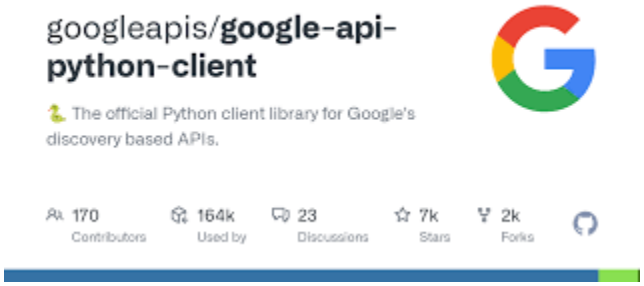
YT Channel Analysis

Unveiling Insights Through Data

In the realm of digital media, the influence of YouTube as a content platform cannot be underestimated. In the pursuit of generating data-driven insights and putting my learnings into a project, I embarked on a data analysis of three distinct YouTube channels. These channels, namely Ken Jee, Tina Huang, and Alex the Analyst, have garnered substantial attention within the data community and I will be using their data for this exploratory analysis project.

Data Collection

The data for this analysis has been obtained using the YouTube Data API.




```
channel_data.head()
```

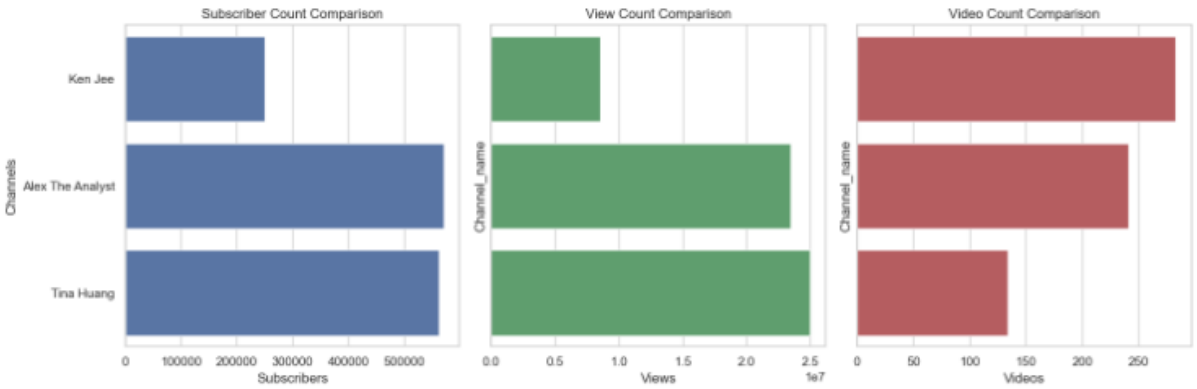
	Channel_name	Subscribers	Views	Total_vidoes	Playlist_id
0	Tina Huang	562000	24961856	134	UU2UXDak6o7rBm23k3Vv5dww
1	Alex The Analyst	571000	23460177	241	UU7cs8q-gJRlGwj4A8OmCmXg
2	Ken Jee	251000	8524446	283	UUiT9RITQ9PW6BhXK0y2jaeg

```
video_data.head()
```

	Channel Owner	Title	Published_date	Duration	Views	Likes	Comments
0	Ken Jee	How to Survive a Down Data Job Market	2023-09-05 10:50:39+00:00	PT9M56S	11498	354	52
1	Ken Jee	The Harsh Reality of the Data Job Market	2023-08-24 11:30:19+00:00	PT6M43S	87339	2202	258
2	Ken Jee	7 Industries AI will Aggressively Disrupt	2023-08-14 11:30:19+00:00	PT10M44S	5016	215	40
3	Ken Jee	What's in My Data Science Travel Bag? (50+ Fli...	2023-08-07 17:57:05+00:00	PT10M12S	3534	114	47
4	Ken Jee	7 Enticing Jobs AI Will Create	2023-07-24 12:00:01+00:00	PT8M24S	4180	185	36

Data Analysis & Visualization

- **Channel Metrics and Significance -** A foundational understanding of each channel's scale - their audience reach and content library size within the YT ecosystem.



