

T1-24-25-DAS 732-Data Visualization

Assignment 1: Analysis of the Twitch Dataset

Udayagiri Narayana Srimanth
IMT2022052
Udayagiri.Srimanth@iiitb.ac.in

Khadarabad Tahir Mohammed
IMT2022100
Khadarabad.Mohammed@iiitb.ac.in

Pradyun Devarakonda
IMT2022525
Devarakonda.Pradyun@iiitb.ac.in

Index Terms—Twitch, Streamers, Views, Followers

I. INTRODUCTION

We worked on analyzing data and creating visualizations of the Top Streamers on Twitch dataset. The dataset contains data on Twitch's top 1000 streamers(based on the number of followers) from August 2020. It has a total of 11 columns.

- 1) Channel Name: Display name of the channel
- 2) Watch Time(Minutes): Total watch time on the channel in that particular year
- 3) Stream time(minutes): Total time streamed on the channel in that particular year
- 4) Peak viewers: Max viewers reached by a stream on the channel in that particular year
- 5) Average viewers: Average viewers across streams on the channel in that particular year
- 6) Followers: The number of followers for each channel
- 7) Followers gained: The number of followers gained for each channel in that particular year
- 8) Views gained: The number of views gained for each channel in that particular year
- 9) Partnered: Whether the channel is a Twitch Partner or not
- 10) Mature: Whether the channel is marked as mature content or not
- 11) Language: The primary language in which the streamer's content is broadcast.

To effectively categorize streamers(based on the number of followers) in the dataset, we divided them into four categories:

- Category 1 - 0 to 100k followers
- Category 2 - 100k to 500k followers
- Category 3 - 500k to 1M followers
- Category 4 - 1M to 10M followers

This categorization serves two main purposes:

- To segment the wide range of follower counts, as the numbers were too large to consider collectively.
- To reflect social perceptions of popularity, where 100k, 500k, 1M, and 10M followers are commonly regarded as significant milestones for streamers.

Of the 11 columns, 2 columns (Channel Name and Language) are of string data type, 2 columns (Partnered and Mature) are of boolean data type, and the remaining 7 columns are of integer data type.

II. TASK DIVISION

We divided our work into three major tasks:

- T1: Overview of Key Metrics
- T2: Growth Trends and Audience Engagement
- T3: Language-Specific Insights

As a first step, we created a correlation matrix of all 10 columns to analyze the relationships between the variables.

	Watch time(Minutes)	Stream time(minutes)	Peak viewers	Average viewers	Followers	Followers gained	Views gained
Watch time(Minutes)	1	-0.119540289	0.650332429	0.650332429	0.428801725	0.428801725	0.276466509
Stream time(minutes)	0.150587901	1	0.349247788	0.349247788	0.352529114	0.352529114	0.244296867
Peak viewers	0.582796449	0.349247788	1	0.349247788	0.352529114	0.352529114	0.244296867
Average viewers	0.476165001	0.349247788	0.349247788	1	0.420937465	0.420937465	0.244296867
Followers	0.518467583	0.352529114	0.352529114	0.420937465	1	0.715618458	0.244296867
Followers gained	0.518467583	-0.158164785	0.470114714	0.420937465	0.715618458	1	0.244296867
Views gained	0.529862014	0.064370002	0.298062634	0.250348872	0.276466509	0.244296867	1

Fig. 1. The Correlation matrix of the Dataset

III. T1: OVERVIEW OF KEY METRICS

This section provides analysis of key performance metrics for the top Twitch streamers, including follower counts, total views, and watch time. By examining these metrics, we aim to understand the correlation between them. The insights gained offer a foundational overview of streaming trends and audience interaction on the platform.

A. Observation 1

The majority of the top 1000 streamers on Twitch fall within Category 2 (100,000 to 500,000 followers), comprising more than 55% of the dataset. This is followed by Category 1 (1,000,000 to 10,000,000 followers), Category 3 (500,000 to 1,000,000 followers), and Category 4 (0 to 100,000 followers).

