Paradigm Academic Press
Studies in Social Science & Humanities
ISSN 2709-7862
JAN. 2023 VOL.2, NO.1

Exploring the Preferences of US YouTube Users and Factors Related to YouTube Uploader's Revenue

Shuangdan Ni[1]

[1] Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, 59200, Malaysia

Correspondence: Shuangdan Ni, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, 59200, Malaysia.

doi:10.56397/SSSH.2023.01.08

Abstract

YouTube is a mainstream media that is popular with the general public. It has a powerful yet special force that attracts almost all young people who come into contact with it. YouTube permeates every aspect of people's lives, education, daily life, work, leisure, and entertainment. Everyone can be a professional uploader. It is very important to know how to be a good uploader and what the important elements are related to generating revenue from YouTube. This article uses the top 1000 channels and the top 200 daily YouTube video's data from September 2020 to January 2022 in the United States to analyze popular categories of channels, and use the YouTube videos and their related data from September 2020 to January 2022 from Ken Jee to analyze uploader characteristics. The dataset of this article is based on YouTube API and the uploader's release. CRISP-DM method is applicated to exploring the data in Jupyter using Python. In addition, Pearson correlation analysis is used to analyze revenue-related metrics. This paper contains both qualitative and quantitative analyses. In this paper, we found that among the top 1000 most popular uploaders, Entertainment, and Music types are the most popular, with 241 and 222 respectively. The Autos & Vehicles and Travel & Events categories are the least represented, with only 1 on the list. This article also finds out Entertainment, Music, People & Blogs and Gaming type of videos users are most attracted to, the core audience, and their interests. Have strong positive relation with revenue. In addition, watch time (hours), subscribers lost, likes, shares, and dislikes are strongly positively correlated with revenues. Comments are moderately positively correlated with revenues. RPM and CPM almost have no correlation with revenue. This study offers a statistical analysis of the content genre based on a number of categories and details of YouTube users' interests. This study also demonstrates the variables that might have an impact on YouTube uploaders' revenue. So that uploaders, researchers, and investors can acquire a rough idea of the impact on YouTube.

Keywords: YouTube uploaders, CRISP-DM, data analyze, social media, YouTube

1. Background

Social networking sites, e-commerce, search engines, and other Internet industries are penetrating each other, integrating and growing together, and having a relatively stable scale of users (Sara Fraccastoroa, Mika Gabrielssona, Ellen, & Bolman Pullinsb, 2021). According to an article published by Statista in 2022, Facebook, YouTube, WhatsApp, Instagram, WeChat, and TikTok are some of the most popular social networking platforms today (S. Dixon, 2022). In this paper, we will focus on YouTube because it is stable and young enough, with a wide variety of channels and user data. YouTube is a video-sharing website of Google, registered on February 14, 2005. It is a huge online social platform based on a combination of various interpersonal elements such as interests, education, beliefs, and professions. The slogan of the site is "Broadcast Yourself" (Wikipedia, YouTube). They have indeed done so. There is a new wave of collaborative studies emerging from various disciplines, ranging from computer science to behavioral sciences (Giglietto, Rossi, & Bennato, 2012). YouTube is not only a sharing platform but also an effective search engine. YouTube has many crazy fans, such as Green and Burgess. They stated in their book, "YouTube is considered to be part of the mainstream media landscape, and a force to be reckoned with in contemporary popular culture." Due to COVID-19, many people are staying home, and the number of people on YouTube has increased dramatically (Yuksel bahar & Cakmak Kubra, 2020). As of July 2022, YouTube has 2.475 billion users worldwide, which means that about 31% of the people on the planet use YouTube, with the most active users from India at 225 million, followed by the U.S. at 197 million. The largest user group is 25-34 years old, with 20.5%, followed by 35~44 years old, accounting for 15.5%. (Source: datareportal.com). YouTube uploaders are the primary players in the YouTube universe (Ding, et al., 2011). It is obvious that YouTube uploaders in the United States have a great opportunity to grow and compete globally in content creation. The data source for this article has three parts. One is the top two hundred daily YouTube videos and their related data from September 2020 to January 2022 in the United States from YouTube API. Another is the YouTube videos and their related data from September 2020 to January 2022 from Ken Jee, an uploader in data science and sports analytics. He released his channel's data on Kaggle. Both have the same duration and the same countries. The third part is the top 1000 accounts with the highest number of followers in 2021 from YouTube API. This article will use python as a tool to explore the elements that will drive traffic and the relationship between elements and revenue.

2. Materials and Methods

This article uses CRISP-DM as the main guidance step from collecting data to getting results. Clean and analyze data using the Python programming language through the Jupyter notebook code editor. Also, using Pearson correlation analysis to calculate the correlation of different indicators, focus on those elements that have a strong correlation with revenue. There are four steps. Data collection, data preparation, data analysis, and getting results. The data for this article came from the open-source Kaggle. The data on the Kaggle come from YouTube API, an official tool deployed by YouTube to scrap data on YouTube.

