

Youtube Dataset Project

By: Mai-Anh Nguyen

[Link to report](#)

Section 1: Introduction

[Link to Dataset](#)

rank	Youtuber	subscribers	video views	category	Title	uploads	Country	Abbreviation	channel_type
1	T-Series	245000000	2.280000e+11	Music	T-Series	20082	India	IN	Music
2	YouTube Movies	170000000	0.000000e+00	Film & Animation	youtubemovies	1	United States	US	Games
3	MrBeast	166000000	2.836884e+10	Entertainment	MrBeast	741	United States	US	Entertainment
4	Cocomelon - Nursery Rhymes	162000000	1.640000e+11	Education	Cocomelon - Nursery Rhymes	966	United States	US	Education
5	SET India	159000000	1.480000e+11	Shows	SET India	116536	India	IN	Entertainment
...
991	Natan por Aĩ	12300000	9.029610e+09	Sports	Natan por Aĩ	1200	Brazil	BR	Entertainment
992	Free Fire India Official	12300000	1.674410e+09	People & Blogs	Free Fire India Official	1500	India	IN	Games
993	Panda	12300000	2.214684e+09	NaN	HybridPanda	2452	United Kingdom	GB	Games
994	RobTopGames	12300000	3.741235e+08	Gaming	RobTopGames	39	Sweden	SE	Games
995	Make Joke Of	12300000	2.129774e+09	Comedy	Make Joke Of	62	India	IN	Comedy

- I chose the "2023 Global Youtube" dataset on Kaggle
- It contains statistics on the top global Youtube videos of 2023, including each video's category, channel rank, title, subscribers, etc.

Section 2: Exploratory Data Analysis

To begin, I wanted to focus on the category and amount of video views. To do this, I grouped the dataset by channel category and aggregated by “mean”

	category	rank	subscribers	video views	uploads	video_views_rank	country_rank	channel_type_rank	video_views_for_the_last_30_days	lowest_month
0	Autos & Vehicles	560.500000	1.785000e+07	7.501729e+09	1898.500000	8.415000e+02	11.000000	45.500000	2.732165e+08	
1	Comedy	523.623188	2.012319e+07	7.972649e+09	1119.753623	4.058789e+05	487.104478	346.863636	1.853106e+08	
2	Education	477.311111	2.654222e+07	1.548032e+10	3142.866667	3.614044e+05	330.785714	703.466667	1.858635e+08	
3	Entertainment	512.551867	2.140332e+07	1.048854e+10	11908.481328	6.167031e+05	450.028037	889.189655	1.796649e+08	
4	Film & Animation	361.869565	2.858478e+07	1.183527e+10	2799.652174	7.925529e+05	811.108108	1252.688889	1.930848e+08	
5	Gaming	538.191489	2.085213e+07	7.634456e+09	4313.414894	6.010272e+05	319.621951	891.434783	7.163692e+07	
6	Howto & Style	594.775000	1.939000e+07	5.045439e+09	1673.350000	4.815907e+05	309.058824	436.432432	5.081039e+07	
7	Movies	241.000000	2.565000e+07	7.691628e+09	3553.000000	8.400000e+02	53.000000	16.500000	1.731005e+08	
8	Music	446.000000	2.571782e+07	1.545286e+10	2335.509901	4.041852e+05	267.815217	627.277778	1.790282e+08	
9	News & Politics	506.038462	2.063077e+07	1.039969e+10	112484.384615	3.250163e+05	457.846154	527.000000	1.607470e+08	
10	Nonprofits & Activism	314.000000	2.775000e+07	5.431456e+09	102912.000000	1.537500e+03	27.000000	1.500000	9.759050e+07	
11	People & Blogs	534.500000	2.105606e+07	9.589327e+09	9190.780303	7.433466e+05	405.548077	1042.171875	1.432201e+08	
12	Pets & Animals	562.750000	1.810000e+07	1.121669e+10	4451.500000	1.013181e+06	70.000000	54.333333	2.006161e+08	
13	Science & Technology	532.000000	1.861765e+07	3.939809e+09	2114.058824	7.128223e+05	632.933333	302.625000	4.654797e+08	

- Average views for auto & vehicle videos: Seven billion five hundred million
- Average views for comedy videos: 8 billion
- Average views for education videos: 15 billion
- Average views for entertainment videos: 10 billion
- Average views for film & animation videos: 12 billion
- Average views for gaming videos: Seven billion six hundred million
- Average views for how-to & style videos: 5 billion
- Average views for movie videos: Seven billion seven hundred million
- Average views for music videos: 15 billion
- Average views for news & politics videos: 10 billion
- Average views for nonprofit & activism videos: Five billion four hundred million
- Average views for people & blog videos: Nine billion six hundred million
- Average views for pets & animals videos: 11 billion
- Average views for science & technology videos: Three billion nine hundred million
- Average views for show videos: 34 billion
- Average views for sport videos: 13 billion
- Average views for trailer videos: 17 billion
- Average views for travel & events videos: Three billion one hundred million

