**CCT416 – Final Project**

**Introduction**

The topic of my data analysis final project is an examination of Felix Arvid Ulf Kjellberg’s YouTube channel, widely known as PewDiePie. PewDiePie has been one of the most prominent gaming YouTubers since the rise of gaming content on YouTube, dating back to the early 2010s when the “Let’s Play” format dominated the platform. Analyzing PewDiePie’s channel provides valuable insight into the evolution of video trends and personal channels on YouTube, as he was arguably the most-watched YouTuber of the 2010s.

**Research Questions**

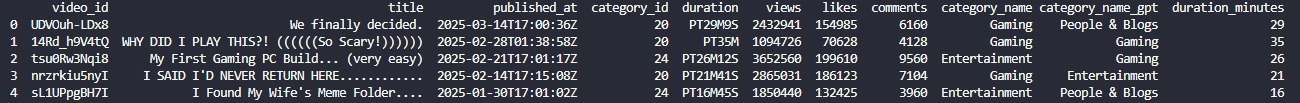
* How have PewDiePie’s video upload patterns—specifically in terms of engagement (views and comments)—changed over time?
* What is the correlation between shifts in video category trends and view counts on PewDiePie’s channel over the years, and how do these changes reflect broader trends in global YouTube engagement?

**Relevancy To Social Data Analytics**

These research questions are relevant to social data analytics because they offer valuable insights into the shifting engagement patterns and content trends of PewDiePie, one of YouTube’s most influential creators. By analyzing changes in video upload patterns, engagement metrics, and category trends over time, this study helps us understand how audience behavior and digital media consumption have evolved. Examining whether PewDiePie's viewers' habits developed into distinct patterns or remained consistent as YouTube engagement grew provides a deeper look into audience dynamics. These insights contribute to a broader understanding of how YouTube channels adapt to changing viewer preferences and engagement trends. By studying PewDiePie’s channel, we can apply these findings to larger patterns in global YouTube trends, offering a meaningful perspective on the evolution of YouTube content production.

**Data Description**

* The data comes from the YouTube Data API, provided by Google Developers tools. Since YouTube is operated by Google, this data source can be considered a primary legitimate source of information.
* The key variables in my dataset **are video\_id, title, category\_id, duration, views, likes, comments, category\_name** (mapped from category\_id), and **category\_name\_gpt.**
  + Why there are two **category\_name** variables is explained in the limitations section.
* In addition, when working on data analysis I manually assigned a new column to DataFrame in main.py named **duration\_minutes** to see duration data in integer format.
* After I made a request from API for PewDiePie’s channel, I written the DataFrame to csv file format. In the file, there are 4604 rows (videos) in total, last video being published on March 14, 2025.



**Rationale for Data Selection**

This dataset is appropriate for my research questions because it provides the necessary tools to evaluate engagement with PewDiePie’s production patterns over the years. The date data allowed me to observe overall production changes over time, while other engagement metrics—such as comments, likes, and view counts—helped me analyze how audience consumption varied across different years. Most importantly, the category column enabled me to categorize these habits into smaller segments for a more detailed analysis. Additionally, it allowed me to incorporate categorical data into a time series analysis.

In terms of understanding how his behavior resonated within the global YouTube landscape, the dataset is particularly useful where it highlights spikes in views, likes, and comments. Analyzing these peak points helped me identify trends that were popular on global YouTube and assess PewDiePie’s influence on them.

**Limitations on The Data**

In my Stepping Stone assignment, I initially expected that YouTube videos from the early 2010s might have data gaps. However, I found that the data provided by the YouTube API contained no gaps across any videos—I had access to all the necessary data.

The most challenging part of this assignment was working with the category name data. When I first began my analysis, nearly all the videos were categorized as gaming videos. I manually checked a few videos to verify their categories, but they were incorrect. Thus, the most significant challenge I faced was cleaning the unclean category data.

To address this, I utilized ChatGPT’s machine learning capabilities to refine the category data. After incorporating ChatGPT’s assistance, I re-examined the categories and found that the results were much improved—videos and categories aligned more accurately.

