Twitch Channel Analysis

By Allen Yu, Lilah Favour

Summary of research questions

1. What is the percent growth rate of each of our favorite channels over the last year? We are interested in this because we want to see how they have grown and compare them to each other.
    a. Every channel we analyzed except one grew by more than 1300% since February 2020
2. How has the pandemic influenced the viewership of the entire Twitch platform? Has there been an increase or decrease, and if so by what magnitude? We are interested because we want to see if there is a dramatic change in the number of viewers on Twitch due to the extra free time many people have.
    a. Since the start of quarantine in March of 2020, the average viewers on Twitch a month has increased significantly in comparison to pre quarantine.
3. Are there games that result in consistently high viewers across multiple channels or does the viewership for a game depend on the streamer? We are curious to see if there is a clear answer in whether the game played or the streamer playing it has greater influence on viewer numbers.
    a. Mincraft and Among Us has the highest viewership

Motivation and background

We are interested in analyzing data about Twitch because Twitch is one of our main sources of entertainment and we spend a lot of time on it. We have grown to love many Twitch streamers and want to view the statistics on their streams and see if there are any trends among them. We believe this data can be useful for friends who are interested in starting to stream with growing their channels.

Dataset

The datasets we will be using are CSV files downloaded for individual twitch channels from https://sullygnome.com as well as data web scraped from https://sullygnome.com. Since on the data it says no scraping, we emailed the owner and received explicit permission to scrape data from sullygnome.

Methodology (algorithm or analysis)

viewers_vs_games(channel_dict): Takes in a dictionary channel_dict and plots games based on average viewers and stream time for each individual channel.

covid_impact(channel_dict): Takes in a dictionary channel_dict and plots the data that shows how the pandemic shapes the entirety of Twitch, and the individual streamers.

custom_time_data(start_m, start_y, end_m, end_y): Takes in 4 integers, start month, start year, end month, and end year. Then scrape the data from a specific time frame and plot both average viewers and peak viewers.

get_soup(url): Takes in a string URL and returns the soup object for web scraping purposes.

yearly_growth_rate(channel_dict): Takes in a dictionary channel_dict and returns the growth rate of average viewers for that channel over the last 12 months. If the channel hasn't been active for that long, it calculates from the earliest month available.

combine_csv(channel_dict): Takes in a dictionary channel_dict and combines all related csv files into one.

agg_columns(csv_path): Takes in a string representing the CSV file path and returns a DataFrame with columns filtered.

get_time_range(start_m, start_y, end_m, end_y): Takes in 4 integers, start month, start year, end month, and end year. Return a list within a range of time.

Results

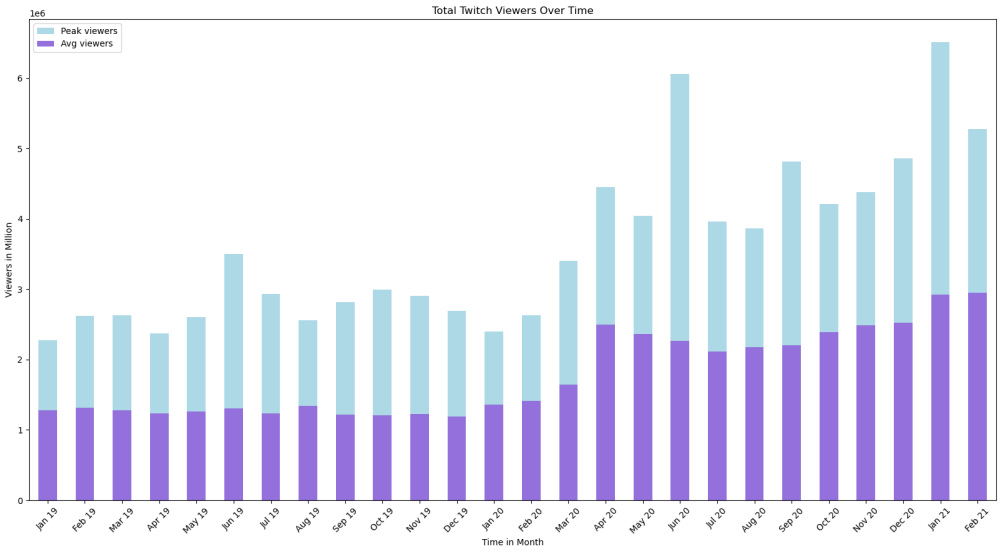
Growth Rate (Research Question 1)
Results:
The growth rate for jacksepticeye from January 2020 to February 2021 is 1352.46%
The growth rate for karljacobs from April 2020 to February 2021 is 152961.9%
The growth rate for Philza from January 2020 to February 2021 is 1789.24%
The growth rate for Pokimane from January 2020 to February 2021 is 48.47%
The growth rate for RanbooLive from June 2020 to February 2021 is 2263033.33%
The growth rate for Sapnap from April 2020 to February 2021 is 30488.93%
The growth rate for Sykkuno from January 2020 to February 2021 is 3712.34%
The growth rate for tommyinnit from January 2020 to February 2021 is 32188.17%
The growth rate for Tubbo from January 2020 to February 2021 is 184459.46%
The growth rate for WilburSoot from January 2020 to February 2021 is 1840.18%