Number of streamers belonging to each category



Count of Channel: 1,000
Streamer Category: Category 1, Category 2, Category 3, Category 4

Fig. 2. Streamers of each category

The distribution of these categories is depicted through a pie chart (Fig.2), with each colour corresponding to one of the categories. The size of each segment illustrates the proportion of the respective category, and the colour scheme was selected to reflect the aesthetic used by Twitch in their annual Twitch Recaps.

At the time of data collection in August 2020, Tfue was the top streamer, with over 8.9 million followers, while the streamer ranked 1000th, voicetv, had 3,660 followers. The range of follower counts within the top 1000 highlights significant variability in audience size. This dataset allows for the examination of correlations between variables such as follower count, watch time, and stream time.

Given that the dataset is of the top 1000 streamers, any trends observed can be used to analyze how the average viewer interacts with users who stream professionally because it excludes the vast majority of accounts, operated by viewers, where stream times and follower counts are virtually zero. This prevents skewing the data, and allows us to apply our insights to the platform as a whole.

B. Hypothesis 1

High-performing streamers are more likely to be Twitch Partners. Successful streamers on Twitch are heavily engaged in the platform's revenue streams, which are primarily derived from advertisements and channel subscriptions. To access these revenue sources, streamers must achieve partnership status, which provides them with a share of the platform's earnings, a direct point of contact at Twitch, better channel customization and special promotional opportunities.

This hypothesis is validated by the data, it shows that 97.8% of the top 1000 streamers are partnered, supporting the conclusion that partnership is nearly universal among this group.

C. Observation 2

We observe that streamers with higher follower counts are more likely to maintain higher average watch time and viewer counts(Fig.3, Fig.4, Fig.5). This phenomenon is logical, as streamers who consistently attract large audiences tend to cultivate a loyal viewer base. These viewers are more inclined to return for subsequent streams, contributing to sustained high viewership and prolonged watch times. Streamers with larger followings often produce content that is more engaging, diverse, or of higher quality, which in turn fosters viewer retention. This feedback loop of increased audience engagement and viewer retention likely reinforces the correlation between high follower counts and elevated average watch time and viewer numbers.

The figure(Fig.3) is a line chart in which average watch time is plotted against the number of followers, excluding Category 4 (1,000,000 to 10,000,000 followers). The exclusion of the top 204 streamers in Category 4 was necessary, as their inclusion would have resulted in a disproportionately steep linear regression slope. This provides a more accurate representation of the correlation between these two variables.

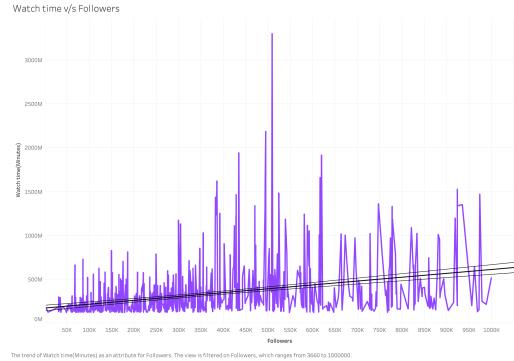


Fig. 3. Watchtime vs Followers

The colour scheme was chosen based on Twitch's colour palette for visual appeal.

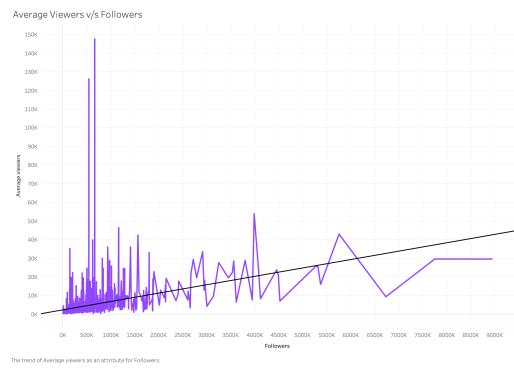


Fig. 4. Average Viewers vs Followers

The figure(Fig.4) is a line plot, where every point on it represents average viewership at a specific follower count. As anticipated, the trend line shows a positive correlation between follower count and average viewership. The colour scheme was similarly selected based on Twitch's colour palette for coherence and visual consistency across figures.

