# Assignment 2 Part 2

**Professor**: Ming Jiang

**Student Name**: Deepak Rajput **Program**: MS Applied Data Science

Dataset Analysis

Please select a publicly accessible dataset, analyse data characteristics, and answer the following questions.

1. What dataset are you studying? What's the link to access the dataset?

   Answer: Dataset is **Most Subscribed 1000 Youtube Channels**
   Link for the same is
   https://www.kaggle.com/datasets/themrityunjaypathak/most-subscribed-1000-youtube-channels

2. What task (e.g., sentiment analysis, image object classification) did data collectors focus on when building this dataset?

   Answer:
   Data collected in the dataset is from 1970 to 2022 and priority in arrangement is given to number of subscribers as of 2022.
   Use API & web scrape to download and build data, then clean data by avoiding null and inappropriate values, then categories it to complete the task.
   Quantitative research methodologies and variables measured for analysis and visualised results through various graphs.
   Primary focus of data collectors is to find out the top subscribed YouTube channels and divide them in categories and subscribers so they can further derive answers for business questions like:
   i) How many hours of video are uploaded every minute?
   ii) Who is the highest earner?
   iii) Influence and importance of YouTube partner program

3. What features does this dataset contain?

   Answer: Data set contains: a) Rank, b) Youtube Channel, c) Subscribers, d) Video Views, e) Video Count, f) Category, g) Started (year)

4. Are any social agents associated with this dataset? How's their relationship with this dataset?

   Answer: Social agents are: a) You tube associates or technician, b) Channel's owner and content creators and their team, c) Subscribers and viewers, d) Ad agencies

Here primary agents are a) You tube associates or technician, b) Channel's owner and content creators which have direct and financial relationship whereas secondary agents are c) Subscribers and viewers, d) Ad agencies serving indirect relationships

5. Is there any possibility of data biases involved in this dataset? Please list some key issues (e.g., data representativeness; biases during data collection).

Answer: Yes, may be some data biases involved in this dataset as some top ranked channels not having appropriate data like number of views and video count table snap 1 is given below. Also, data collectors mentioned about earnings and YouTube partner program but dataset not showing any financial figures so hard to detect the relationship without figures.

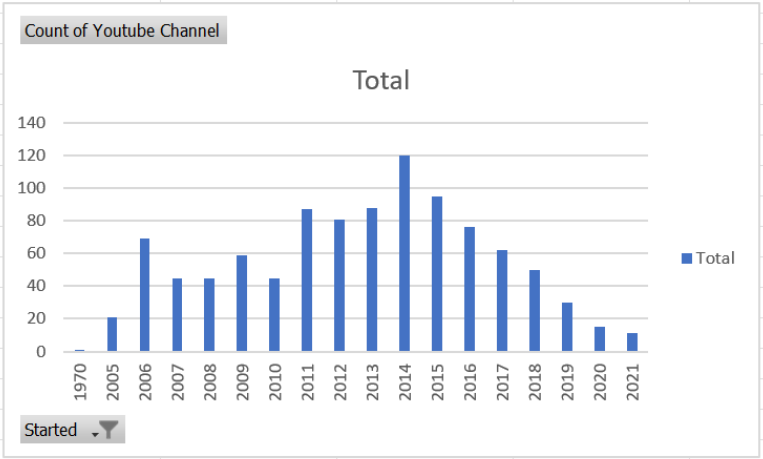
From second snap one can figure out that to be among top ranked channels required several years and continuously need to create new contents and feed it on YouTube.

Categories analysis showing that on YouTube entertainer, music and people & blogs are more effective in getting subscribers and views rather than general category content creator like comedy, sports, trailer, travel and events and education.

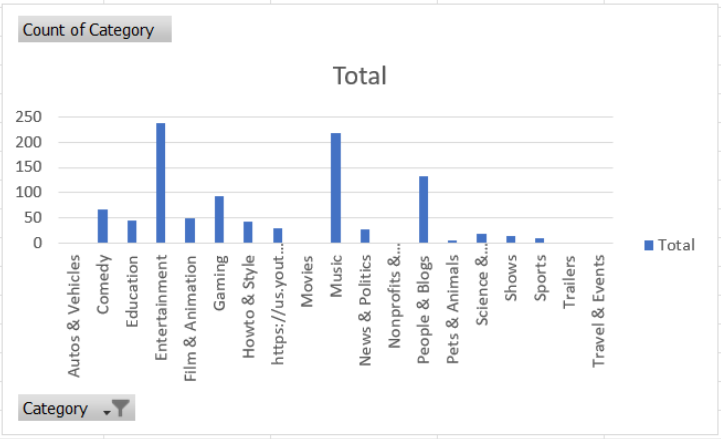
While last 2 snaps showing direct relationship between number of videos and number of views, more video delivers more views, it's showing quantitative relationship.

