORAR...	Silvia en tu Vida	1	Film & Animation	
31	Dx7S34INt6A	2017-11-14	PARA TI 🎁...con todo el amor de mi corazón Ábrelo	Frases de Amor	22	People & Blogs	
32	QSwIPz3ruYA	2017-11-14	Caballo Sereno despide emotivamente a su dueño...	top virar	22	People & Blogs	
33	Nfd31XfFZ94	2017-11-14	GASTE 10,000 DOLARES EN MI NUEVO MUSIC VIDEO	Aaron Echeverria	22	People & Blogs	
34	jRI689eyaWg	2017-11-14	AYÚDAME SEÑOR #Reflexión	lili1623	22	People & Blogs	
35	Zw5jzd4BvBY	2017-11-14	Un día muy movido ツ 🚗	pepenaker	15	Pets & Animals	
36	PpvzVcnICJo	2017-11-14	Miraculous Ladybug La Befana Temporada 2 C...	Alfrely	1	Film & Animation	
37	Ttqo0pFoAAM	2017-11-14	Rezo de Laudes y Misa Coral del Cabildo, 13 de...	Basílica de Guadalupe A.R	27	Education	
38	RtPf3dHF8uM	2017-11-14	La ex Timbiriche Alexa Lozano se siente diva y...	Espectáculo ShowTV	22	People & Blogs	
39	xQW3-CeUV2c	2017-11-14	Decidieron sacrificar a este perro con cara to...	STR Noticias	15	Pets & Animals	
40	dTuYRCf-G1E	2017-11-14	Las Noticias de la mañana, lunes 13 de noviemb...	Noticias Telemundo	25	News & Politics	
41	MGeSKdq1V20	2017-11-14	Iraqi Kurdistan Earthquake live during an inte...	Ahmed Nasradin	22	People & Blogs	
42	Mwsy6T99j04	2017-11-14	Chucky Lozano Puede Ser Estrella Mundial, José...	Estadio Videos Play	17	Sports	
43	8rHbN5F_40I	2017-11-14	RÉVÉLATIONS : CHAUFFARDS SOUS HAUTE SURVEILLAN...	FR REPLAY	2	Autos & Vehicles	
44	49c5C9NCrLc	2017-11-14	LE PETIT YAYA 8 🤖 Avec la participation de Q....	Thegrims TV	23	Comedy	

	video_id	trending_date	title	channel_title	category_id	category
45	rNLTO9ezW4U	2017-11-14	On vous dit tout sur l'album Trône de Booba !	Booska-p.com	24	Entertainment
46	Ti_hmTpucml	2017-11-14	CE DJ AVEUGLE VA VOUS SURPRENDRE !!!	Felckin	10	Music
47	5dCjl2efnyQ	2017-11-14	Episodio #1257 ¿Como el hongo candida ataca el...	MetabolismoTV	27	Education
48	ggzueeHh7Tk	2017-11-14	L'ombre d'un doute - Val de Loire des châteaux...	L'ombre d'un doute	1	Film & Animation
49	Nr3AuidaiSo	2017-11-14	TODO sobre el MISTERIOSO caso de JOHN F. KENNE...	Paulettee	24	Entertainment



In [38]: `publish_day_count = df['publish_day'].value_counts()
print(publish_day_count)`

```
4    26978
3    25624
0    24920
1    23992
2    23868
6    23233
5    22534
Name: publish_day, dtype: int64
```

```
In [39]: def dayofweek(row):
    if row['publish_day'] == 0:
        return 'Monday'
    elif row['publish_day'] == 1:
        return 'Tuesday'
    elif row['publish_day'] == 2:
        return 'Wednesday'
    elif row['publish_day'] == 3:
        return 'Thursday'
    elif row['publish_day'] == 4:
        return 'Friday'
    elif row['publish_day'] == 5:
        return 'Saturday'
    elif row['publish_day'] == 6:
        return 'Sunday'

df.loc[:, 'publish_day'] = df.apply(dayofweek, axis=1)

df.head()
```

Out[39]:

	video_id	trending_date	title	channel_title	category_id	category	publish_date
0	cJon8yBKv3g	2017-11-14	Kingsman : Le Cercle d'or (Mickael J Critique)	MJ - FERMEZ LA	1	Film & Animation	2017-11-12
1	g12KeCPRB4Q	2017-11-14	Vide grenier LIVE - 12 Novembre 2017	Kirby-54	20	Gaming	2017-11-12
2	iriZRH3Jllw	2017-11-14	لقجع يثور في وجه الإتحاد الإيفواري ببسيب الجماء...	أخبار اليوم	25	News & Politics	2017-11-12
3	SnHtzkQZggA	2017-11-14	عروس اسطنبول اعلان الحلقة 25 - مترجم للعربية	تركييتنا - bizim Türkiye	1	Film & Animation	2017-11-13
4	u6EoWM2KMPE	2017-11-14	Portokalli, 12 Nentor 2017 - Policat e postbll...	Top Channel Albania	23	Comedy	2017-11-13