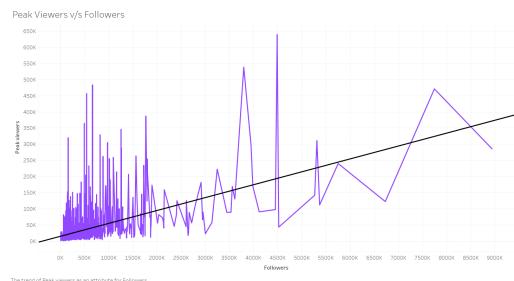


Fig. 5. Peak Viewers vs Followers

The graph(Fig.5) presents a line plot in which each data point represents the peak viewership of a channel at a given follower count. The trend line illustrates a positive correlation between follower count and peak viewership. The colour scheme was chosen to align with Twitch's colour palette.

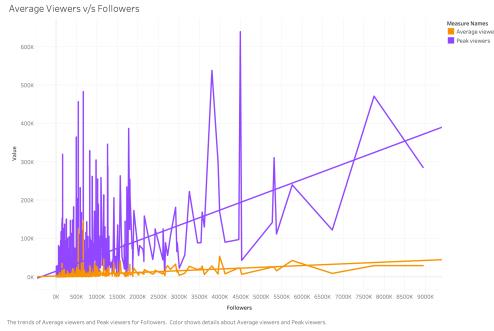


Fig. 6. Peak Viewers and Average Viewers vs Followers

The graph (Fig.6) illustrates that peak views exhibit a higher initial value and increase more rapidly with follower count compared to average views. This observation is supported by the trend lines, where the slope and intercepts of the peak views trend line (purple) are significantly greater than those of the average views trend line (orange).

D. Observation 3

The impact of stream time across various metrics reveals nuanced patterns in the behaviors of streamers and their audience. As streamers increase their follower counts, the average amount of time they stream tends to decrease(Fig.7).

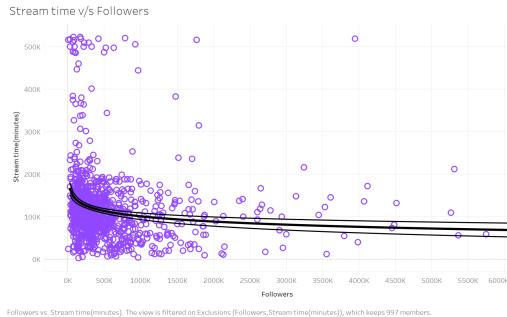


Fig. 7. Stream time vs Followers

Streamers with fewer than 500,000 followers tend to have significantly higher stream times, while those with greater follower counts exhibit shorter stream durations. This trend suggests that streamers with higher follower counts have likely already invested substantial time in growing their audience and now adopt a more comfortable streaming schedule, relying on their established viewership base.

The figure(Fig.7) is a scatter plot, with each point representing the number of hours streamed by a streamer relative to their follower count. A logarithmic curve best fits the data, indicating that streamers with lower follower counts typically invest more hours than those with a larger audience. The colours are aligned with Twitch's colour palette.

We observe that average viewership tends to decrease as stream time increases(Fig.8). This outcome is expected, as extremely long streams, such as those lasting 15 hours, are unlikely to retain viewers for the entire duration. Instead,

viewers tend to rotate in and out, depending on time zones and the content being streamed during particular segments, resulting in fluctuating viewership and lower overall averages. The scatter plot(Fig.8) represents average viewership

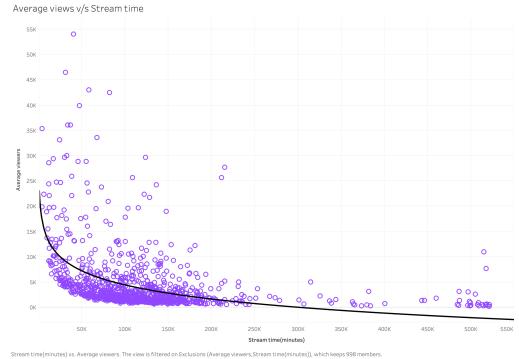


Fig. 8. Average views v/s Stream time

for different stream duration, where each point corresponds to a specific time streamed. A logarithmic trend line best fits the data, showing an initial sharp decline in viewership, which stabilizes after a certain threshold. Two outliers: dota2ti and dota2ti_ru were removed due to their low stream times but unusually high viewership. These channels stream "The International", a yearly Dota 2 tournament that is rapidly growing in viewership every year. The colour scheme follows Twitch's colour palette.-

