Visualisation of top 200 YouTuber Content Creators

Shreya Kathuriya
480291367

***Abstract*—This report discusses the various visualization methods of YouTube data. It also lists various audience for such data and how the data will be useful to them. Some mistakes for representation of the data are also discussed.**

I. INTRODUCTION TO DATA

YouTube is an American internet based video sharing and online entertainment platform. It is owned by Google and is the second most visited website. As of 2022, it has 2.6 billion monthly active users who watch over 1 Billion hours of videos per day. And, over 500 hrs of videos are uploaded per minute. [1]

Now, with the growth of YouTube channels, it has become a platform for businesses to sponsor the channels in return of marketing. YouTube creators also earn through AdSense based on the amount of ads watched by users on their video. [2] A lot of these depend on how many views, subscribers, likes and comments the channel attracts.

The dataset 'top200youtubers', hereafter referred as the dataset, represents top 200 YouTube channels and information about them. The data is taken from Kaggle. It is a cleaned dataset. It is owned by Syed Jaffer and is under CC0: Public Domain License, visible to Public. The source of the data a website called 'socialbook.io' and the method of collection is through scraping and request module. [3]

There are 22 attributes. Some attributes are as follows:

1. Country: Name of the country
2. Channel Name: Name of the channel
3. Category: Category the channel uploads videos about apart from main topic.
4. Main Video Category: Main category the channel uploads videos about
5. Username: Username of the YouTube channel
6. Followers: Number of followers
7. Main topic: Main topic the channel uploads videos about
8. Likes: Total Likes
9. Engagement Rate: Rate of engagement with the users.
10. View: Total Views

II. AUDIANCE

YouTube, YouTube Employees

YouTube can identify their top ranking content creators using this data. YouTube employees can use the data for their algorithm to make recommendations to users depending on country, topic and category of videos the channel uploads

YouTube and Employees can use this data in YouTube Studio where they provide stats to creators about their channel

Youtubers/ Creators

Creators can see their ranking and stats as compared to others channels, helping them identify their position and competition. This data can help them plan and identify how to grow their channel

Google Ads

YouTube uses Google's AdSense program to pay their creators. YouTube monetizes creator's accounts with AdSense. Depending on their subscriber count, country of video upload or live video streaming, viewer count and more, they provide channel memberships, merch self, super chat and stickers and YouTube premium revenue. [4]

Independent Businesses

Independent businesses promote their products and services through YouTube channels with high traction. This data can help them identify which creator to sponsor. Information like country of most viewership, category, subscribers, view counts and more can help them make their decision.

III. DATA TYPES

Typically the data is depicted/conveyed in form of charts, numbers, stats, infographic or maps. Data such as followers, likes, engagement rate, engagement rate 60 days, views, average views, average 1/3/7/14/30/60 day, views and average comments are numerical and represent quantity. These are also continues values. The quality is ratio since there can be an absolute 0 for any of these attributes. This continuous data is best represented as a histogram. [5] This can be made to compare values for different channels, growth of individual channels over the 1/3/7/14/30/60 day period.

Data such as category, main video category, country, main topic and more topics are discrete. These can be used to categorise data and visualise stats for which category or topic is more watched in which countries. A map representation can also be made to depict which category of video is most popular in different countries.

The engagement, subscribers and views data can be represented as ranges making them ordinal. It is a structural database, which makes it easy to visualize.

IV. ANALYSIS QUESTIONS

YouTube, YouTube Employees

a. What video category is most popular in the countries?
b. How many subscribers to channels have?
c. How are views and subscriber count related in engagement?
d. What categories do the channels make videos on?

YouTubers/ Creators

a. What are the engagement rate for their channel over time?
b. What are the competitive channels' engagement rate/ likes/ comments/ followers?
c. What video categories do their competitive channels upload videos on?
d. What are the most viewed categories/ topics of videos in the region they want to expand?

Google Ads
a. What is the engagement rate of the video?
b. How many followers do the accounts have?

Business owners
a. How many followers does the YouTube channel have?
b. What is the main topic the YouTube channel uploads videos about?
c. Which YouTubers around the world are popular in the category the product/service their business sells
d. Does a greater number of subscribers mean more views and engagement?
e. What is the engagement rates of the YouTube channels?