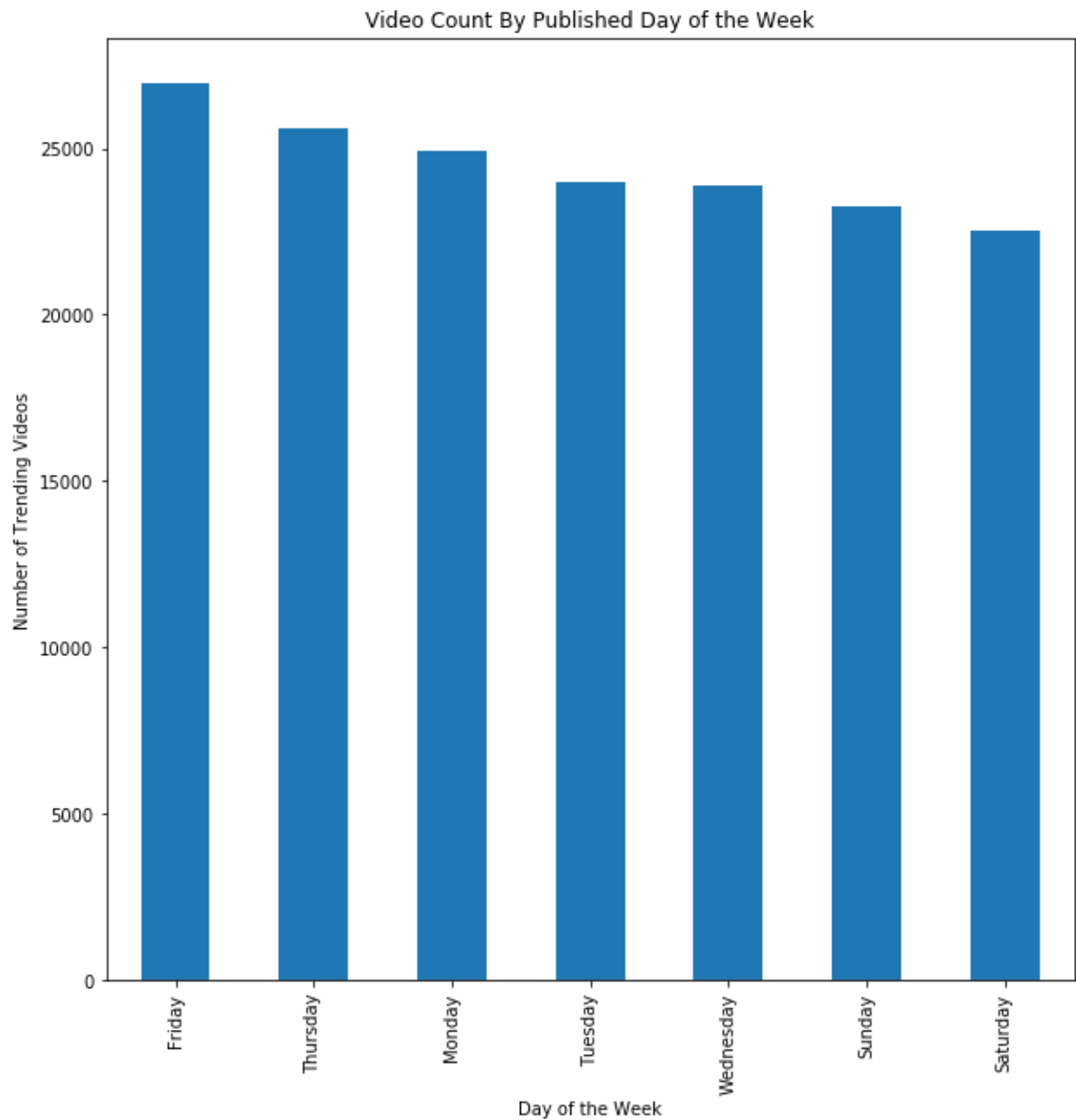

```
In [40]: publish_day_count = df['publish_day'].value_counts()  
print(publish_day_count)
```

```
Friday      26978  
Thursday    25624  
Monday      24920  
Tuesday     23992  
Wednesday  23868  
Sunday      23233  
Saturday    22534  
Name: publish_day, dtype: int64
```