Lastly, we observe a gradual increase in average watch time as stream duration lengthens(Fig.9). Watch time is a key metric for advertisers, who prioritize platforms where viewers are highly engaged for longer periods. This trend highlights why streamers may choose to extend their stream times in an effort to enhance monetization potential through advertising revenue.

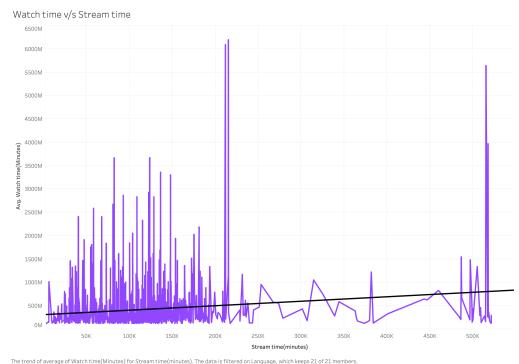


Fig. 9. Average Watch time v/s Stream time

The line plot(Fig.9) illustrates how average watch time evolves with increasing stream duration. The linear regression line shows a slight upward trend, indicating a slow but steady increase in watch time as stream length grows. The colours are based on Twitch's colour palette.

E. Observation 4

According to Twitch, streamers are labeled as mature if their content includes themes such as mature-rated games, sexual content, drug use, violence, excessive profanity, or gambling. Out of the top 1000 streamers, 230 are categorized as mature.

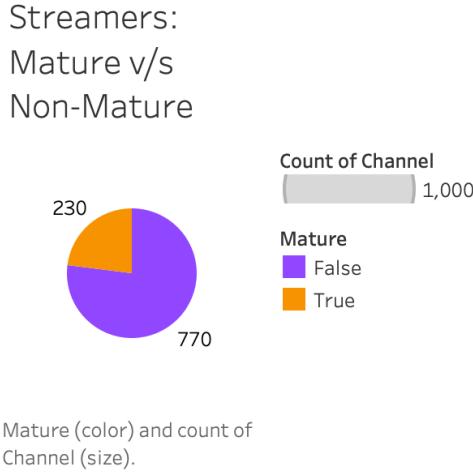


Fig. 10. Number of streamers: mature vs. non mature

The accompanying figure(Fig.10) is a pie chart illustrating the proportion of mature versus non-mature streamers, with colour coding representing each category. The size of the segments corresponds to the number of streamers in each category, and the colour scheme is consistent with Twitch's official palette.

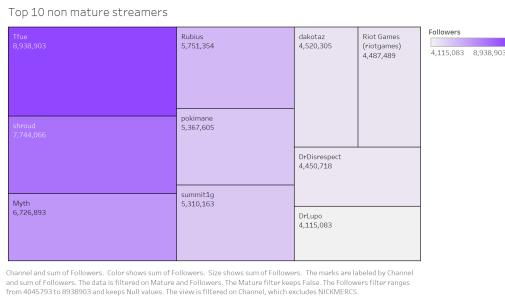


Fig. 11. Top 10 Non-mature streamers(by followers)

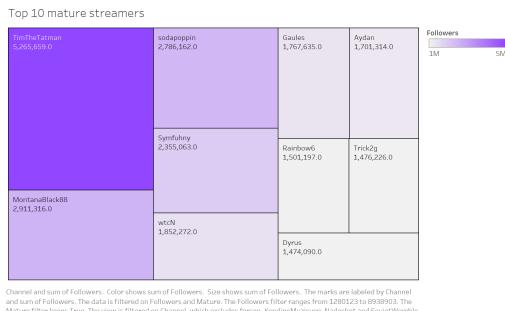


Fig. 12. Top 10 Mature streamers(by followers)

When analyzing the top streamers in both mature and non-mature categories, we find that non-mature streamers such as Tfue, Myth, Rubius, and Pokimane, initially built their following through family-friendly games like Fortnite. Although some later transitioned to games like Valorant or CS:GO, they maintained family-friendly content to retain their younger audience base.