Still focusing on the category and video views columns, I grouped the dataset by channel category again and aggregated by “median” this time. I wanted to see how the video views would compare

	category	rank	subscribers	video views	uploads	video_views_rank	country_rank	channel_type_rank	video_views_for_the_last_30_days	lowest_monthly_earning
0	Autos & Vehicles	560.5	17850000.0	7.501729e+09	1898.5	841.5	11.0	45.5	273216500.0	68300.
1	Comedy	530.0	17100000.0	5.863457e+09	554.0	1568.0	62.0	36.0	41672000.0	8300.
2	Education	443.0	18800000.0	7.876741e+09	847.0	687.0	59.0	28.0	103150000.0	20900.
3	Entertainment	531.0	17100000.0	7.886440e+09	965.0	956.0	52.5	125.0	59184000.0	11400.
4	Film & Animation	317.0	22200000.0	8.660225e+09	1266.0	661.5	50.0	51.0	68099000.0	14550.
5	Gaming	580.5	16300000.0	6.318769e+09	1708.5	1375.5	26.0	49.0	40425000.0	9800.
6	Howto & Style	651.0	15300000.0	3.690895e+09	889.5	2752.5	105.5	30.0	23977000.0	5700.
7	Movies	241.0	25650000.0	7.691628e+09	3553.0	840.0	53.0	16.5	173100500.0	28400.
8	Music	419.5	19500000.0	1.244479e+10	252.5	385.0	39.5	107.0	90450000.0	22100.
9	News & Politics	560.5	16550000.0	1.043626e+10	86267.0	759.5	63.0	16.0	123330500.0	30850.
10	Nonprofits & Activism	314.0	27750000.0	5.431456e+09	102912.0	1537.5	27.0	1.5	97590500.0	24400.
11	People & Blogs	509.5	17550000.0	6.456770e+09	681.0	1419.5	74.5	50.0	48947000.0	11250.
12	Pets & Animals	569.0	17550000.0	9.824589e+09	875.5	935.0	81.0	2.0	23701500.0	5551.
13	Science & Technology	520.0	17300000.0	3.606912e+09	1022.0	2583.0	84.0	11.5	49404500.0	11500.
14	Shows	294.0	23100000.0	1.393002e+10	3091.0	257.0	14.5	26.0	204624000.0	33800.

- Median views for auto & vehicle videos: Seven billion five hundred million
- Median views for comedy videos: Five billion nine hundred million
- Median views for education videos: Seven billion nine hundred million
- Median views for entertainment videos: Seven billion nine hundred million
- Median views for film & animation videos: Eight billion seven hundred million
- Median views for gaming videos: Six billion three hundred million
- Median views for how-to & style videos: Three billion seven hundred million
- Median views for movie videos: Seven billion seven hundred million
- Median views for music videos: 12 billion
- Median views for news & politics videos: 10 billion
- Median views for nonprofit & activism videos: Five billion four hundred million
- Median views for science & technology videos: Three billion six hundred million
- Median views for show videos: 14 billion
- Median views for sport videos: Seven billion one hundred million
- Median views for trailer videos: 17 billion
- Median views for travel & events videos: Three billion one hundred million

Section 3: Exploratory Data Visualization

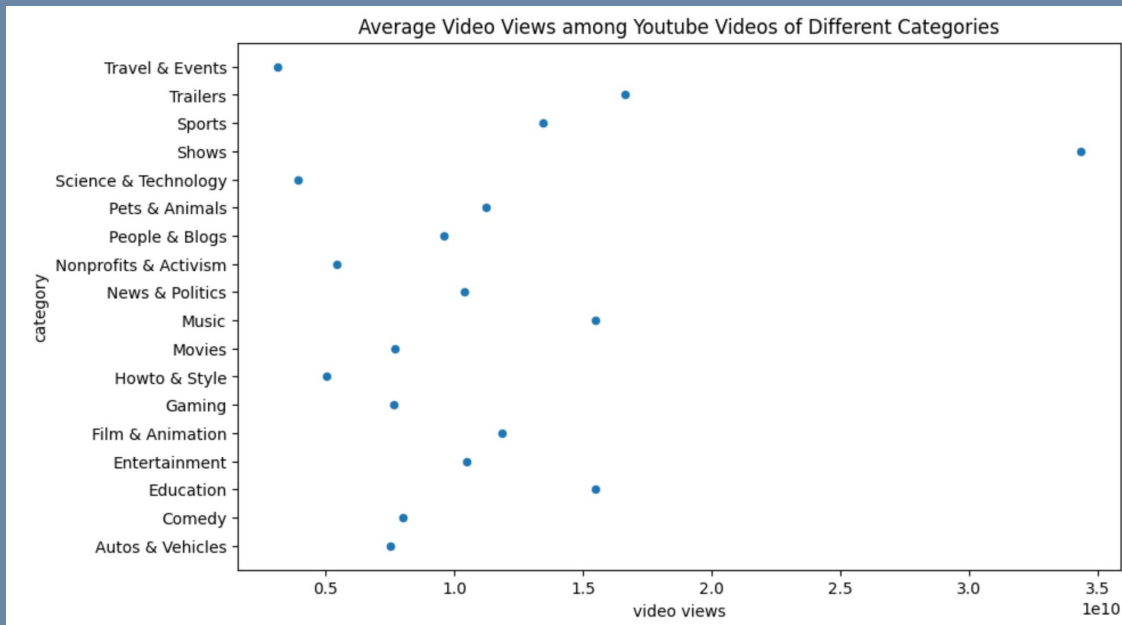
What this graph shows:

This graph shows the amount of video views each channel category got on average in 2023

Please note that although video views is the dependent variable and should be on the y-axis, it's on the x-axis due to spacing

Observations

- The “shows” category has the highest average number of video views of 34 billion
- The “trailers” category has an average number of video views of 17 billion



Section 4: Data Science

Question: Is the ranking of channel types correlated with the number of video views?

1.

I decided to find the correlation coefficient since it measures the amount of the linear association between the x and y variables.

2.

Results:

Negative correlation coefficient of -0.08

3.

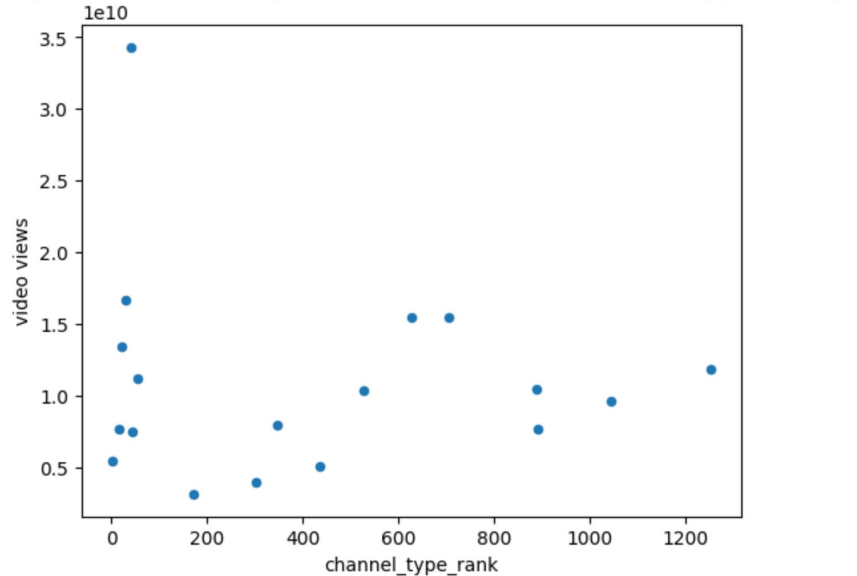
The video views for the "shows" category appeared to be an outlier as it seemed to have too great of an effect on the correlation coefficient
→ False data?

4.

I filtered the data frame to remove the video views for the "shows" category

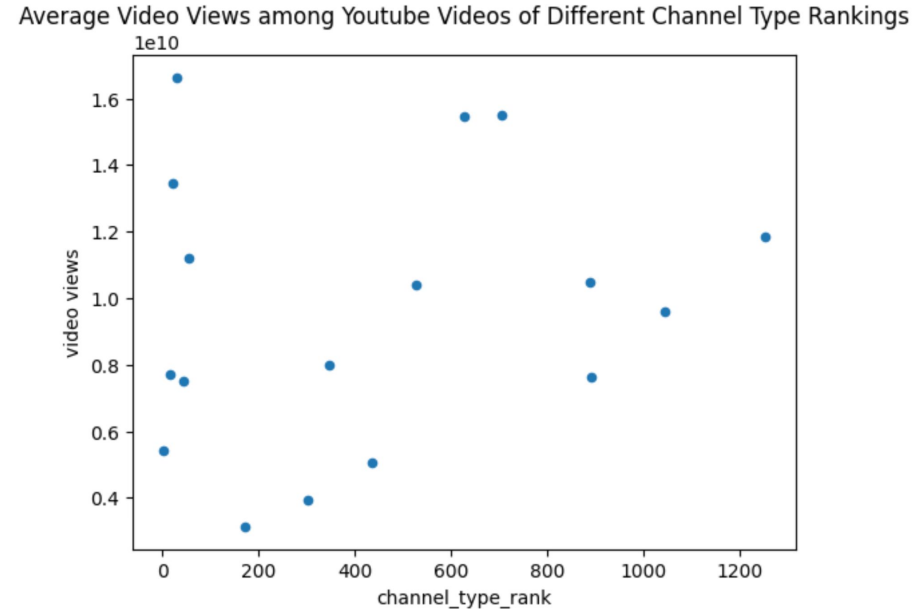
New correlation coefficient: 0.2

With the “show” channel type data included in the data frame & graph



Without the “show” channel type data included in the data frame & graph

- Data is closer together



Section 5: Conclusion

- While each test had different r-values, both findings showed little correlation between channel type rank and video views
 - You cannot predict x based on y
- Both numbers were close to 0 rather than -1 or 1 & showed no linear pattern



Thank you!

Any questions?