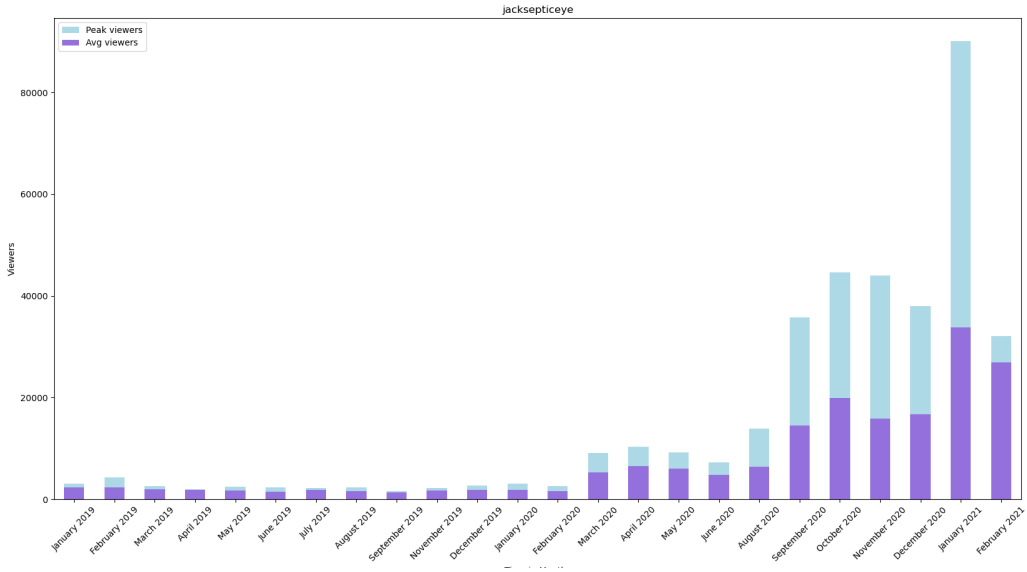
We found that the growth rate for every channel except Pokimane had a growth rate of over 1300% in the past year. Pokimane only grew by 48.47%. It is interesting to note that Pokimane is the only female streamer on this list, and the biggest one on Twitch, being the only female streamer in the top 50 twitch channels ranked by average viewers, according to https://twitchstats.net.

