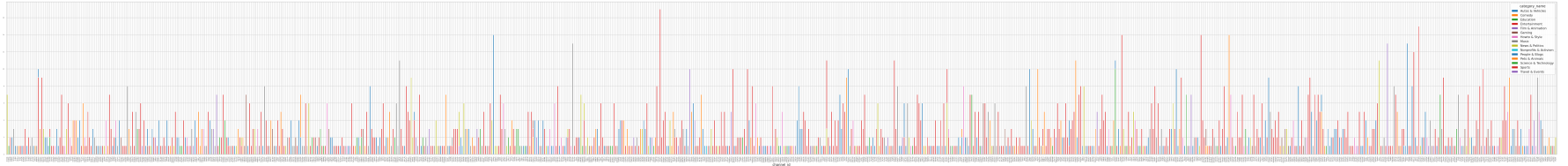


# Youtube Trending Video Data EDA

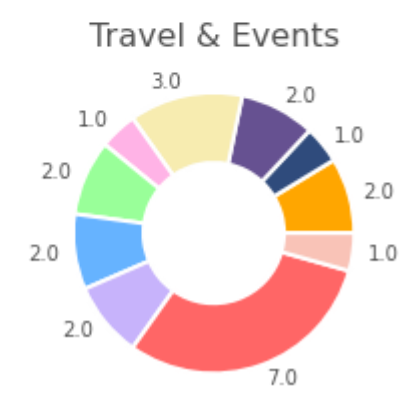
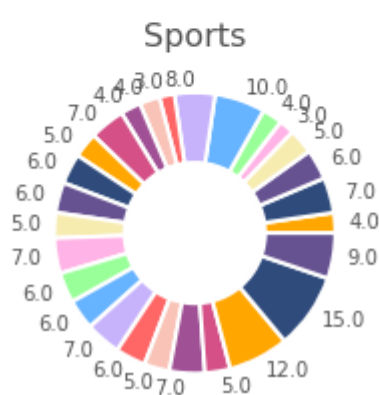
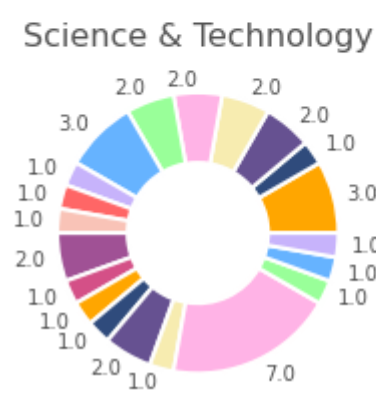
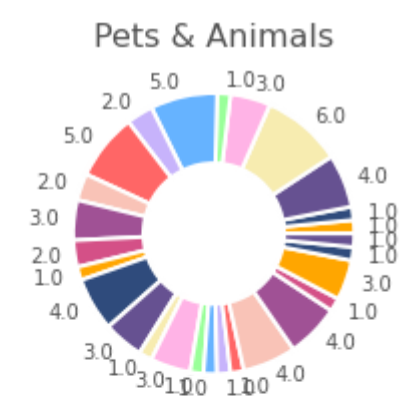
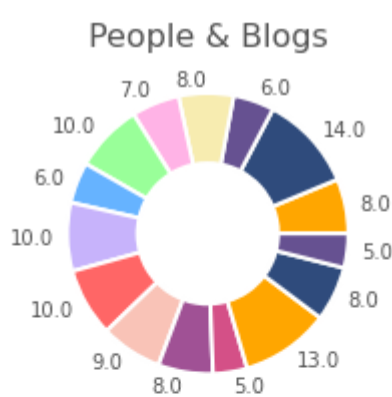
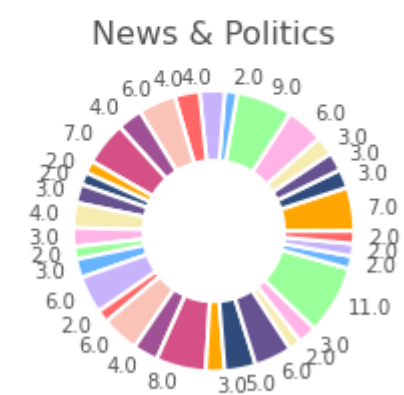