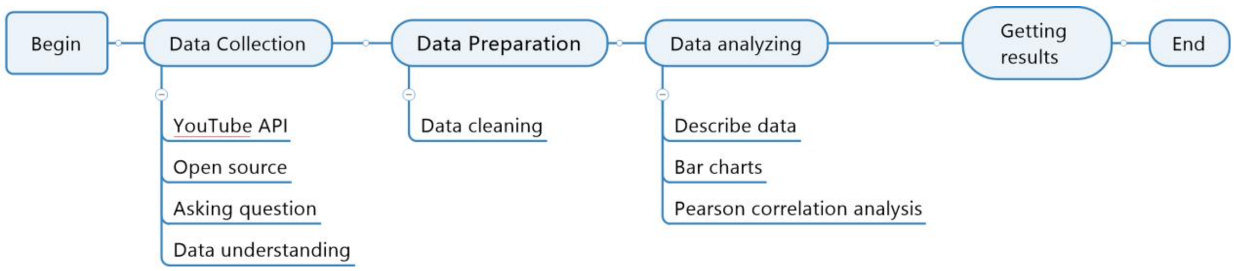

Figure 1. The step of dealing with data

Source: author's analysis

2.1 CRISP-DM

CRISP-DM was published in 1999 to standardize data mining processes across industries and has since become the most common methodology for data mining, analytics, and data science projects (NICK HOTZ, 2022). CRISP-DM is an industry-independent process model for data mining. It is the de-facto standard and an industry-independent process model for applying data mining projects. (Schroeer Christoph, Kruse Felix, & Gomez Jorge Marx, 2021). It has six sequential phases, from business understanding to deployment. (Nick Hotz last, 2022)

Table 1. The step of CRISP-DM

Sequential phases	Main work of this step	Tasks
Business understanding	Understand the objectives and requirements of the project	Determine business objectives, assess situation, determine data mining goals, produce project plan
Data understanding	Identify, collect, and analyze the data sets	Collect initial data, describe data, explore data, verify data quality
Data preparation	Prepare the final data set(s) for	Select data, clean data, construct data, integrate

	modeling	data, format data
Modeling	Build and assess various models based on several different modeling techniques	Select modeling techniques, generate test design, build model, assess model
Evaluation	Evaluation phase looks more broadly at which model best meets the business and what to do next	Evaluate results, review process, determine next steps
Deployment	Access its results	Plan deployment, plan monitoring and maintenance, produce final report, review project

Source: https://www.datascience-pm.com/crisp-dm-2/

From the KDnuggets polls in 2002, 2004, 2007, 2014, and 2022, the question "What is the main methodology you are using for data mining?" CRISP-DM is the most popular method used in data mining. The percentage is 49% in 2022. (Jeff Saltz, 2022)

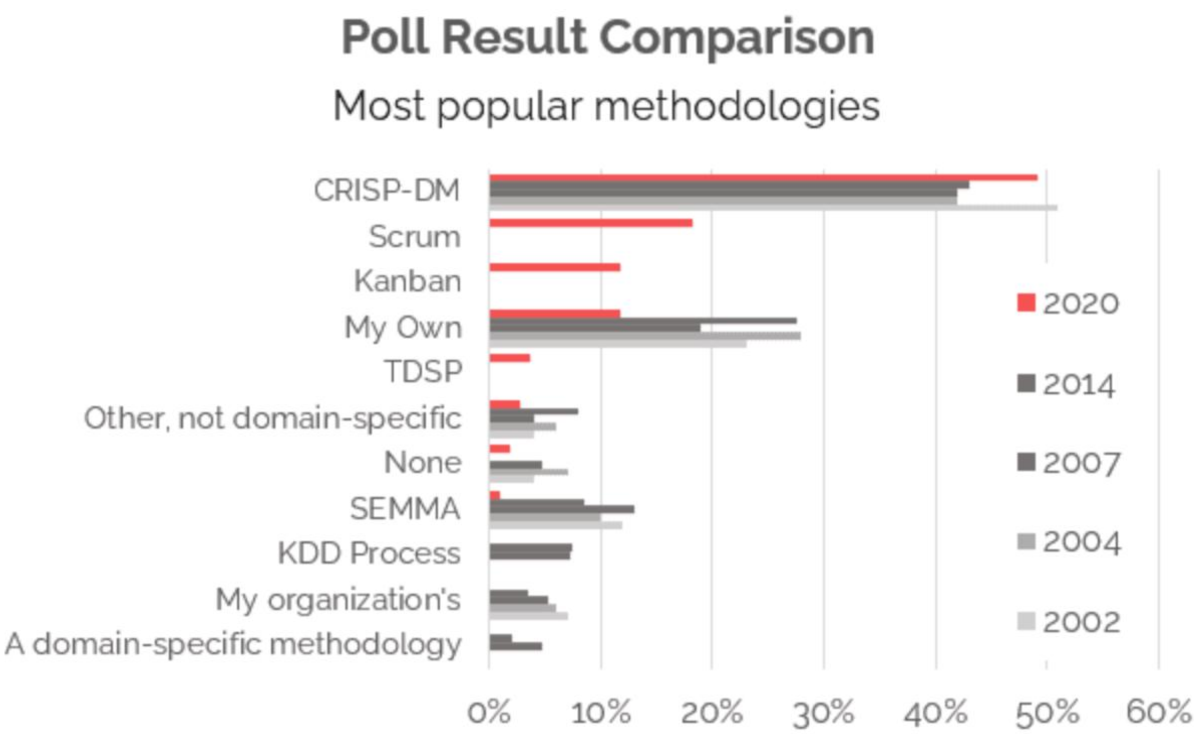

Figure 2. Poll result comparison

Source: https://www.datascience-pm.com/crisp-dm-still-most-popular/

2.2 YouTube API

Twitter and YouTube, for example, provide a plethora of datasets that can be accessed via APIs for research purposes (Joseph Kready, Shishila Awung Shimray, Muhammad Nihal Hussain, & Nitin Agarwal, 2017). To conduct data mining for this article, the data from the YouTube API was downloaded as a CSV file and imported into Jupyter Notebook.

