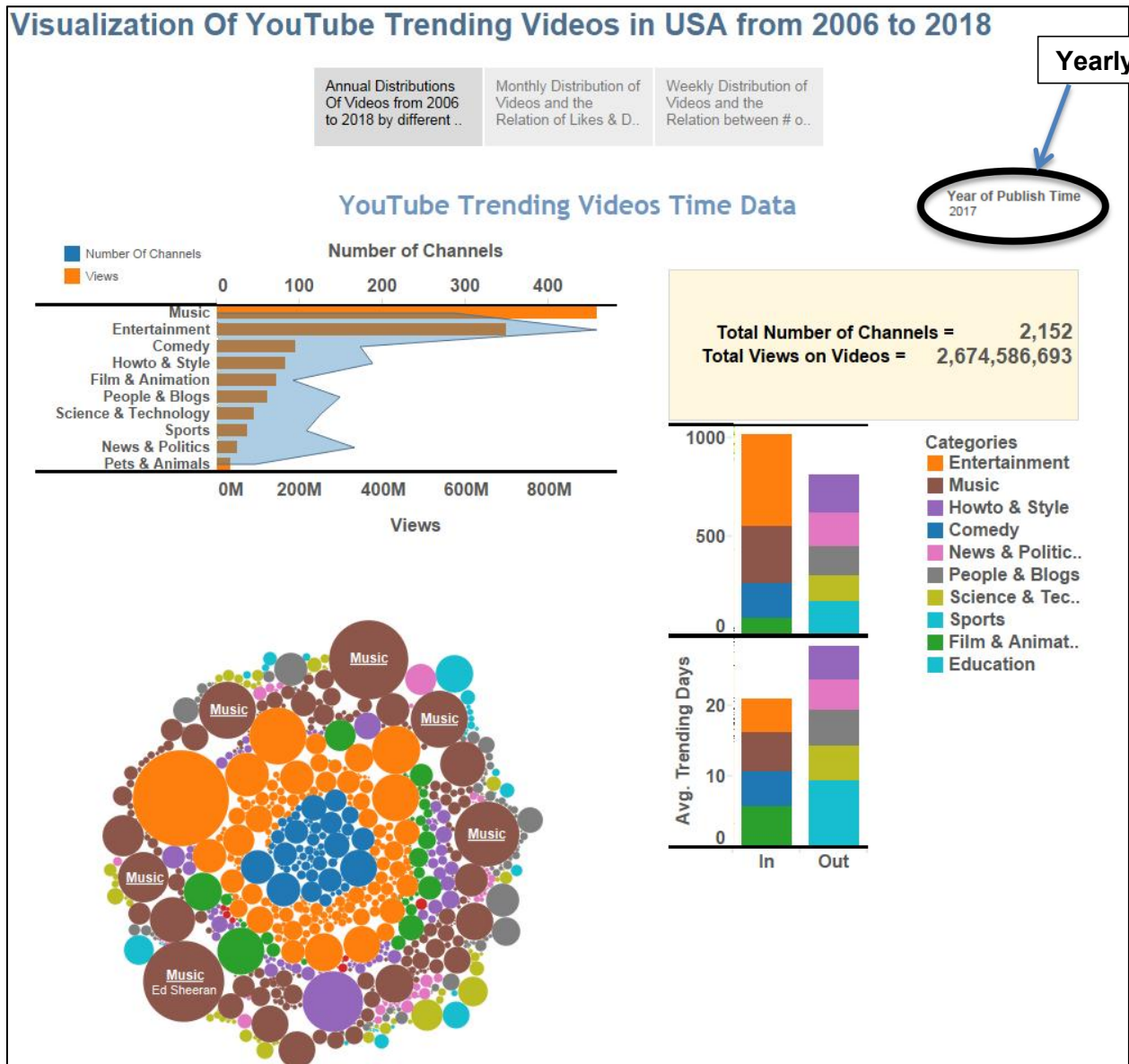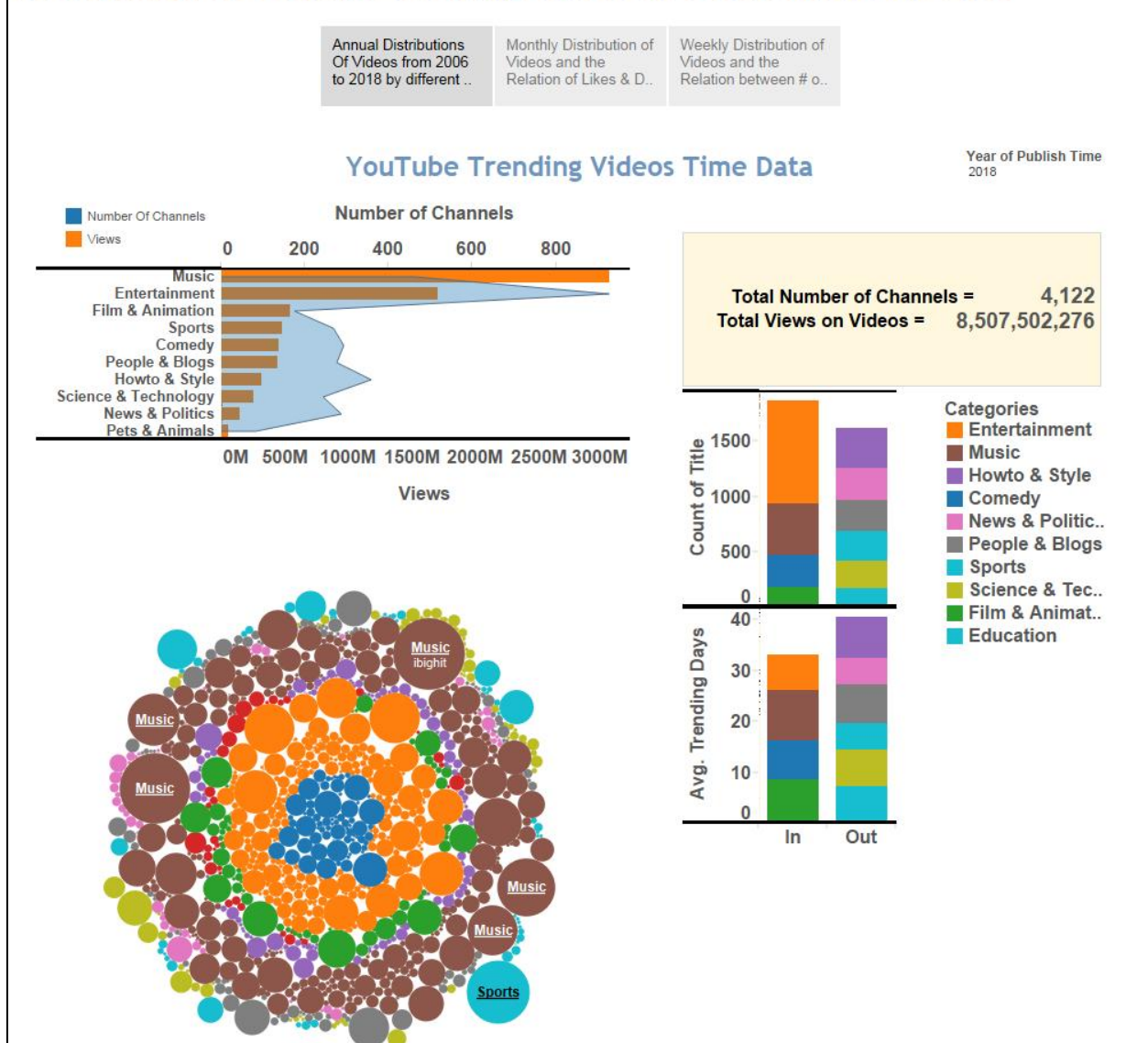# Tableau project - YouTube Trending Videos of USA from 2006 to 2018