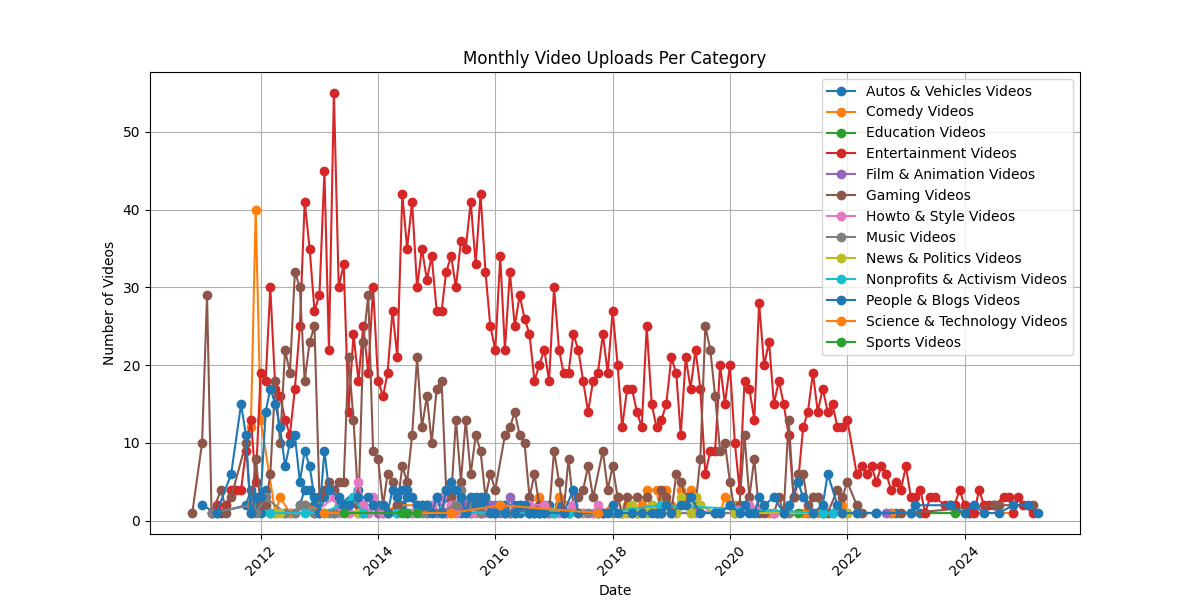
However, I am still not fully satisfied with the results. PewDiePie’s video titles are often very random, which can confuse machine learning models, as troll titles do not provide reliable categorization data. For future research, a more advanced machine learning approach that categorizes videos by analyzing their thumbnails could yield better results.