2.3 Pearson Correlation Analysis

Pearson's Correlation is a measure of the linear relationship between two random variables - X and Y. It can be computed in Python using a method from numpy (Mehreen Saeed, 2022). This article uses Pearson correlation analysis to explore the correlation among variables.

3. Results and Discussion

methodology and using Jupyter Notebook as a tool exploring data to find user's preferences and the factors related to. This article uses three datasets, data from the top 1000 most popular uploaders' channels, the top 200 daily YouTube video data from September 2020 to January 2022 in the United States, and data from September 2020 to January 2022 from Ken Jee's channel. Using CRISP-DM as the main o the uploader's income.

3.1 The Analysis of Top 1000 Channels

Exploring features of the top 1000 channels can help us to know the key category of channels that drives more traffic.

Firstly, analyzing the number of different categories of channels, can help us understand what most people are doing.

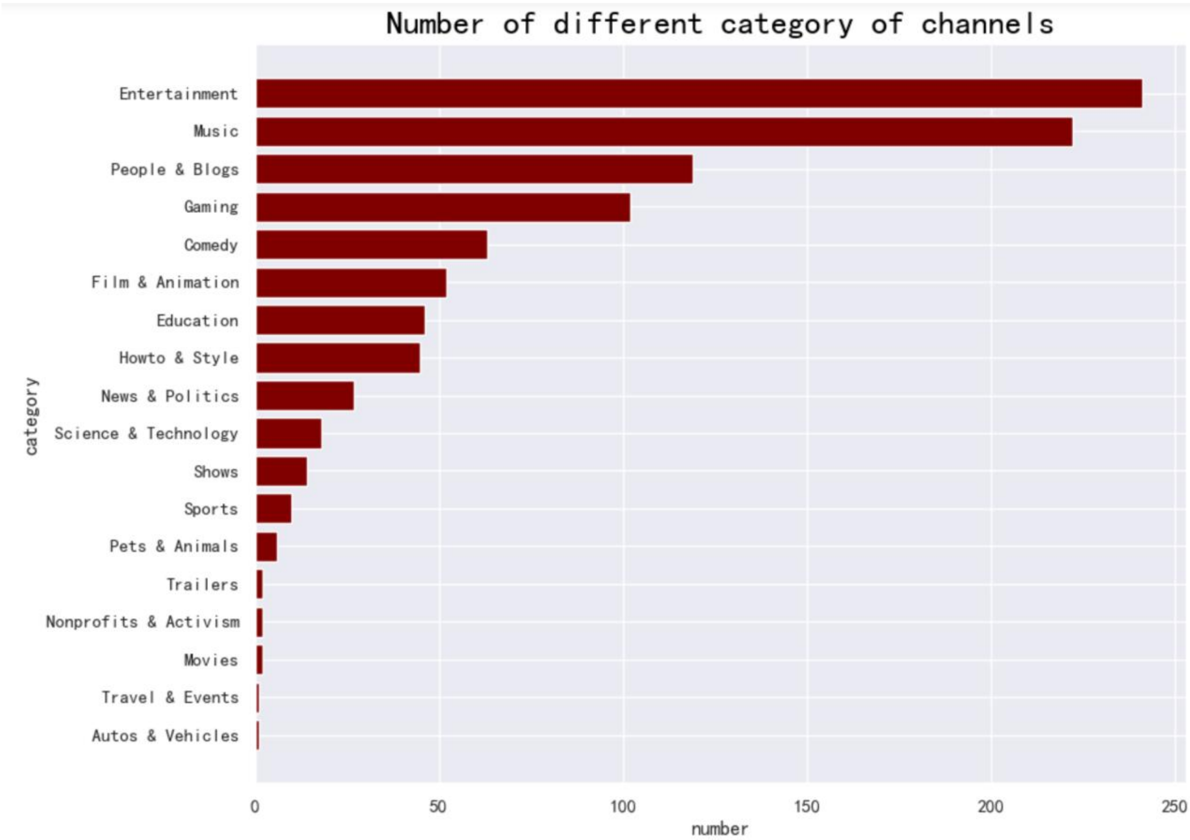

Figure 3. Number of different category of channels

Source: author's analysis using Jupyter Notebook

From the top 1000 channels on the planet, the top three categories with the highest numbers are entertainment, music, and people & blogs. While the movies, travel & events, and auto & vehicles categories have nearly no uploaders. It can be figured that most uploaders are attracted to entertainment, music, and people & blogs but dislike movies, travel & Events, anto & vehicles. However, it doesn't mean the entertainment, Music, and People & Blogs categories are more popular. It still needs more analysis to explore whether people are attracted to these types of videos.

In addition, exploring the most subscribers of different categories of channels to see whether people are attracted by Entertainment, Music, and People & Blogs.