V. VISUALIZATION

A. *Some Common Mistakes*

Some mistakes that can be made with the visualizations are as follows:

- Using the wrong type of charts to demonstrate data.
- Using too many variables in one graph. An implication of this is that the visualisation leads to a long legend which makes it complicated to read the data.
- It is very messy to chart all 200 YouTube channels in one graph to compare their stats. It will be better to look at smaller groups of channels like top 10 or bottom 10, category or country wise etc.
- While making comparisons, the scale should be common for both. For example, engagement rate vs view counts. Engagement rate has values between 0 and 1 whereas view counts has whole number values. If a grouped plot is made, the difference in values will make the engagement rate very tiny bin size compared to views.
- Not adhering to KPI and going out of scope to show more data will overwhelm the audience, defeating the purpose of visualization.

The above mentioned can make it hard to read the visualised data. This can further lead to wrong conclusions and business decisions.

B. *Symbolic representation*

The KPI is so to depict the contrast between followers and views of the top channels based on their engagement rates along with the main category the channels make videos in.

For this, the imposition will be a diagram. 3 of them can be stacked, depicting engagement, followers and views each. It will be an orthogonal construction. It was decided to not make a grouped graph since that will take more space to depict one channel and less channels will be able to be seen in the visualization. It will also make it hard to compare just the views or engagement or followers of the channel. Having different graphs for each will allow that as well as comparing them within the channel itself.

Channel names are differentiative and qualitative where are the number of views, followers and engagement rate is quantitative. The graph will be ordered based on engagement rate. They will also be the retinal variables.

Color will be the visual variable, representing the main video category for each channel.

This visualization uses the following attributes:

1. Engagement rate
2. Followers
3. Views
4. Channel names
5. Main video category

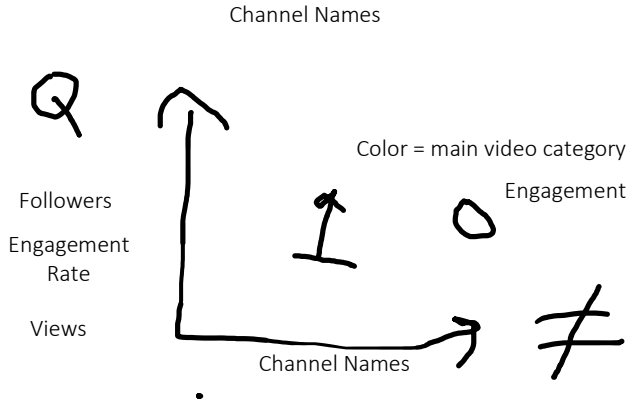

Figure 1

B. *Visualizationymbolic representation*

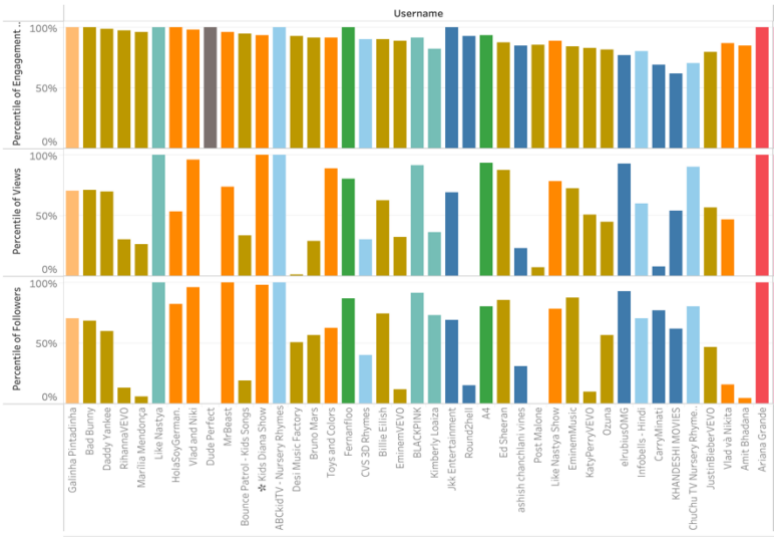

Figure 2.1

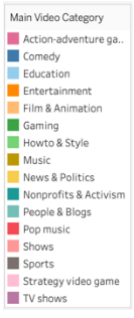

Figure 2.2

Figure 2.1 represents the visualization derived from the symbol in Figure 1. Figure 2.2 is the legend for that visualization.