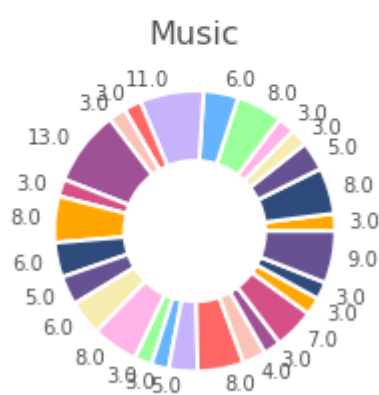
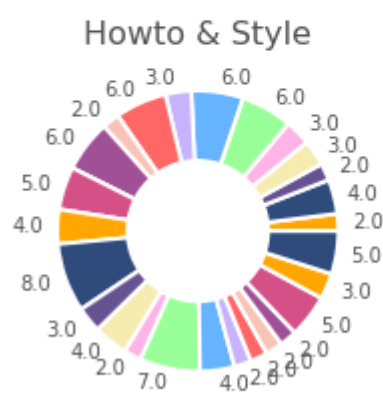
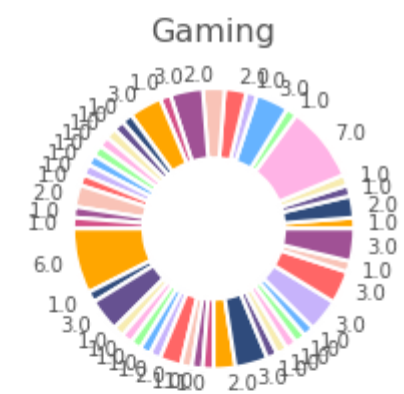
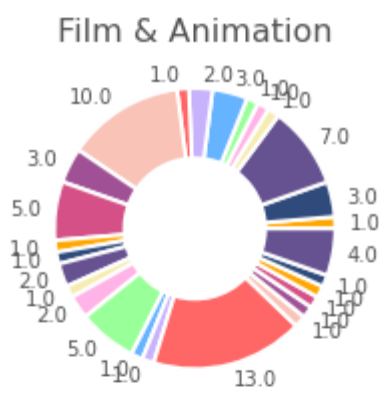
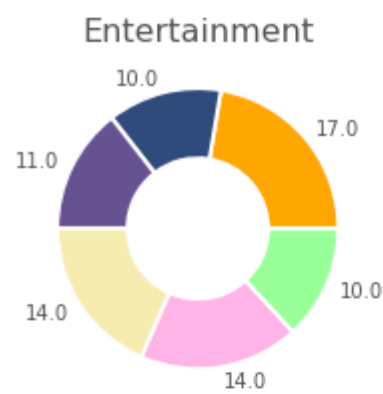
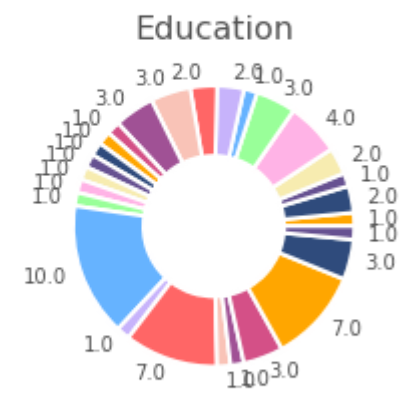
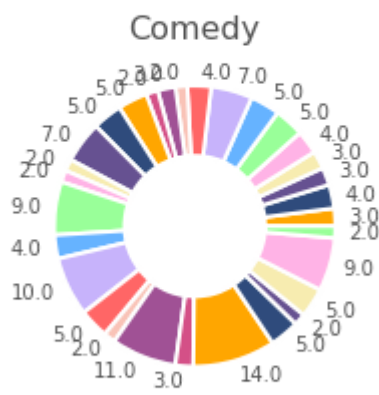
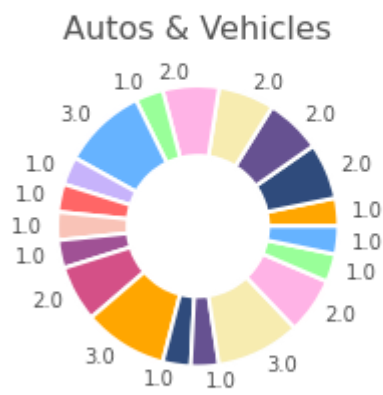
## ▼ 1.Visualization