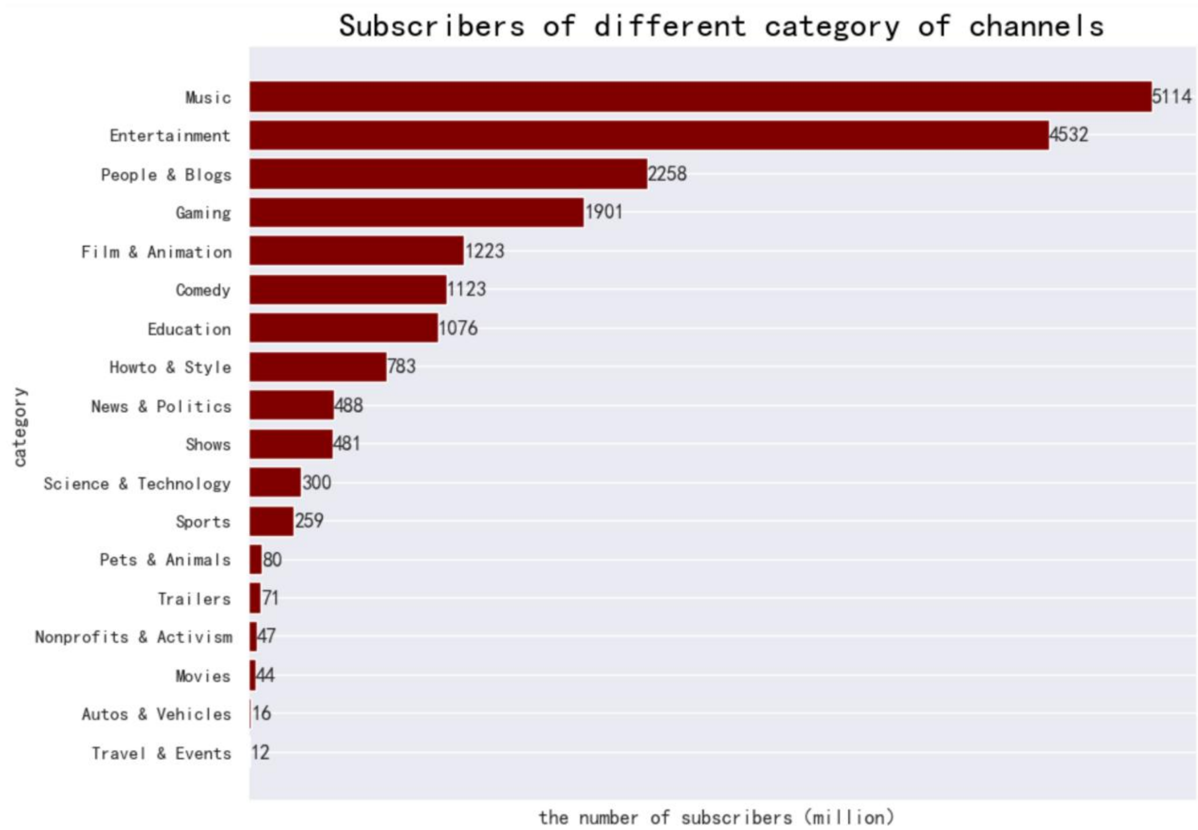

Figure 4. Subscribers of different category of channels

Source: author's analysis using Jupyter Notebook

The image above shows the top three categories: music, entertainment, and people & blogs. Entertainment has fewer subscribers than music, although it has the most subscribers in the top 1000 channels. People prefer music channels to entertainment. Movies, travel & events, auto & vehicles have few subscribers. Anyway, we can conclude that music, entertainment, and people & blogs are done by more uploaders and have more subscribers than other categories. Movies, travel and events, and automobiles and vehicles with little traffic fail to attract both uploaders and subscribers.

3.2 The Analysis of the Top Two Hundred Daily YouTube Videos in the US

This part of the analysis shows aggregated indicators for American users. Likes, dislikes, view counts, and comment counts are the indicators to evaluate traffic and people's interests. When people live a video, they will click the Like button, or share it with their friends. If they have something to say, they will leave their comments. The table below shows those indicators' summary of the top two hundred daily YouTube videos from September 2020 to January 2022 in the US group by likes because likes best reflect people's liking for videos.

	category	likes	dislikes	view_count	comment_count
7	Music	5095894520	86056709	77026145111	547567356
3	Entertainment	3502764383	58686590	68403068887	218820718
5	Gaming	2239901239	36285457	40681510077	174611388
10	People&Blogs	1146370100	29225129	19732837611	74154895
1	Comedy	839350171	10739690	11837321488	39725815
13	Sports	567147897	15083534	22521923660	47169634
4	Film&Animation	449340970	6348743	9599775434	32362873
12	Science&Technology	431037164	8980967	10004393834	25047438
2	Education	255490798	3941896	4339407580	16325845
6	Howto&Style	239846169	5936099	4721449662	17099981
0	Autos&Vehicles	92813182	1347517	2059134786	8016841
8	News&Politics	76768666	9184985	6008435218	21776410
14	Travel&Events	34774667	438087	584621450	1616071
11	Pets&Animals	31097428	381531	657660334	2152685
9	Nonprofits&Activism	10395697	110185	126806653	374099

Figure 5. The summary indicators of different categories

Source: author’s analysis using Jupyter Notebook

From this table, we can see that music, entertainment, and gaming have more likes, dislikes, views, and comments than other categories. While people’s & blogs’ indicators are lower than gaming’s. Music is more popular than entertainment, and gaming is more popular than people & blogs. Travel & Events, Pets & Animals, and Nonprofits & Activism have few likes, dislikes, views, and comments. I love pets & animals very much. I thought many people would like this category. Anyway, data mining always reveals the truth. This is the charm of data science. From the table above, we can claim, there is a correlation between different factors.