Here are the main takeaways from the chart:

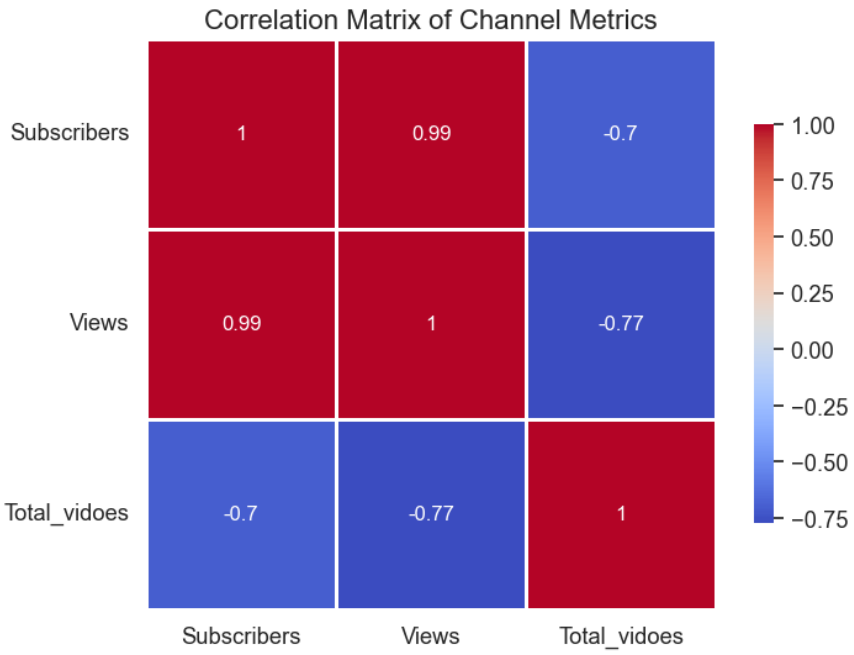
1. Alex commands the largest subscriber base among the three, but interestingly, his videos tend to receive fewer views compared to Tina, who maintains a subscriber count almost equal to Alex's, but her content library is notably smaller. On the contrary, Ken has the lowest number of subscribers but uploads the most videos, and yet his videos garner the least number of views.

2. This chart underscores a crucial point: a higher volume of video uploads doesn't necessarily translate into more subscribers or views. It suggests that within the YouTube ecosystem, there are other parameters as well that play a role in achieving broader reach and engagement.

3. Initially, when glancing at the channel metrics, Tina appears to be the top performer, followed by Alex, and then Ken.

4. Assuming that all three channel owners started their YouTube journeys around the same time, it's plausible that Ken uploads most frequently, followed by Alex, and then Tina, who maintains the lowest count of video uploads. The significant disparity in video counts across their channels may also stem from variations in the timing of when they initiated their channels, with the possibility that Ken began first.

➢ **Metrics Interplay and Correlations -** Exploring the correlations and relationships among engagement metrics - views, likes and comments.



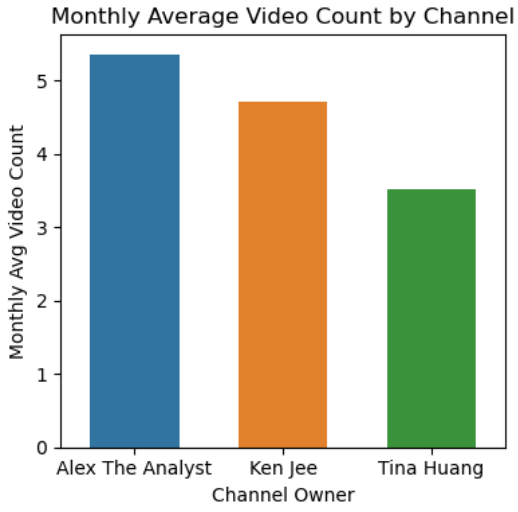
In summary, based on the correlation coefficients of the above heatmap, here is what can be interpreted:

1. Subscribers have an extremely strong positive correlation with views, indicating that a larger subscriber base is closely associated with higher view counts.

2. Subscribers have a strong negative correlation with the total number of videos, which may suggest a focus on quality or more engaging content as your subscriber base grows.