### ▼ 1-1. Categorical > By Channel > Video Count



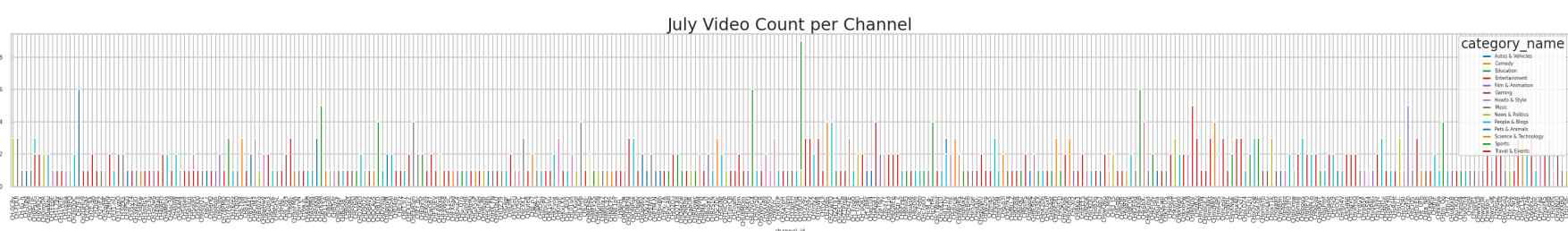
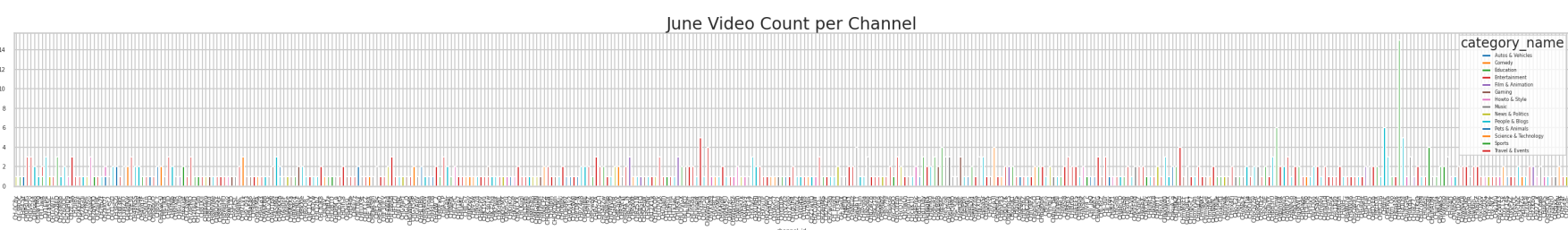
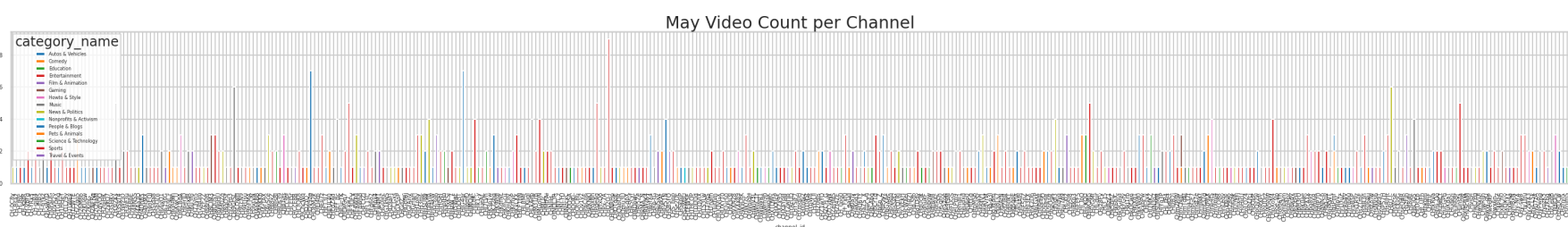
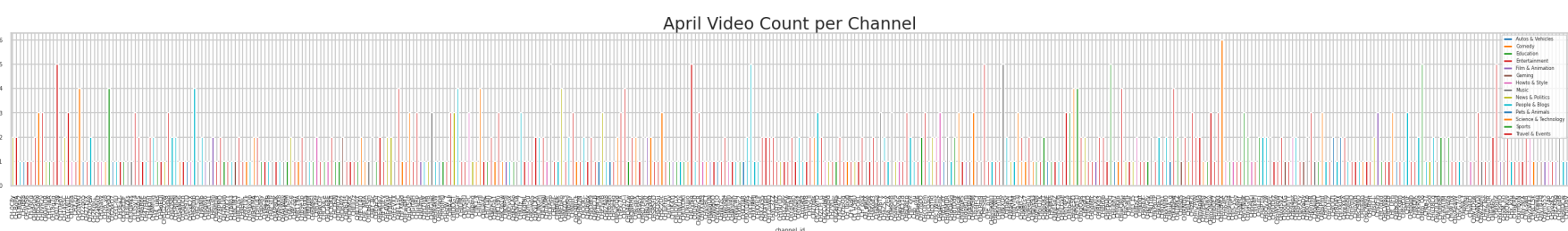
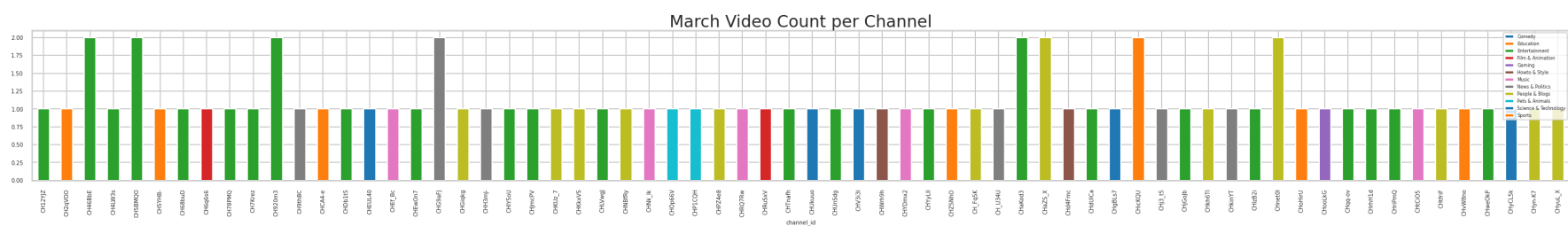
- Please magnify the above plot to see video count by channel for all categories.
- Each color represents the category. Please refer to the legend on the upper right for description.
- CHECKPOINTS
  - There are a few channels that upload in different categories;(eg. CH2DHU9) most upload in one category.
- The below plot is also another way to show video count by channel.