1. Insight no.1
- **Link -** I worked on Tableau Desktop. Here I attached Screen Shots of Dashboard 1.

Visualization Of YouTube Trending Videos in USA from 2006 to 2018

Yearly Analysis

Annual Distributions Of Videos from 2006 to 2018 by different ..

Monthly Distribution of Videos and the Relation of Likes & D..

Weekly Distribution of Videos and the Relation between # o..

Year of Publish Time
2017

YouTube Trending Videos Time Data

Number of Channels

- Number Of Channels
- Views

Total Number of Channels = 2,152
Total Views on Videos = 2,674,586,693

Categories
- Entertainment
- Music
- Howto & Style
- Comedy
- News & Politic..
- People & Blogs
- Science & Tec..
- Sports
- Film & Animat..
- Education

Visualization Of YouTube Trending Videos in USA from 2006 to 2018

| Annual Distributions Of Videos from 2006 to 2018 by different .. | Monthly Distribution of Videos and the Relation of Likes & D.. | Weekly Distribution of Videos and the Relation between # o.. |

YouTube Trending Videos Time Data

Year of Publish Time
2018

Number of Channels

- ■ Number Of Channels
- ■ Views

Number of Channels: 0 200 400 600 800

Music
Entertainment
Film & Animation
Sports
Comedy
People & Blogs
Howto & Style
Science & Technology
News & Politics
Pets & Animals

Views: 0M 500M 1000M 1500M 2000M 2500M 3000M

Views

Total Number of Channels =   4,122
Total Views on Videos =   8,507,502,276

Count of Title: 1500 1000 500 0

Categories
- ■ Entertainment
- ■ Music
- ■ Howto & Style
- ■ Comedy
- ■ News & Politic..
- ■ People & Blogs
- ■ Sports
- ■ Science & Tec..
- ■ Film & Animat..
- ■ Education

Avg. Trending Days: 40 30 20 10 0

In   Out

Music ibighit

Music

Music

Music

Music

Sports

Summary - "Visualization of the yearly behavior of Trending Youtube videos grouped into the Categories of Channels along with the total views they got"

My first Dashboard consist of 4 different analysis that could be vary yearly by using dropdown button. I would like to take your attention to the Upper-Right Table that shows the Total Number of Youtube Channels and the Total Views on Videos in the given year.

Upper-Left Bar Plot represents the famous categories of Youtube channels and the number of Views their videos get. Here we can witness the

importance of Music, Entertainment, Comedy, and Film & Animation are the main categories according to their views and number of Channels they have.

Lower-Left is the Bubble Packed Plot that illustrates the importance of individual Youtube channel accumulated by the categories (colored bubbles) and the views (Sized of bubbles).

Middle-Right Bar Plot shows the behavior of Views on videos in different categories along with the total trending days of those videos shown by color. The overlap black-line represents the percent the uploaded videos have been removed by the youtube authority. It can be calculated as,

Created Calculated Field named Videos Removed From YouTube=

if [Video Error Or Removed] = TRUE then
1
END

Created Calculated Field named Videos removed per Videos upload =

SUM([Videos Removed From YouTube])/COUNT([Channel Title]) *100

- **Design -**

Under my analysis, I have found Bar-Plots are the best representation of channel categories. Although it could be a good alternative instead of Bubble Packed. Here my idea was, no one is interested to see the details of individual channels in Youtube those are in Thousands. StakeHolders or concerned members are only interested to see the Best Youtube channel that has maximum views and its related category. In this way, Bubble Packed should be best representation of all Data.
Also some data analyst prefer to plot time series to see the Total number of youtube channels and their views. But I recommend Chart is even more better since exact value could be more important to be known instead of trend.