3. Views also have a strong negative correlation with the total number of videos, indicating that highly viewed videos may lead to a reduction in the production of new content.

➢ **Content Consistency & Frequency -** Investigating the monthly average uploads for each channel to reveal the pattern.



Here are the key insights from this chart:

1. On average, Alex consistently uploads approximately 5-6 videos each month. Ken Jee follows closely with an average of 4-5 monthly uploads, while Tina maintains a lower average of 3-4 videos per month.

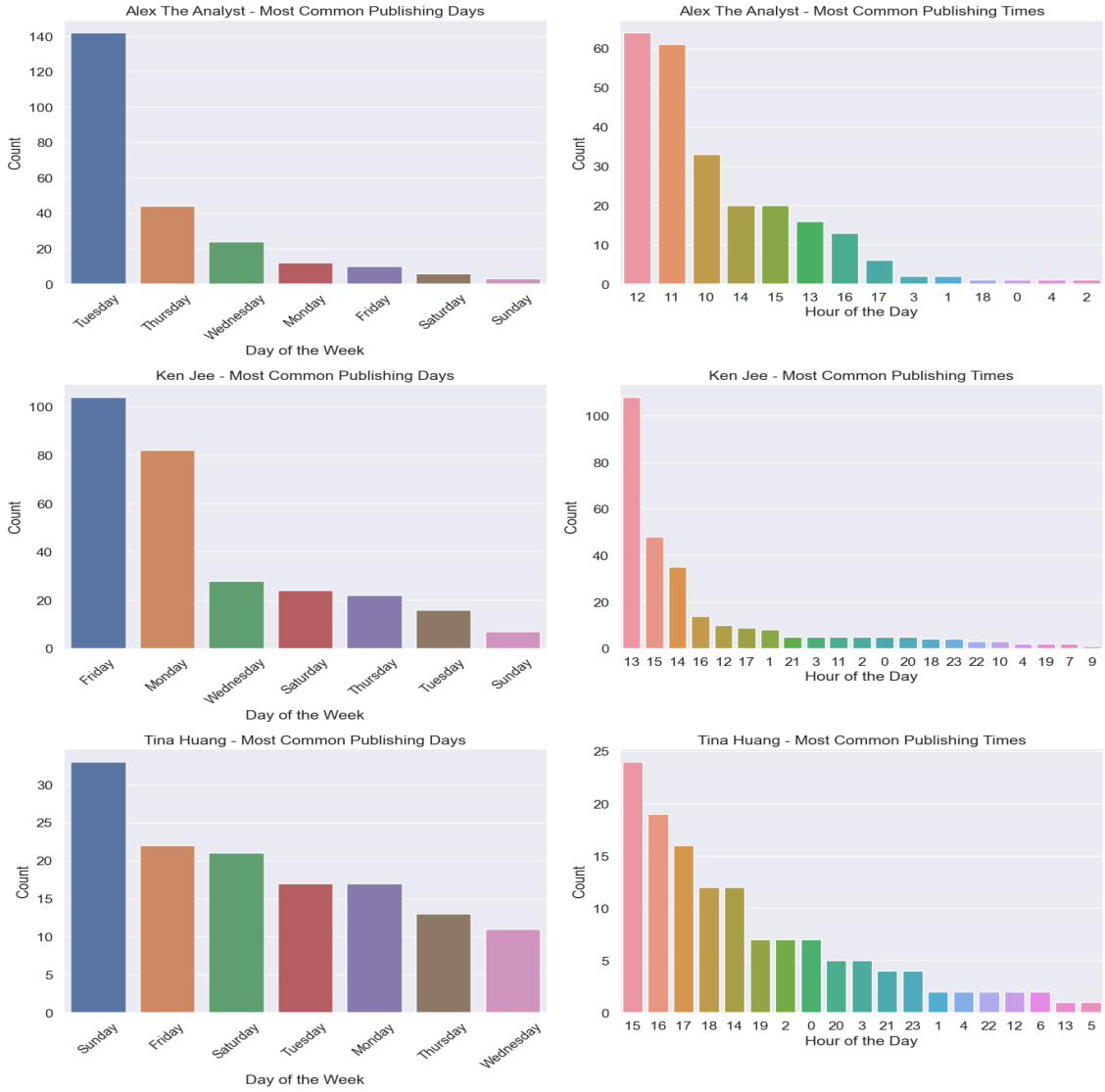
2. The bar chart reflects that Tina's content library is relatively smaller than the others, contributing to her lower monthly average and vice versa.