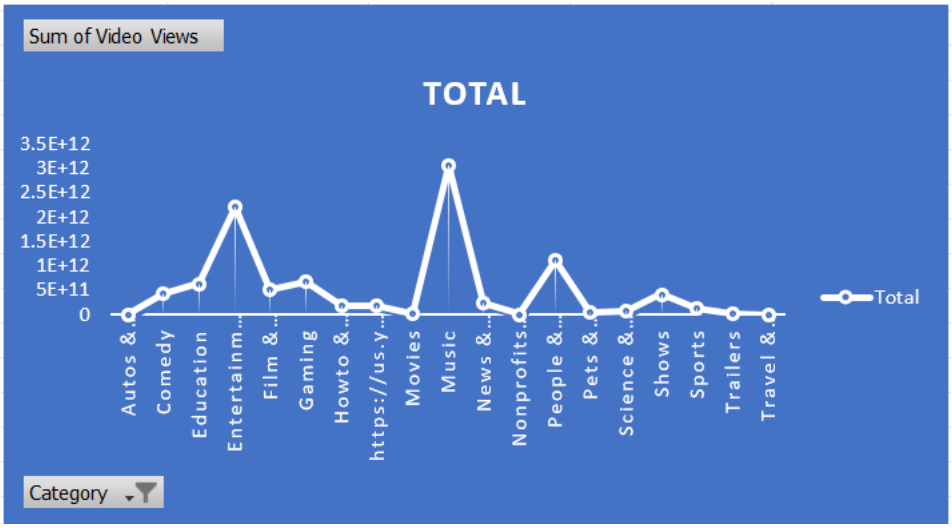
| Rank | Youtube Channel | Subscribers | Video Views | Video Count | Categor | Started |
|---|---|---|---|---|---|---|
| 2 | YouTube Movies | 16,10,00,000 | 0 | 0 | Film & Ani | 2015 |
| 6 | Music | 11,80,00,000 | 0 | 0 | https://us | 2013 |
| 10 | Gaming | 9,33,00,000 | 0 | 0 | https://us | 2013 |
| 18 | Sports | 7,51,00,000 | 0 | 0 | https://us | 2013 |
| 92 | News | 3,63,00,000 | 0 | 0 | https://us | 2013 |
| 155 | Popular on YouTube | 2,93,00,000 | 0 | 0 | https://us | 2013 |
| 358 | Minecraft - Topic | 1,95,00,000 | 0 | 0 | https://us | 2013 |
| 551 | Live | 1,57,00,000 | 0 | 0 | https://us | 2015 |
| 972 | Machinima | 1,16,00,000 | 0 | 0 | Film & Ani | 2006 |

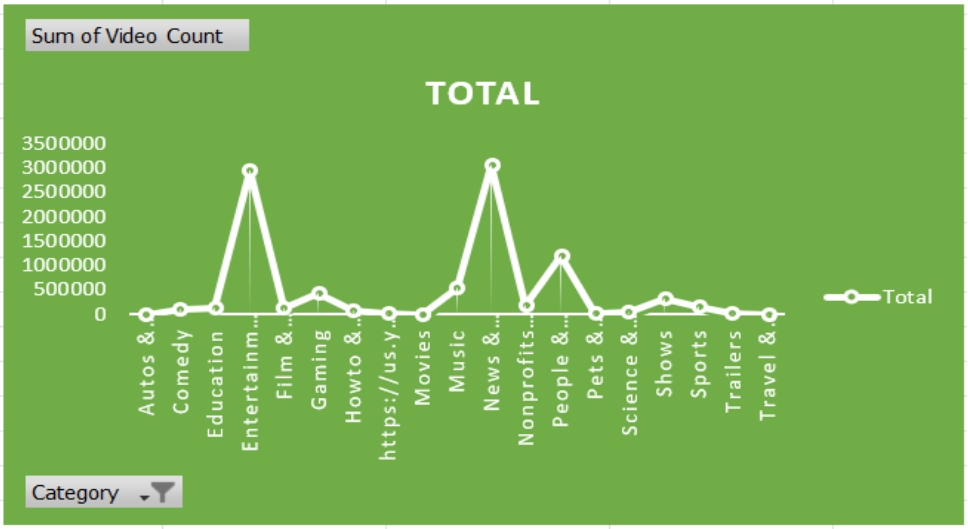
| Row Labels | Count of Youtube Channel |
|---|---|
| 1970 | 1 |
| 2005 | 21 |
| 2006 | 69 |
| 2007 | 45 |
| 2008 | 45 |
| 2009 | 59 |
| 2010 | 45 |
| 2011 | 87 |
| 2012 | 81 |
| 2013 | 88 |
| 2014 | 120 |
| 2015 | 95 |
| 2016 | 76 |
| 2017 | 62 |
| 2018 | 50 |
| 2019 | 30 |
| 2020 | 15 |
| 2021 | 11 |
| **Grand Total** | **1000** |




| Row Labels | Count of Category |
|---|---|
| Autos & Vehic | 1 |
| Comedy | 68 |
| Education | 45 |
| Entertainment | 238 |
| Film & Animat | 50 |
| Gaming | 94 |
| Howto & Style | 43 |
| https://us.you | 30 |
| Movies | 2 |
| Music | 217 |
| News & Politic | 27 |
| Nonprofits & / | 2 |
| People & Blog | 132 |
| Pets & Animal | 6 |
| Science & Tecl | 18 |
| Shows | 14 |
| Sports | 10 |
| Trailers | 2 |
| Travel & Event | 1 |
| **Grand Total** | **1000** |




Sum of Video Views
TOTAL
3.5E+12
3E+12
2.5E+12
2E+12
1.5E+12
1E+12
5E+11
0
Autos &...
Comedy
Education
Entertainm...
Film &...
Gaming
Howto &...
https://us.y...
Movies
Music
News &...
Nonprofits...
People &...
Pets &...
Science &...
Shows
Sports
Trailers
Travel &...
Total
Category


Sum of Video Count
TOTAL
3500000
3000000
2500000
2000000
1500000
1000000
500000
0
Autos &...
Comedy
Education
Entertainm...
Film &...
Gaming
Howto &...
https://us.y...
Movies
Music
News &...
Nonprofits...
People &...
Pets &...
Science &...
Shows
Sports
Trailers
Travel &...
Total
Category