Report

Introduction to the Conclusions:

The analysis conducted on the data has provided valuable insights into various aspects of channel success on YouTube. By examining different graphs, we have gained a deeper understanding of the factors that contribute to a channel's popularity and engagement. In this section, we will summarize the key conclusions derived from the findings, shedding light on the relationship between video categories, subscriber count, view count, elapsed time, likes, dislikes, and comments. These conclusions highlight important patterns and trends that can guide YouTube channel owners in their efforts to enhance their channel's success and maximize audience engagement.

The analysis conducted on the data has provided us with two important graphs, Graph 13 and Graph 14, which offer insights into the relationship between elapsed time and video view count. These graphs are based on a sample of 2000 data points and provide visual representations of the data, allowing us to observe trends and patterns.

In Graph 13, we see a linear regression analysis illustrating how the view count changes over time. Additionally, we have incorporated measures of central tendency, such as the mean, mode, and median, to provide a more comprehensive understanding of the data distribution.

Similarly, Graph 14 showcases a polynomial regression analysis for the same sample, presenting a different perspective on the relationship between elapsed time and video view count. Again, we have included measures of central tendency to enhance our understanding of the data's characteristics and trends.

Conclusions:

Clear conclusions:

1. Graphs 1-3: Certain video categories appear to be much more popular than others. This conclusion is based on three graphs: video category ID to the number of subscribers, video category ID to the view count, and video category ID to the number of channel views.
2. Graphs 4-6: Channels that are popular in terms of the number of subscribers, view count, or channel view count will likely receive an average of around 1,500 likes. Having more than 3,000 likes seems to be an ineffective method of increasing popularity.
3. Graphs 7-9: Channels that are popular in terms of the number of subscribers, view count, or channel view count will likely receive an average of about 61 comments. Having more than 110 likes seems to be an ineffective method of increasing popularity in the case of graph 7, for example.
4. Graphs 10-12: Channels that are popular in terms of the number of subscribers, view count, or channel view count will likely have an average elapsed time ranging from 40,000 seconds or more up to about 76,000 seconds.
5. As expected, the more popular a channel is in terms of the number of subscribers, the fewer dislikes/views it receives. Sections of this type do not generate any new information; they simply confirm already known knowledge.

Unclear conclusions:

1. There is no clear correlation between channel elapsed time and total channel views.
2. There is no clear correlation between channel elapsed time and dislikes per subscriber.
3. There is no clear correlation between comments per view and the number of subscribers.
4. There is no clear correlation between comments per view and total channel views.
5. There is no clear correlation between channel elapsed time and total channel views.
6. There is no clear correlation between likes per view and the number of subscribers, or between likes per dislike and total channel views.
7. This is contrary to what was expected; a successful channel is expected to receive a large number of likes compared to its number of views or its likes-to-dislikes ratio.

Final conclusions:

Based on the clear conclusions derived from the analysis:

1. Certain video categories exhibit significantly higher popularity compared to others, as observed in Graphs 1-3, where video category ID correlates with the number of subscribers, view count, and channel views.
2. Channels that have a high number of subscribers, view count, or channel view count tend to receive an average of around 1,500 likes, suggesting a potential saturation point beyond 3,000 likes, as shown in Graphs 4-6.
3. For channels with a substantial number of subscribers, view count, or channel view count, the average number of comments hovers around 61. However, surpassing 110 likes may not effectively contribute to increasing popularity, as indicated by Graphs 7-9.
4. Channels that enjoy popularity in terms of subscribers, view count, or channel view count typically have an average elapsed time ranging from 40,000 seconds or higher up to approximately 76,000 seconds, as demonstrated in Graphs 10-12.
5. Linear regression analysis on a sample of 2000 data points (Graph 13) and polynomial regression analysis on another sample of 2000 data points (Graph 14) illustrate the relationship between elapsed time and video view count. These analyses provide regression equations that can be used to estimate the effectiveness of broadening the channel's success by inputting an elapsed time value into the equation.
6. As anticipated, channels with a higher number of subscribers tend to receive fewer dislikes/views, indicating a negative correlation between popularity and dislikes/views.
7. In summary, these findings provide insights into the dynamics of channel success on YouTube. It is crucial to choose video categories wisely, be mindful of saturation points for likes, and recognize that excessive likes may not contribute significantly to popularity.

The code template used to create the graphs using Jupyter Notebook, Pandas, 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()