Conversely, top mature streamers primarily focus on mature-rated games such as Call of Duty and Rainbow Six Siege, attracting older audiences. Some of these streamers also stream Valorant or CS:GO, but their target demographic is not younger viewers, which allows them more flexibility in content maturity.

Further analysis of average viewers versus followers(Fig.13) shows that around the 800,000-follower mark, the trend lines for mature and non-mature content intersect. Below this threshold, mature streamers tend to have fewer average viewers compared to their non-mature counterparts. However, once this threshold is surpassed, content maturity has little impact, and average viewership increases across the board for both categories. The accompanying scatter plot represents

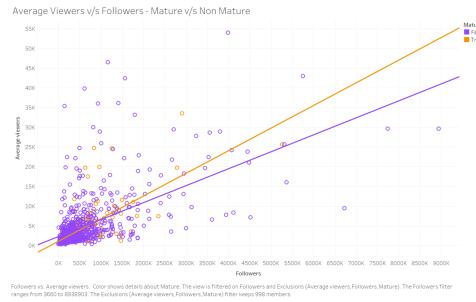


Fig. 13. Average views vs. Followers: mature vs. non mature

streamers' follower counts and their corresponding average viewer counts, with separate colours indicating mature and non-mature categories. Trend lines for both groups are shown, with two outliers (dota2ti and dota2ti_ru) excluded to prevent skewing. The colour palette adheres to Twitch's colours.

IV. T2: GROWTH TRENDS AND AUDIENCE ENGAGEMENT

This section examines growth trends and audience engagement metrics to uncover patterns and insights into streamer performance and viewer behavior.

A. Observation 5

The regression line in the plot of the percentage of followers(Fig.14) gained versus the number of followers indicates that smaller channels are more likely to experience a higher percentage increase in followers compared to larger channels. This suggests a saturation effect, where follower growth becomes more gradual as a streamer accumulates a larger audience. Although the general trend is downward, the distribution of data points reveals that growth stagnation is possible at any follower count. This suggests that follower gain is not necessarily guaranteed for larger channels, just as smaller streamers may still experience rapid growth.

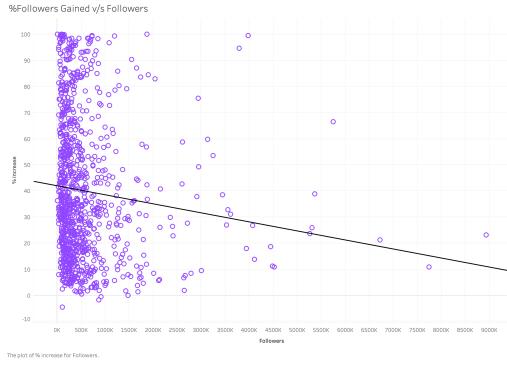


Fig. 14. Percentage Followers Gained vs Followers

The graph is a scatter plot, where each point represents a streamer's follower count and the percentage increase in followers over a year. The colour palette adheres to Twitch's official branding.

Contrary to expectations, the trend for average followers gained versus stream time(Fig.15) is also downward. This suggests that streaming for longer durations does not necessarily correlate with an increase in followers. From prior analysis, we know that smaller streamers tend to stream for longer hours. However, follower growth depends more on expanding the viewer base than on simply increasing stream time.

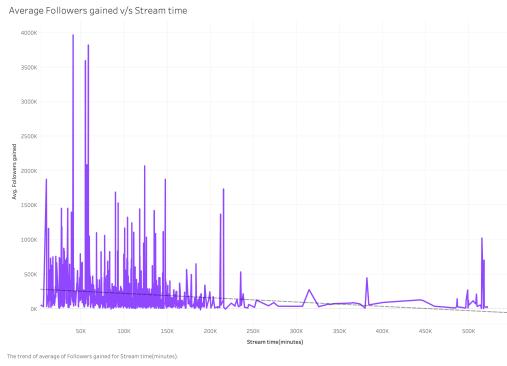


Fig. 15. Average Followers Gained vs Stream Time

This line chart represents the average number of followers gained for specific stream times. The colour scheme is based on Twitch's palette.