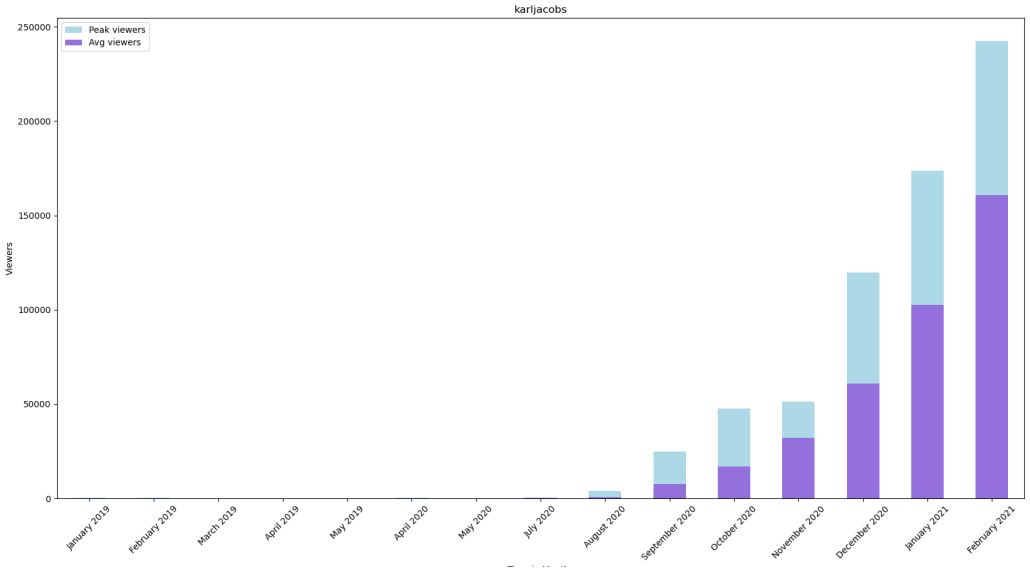
Many of the twitch channels we analyzed experienced astronomical growth in 2020, with the most being RanbooLive with over a 2 million percent growth rate. Looking at the numbers produced, the way we chose to analyze growth rate is accurate but makes established, big channels seem small in comparison to channels that might be around the same size but came into popularity at a much more rapid pace. Most of these channels have a similar average

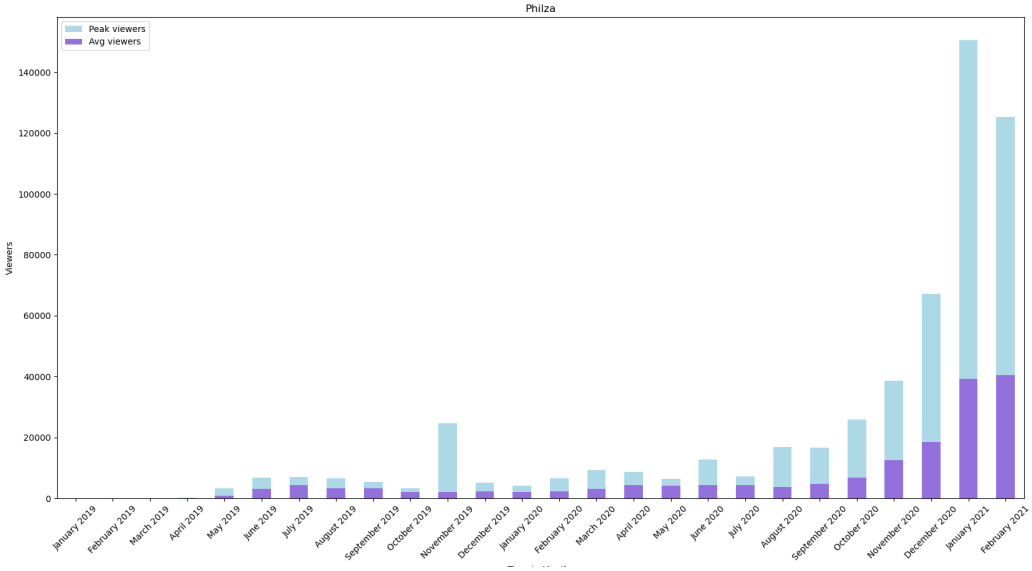
viewer count but some gained that following over a much shorter time period that others, as apparent by the larger growth rates.