### Video Count of channels by Category

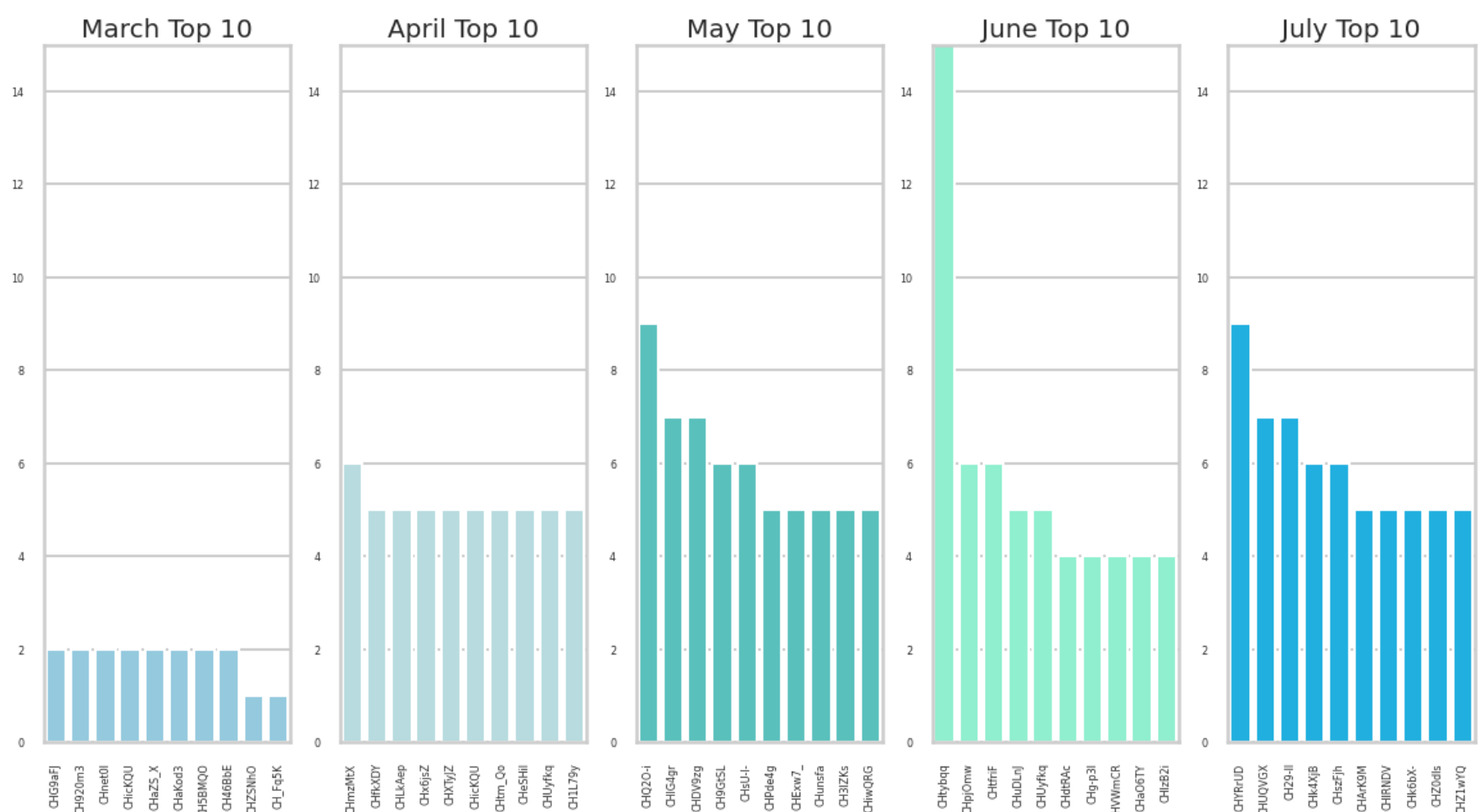


▼ 1-2. Monthly > Categorical > By Channel > Video Count

- There are total 5 months(March~ July), resulting in 5 graphs.
- Please magnify the plots for details.