B. Hypothesis 2

We hypothesize that viewers do not spend significantly more time on channels with higher follower counts. This may be because larger channels often have a broader and more transient audience, with viewers cycling in and out more frequently. Additionally, smaller channels might cultivate a more engaged and loyal viewer base, leading to longer watch times per viewer. Factors such as highly interactive content or personal connections with the audience in smaller channels could contribute to this trend, whereas larger streamers may experience more passive engagement due to the size and diversity of their audience.

In fact, the data(Fig.16) shows that average watch time per viewer decreases steadily as follower numbers increase. We can also see the distribution at any number of followers is quite wide, meaning engaging content is agnostic to follower counts.

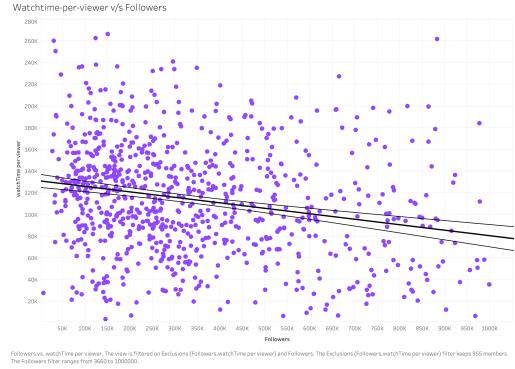


Fig. 16. Average Watch time per viewer in a Year vs Followers

Average watch time per viewer was calculated by dividing the average watch time by the average number of viewers. This data was plotted on a scatter plot, with the dots representing individual channels. The colour scheme, based on the Twitch palette, enhances visual clarity. Forty-four outliers with exceptionally high per viewer watch times were removed from the dataset to avoid skewing results. These outliers could be due to several factors like:

- Exceptionally engaging or interactive content
- Viewer incentives, such as Twitch Drops
- Alternative strategies that, while compliant with Twitch's Terms of Service, do not reflect typical viewer behavior

Similar to Observation 2, Category 4 streamers were excluded to prevent the plot from being skewed. The trend observed is likely to be consistent across the platform.

C. Observation 6

The plot examining the relationship between average views gained and stream time(Fig.17) reveals little to no significant correlation between the two variables. This suggests that merely increasing the number of hours a streamer broadcasts does not directly result in an increase in viewership. This insight highlights the limitation of relying solely on stream duration as a strategy for growth, implying that factors other than stream time, such as content quality, audience engagement, and external promotion play a more critical role in driving viewership. The graph is a line plot, with the line representing the average views gained corresponding to specific stream durations. One outlier, Fextralife, was excluded due to its anomalous data: 670,137,548 views gained with only 147,885 minutes streamed. This channel has faced accusations of inflating its viewership by embedding its Twitch stream within its external website, which features wikis, guides, and reviews for popular games. This contributed to an inflated view count that skews statistical analysis. The colour palette for the plot aligns with Twitch's colours.

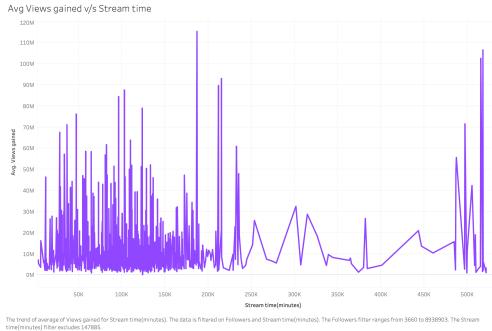


Fig. 17. Average Views Gained vs Stream time

V. T3: LANGUAGE-SPECIFIC INSIGHTS

The content viewers engage with on Twitch varies significantly depending on the language of the stream. Language plays a crucial role in shaping audience preferences, engagement levels, and community dynamics. This section explores language-specific insights, highlighting key trends and differences in viewer behavior across various language groups on the platform.