- **Resources -** N/A

2. Insight no.2
- **Link -** I worked on Tableau Desktop. Here I attached Screen Shots of Dashboard 2.

Visualization Of YouTube Trending Videos in USA from 2006 to 2018

Grouping Channels

YouTube Trending Videos Categorized by Channels

Categories_Main (group)
Comedy, Entertainment,

In / Out of Main YouTube Categories According to Views

Categories	In
Comedy	646M
Entertainment	2,406M
Film & Animation	690M
Music	3,974M

Total Trending Days
1.00 — 41.00

LOG Transform Dislikes

Average

Average

LOG Transform Likes
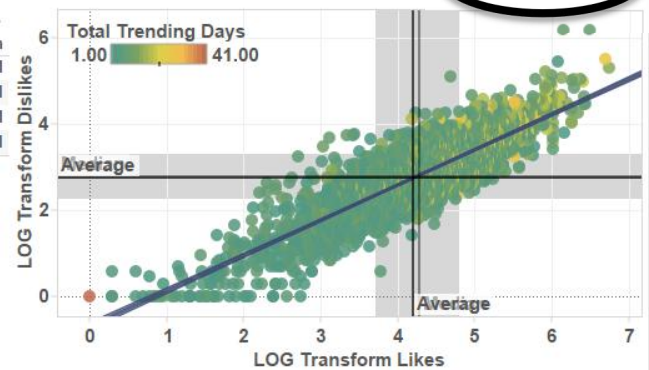
Categories
- Comedy
- Film & Animati..
- Entertainment
- Music
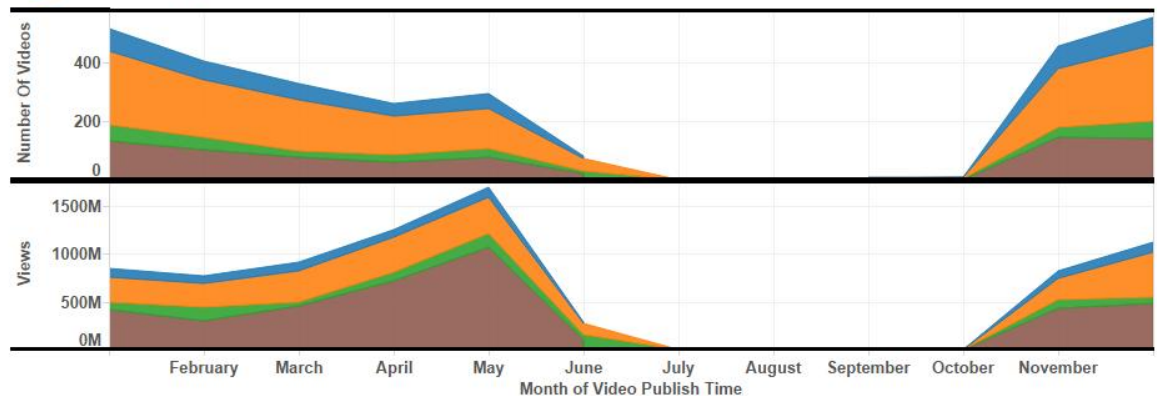
Number Of Videos
400
200
0

Views
1500M
1000M
500M
0M

February   March   April   May   June   July   August   September   October   November
Month of Video Publish Time