3. Interestingly, despite uploading fewer videos compared to Alex, Ken Jee has the highest total number of uploads among the channel owners. This could suggest that Ken might have started his channel before Alex.

4. Alex stands out as the most prolific uploader, with the potential to catch up to Ken's content library size if this level of consistency continues.

5. One intriguing aspect to explore further is why Ken's videos receive fewer views despite 4-5 monthly uploads. This raises questions about viewer engagement and the factors influencing video viewership, especially considering Ken's prominence in the data community.

➢ **Video Publishing Trends -** Revealing the strategic decisions behind the channels' preferred days and times for content distribution. This analysis highlights the synchronization between publishing schedules and audience behavior.



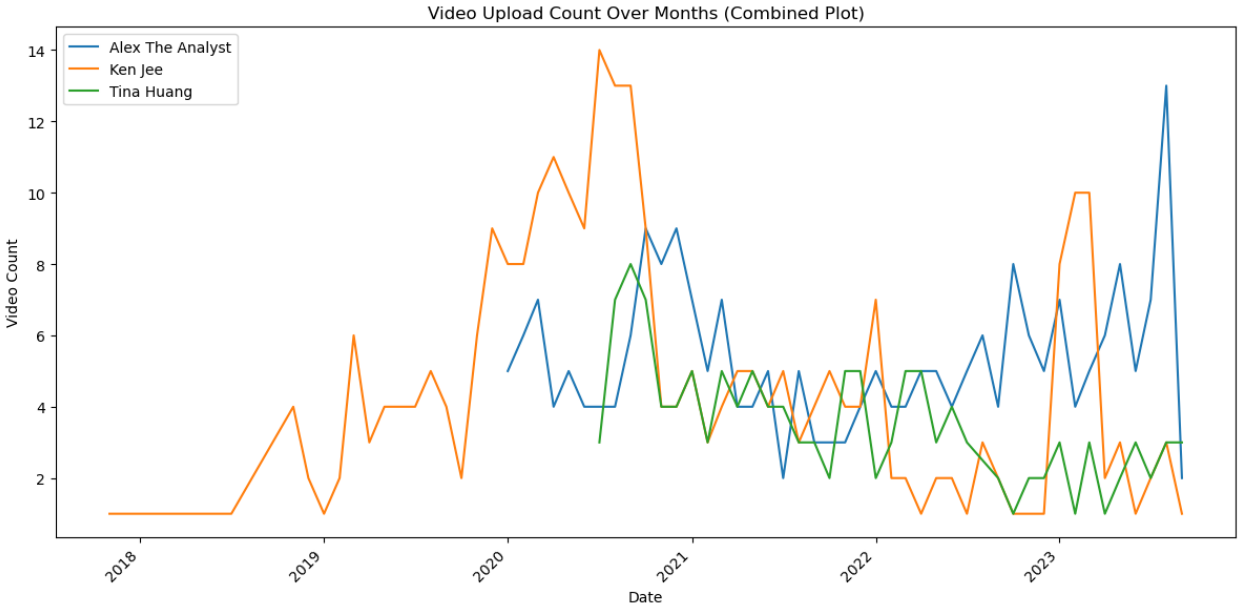
Here's what can be interpreted :

1. **Alex**: Alex's preferred posting time is typically between 11 AM and 12 PM on Tuesdays.

2. **Ken**: Ken tends to favor posting his videos around 1 PM, and he alternates between Fridays and Mondays for these uploads.

3. **Tina**: Tina's preferred video upload time falls between 3 PM and 4 PM, and she consistently chooses Sundays for her uploads.

➢ **Video Upload Trends** - Identifying a panoramic view of the evolution of the video uploads and a visual narrative of content consistency and strategic timing on a macroscopic scale.