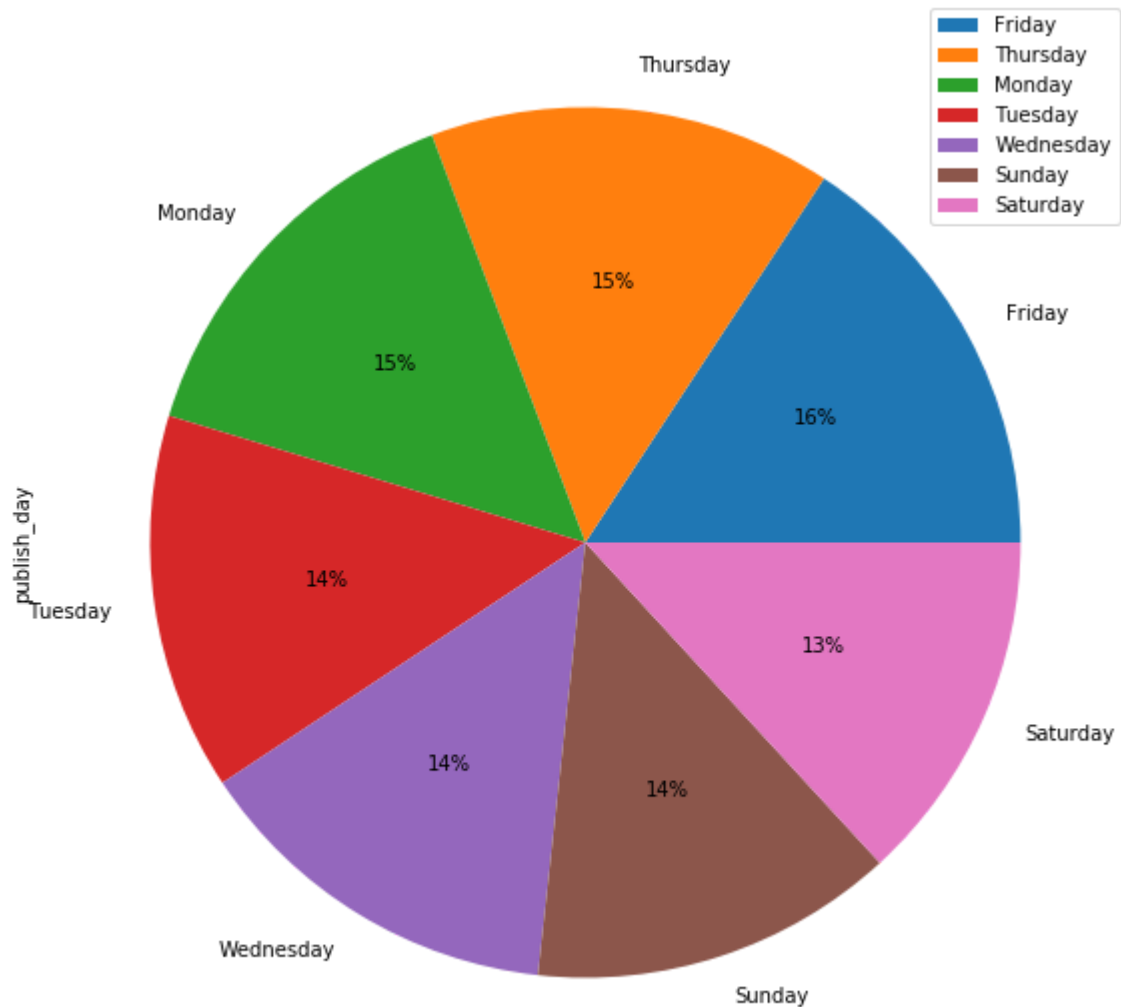
```
In [41]: # Bar Plot by Publish Date of the week
# 0 = Mon, 1 = Tues, 2 = Wed, 3 = Thurs, 4 = Fri, 5 = Sat, 6 = Sun

fig = plt.figure(figsize=(10,10))
ax = fig.gca()
publish_day_count.plot(kind = 'bar', ax = ax)
ax.set_title('Video Count By Published Day of the Week')
ax.set_xlabel('Day of the Week')
ax.set_ylabel('Number of Trending Videos')
```

Out[41]: Text(0, 0.5, 'Number of Trending Videos')



```
In [42]: df5 = pd.DataFrame(publish_day_count)
plot = df5.plot.pie(subplots = True, figsize=(10, 10), autopct='%1.0f%%')
```



... TRENDING DATE ... (MIGHT SCRAP FINDINGS)

- The day of the week is pretty evenly split so the difference between days is not drastic. However... when placed in video count order, there seems to be trend that videos are least likely to reach trending on Sunday and Monday which seems to be in line with videos being less likely to be published Saturday and Sunday.
- Tuesday was the most popular day for a video to reach trending. Does this say anything about suggested patterns for creators?

```
In [43]: # added column for trending day of the week

df['trending_day'] = None
df['trending_day'] = df['trending_timestamp'].dt.dayofweek
df.head()
```

Out[43]:

	video_id	trending_date	title	channel_title	category_id	category	publish_date
0	cJon8yBKv3g	2017-11-14	Kingsman : Le Cercle d'or (Mickael J Critique)	MJ - FERMEZ LA	1	Film & Animation	2017-11-12
1	g12KeCPRB4Q	2017-11-14	Vide grenier LIVE - 12 Novembre 2017	Kirby-54	20	Gaming	2017-11-12
2	irIZRH3Jllw	2017-11-14	لقبح بثور في وجه الإتحاد الإيفواري بسبب الجماء...	أخبار اليوم	25	News & Politics	2017-11-12
3	SnHtzkQZggA	2017-11-14	عروس اسطنبول اعلان الحلقة 25 - مترجم للعربية	تركييتنا - bizim Türkiye	1	Film & Animation	2017-11-13
4	u6EoWM2KMPE	2017-11-14	Portokalli, 12 Nentor 2017 - Policat e postbll...	Top Channel Albania	23	Comedy	2017-11-13

```
In [44]: trending_day_count = df['trending_day'].value_counts()
print(trending_day_count)
```

```
1    25968
4    25095
3    24971
2    24648
5    24479
0    23033
6    22955
Name: trending_day, dtype: int64
```

```
In [45]: def dayofweek(row):
    if row['trending_day'] == 0:
        return 'Monday'
    elif row['trending_day'] == 1:
        return 'Tuesday'
    elif row['trending_day'] == 2:
        return 'Wednesday'
    elif row['trending_day'] == 3:
        return 'Thursday'
    elif row['trending_day'] == 4:
        return 'Friday'
    elif row['trending_day'] == 5:
        return 'Saturday'
    elif row['trending_day'] == 6:
        return 'Sunday'

df.loc[:, 'trending_day'] = df.apply(dayofweek, axis=1)

df.head()
```

Out[45]:

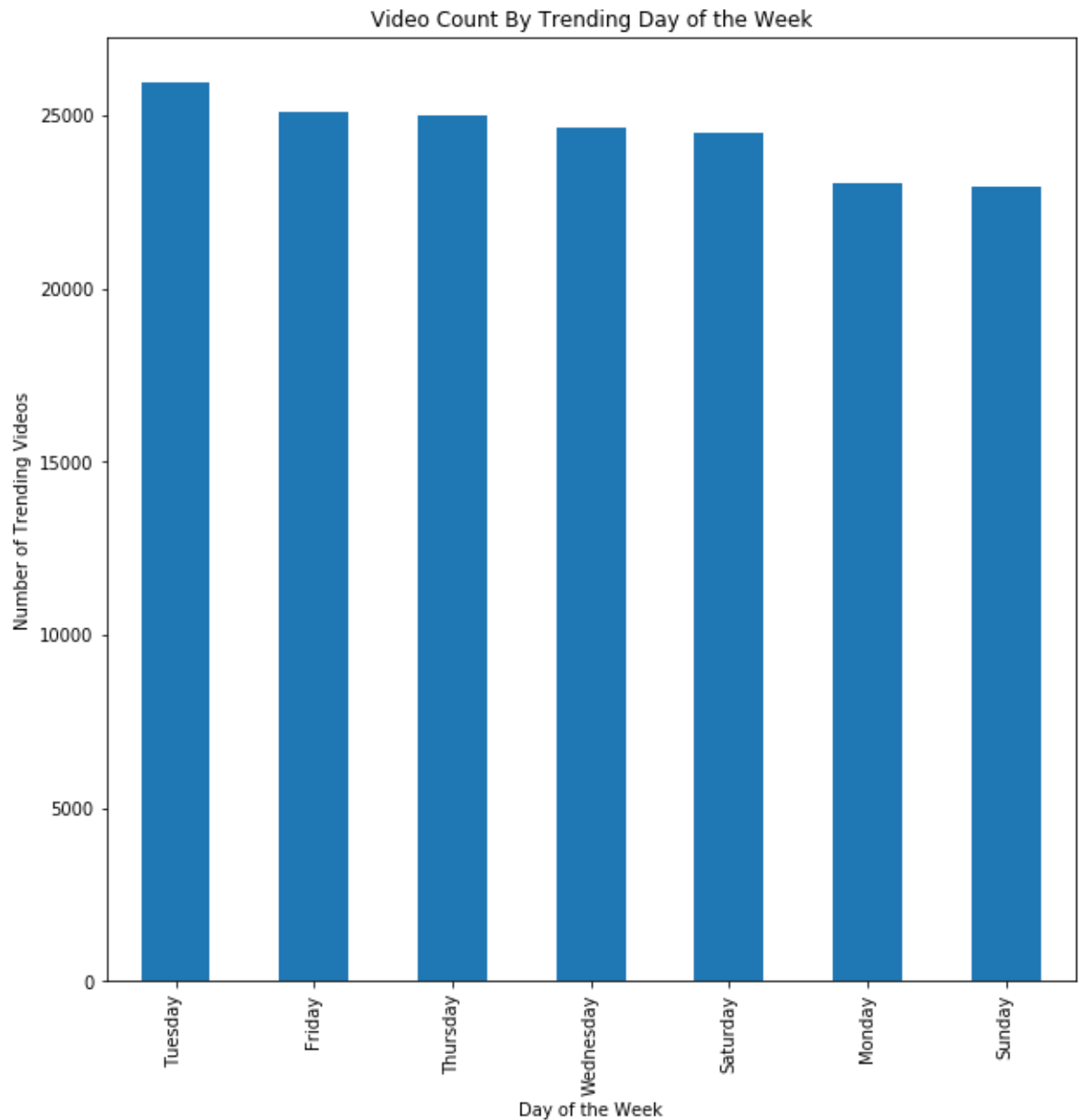
	video_id	trending_date	title	channel_title	category_id	category	publish_date
0	cJon8yBKv3g	2017-11-14	Kingsman : Le Cercle d'or (Mickael J Critique)	MJ - FERMEZ LA	1	Film & Animation	2017-11-12
1	g12KeCPRB4Q	2017-11-14	Vide grenier LIVE - 12 Novembre 2017	Kirby-54	20	Gaming	2017-11-12
2	iriZRH3Jllw	2017-11-14	لقجع يثور في وجه الإتحاد الإيفواري ببسيب الجماء...	أخبار اليوم	25	News & Politics	2017-11-12
3	SnHtzkQZggA	2017-11-14	عروس اسطنبول اعلان الحلقة 25 - مترجم للعربية	تركييتنا - bizim Türkiye	1	Film & Animation	2017-11-13
4	u6EoWM2KMPE	2017-11-14	Portokalli, 12 Nentor 2017 - Policat e postbll...	Top Channel Albania	23	Comedy	2017-11-13

```
In [46]: trending_day_count = df['trending_day'].value_counts()  
print(trending_day_count)
```