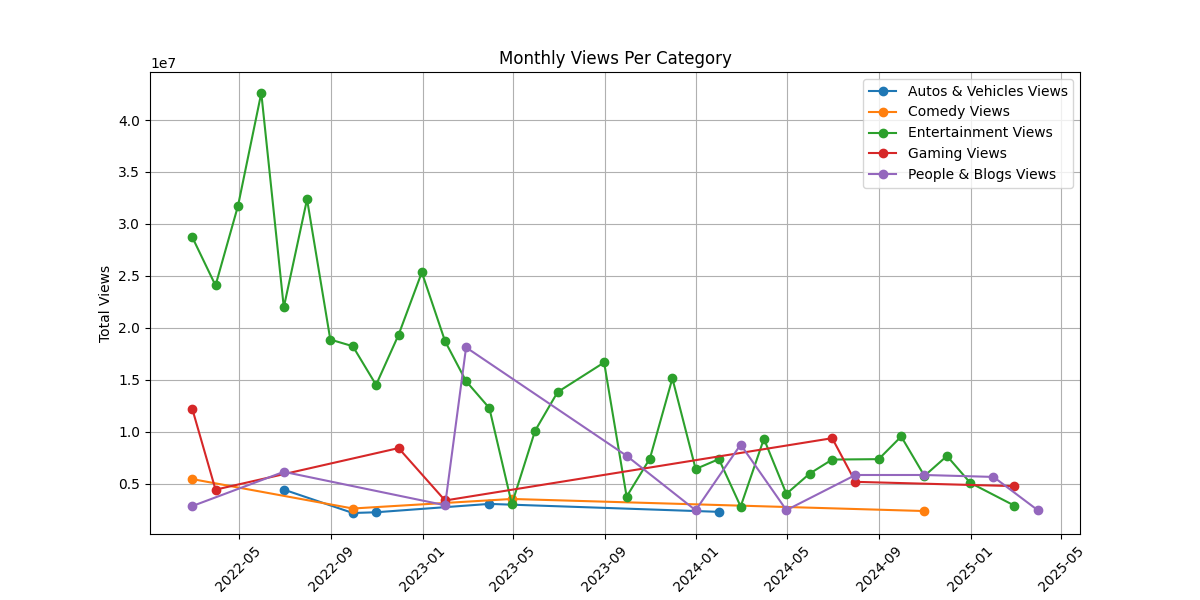
**Methodology**

* **Data Visualization:** The main analysis method I relied on for this project was visualization through Time Series analysis. I chose this method because it allowed me to use multivariate visualization, which was essential for comparing date, views, and categories simultaneously. The purpose of my research questions is to identify patterns in PewDiePie's engagement data over the years, and Time Series analysis enabled me to detect trends, track changes in key metrics over time, and smooth data by reducing noise, such as grouping videos by month. Additionally, Time Series analysis effectively highlighted anomalies, revealing the broader global impact of PewDiePie's content on YouTube, particularly during periods when his views peaked.
* **Descriptive Analysis:** Even though descriptive analysis is challenging to apply to PewDiePie’s channel due to his generally high view counts and the presence of videos that cause the standard deviation to return extremely high values, I used descriptive analysis specifically for the period after 2022. This is because, after 2022, there was a noticeable shift in PewDiePie’s production patterns, resulting in a more compact dataset. By focusing on this period, I was able to apply descriptive analysis to gain insights into how his audience's consumption behavior changed after he began producing fewer videos. When using this method, I focused on measures of location and measures of spread to understand the general trend of video categories and assess the reliability of these trends by examining how dispersed the data was after 2022.

**Results and Discussion**

1. ***PewDiePie’s Upload Patterns and Audience Engagement***

****

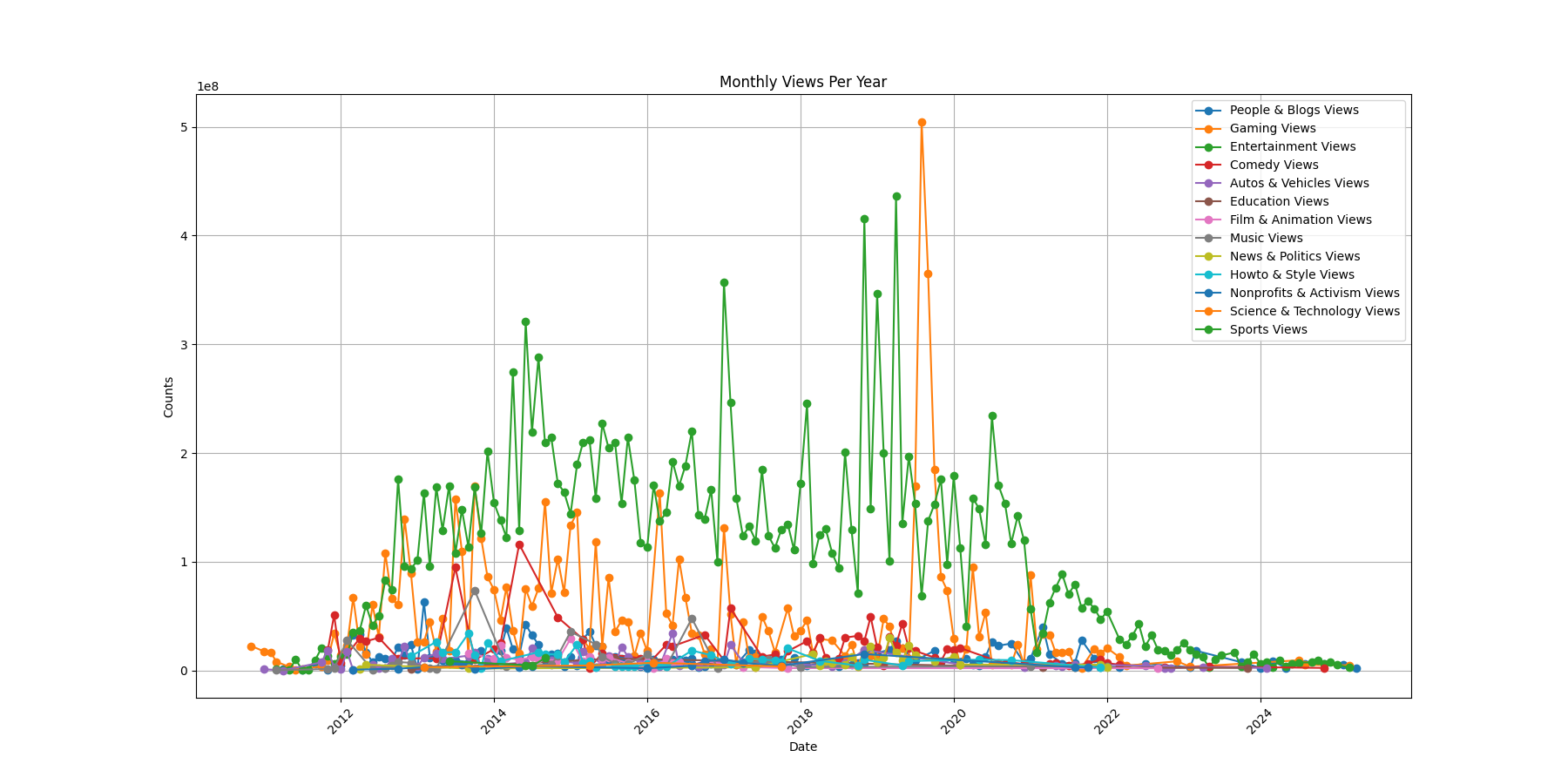
This time series plot illustrates the number of video categories uploaded to PewDiePie's channel over time. The spike in 2013 indicates a peak in uploads, after which PewDiePie gradually reduced the number of videos he uploaded each year. Notably, after 2022, there was no month in which he uploaded 10 or more videos. Despite this decline in uploads, his content production remained primarily focused on the entertainment category to current time, while his gaming-related uploads significantly decreased. This suggests that something in 2022 influenced PewDiePie’s upload patterns. To better understand this shift, we can examine audience engagement metrics—specifically, view counts—after 2022. For this analysis, I selected the five most frequently appearing categories, as less common categories introduced noise and disrupted the clarity of the visual analysis.