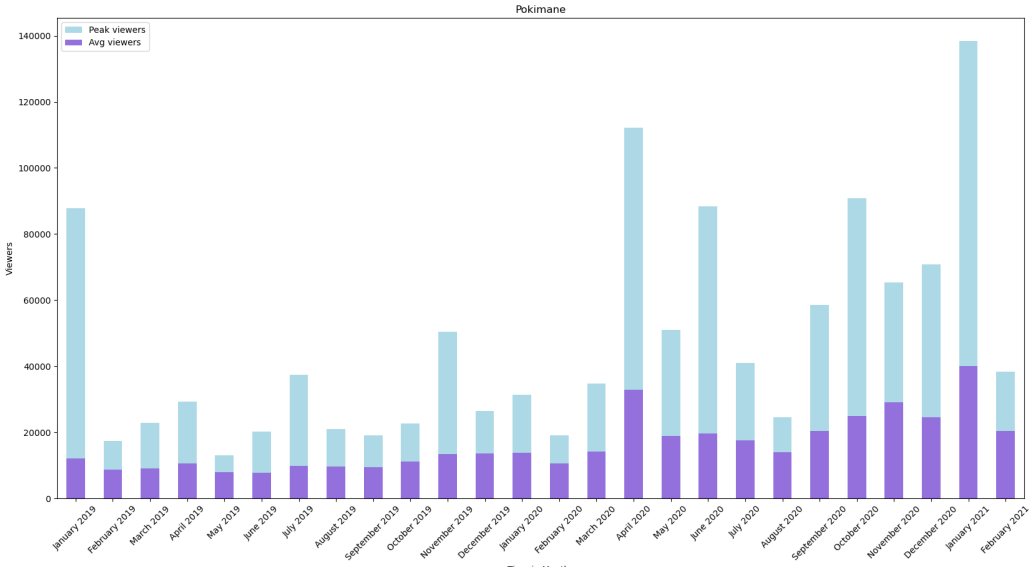
Viewership in the Pandemic (Research Question 2)