This article uses Pearson correlation analysis to explore data. The Pearson correlation coefficient varies from -1 to 1. A coefficient value of one indicates that X and Y can be well described by a linear equation and that all data points fall well on a straight line. A coefficient greater than 0 indicates that Y grows as X grows. A coefficient value of -1 indicates that all of the data points fall on a straight line and decrease as the coefficient increases. A coefficient less than 0 indicates that as X increases, Y decreases. A value of 0 indicates that no linear relationship exists between the two variables. A positive correlation exists when the absolute value of the coefficient is greater than 0.75. In general, it is between 0.25 and 0.75 when the absolute value of the coefficient is between 0.25 and 0.75.

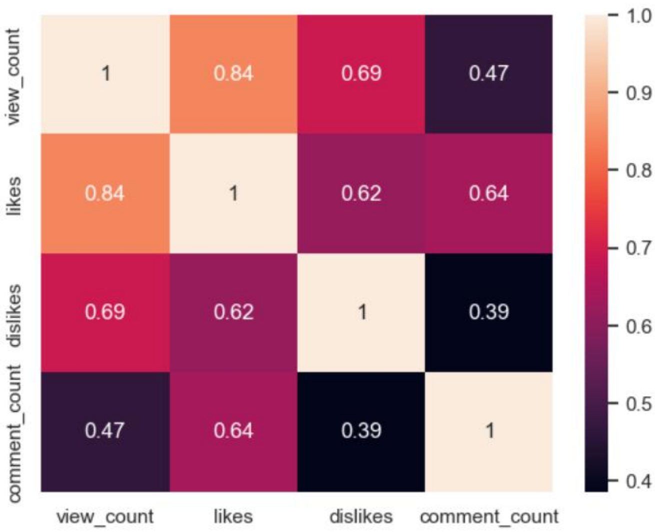

Figure 6. Pearson Correlation of each variable of US data

Source: author's analysis using Jupyter Notebook

From the picture above, we can see some of those indicators have a correlation with each other while some don't have any correlation. View count and likes have a strong positive correlation, the correlation coefficient is 0.84. Dislike and view count have an intermediate correlation, the correlation coefficient is 0.69. The correlation coefficient between likes and dislikes is 0.62. They are all related to view count. Comment count has a positive coefficient with likes and view count, the coefficients are 0.64 and 0.47. Therefore, the view count is important to attracting traffic. Uploaders should try to add view count. The more view count, the more traffic.

The following figure shows the correlation between the two variables.

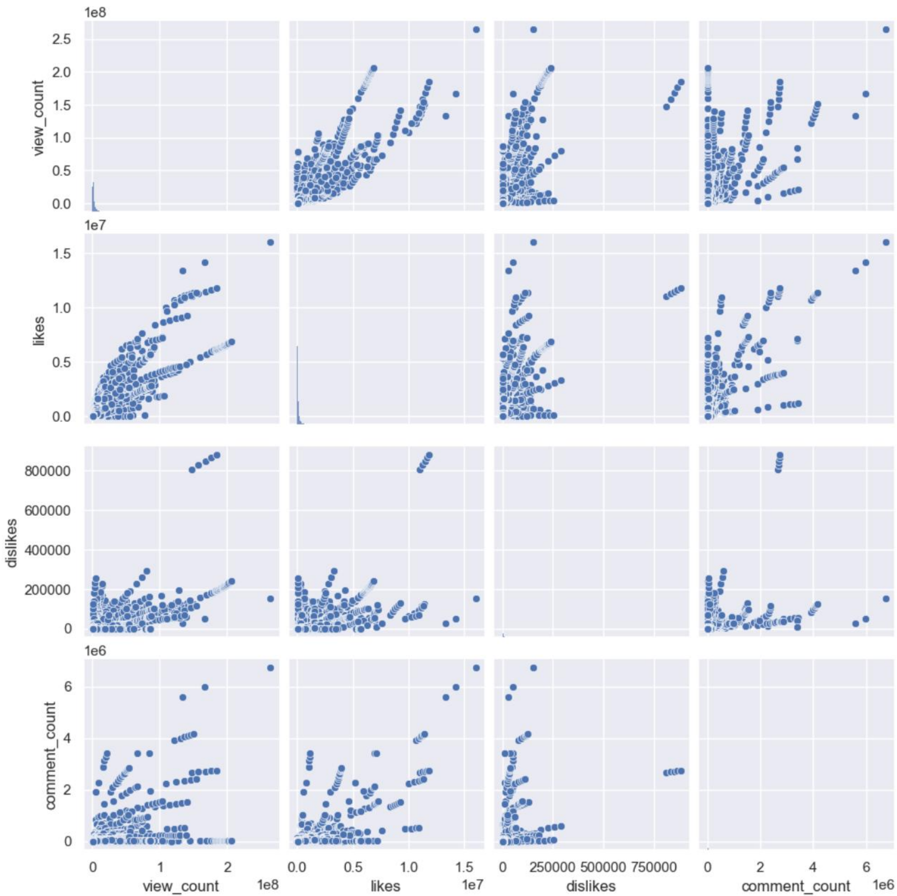

Figure 7. The correlation between two variables

Source: author's analysis using Jupyter Notebook

View counts have a clear positive correlation with likes and comments counts. Likes are strongly related to the number of views and comments. Likes and dislikes have a positive correlation with each other too, but it's not very obvious. The number of views, likes, and dislikes grows as the comment count grows.

