Report

One of the main assumptions made on this report is that the total channel views is the most important factor reflecting channel success, more important than for example amount of subscribers.

Conclusions:

Clear conclusions:

* Graph 1-3: There are specific video categories that seems to be a lot more popular than others. This conclusion is based on three graphs video category id to amount of subscribers, video category id to of view count and video category id to amount of channel views.
* Graph 4-6: The more a channel is popular in terms of amount of subscribers, view count or channel view count he will possibly get between about on average 1500 likes, more than 3000 likes seems to be an ineffectual method of increasing popularity.
* Graph 7-9: The more a channel is popular in terms of amount of subscribers, view count or channel view count he will possibly get between about an average 61 comments, more than 110 likes seems to be an ineffectual method of increasing popularity for graph 7 for example.
* Graph 10-12: The more a channel is popular in terms of amount of subscribers, view count or channel view count he will possibly get between about an average from 40,000 seconds of elapsed time or above until it reaches about 76,000 seconds of elapsed time.
* As expected the most a channel is popular in terms of amount of subscribers the less dislikes/views he gets.

Section of this type doesn't generate any new information they just confirm already known knowledge.

Unclear conclusions:

* There is no clear correlation between channel elapsed time and total channel views.
* There is no clear correlation between channel elapsed time and dislike/subscriber.
* There is no clear correlation between comments/views and amount of subscribers.
* There is no clear correlation between comments/views and total channel views.
* There is no clear correlation between channel elapsed time and total channel views.
* There is no clear correlation between likes/views and amount of subscribers or between likes/dislikes and total channel views.

This is contrary to what was expected a successful channel is expected to get a large amount of likes compared to his number of views or compared to his likes/dislikes ratio.

Final conclusion

Although it was expected that the data will bring positive results regarding improving YouTube channels, the only definitive advice that can be given to a YouTube channel owner is that specific few video categories are a lot more popular by a substantial margin compared to the remaining video categories.

The Code template used to create the graphs using jupyter notebook, panda, numpy

and Matplotlib libraries is shown below.

import pandas as pd

import numpy as np

from matplotlib import pyplot as plt

# Load data from CSV file

sample = pd.read\_csv('you.csv')

# Find the 2.5th and 97.5th percentiles of the data

x\_min = np.percentile(sample['videoLikeCount'], 2.5)

x\_max = np.percentile(sample['videoLikeCount'], 97.5)

y\_min = np.percentile(sample['videoViewCount'], 2.5)

y\_max = np.percentile(sample['videoViewCount'], 97.5)

plt.figure(figsize=(8, 6)) # Adjust the size of the figure

# Plot the scatter plot with larger data points and a plus sign marker

plt.scatter(sample['videoLikeCount'], sample['videoViewCount'], s=50, marker='+')

# Set the x-axis and y-axis limits to show only the middle 95% of the data

plt.xlim(left=x\_min, right=x\_max)

plt.ylim(bottom=y\_min, top=y\_max)

# Set the x-axis and y-axis labels

plt.xlabel('videoLikeCount')

plt.ylabel('videoViewCount')

plt.grid(axis='y')

plt.show()