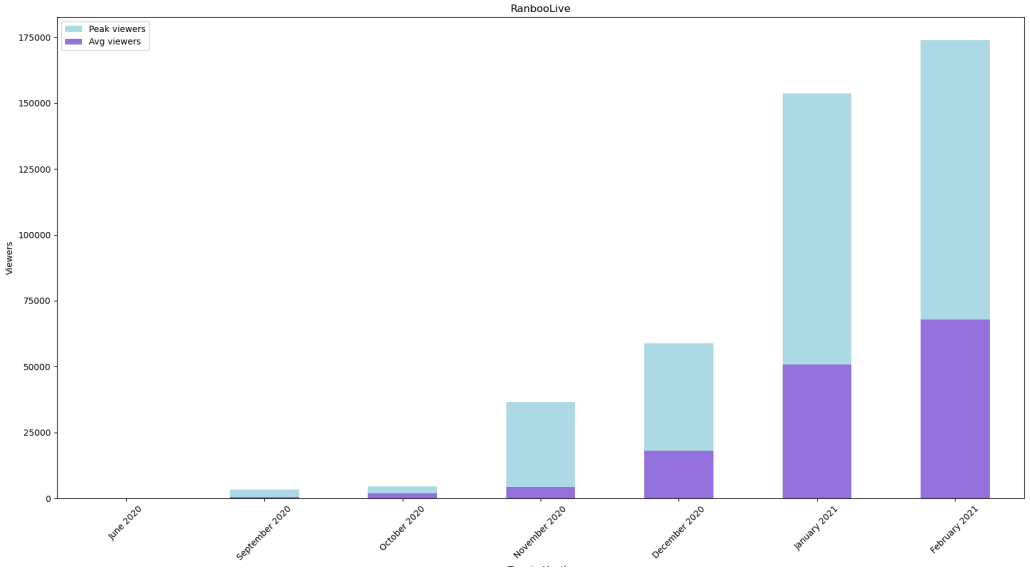


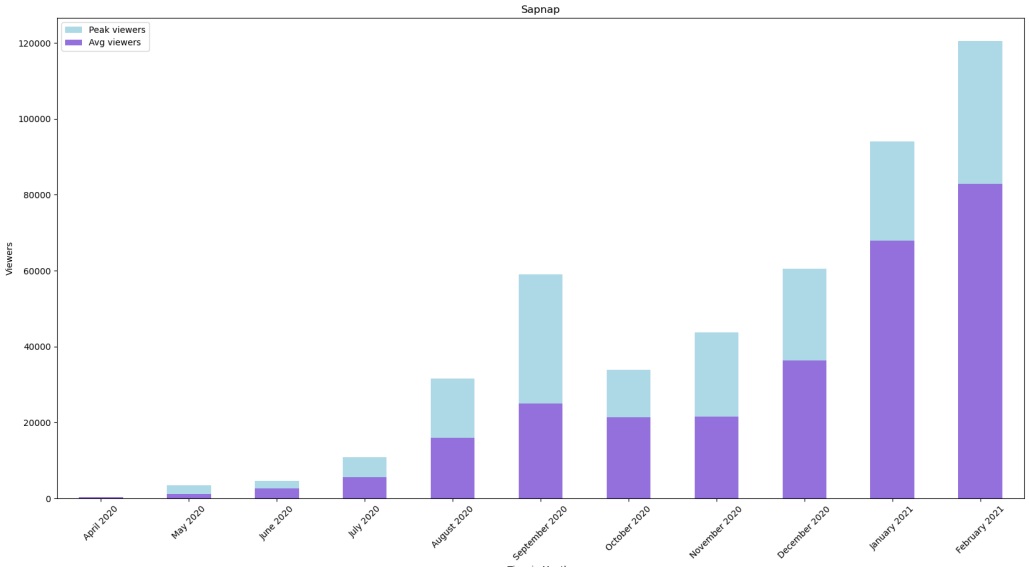


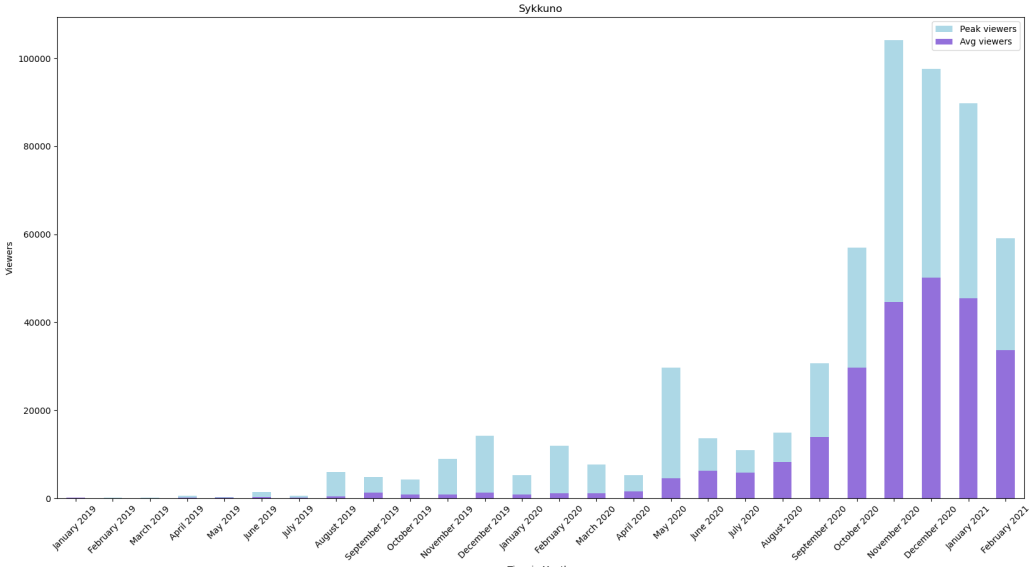

karljacobs
Peak viewers
Avg viewers
Viewers
250000
200000
150000
100000
50000
0
January 2019
February 2019
March 2019
April 2019
May 2019
April 2020
May 2020
July 2020
August 2020
September 2020
October 2020
November 2020
December 2020
January 2021
February 2021