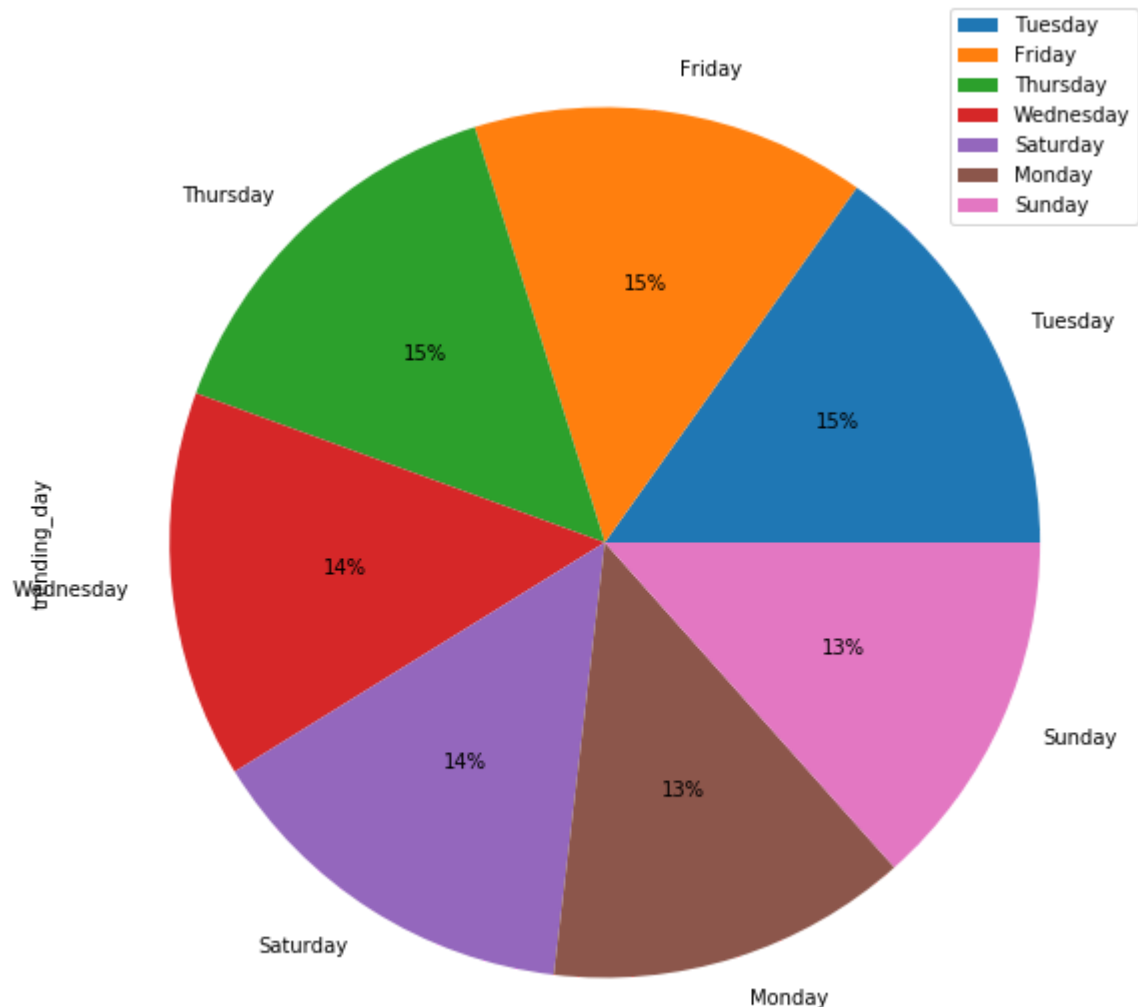
```
Tuesday      25968  
Friday       25095  
Thursday     24971  
Wednesday   24648  
Saturday     24479  
Monday       23033  
Sunday       22955  
Name: trending_day, dtype: int64
```

```
In [47]: fig = plt.figure(figsize=(10,10))
ax = fig.gca()
trending_day_count.plot(kind = 'bar', ax = ax)
ax.set_title('Video Count By Trending Day of the Week')
ax.set_xlabel('Day of the Week')
ax.set_ylabel('Number of Trending Videos')
```

```
Out[47]: Text(0, 0.5, 'Number of Trending Videos')
```



```
In [48]: df6 = pd.DataFrame(trending_day_count)
plot = df6.plot.pie(subplots = True, figsize=(10, 10), autopct='%1.0f%%')
```



... DAYS ON THE PLATFORM ... (COMPLETE)

- The overwhelming majority of trending videos have been on the platform for less than a day 56%, with 85% for less than 3 days.
- This shows a clear correlation between the publish date and trending day. That videos tend to trend within days if not hours from the date published.
- In this perhaps an indication of viewer's attention span?

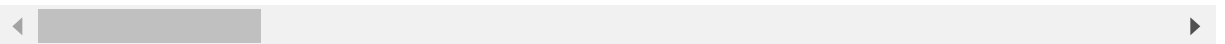
```
In [49]: df['Time_Live'] = df['trending_timestamp'] - df['publish_timestamp']
```


In [62]: `df.head(100)`

Out[62]:

	video_id	trending_date	title	channel_title	category_id	category	publis
0	cJon8yBKv3g	2017-11-14	Kingsman : Le Cercle d'or (Mickael J Critique)	MJ - FERMEZ LA	1	Film & Animation	201'
1	g12KeCPRB4Q	2017-11-14	Vide grenier LIVE - 12 Novembre 2017	Kirby-54	20	Gaming	201'
2	irIZRH3Jllw	2017-11-14	لقجع بثور في وجه الاتحاد الإفريقي... يسبب الجماه	أخبار اليوم	25	News & Politics	201'
3	SnHtzkQZggA	2017-11-14	عروس اسطنبول اعلان الحلقة 25 - مترجم للعربية	bizim Türkiye	1	Film & Animation	201'
4	u6EoWM2KMPE	2017-11-14	Portokalli, 12 Nentor 2017 - Policat e postbl...	Top Channel Albania	23	Comedy	201'
...
95	-bqyTKrGq18	2017-11-14	DÉFIS TECHNIQUE INCROYABLE ! FOOTBALL SKILLS	Foot Familya	17	Sports	201'
96	LucUgZucG9M	2017-11-14	JE REFAIS DES PHOTOS ! #DÉFI	GREED	24	Entertainment	201'
97	1Cat-Fxks_s	2017-11-14	La cuisson des œufs à la coque différente au M...	Les Grosses Têtes	23	Comedy	201'
98	w9_TuLo3Ao0	2017-11-14	ملخص مباراة المغرب وكوت ديفوار 0-2 شاشة كاملة ...	عيونك	17	Sports	201'
99	k_73Get4zFA	2017-11-14	Breaking news politiques - Morin a fait un rêve	France Inter	23	Comedy	201'