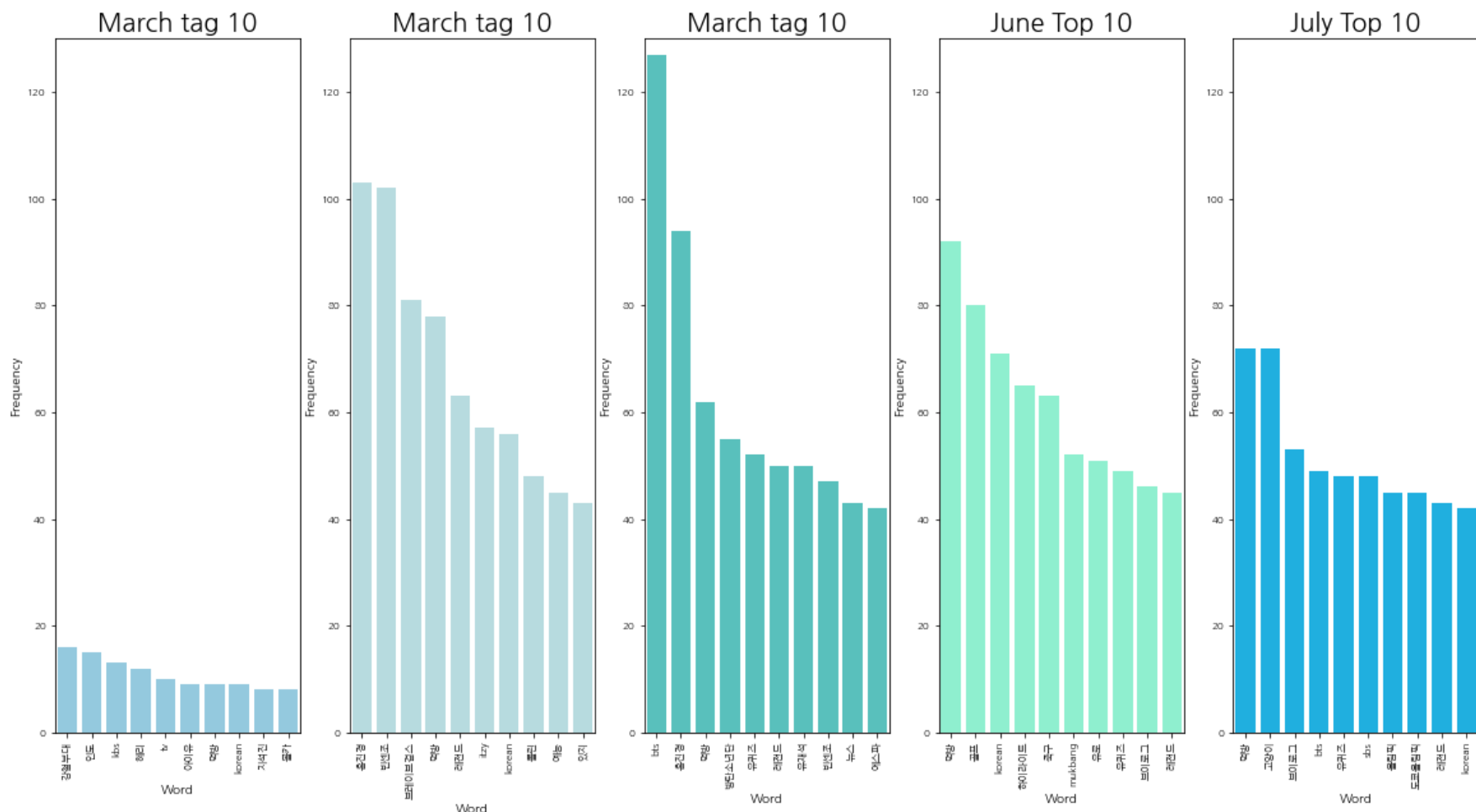
### ▼ 1-3. Monthly Top 10 Channels



### ▼ 1-4. Weekly Top 5 Channels



## ▼ 1-5. Monthly Tag Rank

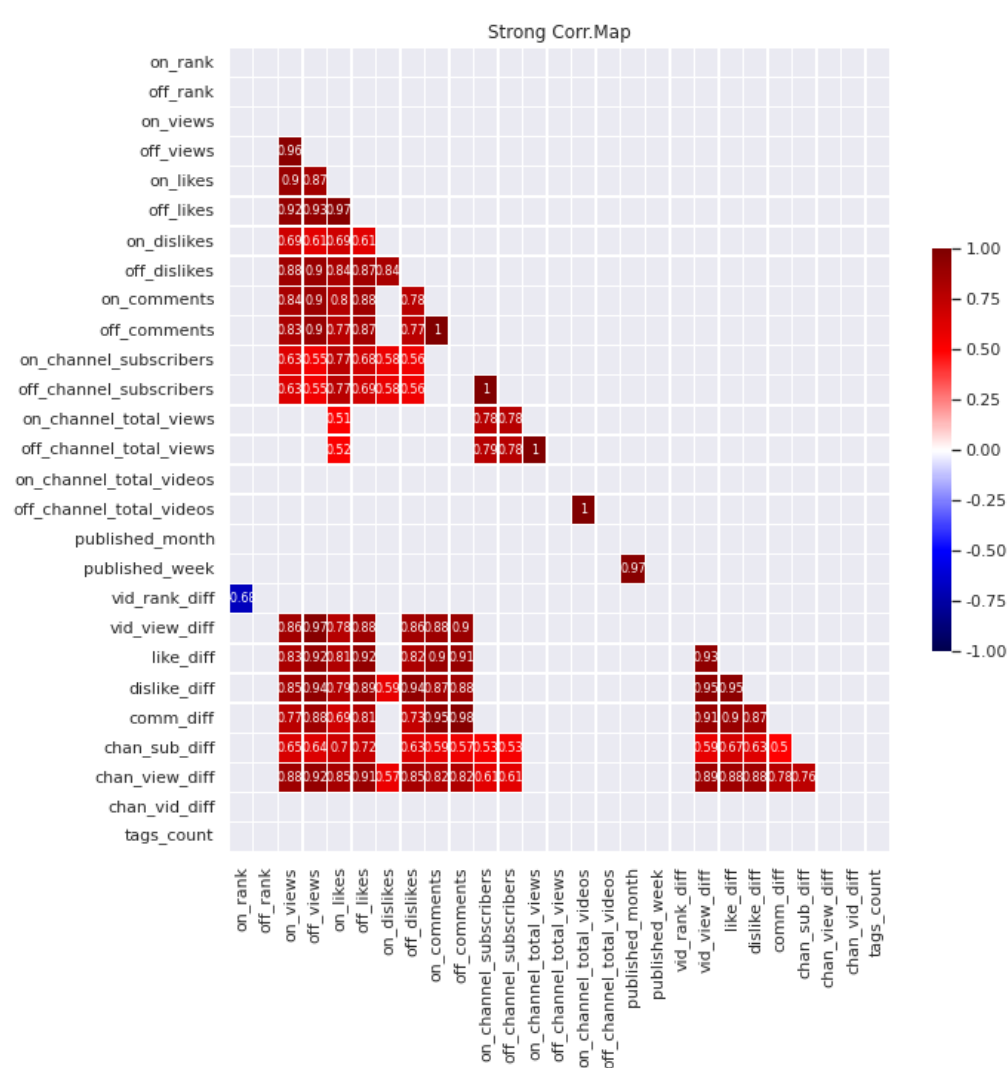
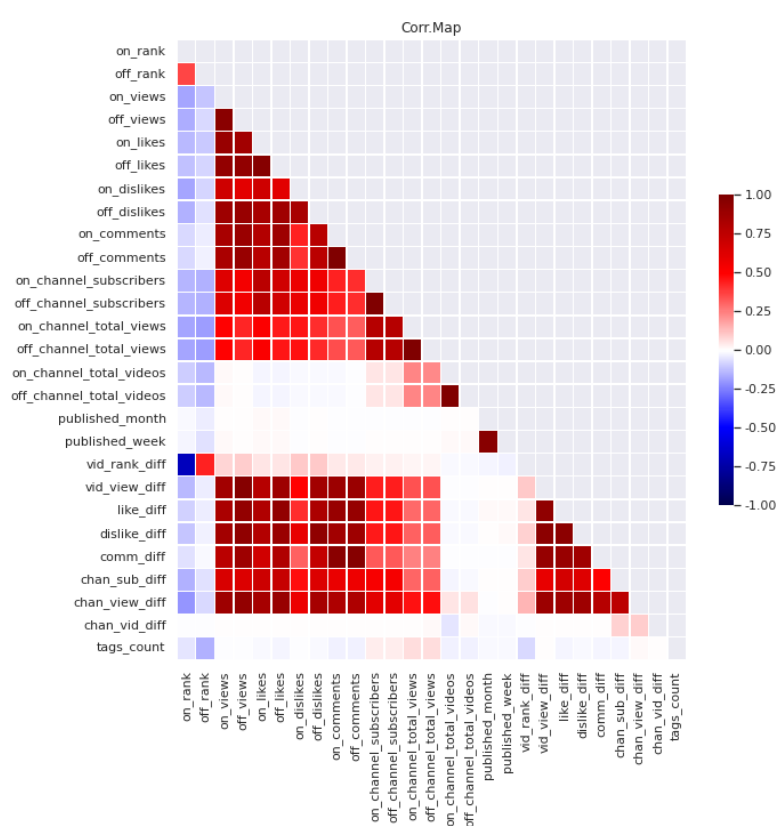


- The monthly Top 10 tags, show that the youtube videos follow the trend of the month. Combining the korean-english word would definitely clarify the Top 10 words, as there are duplicates such as 있지 = Itzy(k-pop group).