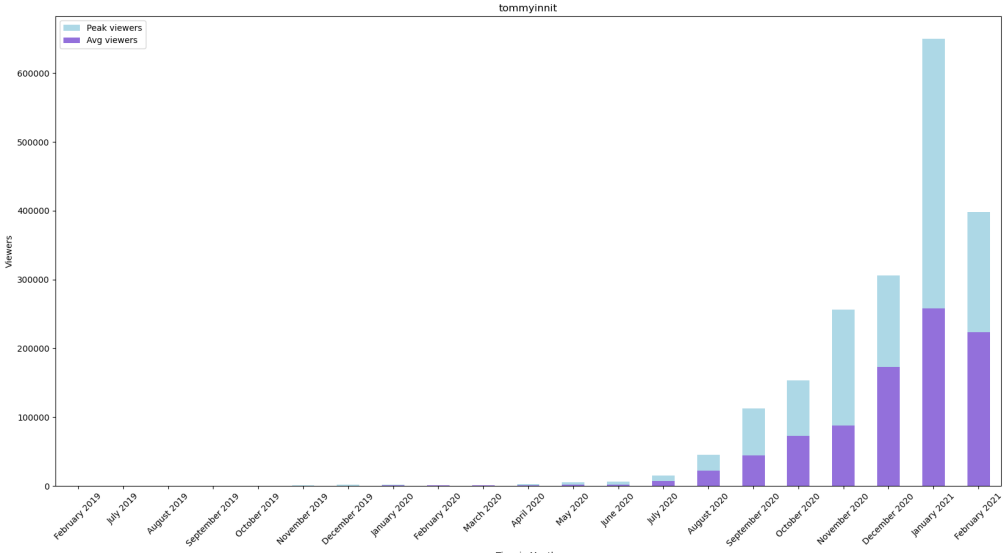

Philza
Peak viewers
Avg viewers
Viewers
140000
120000
100000
80000
60000
40000
20000
0
January 2019
February 2019
March 2019
April 2019
May 2019
June 2019
July 2019
August 2019
September 2019
October 2019
November 2019
December 2019
January 2020
February 2020
March 2020
April 2020
May 2020
June 2020
July 2020
August 2020
September 2020
October 2020
November 2020
December 2020
January 2021
February 2021