In terms of consistency, we define it as the regularity or stability of video uploads over time. In our analysis spanning multiple years, we observed the following patterns:

1. **Alex**: Alex exhibits the most consistent video upload trajectory, characterized by a steady trend in video uploads throughout the analyzed period. This consistency is evident in their year-over-year growth in video uploads except in 2021 where a major dip can be witnessed that he eventually rose from.

2. **Tina**: Tina's video upload trajectory is fairly good, with consistent growth over time. While not as pronounced as Alex's trajectory, Tina maintains a steady increase in uploads with relatively minor fluctuations.

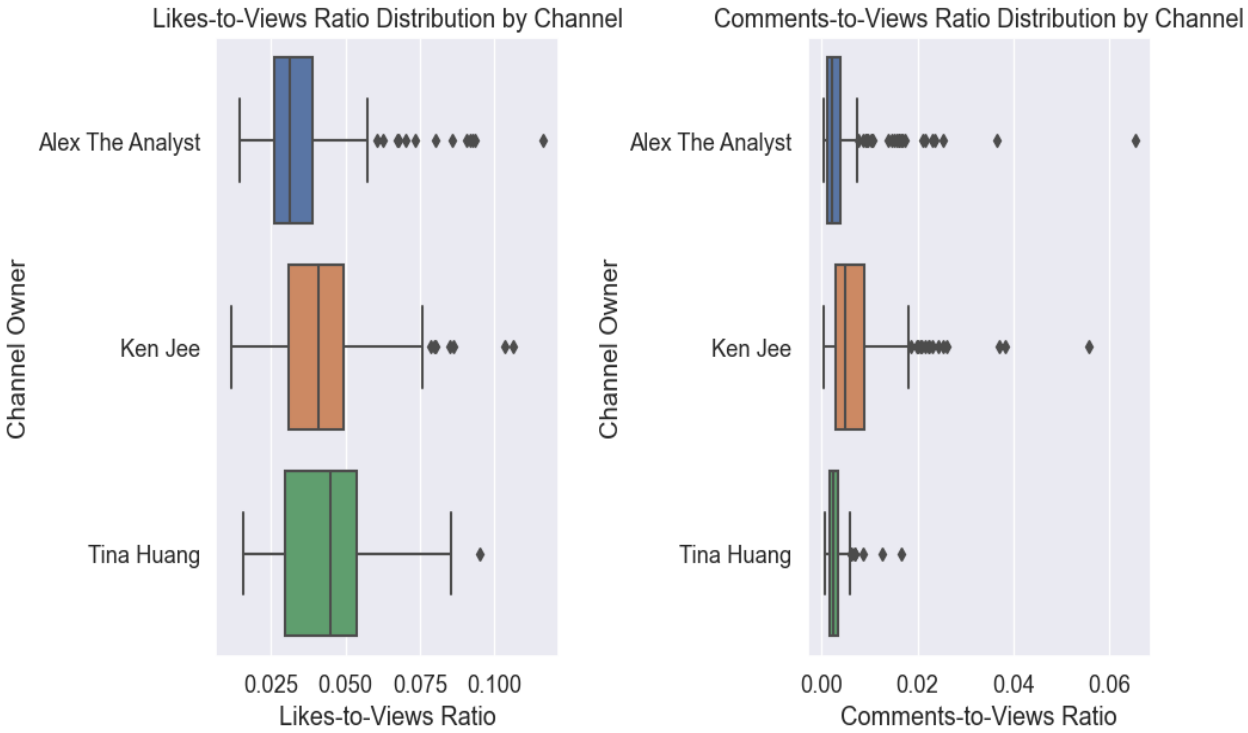
3. **Ken**: Ken's trajectory, on the other hand, displays a notable degree of variability. He has experienced several periods of both increased and decreased video uploads. Notably, 2020 marked a significant downfall in consistency for Ken, coinciding with the year when he uploaded the most videos. This raises questions about what factors might have contributed to this drop in consistency during that year.

➢ **Top Performing Videos Across Channels -** Identifying and ranking the top-performing videos from each channel based on a combined evaluation of views, likes, and comments.