100 rows × 30 columns



```
In [63]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 171149 entries, 0 to 171148
Data columns (total 30 columns):
video_id          171149 non-null object
trending_date     171149 non-null object
title            171149 non-null object
channel_title     171149 non-null object
category_id       171149 non-null int64
category          171149 non-null object
publish_date      171149 non-null object
publish_time      171149 non-null object
tags             171149 non-null object
views            171149 non-null int64
likes            171149 non-null int64
dislikes         171149 non-null int64
comment_count     171149 non-null int64
thumbnail_link    171149 non-null object
comments_disabled 171149 non-null bool
ratings_disabled  171149 non-null bool
video_error_or_removed 171149 non-null bool
description       171149 non-null object
country          171149 non-null object
publish_timestamp 171149 non-null datetime64[ns]
trending_timestamp 171149 non-null datetime64[ns]
log_views         171149 non-null float64
log_likes         171149 non-null float64
log_dislikes      171149 non-null float64
log_comment_count 171149 non-null float64
tag_count         171149 non-null int64
publish_day       171149 non-null object
trending_day      171149 non-null object
Time_Live         171149 non-null timedelta64[ns]
Days_Live         171149 non-null int64
dtypes: bool(3), datetime64[ns](2), float64(4), int64(7), object(13), timedel
ta64[ns](1)
memory usage: 35.7+ MB
```

```
In [53]: df["Days_Live"] = None
```

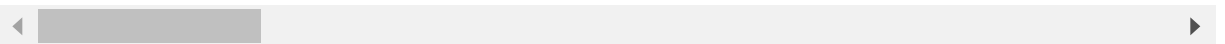
```
In [54]: df["Days_Live"] = df["Time_Live"].dt.days
```

In [64]: df.head(100)

Out[64]:

	video_id	trending_date	title	channel_title	category_id	category	publis
0	cJon8yBKv3g	2017-11-14	Kingsman : Le Cercle d'or (Mickael J Critique)	MJ - FERMEZ LA	1	Film & Animation	201
1	g12KeCPRB4Q	2017-11-14	Vide grenier LIVE - 12 Novembre 2017	Kirby-54	20	Gaming	201
2	irIZRH3Jllw	2017-11-14	لقجع بثور في وجه الاتحاد الإفريقي... يسبب الجماه	أخبار اليوم	25	News & Politics	201
3	SnHtzkQZggA	2017-11-14	عروس اسطنبول اعلان الحلقة 25 - مترجم للعربية	bizim Türkiye	1	Film & Animation	201
4	u6EoWM2KMPE	2017-11-14	Portokalli, 12 Nentor 2017 - Policat e postbl...	Top Channel Albania	23	Comedy	201
...
95	-bqyTKrGq18	2017-11-14	DÉFIS TECHNIQUE INCROYABLE ! FOOTBALL SKILLS	Foot Familya	17	Sports	201
96	LucUgZucG9M	2017-11-14	JE REFAIS DES PHOTOS ! #DÉFI	GREED	24	Entertainment	201
97	1Cat-Fxks_s	2017-11-14	La cuisson des œufs à la coque différente au M...	Les Grosses Têtes	23	Comedy	201
98	w9_TuLo3Ao0	2017-11-14	ملخص مباراة المغرب وكوت ديفوار 0-2 شاشة كاملة ...	عيونك	17	Sports	201
99	k_73Get4zFA	2017-11-14	Breaking news politiques - Morin a fait un rêve	France Inter	23	Comedy	201