Using Tableau, the columns represent usernames of YouTube channels and rows to depict the quantitative data. The data was ordered according to engagement rate. The retinal variable was main category of videos. Upon completion of thus, it was noticed that the scale for the 3 quantitative variables was very different. So, the visualized data was unfairly depicted. To rectify that, the values were changed to percentile, making it fair across the three measurements.

C. Alternative Symbolic representation

In the alternate visualization, it was chosen to make to be Engagement rate, Followers and Views, the retinal variables. The imposition is still Diagram where X axis has main video category and channel name. On the X axis, channel names are to be color coded according to their main video category and grouped according to that. The data being represented and the KPI is the same. Engagement rate, Followers and Views can still be compared, and the channel's main video category can still be seen.

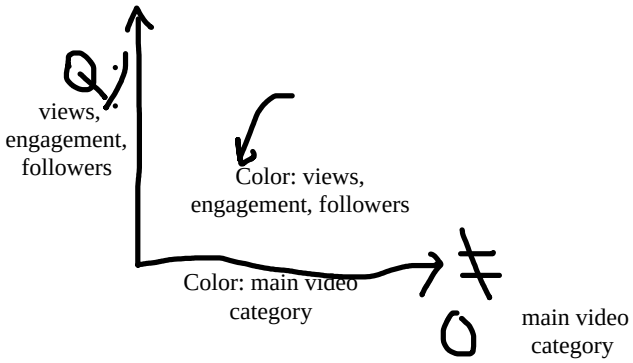

Figure 3

D. Alternative Visualization

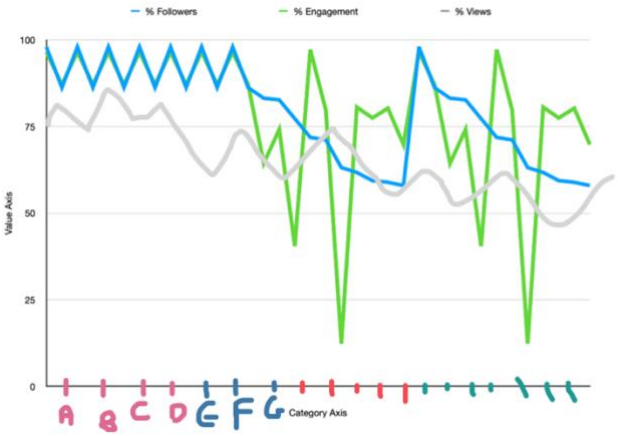

Figure 4

Figure 4 represents the visualization derived from the symbol Figure 3. Figure 2.2 is the legend for that visualization.

X axis was chosen to represent the YouTube channel name as well as the main video category. This video category is represented by numbers. Y axis shows the quantities in percentage, which makes it common measure for all three measurable values. XY plane hence represents the % followers, % engagement and % views for all channels. These are also differentiated by color. In the X axis, the idea is to group each category together and show the YouTube channels (A, B, C,D/ E,F,G/ and so on).

REFERENCES

[1] M. Mohsin, "Oberlo," 17 05 2022. [Online]. Available: https://au.oberlo.com/blog/youtube-statistics. [Accessed 11 09 2022].

[2] "AdSense for YouTube," YouTube, [Online]. Available: https://support.google.com/youtube/answer/11602441?hl=en#:~:text=AdSense%20is%20Google's%20program%20that,account%20from%20within%20YouTube%20Studio.v. [Accessed 13 08 2022].

[3] S. Jafer, "Top 200 Youtubers Data (cleaned)," Kaggle, [Online]. Available:https://www.kaggle.com/datasets/syedjaferk/top-200-youtubers-cleaned. [Accessed 13 08 2022].

[4] "How to earn money on YouTube," Google Adsense, [Online]. Available: https://support.google.com/adsense/answer/72857?hl=en. [Accessed 2022 08 15].

[5] J. Frost, "Guide to Data Types and How to Graph Them in Statistics," [Online]. Available: https://statisticsbyjim.com/basics/data-types/. [Accessed 17 08 2022].