A. Observation 7

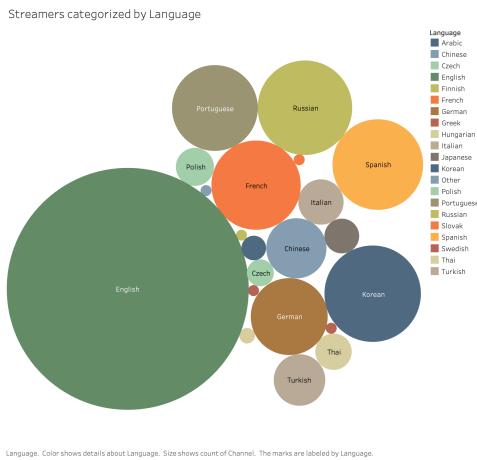


Fig. 18. Streamers: Categorized by Language

The majority of streamers, accounting for 48.5%, conducted their broadcasts in English, followed by Korean at 7.7%, Russian at 7.4%, and other languages as depicted in the diagram (Fig.18). To visually represent the distribution of languages, we employed a bubble chart, where the size of each bubble corresponds to the proportion of streamers using that language. The colour scheme for the chart was selected from the Tableau colour palette to ensure clarity and visual consistency.

B. Observation 8

Let us examine the top ten streamers by followers within each language category to better understand viewing preferences across different linguistic demographics.

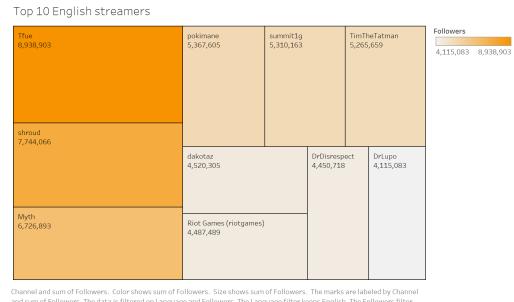


Fig. 19. Top 10 Streamers: English

In the English-speaking community, the leading channels predominantly feature streamers who gained prominence through Fortnite. Additionally, League of Legends enjoys significant popularity, with streamers of Valorant and CS:GO also ranking highly.



Fig. 20. Top 10 Streamers: Korean

In the Korean market, Faker, a renowned professional League of Legends player, holds a commanding lead in follower count. The top channels also include Just Chatting streamers, many of whom engage with League of Legends content.

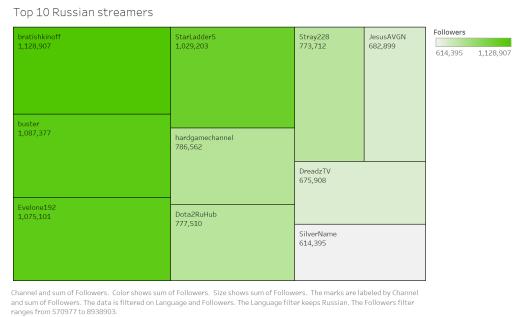


Fig. 21. Top 10 Streamers: Russian

For the Russian audience, Counter-Strike streamers dominate the top ranks, with Just Chatting content also holding substantial viewer interest. Dota 2 remains a significant draw within this demographic.

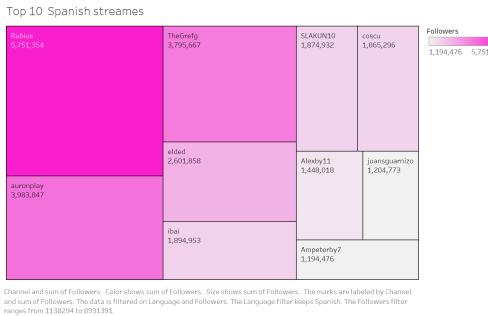


Fig. 22. Top 10 Streamers: Spanish

In the Spanish-speaking community, Fortnite and Minecraft are the most popular among the top streamers. Additionally, content related to League of Legends and Just Chatting also garners considerable attention.

C. Observation 9

We plotted the 'loyalty' of followers of different content(Fig.23,24). The loyalty of followers is higher if they have a higher average number of views for the same number of followers. This loyalty could be compared by comparing the slopes(and the intercepts) of the followers vs average views.