100 rows × 30 columns

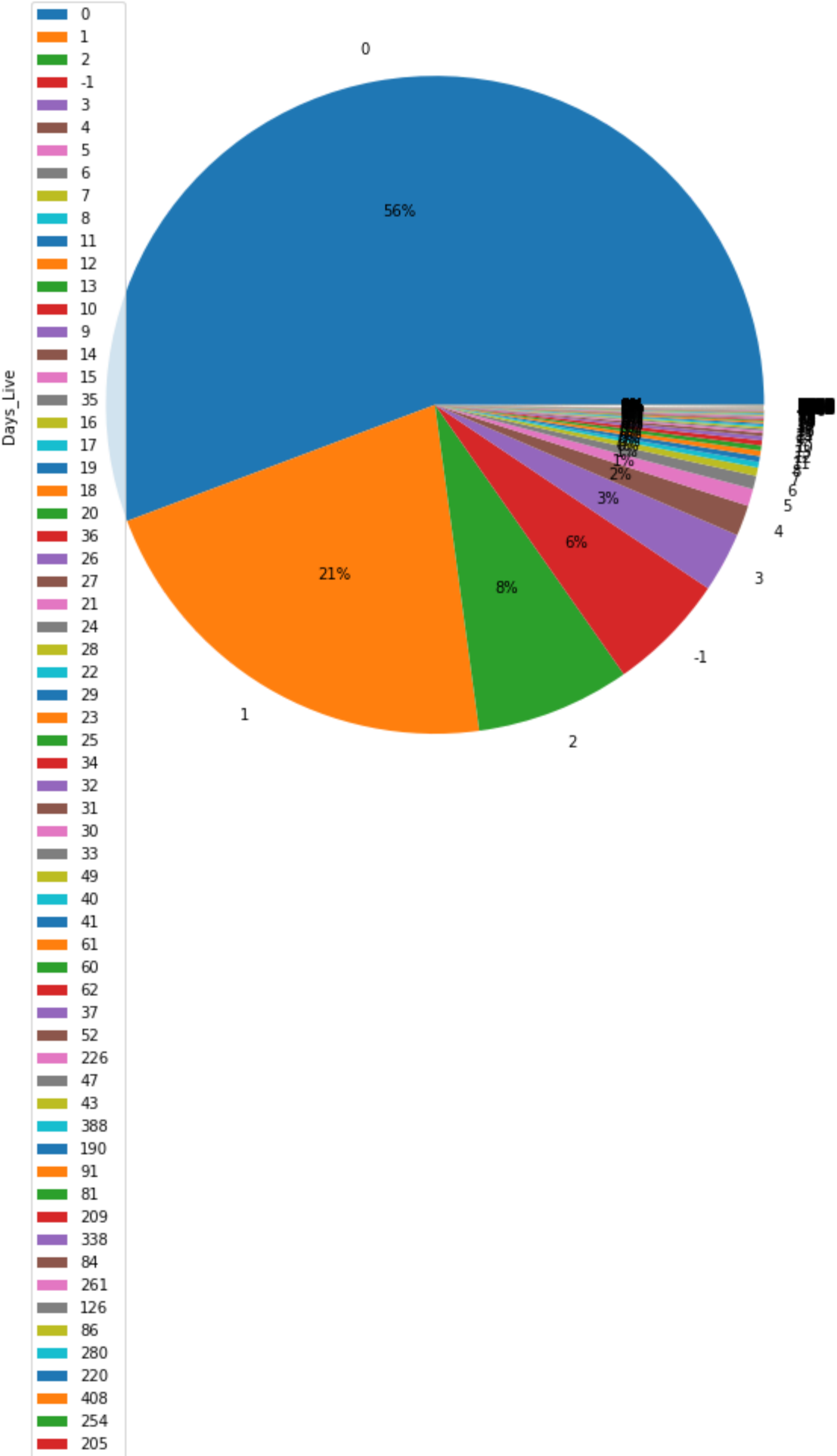


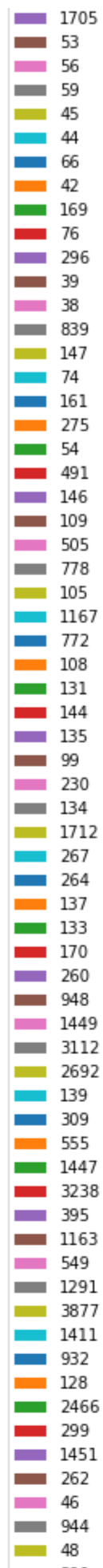
```
In [68]: days_live_count = df['Days_Live'].value_counts()  
print(days_live_count)
```

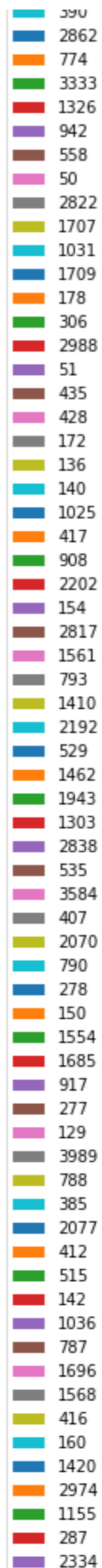
```
0      95410  
1      36631  
2      12972  
-1     10024  
3       5021  
...  
1356      1  
77         1  
333        1  
461        1  
698        1
```

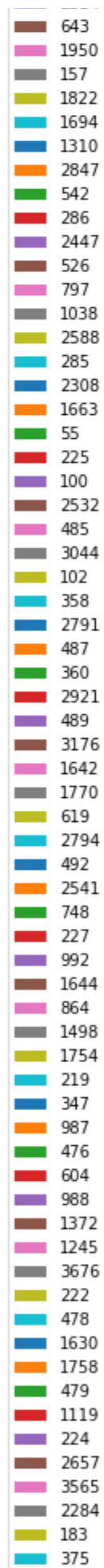
Name: Days_Live, Length: 394, dtype: int64

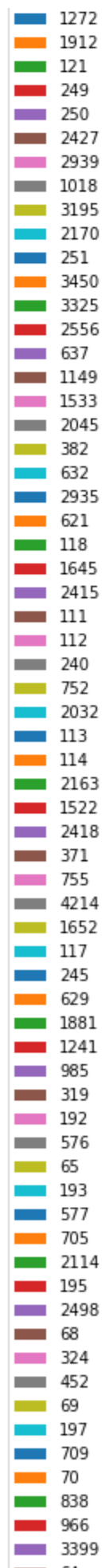
```
In [71]: df7 = pd.DataFrame(days_live_count)
plot = df7.plot.pie(subplots = True, figsize=(10, 10), autopct='%1.0f%%')
```

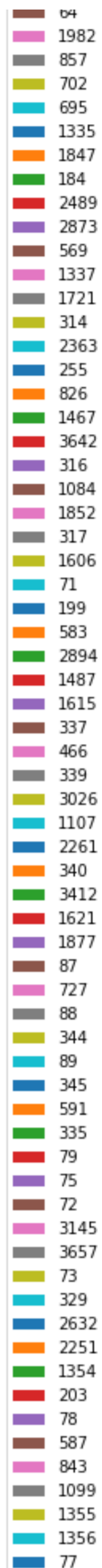


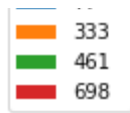












... MULTIPLE TRENDERS ...