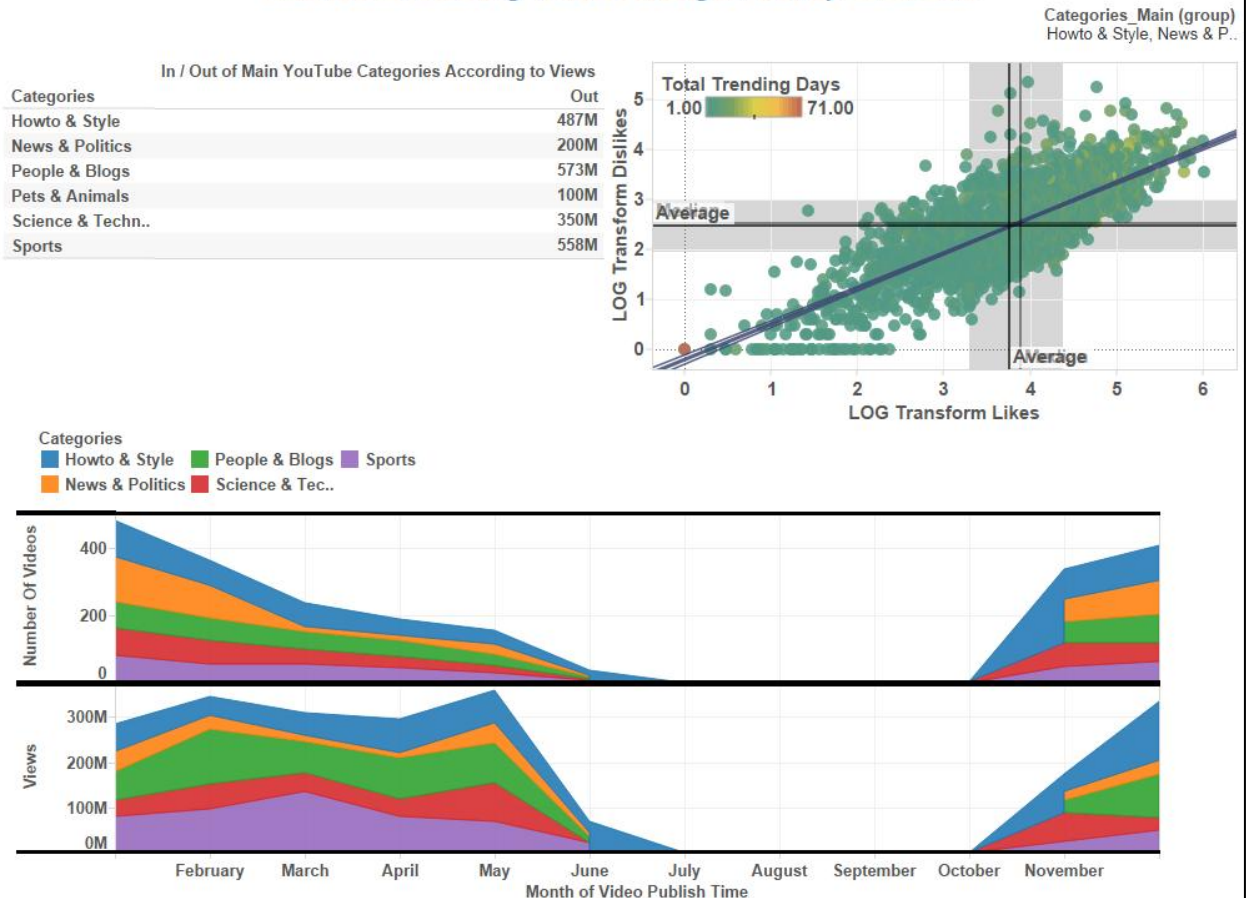## Visualization Of YouTube Trending Videos in USA from 2006 to 2018

Annual Distributions Of Videos from 2006 to 2018 by different ..	Monthly Distribution of Videos and the Relation of Likes & D..
	Weekly Distribution of Videos and the Relation between # o..

YouTube Trending Videos Categorized by Channels

Categories_Main (group)
Howto & Style, News & P..

In / Out of Main YouTube Categories According to Views

Categories	Out
Howto & Style	487M
News & Politics	200M
People & Blogs	573M
Pets & Animals	100M
Science & Techn..	350M
Sports	558M

Total Trending Days
1.00 — 71.00

LOG Transform Dislikes / LOG Transform Likes

Categories
- Howto & Style
- People & Blogs
- Sports
- News & Politics
- Science & Tec..

Number Of Videos / Views / Month of Video Publish Time

- **Summary -** "Monthly Distribution of Trending Videos and the Relation between Likes & Dislikes groupby Channel Categories"

I have split the categories into two according to their Views. Upper-Left Table is the two different categories named (In/Out). "In & Out" consist of 4 and 11 sub-categories respectively. An idea behind categories is to define the thousands of Youtube channel into their main contexts. I came up to this data by using the Python Script.