In this time series analysis plot, each point represents a video. We can observe that the most prominent video category in terms of audience interaction remains entertainment, followed by people & blogs in second place. When applying descriptive analysis to these two most frequently uploaded categories after 2022, we find that the standard deviation for entertainment is 9,613,762, while for people & blogs, it is 4,274,794. The high spread in view counts is largely due to PewDiePie’s overall high average views; for instance, the mean view count for entertainment videos is 6,549,466. Because of this wide variation, views alone may not be the most accurate measure of audience engagement. Instead, we need a more refined metric that reflects intimate audience engagement, rather than a passive engagement metric like views.

For this analysis, I chose the number of comments as a metric, as comments represent a more intimate form of audience engagement, reflecting interaction from viewers who actively share their thoughts. After 2022, the mean number of comments for the entertainment category is 23,754, while for people & blogs, it is 13,836. Correspondingly, their standard deviations are 54,269 and 17,929, respectively. This indicates that for entertainment videos, highly engaging videos significantly raise the average comment count, but not all videos receive the same level of interaction, making audience engagement in this category less stable. On the other hand, while people & blogs videos have a lower average number of comments, engagement within this category is more consistent, with a more interactive audience overall. This suggests that while PewDiePie's entertainment videos remain dominant, he has also established a reliable secondary category where his audience engages more meaningfully with his content.

I conducted social research to understand why the people & blogs category became more prominent in PewDiePie's production habits. The answer can be found in the data—specifically, in a people & blogs video uploaded in February 2023 titled “We're having a baby!”. In this video, PewDiePie shares moments from different stages of his wife's pregnancy, which suggests a shift in his focus toward his personal life even months before the video’s upload. This change is reflected in his video upload patterns, indicating that he may have prioritized spending time with his family over producing YouTube content. Following the birth of his child, he continued to create more family-oriented videos under the people & blogs category, further reinforcing this shift in his content style.

1. ***Video Category Trends and Echoes in Global YouTube***



When we apply measures of central tendency in descriptive analysis, we can see that the most-watched video on PewDiePie’s channel falls under the entertainment category: a video titled "Bitch Lasagna" with 338 million views.