- If you aggregate the total number of videos trending by channel, almost 50% of those channels had more than one video trending.
- Viewers tend to see videos from the same channel close to each other OR presumably, this property feeds into suggestion algorithms and viewers that see 1 video from a channel may be suggested to see another.
- This presents an opportunity for the Youtube Creator to have more than one video trending at the moment and for Youtube to encourage more trending videos.
- If a Creator achieves trending with 1 video, two things may happen: there seems to be a correlation that either one of his/her existing videos can increase in viewership rapidly OR if the creator posts another video soon after, that new video may benefit from the popularity of the previous.
- Assuming the creator does not proactively see this connection, Youtube (who is most likely already suggesting to viewers to see videos from the same channel) may now encourage via prompts or incentives more video uploads in a short period of time from the same creator to build on momentum

```
In [57]: channel_count_series = df['channel_title'].value_counts()
         print(channel_count_series)
```

```
SET India                253
SAB TV                   248
VikatanTV                240
The Late Show with Stephen Colbert  230
Анатолий Шарий          224
...
MCDAVO                   1
Technical Patel          1
Shizarium                1
Deejay Maquina           1
Гельмут Вайссвальд      1
Name: channel_title, Length: 33965, dtype: int64
```

```
In [58]: channel_count_series.describe()
```

```
Out[58]: count    33965.000000
         mean       5.038981
         std       12.047872
         min        1.000000
         25%        1.000000
         50%        1.000000
         75%        4.000000
         max       253.000000
         Name: channel_title, dtype: float64
```

```
In [59]: channel_count = pd.DataFrame(df.channel_title.value_counts().reset_index().values, columns=["channel_title", "VideoCount"])
channel_count
```

Out[59]:

	channel_title	VideoCount
0	SET India	253
1	SAB TV	248
2	VikatanTV	240
3	The Late Show with Stephen Colbert	230
4	Анатолий Шарий	224
...
33960	MCDAVO	1
33961	Technical Patel	1
33962	Shizarium	1
33963	Deejay Maquina	1
33964	Гельмут Вайсвальд	1

33965 rows × 2 columns

```
In [60]: multiple_trender = channel_count['VideoCount'] > 1
channel_count[multiple_trender]
```

Out[60]:

	channel_title	VideoCount
0	SET India	253
1	SAB TV	248
2	VikatanTV	240
3	The Late Show with Stephen Colbert	230
4	Анатолий Шарий	224
...
16399	Sólo Ellas 4	2
16400	Коллекция Рецептов	2
16401	اعلانات مسلسل تركي	2
16402	메이리우 Korean Sister	2
16403	Mehmed Bir Cihan Fatihi	2

16404 rows × 2 columns

```
In [61]: # From the channels with trending videos, almost 50% are concurrently trending
         # more than one video

         # From Previous calculations
         Total_VideoCount_DF = 171149          # df.describe()
         Total_ChannelCount_DF = 33965         # channel_count_series.describe()
         Multiple_TrenderCount_Df = 16404      # multiple_trender view

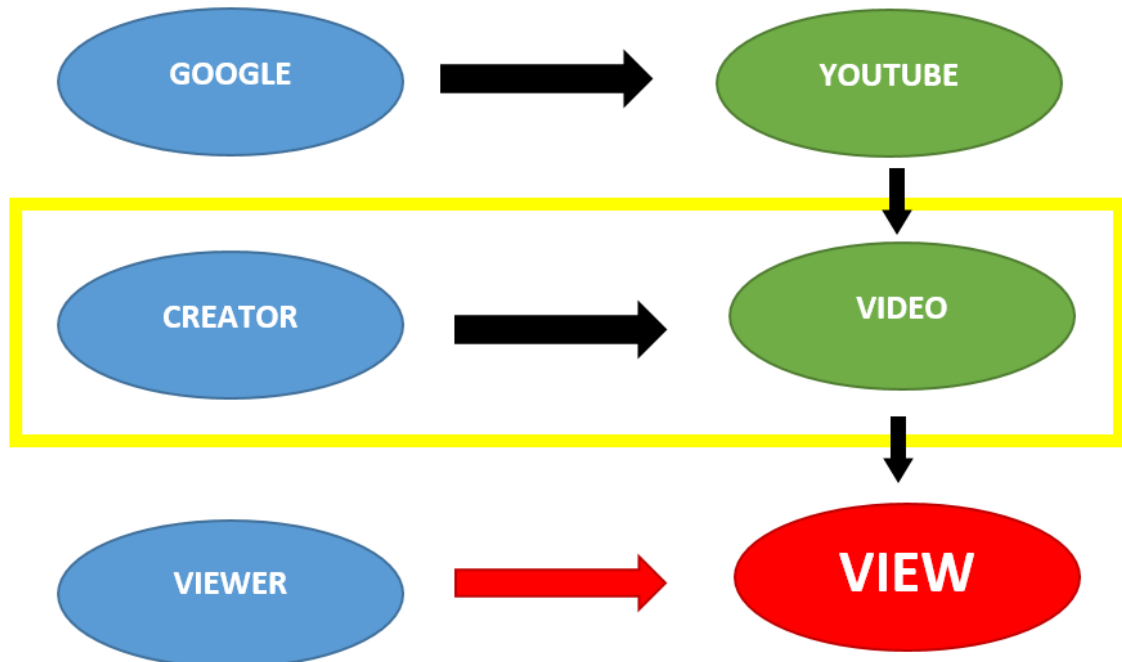
         Multiple_TrenderCount_Df/Total_ChannelCount_DF
```

Out[61]: 0.48296776093036947

CONCLUSION

```
In [75]: from IPython.display import Image
         Image(filename='Youtube Relationship - Diagram.PNG')
```

Out[75]:



QUESTIONS TO EXPLORE

*** Are there any correlations between factors that stem fromt the Youtube Creator and trendability?**

YES!

*** For Creators, are there any opportunities and or suggested behavior to encourage trendability?**

YES!

*** For Youtube, is it worthwhile to spend additional resources on the Youtube Creator? Are there any suggested changes for existing platform?**

YES!