'''Python 3.0 using Jupyter Notebook'''
import pandas as pd

```python
import numpy as np
from statistics import mode
import os
import pafy

#Open and Read the Data File
data = pd.read_csv(path+"\\USvideos.csv")
Unique_videoid = data.video_id.unique()
print (len(Unique_videoid))

DD= pd.DataFrame(index = np.arange(0,len(Unique_videoid)),columns =
('video_id',      'trending_date',      'title', 'channel_title',      'category_id',
   'publish_time',      'tags', 'views',      'likes', 'dislikes',
   'comment_count',    'thumbnail_link',      'comments_disabled',
   'ratings_disabled',   'video_error_or_removed', 'description',
'total_trending_date', 'Categories'))

for i in range(0,len(Unique_videoid)):
   if Unique_videoid[i][0] != "=":
      rslt_df = data[data['video_id'] == Unique_videoid[i]]
      rslt_df['total_trending_date'] = len(rslt_df)
      url = "https://www.youtube.com/watch?v="+Unique_videoid[i]
      # getting video

      try:
         video = pafy.new(url)

         # getting category of the video
         value = video.category

         rslt_df['Categories'] = value
         df_shape = rslt_df[rslt_df['trending_date'] ==
                     rslt_df['trending_date'].max()].shape
         if df_shape[0] != 2:
            DD.iloc[i] = rslt_df[rslt_df['trending_date'] ==
                  rslt_df['trending_date'].max()]
         else:
            continue
```

```
        except OSError:  # OSError appeared due to the restriction applied
by the Owners of some Youtube channels
            rslt_df['Categories'] = "Inaccessible Data"
            df_shape = rslt_df[rslt_df['trending_date'] ==
                            rslt_df['trending_date'].max()].shape
            if df_shape[0] != 2:
                DD.iloc[i] = rslt_df[rslt_df['trending_date'] ==
                    rslt_df['trending_date'].max()]
            else:
                continue
        except KeyError:  # KeyError appreared because the videos have
been removed by Youtube Authority
            continue

    else:
        continue

DD.to_csv('XXX.csv', index=False)
```

Upper-Right Scatter Plot is the relationship between Likes & Dislikes. Since the values are in large number, I have transform the original data into the LOG_10 scales. Here I also implement the Average with 95% confidence and the trending Line.

Lower-panel is the Line Plot (area under the curve). Here I evaluated the number of Views and Trending Videos in monthly basis along with the categories of channels. I have found a strong declined in the number of trending videos and the Views in the third quartile of year (I.e., July, August, and September).