3.3 The Analysis of Ken Jee's Channel

From Ken Jee's channel's data, we can see the mean value, the standard deviation, the minimum value, and the maximum value.

Table 2. Description of each variable

	Comments added	Shares	Dislikes	Likes	Views	Estimated revenue (USD)
mean	63.65470852	177.7533632	17.49775785	1008.865471	24968.6009	130.3485785
std	91.55962003	733.3311735	69.57851709	3577.895609	89609.56137	554.5035156
min	0	0	0	1	60	0
max	907	9583	942	46903	1253559	7959.533

Source: author's analysis using Jupyter Notebook

The mean value of like in each video is 1008.87; the mean revenue of each video is $130.34; the mean value of shares is 177.75; the mean value of comments is 63.65; the mean value of dislike is 17.50; and the mean value of views is 24968.60. If we only consider the metrics after the video is sent, those descriptive values mean if an uploader wants to earn 130 dollars in a video, this video almost needs 24969 people to view it, the video needs to obtain 1009 likes, and 178 people need to share it to their account. Also, maybe there will be 18 people who click the "dislike" button when they view the video.

We know some of those indicators are correlation in the previous analyse of daily YouTube videos in the US. Let's explore the correlation between revenue and other elements.

There are some variables that cannot be seen directly, so here are some explanations.

- RPM (USD) means Revenue Per Mille - Amount earned per 1000 views.
- CPM (USD) means Cost Per Mille - Amount charged to advertisers by 1000 views.
- Impression means count of impressions per video.
- Impressions click-through rate (%) means percent of the time impressions are clicked

From this chart, we can see RPM, CPM, average percentage viewed (%) and impression click-through rate (%) has a small correlation with other elements while other elements are correlation.

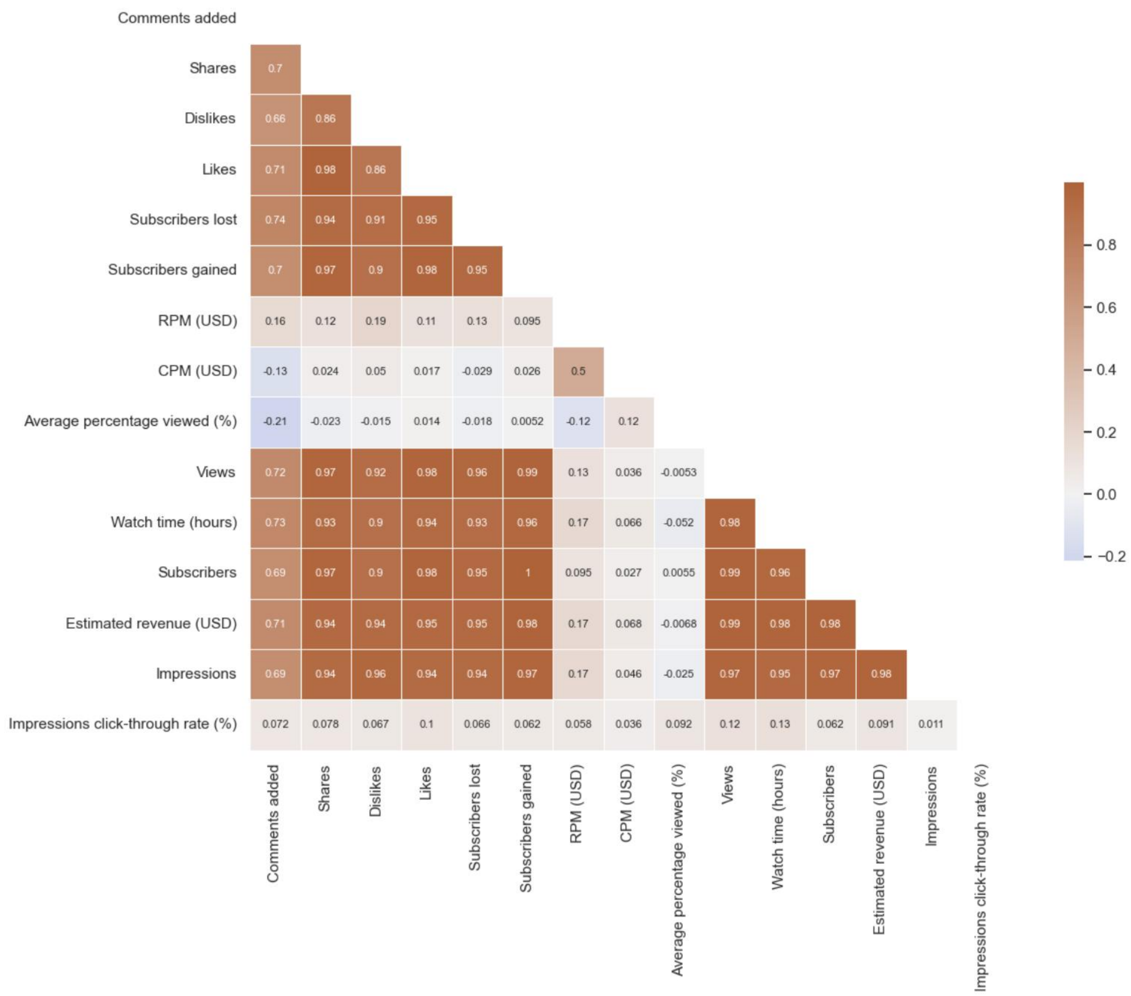

Figure 8. Pearson Correlation of each variable of Ken Jee's channel's data

Source: author's analysis using Jupyter Notebook

The correlation coefficients of these variables with income are shown in the following table:

Table 3. The correlation coefficient of different indicator and revenue

Indicators	Correlation coefficient of revenue
Views	0.990389214
Subscribers gained	0.982537125
Subscribers	0.982360169
Impressions	0.980933365
Watch time (hours)	0.979735466
Subscribers lost	0.952500139
Likes	0.951989776
Shares	0.944897697
Dislikes	0.93660882
Comments added	0.711200907
RPM (USD)	0.171507294
Impressions click-through rate (%)	0.090718499
CPM (USD)	0.068262096

Source: author's analysis using Jupyter Notebook

The element that is most correlative with Estimated revenue (USD) is Views, the correlation coefficient is 0.99. Subscribers gained, Subscribers, Impressions followed, the correlation coefficient is 0.98. Other strong positive correlation elements are Watch time (hours), Subscribers lost, Likes, Shares, and Dislikes. While Comments added and Estimated revenue (USD) have an intermediate correlation. Therefore, if uploaders want to have more revenue, they should pay attention to the number of Views, the number of subscribers, the new subscribers, likes, dislikes, shares, and watch time.

From the analysis result above, we can conclude that people are more like the category of channel that can make them feel relaxed, such as entertainment, music, gaming, and people & blog. Those categories have some same features. They are output items and do not require people to think, and they make people happy. People are generally willing to accept videos that will give them short-term pleasure. Some other ones about vehicles, travel, and pets are the hobbies of a small number of people. Views, the number of subscribers, new subscriber, subscriber lost, impressions, watch time, likes, shares, dislikes, and comment all have a strong positive correlation with revenue. When these indicators grow, revenue grows.

4. Conclusion

4.1 Conclusion of This Article

This article uses three datasets to explore users' preferences and the indicators that have a strong correlation with revenue. CRISP-DM is the guiding idea, while Jupyter Notebook is the analysis tool. Using the histogram to display engaging video types with sorted bar charts. Using Pearson correlation analysis to show the correlation coefficient between two variables and describing them through a correlation matrix (few variables) and correlation triangle array (many variables).

From the analysis of the top 1000 channels and daily US treading videos, the most attractive category for all users and Americans is the same. They are entertainment, music, people & blogs, and gaming, which have the most subscribers, views, comments, likes, and dislikes. While gaming has fewer subscribers and uploaders than people & blogs, it has better aggregate indicators. From the analysis of Ken Jee's channel, views whose correlation coefficient with revenue is 0.99 has the strongest correlation with estimated revenue. Watch time (hours), subscribers lost, likes, shares, and dislikes are followed. While impressions' click-through rate (%), RPM, and CPM have weak correlations with revenue, their correlation coefficients are 0.09, 0.17, and 0.06. Why are entertainment channels popular? What are the drivers, and what makes people go to an entertainment channel and click the like button, leave comments, and share with friends? These will serve as future research.

5. Contribution

YouTube provides an open platform on which everyone can post content and share their daily routine. Researchers and teachers post a tutorial on it to guide their students. Celebrities use it to promote themselves and increase their influence by posting their works and trips. Commercial companies use uploader videos to promote their products or post videos themselves to sell their products. Everyone can view videos posted by others that are of interest to them and post their own videos on this free and equal content-rich platform. Therefore, this article mainly contributes to three kinds of people. First, uploaders can find the type of content that people engage with the most, the estimated revenue, and its aggregate indicator of a video. Also, the elements which have a strong positive correlation with revenue. In addition, through scientific digital analysis, users can also see the signs that the YouTube platform plays in their daily lives. Which could satisfy their curiosity about the platforms they use. Finally, marketing companies can smell the business opportunity to put ads in videos related to their products, which can have good sales revenue.

Ethics Approval and Consent to Participate

Not applicable.

Availability of Data and Materials

The datasets analyzed during the current study are available on the Kaggle.

References

Mafalda moutinho, Maria rodrigues, & Anabela ribeiro. (2022). Determinants and Consequences of YouTube Advertising Value. *International Journal of Marketing, Communication and New Media*, *10*(18). https://doi.org/https://doi.org/10.54663/2182-9306.2022.v10.n18.50-70.

Ini vanwesenbeeck, Liselot hudders, & koen ponnet. (2020). Understanding the YouTube Generation: How

Preschoolers Process Television and YouTube Advertising. *Cyberpsychology Behavior and Social Networking, 23*(6), 426–432. https://doi.org/https://doi.org/10.54663/2182-9306.2022.v10.n18.50-70.

Kay barker, kristin. (2020). Lay Pharmacovigilance and the Dramatization of Risk: Fluoroquinolone Harm on YouTube. *Journal of Health and Social Behavior, 60*(4), 509–524. https://doi.org/10.1177/0022146519888242.

Schroeer, christoph, Kruse, felix, & Gomez, jorge marx. (2021). A Sysmatic Literature Review on Applying CRISP-DM Process Model. *International Conference on Enterprise Information Systems, 181*, 526–534. https://doi.org/10.1016/j.procs.2021.01.199.

Nick hotz last. (2022, November 13). *What Is CRISP DM?* Data Science Process Alliance. https://www.datascience-pm.com/crisp-dm-2/.