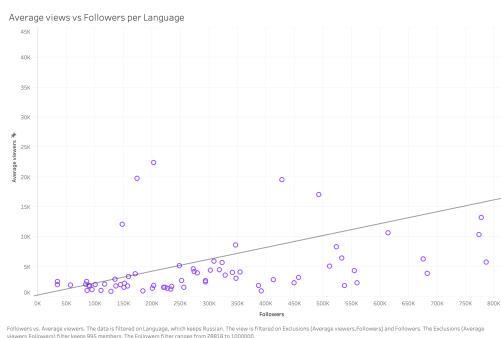


Fig. 23. Average Views vs Followers: Russian

The graph(Fig.23) illustrates that Russian streamers exhibit a steeper slope compared to their counterparts in other languages, suggesting a greater degree of viewer loyalty and engagement within the Russian-speaking audience.

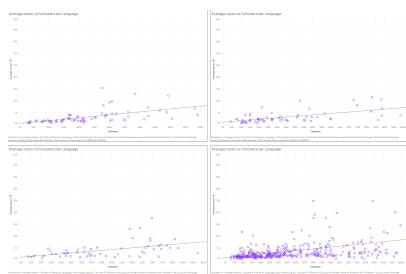


Fig. 24. Top Left - Korean, Top Right - French, Bottom Left - Spanish, Bottom Right - English

D. Observation 10

In the top 5 most popular languages(based on the number of streamers) Fig.25 shows that English content has more mature streamers than any other content, followed by French, Spanish, Russian, and Korean.

Korean content has fewer mature streamers than any other content.

Mature and Non-mature content Streamers: per Language

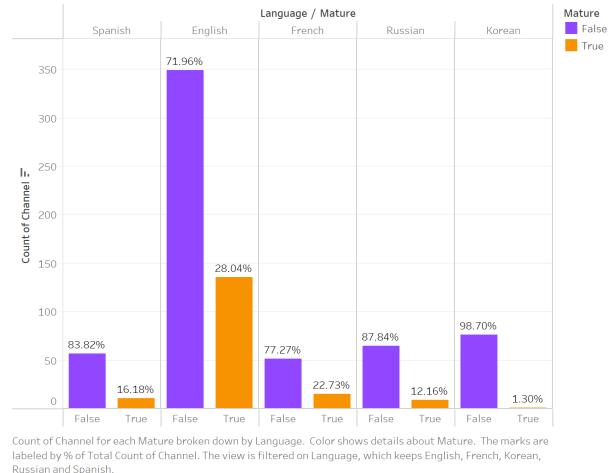


Fig. 25. Percentage Of Mature vs Non-mature streamers: categorized by language

E. Observation 11

From(Fig.26) we can tell that Spanish followers grew the most (concerning the percentage followers gained), followed by Italian followers and Portuguese followers.

Fig.26 also depicts the percentage followers gained concerning mature and non mature content per language.

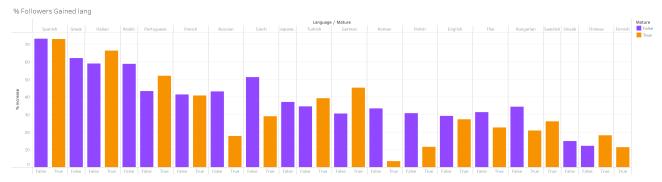


Fig. 26. Percentage Followers gained: categorized by Language

VI. VISUALIZATIONS USED

We've used plenty of visualizations in the proper context.
The visualizations used are:

- 1) Scattered Plots
- 2) Line Plots
- 3) Tree Maps
- 4) Pie charts
- 5) Bar Graphs
- 6) Bubble Graphs

VII. CONTRIBUTIONS

The division of tasks was collaboratively decided, and responsibilities were allocated as follows:

- 1) Information and General Formatting: This task was undertaken collaboratively by all members to ensure consistency and coherence throughout the document.
- 2) T1 & T2: Tahir and Pradyun jointly conducted the analysis and visualizations for these sections. Our exploration of the data made more sense to do collaboratively, as they were closely linked.
- 3) T3: Narayana focused on this task independently. The insights from this section were integrated with the rest of the project to form a cohesive final report.