- **Design -**

In this Dashboard, initially, there is no other alternative of scatter plot to see the relation among two quantitative attributes along with colorbar, that could be readable by color blind as well (I am one of them :)). Secondly, I tried to emphasized the importance of "In/Out" categories of channel based on their Views. Therefore, chart is the good representation that I came upto. Finally, the Area Plot could be replaced by the Line Plot but the disadvantage is that it would be harder to follow individual line for each month.
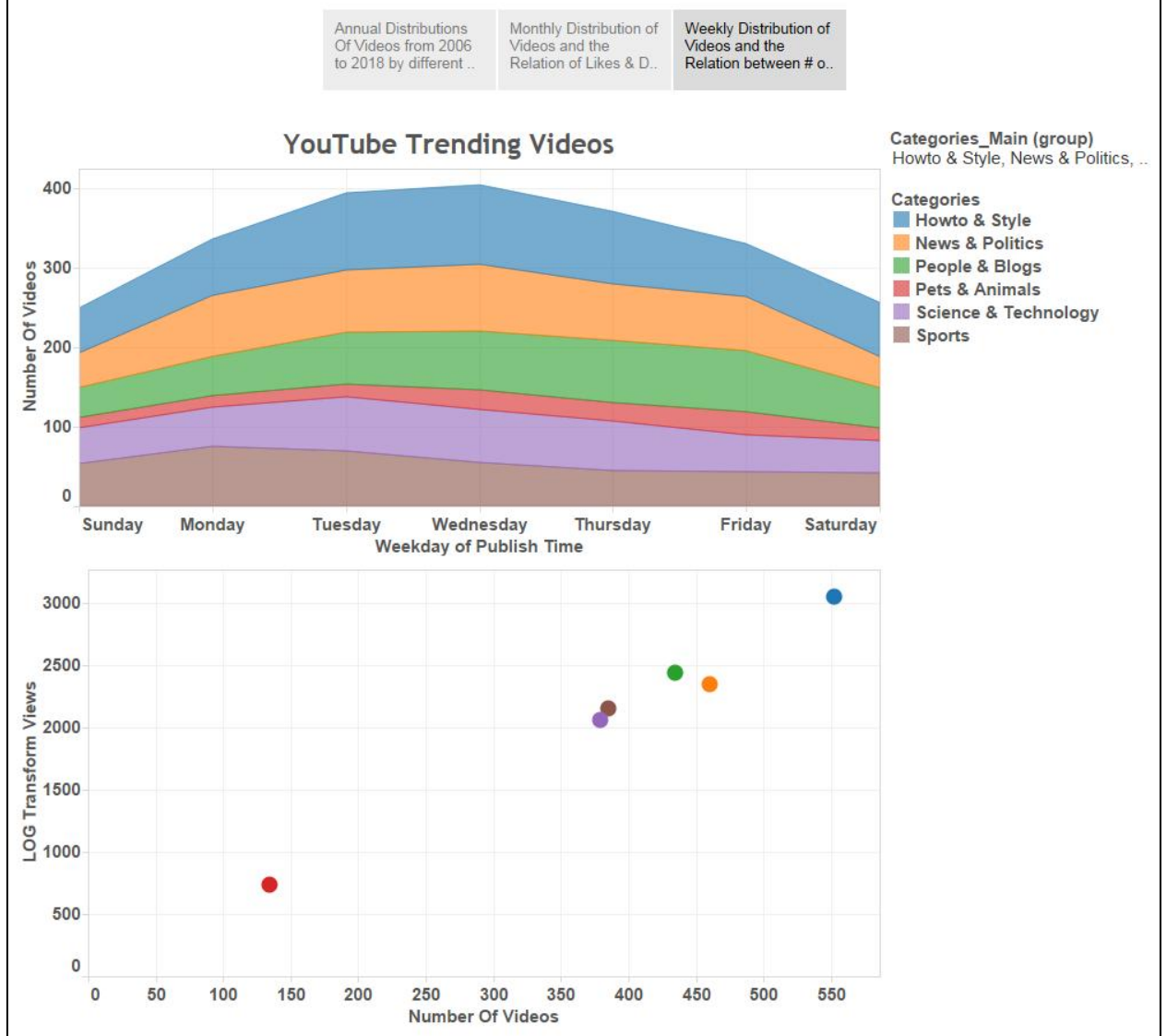
- **Resources -** N/A