Since the video was uploaded at the end of 2018, it cannot be considered the initiator of the diss track trend on YouTube, which gained popularity around 2017. However, we can say that PewDiePie created one of the most prominent examples of the diss trend. Additionally, this video holds historical significance in global YouTube culture as a key moment in the rivalry between independent YouTube creators and corporate channels, particularly since T-Series and PewDiePie were competing for the highest subscriber count during that time (“PewDiePie vs. T-Series,” 2025).

Lastly, when we look at the peak of the monthly views time series plot, we can see that the gaming category experienced a huge surge in views on June 21, 2019, which marks the most-watched month in the channel's history. During this time, PewDiePie began playing Minecraft, and his first episode, "Minecraft Part 1," had 58 million views. With this, PewDiePie sparked a global trend among major channels to start playing Minecraft in 2019, just weeks after he uploaded his first Minecraft video. For example, Jacksepticeye uploaded his first Minecraft video on July 28, 2019. Looking further abroad, we see that the most subscribed Turkish gaming YouTuber, Enes Batur, uploaded his first Minecraft episode on July 31, 2019. We can even observe similarities between the thumbnails of PewDiePie’s and Enes Batur’s (2019) first Minecraft videos. In July 2019, the monthly gaming view was over 500 million.

PewDiePie – Minecraft Part 1 Enes Batur – Minecraft Part 1

A person wearing headphones and holding a sword

AI-generated content may be incorrect. 

The implementation of my findings through descriptive and visual data analysis suggests that PewDiePie’s actions resonate across global YouTube, whether by popularizing gameplay playthroughs or sparking online rivalries between channels through diss tracks. However, this does not imply that he is the sole originator of these trends. To make such a claim, I would need to expand this dataset to include all YouTube gaming uploads from the summer of 2019, not just PewDiePie’s uploads. The data presented about the initiation of trends on global YouTube would need to be backed by an analysis of video upload patterns across the entire YouTube gaming category. Within the scope of my research, however, I will suggest that PewDiePie played a key role in the success of these trends.

**Conclusion (0.5 points)**

In summary, regarding the first question, our findings show that while he still primarily produces entertainment content, he has reduced his overall output and shifted his secondary focus to sharing memorable moments from his family life. His audience is more actively engaged with this content, fostering greater interaction. For the second question, we can see that his channel’s top video is a trending video with exceptionally high views, even by global YouTube standards. Additionally, his ability to consistently generate hundreds of millions of views per month has influenced other YouTubers to adopt similar content strategies. This suggests that PewDiePie plays a role in amplifying the popularity of trends.

Regarding the limitations of answering research questions, the unclean category data should be reviewed to improve the results. In its current form, ChatGPT provided a cleaner output than the initial version after retrieving data from the YouTube API. However, a more effective approach would be training our own AI model for more accurate categorization. Additionally, we cannot definitively say that PewDiePie initiated the trends he made popular videos about, as this would require a comprehensive analysis of the trends within global YouTube data. This research was limited to videos on PewDiePie’s channel. For further studies on trend analysis in relation to PewDiePie, we should examine other spikes in his monthly view chart and compare them to global YouTube trends. For brevity, I focused only on the two most prominent spikes in this social data analysis.

Examining PewDiePie's production patterns, audience engagement, and influence on trends provides a deeper understanding of how first-generation personal YouTube channels have evolved over time. This analysis can also be applied to other major personal channels to anticipate future trends for independent content creators. Economically, his role in shaping trends contributes to the broader YouTube ecosystem, influencing audience reach and marketing value. His ability to dictate what millions watch highlights the power of personal YouTube channels, reinforcing their significance in the digital media landscape. Understanding these dynamics not only sheds light on PewDiePie's impact but also offers valuable insights into the future of online content creation.

**References**

Batur, Enes. (Jul 31, 2019). *Minecraft Part 1 (Yıl 2013)*. YouTube. https://www.youtube.com/watch?v=JkyrAguXSM0

jacksepticeye. (July 28, 2019). *Playing Minecraft For The Very FIRST Time - Part 1*. YouTube. https://www.youtube.com/watch?v=XhqsltLnzBM

PewDiePie vs T-Series. (2025). In *Wikipedia.* https://en.wikipedia.org/wiki/PewDiePie\_vs\_T-Series