## ▼ 2. Creating Viewer Engagement Features & Analysis

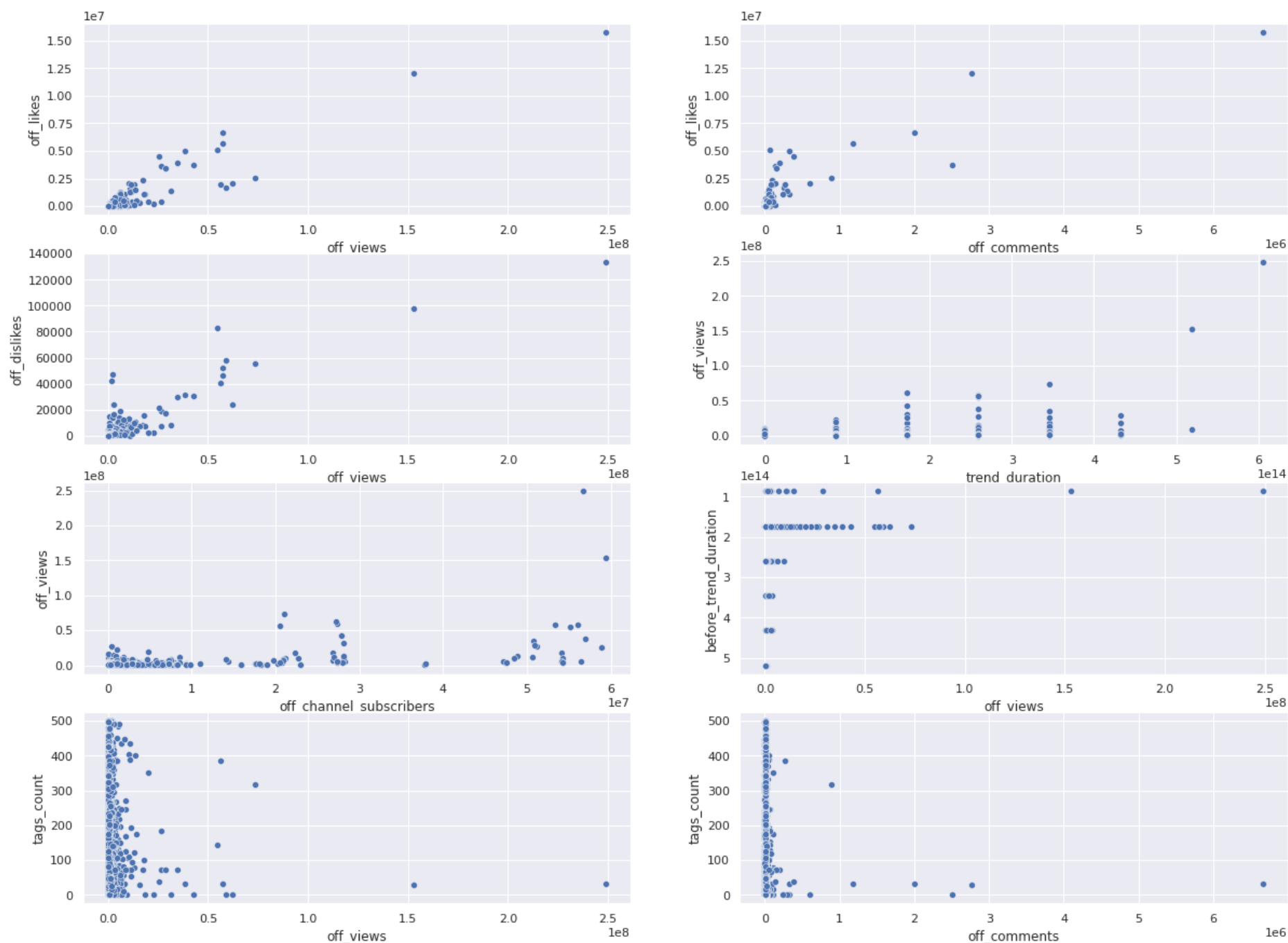
## Created features

- 'vid\_rank\_diff' : the rank difference of on/off trending
- 'vid\_view\_diff' : the view difference of on/off trending
- 'like\_diff' = the like count difference of on/off trending
- 'dislike\_diff' = the dislike count difference of on/off trending
- 'comm\_diff' = the comment count difference of on/off trending
- 'chan\_sub\_diff' = the subscriber count difference of on/off trending
- 'chan\_view\_diff' = the channel total view count difference of on/off trending
- 'chan\_vid\_diff' = the channel total video count difference of on/off trending
- **'trend\_duration'** = the duration of trending
  - It could be generally thought the longer trend duration would mean more views, which could result with high correlation between the two features.
- 'before\_trend\_duration' = the duration before trending
- **'tags\_count'** = the number of tag words
  - It is important to optimize the meta data for better search & visibility of videos, and Tags count would be one of the strongest tools to enhance both.



**From the Correlation Map we could interpret as follows:**

## Scatter Plot of important features



### From the Scatterplot we could interpret as follows:

1. Likes are more correlated with Trending. Dislikes have a slightly lower correlation with the overall features. Refer to 1st, 3rd scatter plot.
2. The channels subscribers & total views show low correlation. Refer to 5th plot.
3. Surprisingly the custom feature 'trend\_duration' has less correlation with views.
4. Contrary to expectations, Tags\_count was also not a strong correlated feature.