3. Insight no.3
- **Link -** I worked on Tableau Desktop. Here I attached Screen Shots of Dashboard 3.

Visualization Of YouTube Trending Videos in USA from 2006 to 2018
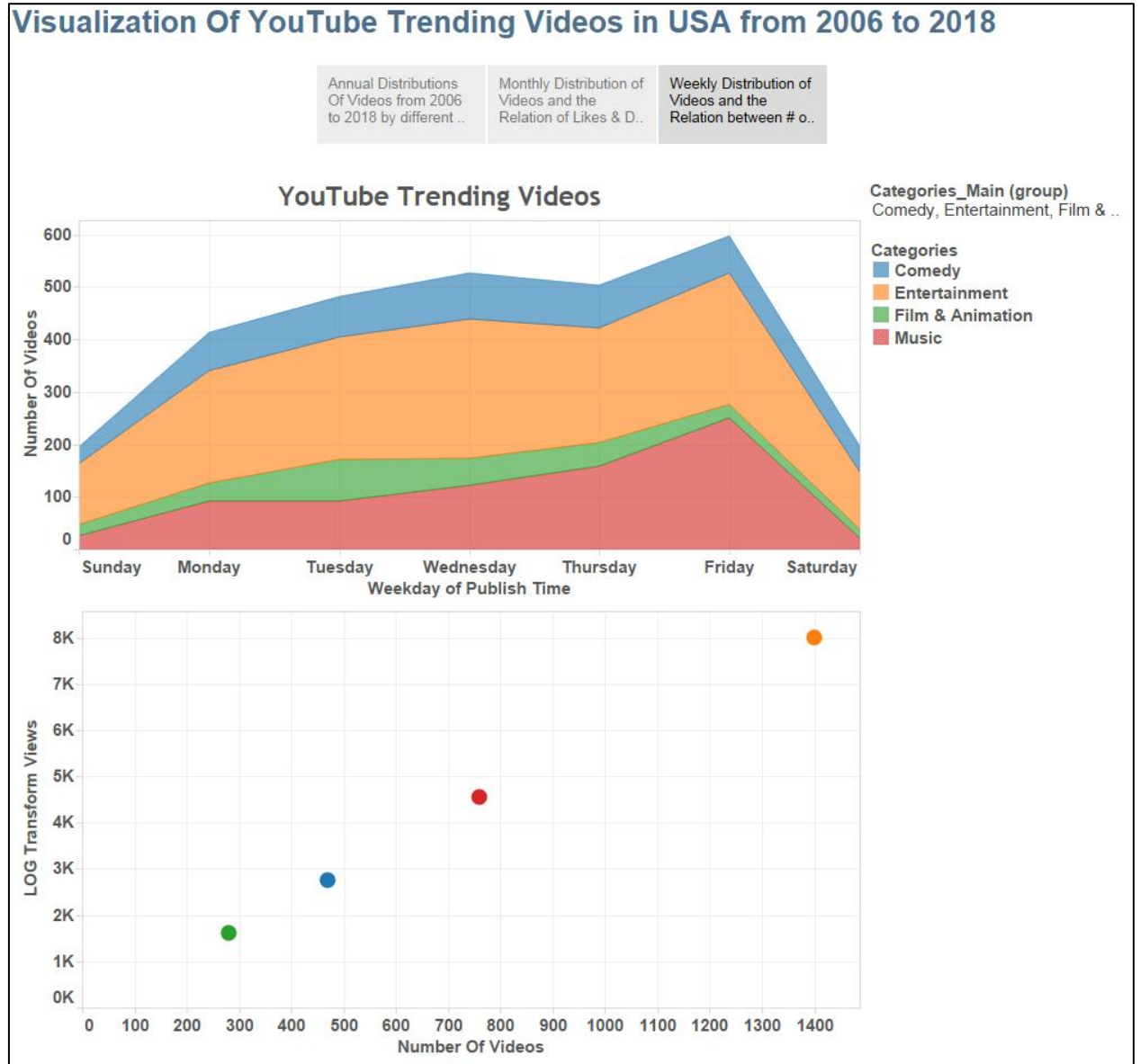
Visualization Of YouTube Trending Videos in USA from 2006 to 2018

- **Summary - "Weekly Distribution of Trending Videos and the Relation between # of videos & # of Views groupby Channel Categories"**

Upper-panel is the Weekly distribution of the trending videos according to the channel categories. I have found the Music, Entertainment, Comedy, and Film&Animation videos are at peak in Friday. Although Education videos get hipe in Monday and Thursday and rest of categories are complex.

Lower-panel is the scatter plot of the relationship between number of trending videos versus number of views. I have found very strong linear relation.

- **Design -**

Same strategies have been applied in Dashboard 3, for the Area-Plot as well as Scatter Plot.

- **Resources -** N/A