	Channel Owner	Title	Views	Likes	Comments
0	Alex The Analyst	Data Analyst Portfolio Project | SQL Data Expl...	1415412	27231	3212
1	Alex The Analyst	Top 5 Reasons Not to Become a Data Analyst	909828	19625	1053
2	Alex The Analyst	SQL Basics Tutorial For Beginners | Installing...	913954	13151	1478
3	Alex The Analyst	What Does a Data Analyst Actually Do?	816358	16422	446
4	Alex The Analyst	Data Scientist vs Data Analyst | Which Is Righ...	741182	23339	843
5	Ken Jee	How I Would Learn Data Science (If I Had to St...	1393196	49026	1998
6	Ken Jee	The Best Free Data Science Courses Nobody is T...	345099	21173	622
7	Ken Jee	How I Would Learn Data Science in 2022 (If I H...	339940	14972	539
8	Ken Jee	3 Proven Data Science Projects for Beginners (...	324016	17691	525
9	Ken Jee	Beginner Kaggle Data Science Project Walk-Thro...	323648	7217	423
10	Tina Huang	How I would learn to code (if I could start over)	3290164	183683	2452
11	Tina Huang	How I consistently study with a full time job ...	2535286	110805	1985
12	Tina Huang	How to self study technical things	2271650	80284	1731
13	Tina Huang	How to learn to code FAST using ChatGPT (it's ...	1914569	68528	1685
14	Tina Huang	How to Completely Transform Your Life in 6 Mon...	887768	45415	645

The data pertaining to the high-performing videos of each channel offers valuable insights into their reach and impact. Although the scope of viewership is generally influenced by the channel's subscriber base, a deeper analysis of the engagement-to-reach ratio could reveal more nuanced aspects of each video's performance.

➢ **Audience Engagement Spectrum -** A comparative analysis, employing box plots to characterize the spectrum of audience engagement.

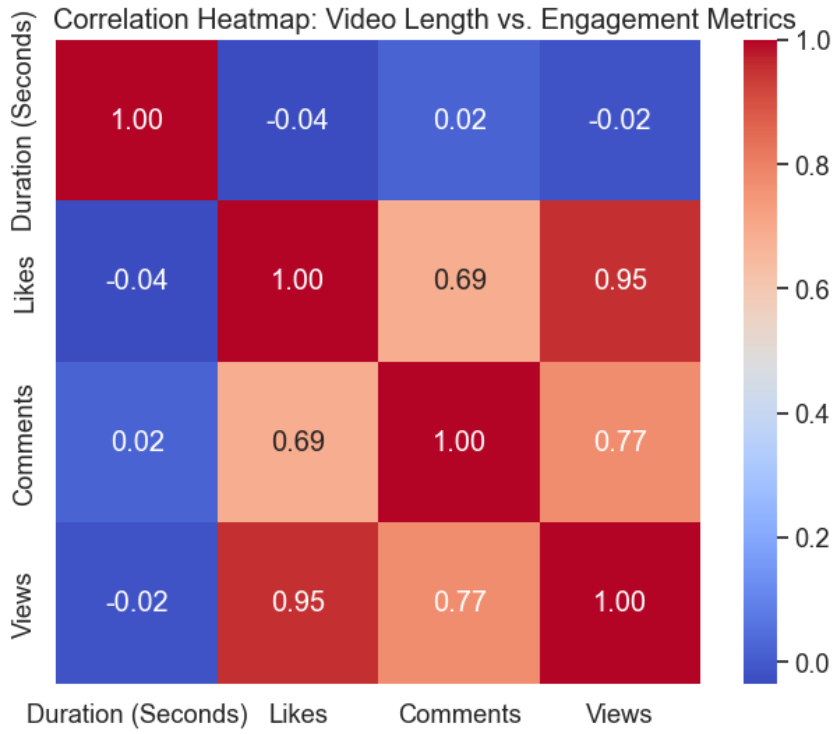


This analysis focuses on comparing three groups of data related to the likes-to-views and comments-to-views ratios.The median position for all three data groups indicates a skewed distribution of data. The imbalanced length of the whiskers on either side of the boxes and the presence of outliers only towards the right side further confirm it.