Tolkach denis, & Pratt stephen. (2021). Travel Professors: A YouTube Channel about Tourism Education & Research. *Journal of Hospitality Leisure Sport & Tourism Education*, *28*. https://doi.org/10.1016/j.jhlste.2021.100307.

Yuksel bahar, & Cakmak kubra. (2020). Healthcare Information on YouTube: Pregnancy and COVID-19. *International Journal of Gynecology & Obstetrics, 150*(2), 189–193. https://doi.org/10.1002/ijgo.13246.

Jl. scientia boulevard, gading serpong, tangerang. (2021). Mapping the ASEAN YouTube Uploaders Pemetaan Pengunggah Video YouTube Dari Negara ASEAN. *Journal ASPIKOM, 6,* 1. https://doi.org/https://wordvice.cn/citation-generator/article-journal/apa7.

Mehreen saeed. (2022, July 21). *Calculating Pearson Correlation Coefficient in Python with Numpy*. Stark Abuse. https://stackabuse.com/calculating-pearson-correlation-coefficient-in-python-with-numpy/.

Akhremenko, a. s., Stukal, d. k., & Petrov, a. p. (2020). Network Vs Message in Protest Diffusion on Social Media_Theoretical and Data Analytics Perspectives. *Polis-Politicheskiye Issledovaniya, 2*, 73–91. https://doi.org/10.17976/jpps/2020.02.06.

Amarasekara, inoka, & Grant, will j. (2019). Exploring the YouTube Science Communication Gender Gap: A Sentiment Analysis. *Public Understanding of Science, 28*(1), 68–84. https://doi.org/10.1177/0963662518786654.

Yang zhou, Zubair ahmad, Hassan alsuhabi, M. yusuf, Ibrahim alkhairy, & Ibrahim alkhairy. (2021). Impact of YouTube Advertising on Sales with Regression Analysis and Statistical Modeling: Usefulness of Online Media in Business. *Computational Intelligence and Neuroscience, 10.* https://doi.org/https://doi.org/10.1155/2021/9863155.

Sara fraccastoro, Mika gabrielsson, & Ellen bolman pullins. (2021). The Integrated Use of Social Media, Digital, and Traditional Communication Tools in the B2B Sales Process of International SMEs. *International Business Review, 30*(4). https://doi.org/10.1016/j.ibusrev.2020.101776.

S. dixon. (2022, July 26). *Most Popular Social Networks Worldwide as of January 2022, Ranked by Number of Monthly Active Users.* Statista. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.

Jeff saltz. (2022, May 2). *CRISP-DM Is Still the Most Popular Framework for Executing Data Science Projects*. Data Science Process Alliance. https://www.datascience-pm.com/crisp-dm-still-most-popular/.

Jane e. klobas, Tanya j. mcgill, & Sedigheh moghavvemi and tanuosha paramanathan. (2018, August 11). *Online Information Review Problematic and Extensive YouTube Use: First Hand Reports.* Emerald Insight. www.emeraldinsight.com/1468-4527.htm.

Janarthanan balakrishnan, & Mark d.griffiths. (2017). Social Media Addiction: What Is the Role of Content in YouTube? *Journal of Behavioral Addictions*, 1–2. https://doi.org/10.1556/2006.6.2017.058.

Erwin halim, Rena anindya, & Marylise hebrard. (2020). The Impact of Motivation to Watch YouTube, Subjective Norms, Behavior Control, Information Success Model to Watching YouTube Engagement. 2020 International Conference on Information Management and Technology (ICIMTech), 1–3.

Ismail selvi, & Numan baydilli. (2021). An Analysis of Misleading YouTube Videos on Urological Conditions: What to Do about the Danger of Spreading Misinformation of the YouTube Videos. *World Journal of Urology,* 1–3. https://doi.org/https://doi.org/10.1007/s00345-021-03623-7.

Rosenthal, S. (2018), Motivations to seek science videos on YouTube: free-choice learning in a connected society, *International Journal of Science Education, Part B, 8(*1), pp. 22-39.

DIGITAL 2022: JULY GLOBAL STATSHOT REPORT. (2022, July 21). DATA REPORTAL. https://datareportal.com/reports/digital-2022-july-global-statshot#:~:text=The%20global%20state%20of%2

0digital,-View%20fullsize&text=The%20global%20mobile%20population%20has,5.03%20billion%20in%20July%202022.

Jack shepherd. (2022, October 5). 22 Essential YouTube Statistics You Need to Know in 2022. Social Shepherd. https://thesocialshepherd.com/blog/youtube-statistics#:~:text=YouTube%20has%202.1%20billion%20monthly,some%20pretty%20impressive%20YouTube%20.

Joseph kready, shishila awung shimray, muhammad nihal hussain and nitin agarwal. (2017). Statistical Migration of API Usages. *IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C),* 47–50. https://doi.org/10.1109/ICSE-C.2017.17.

Appendix A

The Datasets

Dataset	Introduction of the dataset	Link
Ken Jee YouTube Data	Dataset for analyzing Ken Jee's YouTube Channel. Data by video and over time	https://www.kaggle.com/datasets/kenjee/ken-jee-youtube-data
YouTube Trending Video Dataset (updated daily)	YouTube Trending Video data-set which gets updated daily.	https://www.kaggle.com/datasets/rsrishav/youtube-trending-video-dataset
Most Subscribed YouTube Channels	Analysis of Top YouTube channels	https://www.kaggle.com/datasets/surajjha101/top-youtube-channels-data