Likes-to-Views Ratio: The interquartile range (IQR) for Tina's data is the widest among the three groups, suggesting a greater variation in the ratio values which are predominantly gathered at the upper end, creating a short and tight section in that region and indicating a negatively skewed data distribution. She also has the highest likes-to-views ratio, as evidenced by the upper extreme value in her data. Ken's ratio is relatively smaller, while Alex's ratio is the smallest among the three groups.

Comments-to-Views Ratio: The interquartile range (IQR) for Ken's data is the widest among the three groups, suggesting a greater variation in the ratio values which are predominantly gathered at the lower end, creating a short and tight section in that region and indicating a positively skewed data distribution. Ken also has the highest comments-to-views ratio, as evidenced by the upper extreme value in her data.Alex's ratio is relatively smaller, while Tina's ratio is the smallest among the three groups.

➢ **Video Length Impact on Viewer Engagement -** Exploring the relationship between video duration and engagement metrics to evaluate the implication of content duration and its effects on viewer likes and comments.



1. The correlation coefficients for Duration to Views (-0.02), Duration to Comments (0.02), and Duration to Likes (-0.04) reveal negligible correlations, indicating that, within this data set, video duration has a minimal impact on the metrics of views, comments, and likes.

2. A high correlation exists between video metrics, with a 0.95 correlation coefficient for views to likes and a 0.77 coefficient for views to comments, indicating that videos with higher views not only garner more likes but also attract a greater number of comments.

- **Descriptive Statistics for the engagement metrics -** Calculating descriptive statistics (mean, median, standard deviation, etc.) for key engagement metrics such as likes, comments to draw a summary of the central tendencies and variability of these metrics, and to understand the typical performance of each channel.

Descriptive Statistics for Likes (by Channel):

	count	mean	std	min	25%	50%	75%	max
Alex The Analyst	241.0	2734.141079	4475.401655	93.0	521.00	1047.0	2838.00	27236.0
Ken Jee	283.0	1117.667845	3524.020122	18.0	186.50	391.0	854.50	49026.0
Tina Huang	134.0	8375.597015	21032.836446	73.0	724.75	2624.5	6856.25	183685.0

Descriptive Statistics for Comments (by Channel):

	count	mean	std	min	25%	50%	75%	max
Alex The Analyst	241.0	188.477178	374.538105	6.0	38.00	84.0	194.0	3813.0
Ken Jee	283.0	98.586572	154.211937	1.0	28.50	60.0	115.5	1998.0
Tina Huang	134.0	267.007463	390.462315	6.0	59.25	145.0	260.0	2452.0

Summary and Key Achievements

This meticulous analysis scrutinized three YouTube channels to understand the nuanced variables affecting their performance, content strategy, and audience engagement. Notable findings include:

Audience Reach vs. Engagement: Tina's channel stands out for having high views despite a smaller content library, pointing to superior audience engagement. On the other hand, Ken's channel, despite its frequent uploads, struggles with lower viewership.

Correlation Insights: Strong correlations were discovered between subscriber count and views (positive) as well as subscriber count and total number of videos (negative). The data implies that higher subscriber bases contribute to more views but potentially fewer, higher-quality uploads.

Content Frequency: Alex's consistent uploading schedule positions him as a potential leader in content volume, suggesting future growth potential.

Video Length Impact: Contrary to popular belief, the duration of videos showed negligible correlation with engagement metrics like views, comments, and likes, debunking the notion that video length significantly impacts these metrics.

Audience Engagement Metrics: A granular view into likes-to-views and comments-to-views ratios revealed asymmetrical data distributions for all three creators, with Tina having the highest likes-to-views ratio and Ken having the highest comments-to-views ratio.

Video Publishing Trends: Each creator has a preferred time and day for uploads, potentially optimizing their reach based on their audience's viewing habits.

The study has successfully uncovered valuable insights that not only serve as performance indicators but also contribute to a broader understanding of content strategy and audience behavior in the YouTube ecosystem. Future work could extend this analysis to include other factors such as video genres, audience demographics, and seasonal trends, to name a few.