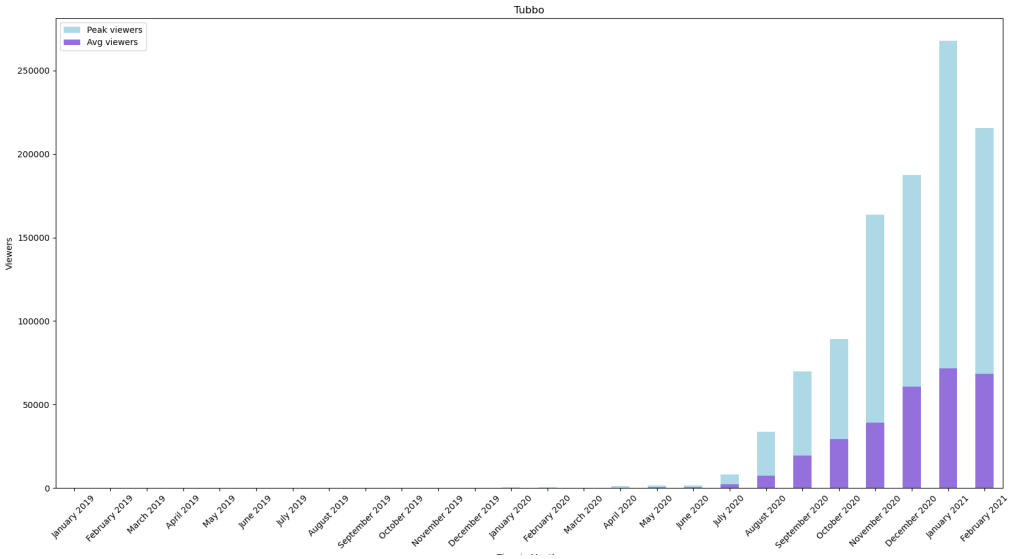

Pokimane
Peak viewers
Avg viewers
Viewers
0
20000
40000
60000
80000
100000
120000
140000
January 2019
February 2019
March 2019
April 2019
May 2019
June 2019
July 2019
August 2019
September 2019
October 2019
November 2019
December 2019
January 2020
February 2020
March 2020
April 2020
May 2020
June 2020
July 2020
August 2020
September 2020
October 2020
November 2020
December 2020
January 2021
February 2021

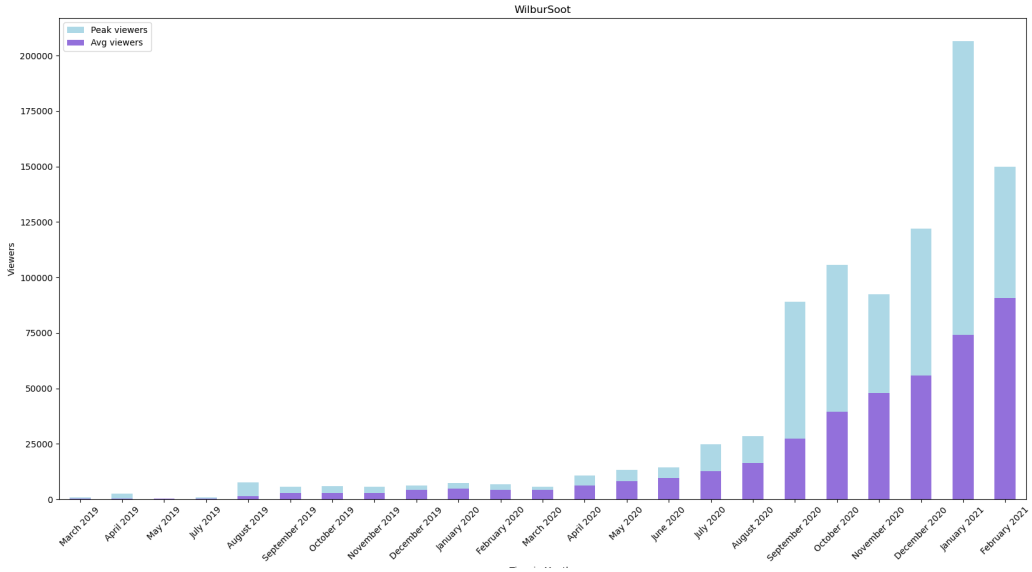

RanbooLive
Peak viewers
Avg viewers
Viewers
0
25000
50000
75000
100000
125000
150000
175000
June 2020
September 2020
October 2020
November 2020
December 2020
January 2021
February 2021


Sapnap
Peak viewers
Avg viewers
Viewers
0
20000
40000
60000
80000
100000
120000
April 2020
May 2020
June 2020
July 2020
August 2020
September 2020
October 2020
November 2020
December 2020
January 2021
February 2021


Sykkuno
Peak viewers
Avg viewers
Viewers
0
20000
40000
60000
80000
100000
January 2019
February 2019
March 2019
April 2019
May 2019
June 2019
July 2019
August 2019
September 2019
October 2019
November 2019
December 2019
January 2020
February 2020
March 2020
April 2020
May 2020
June 2020
July 2020
August 2020
September 2020
October 2020
November 2020
December 2020
January 2021
February 2021

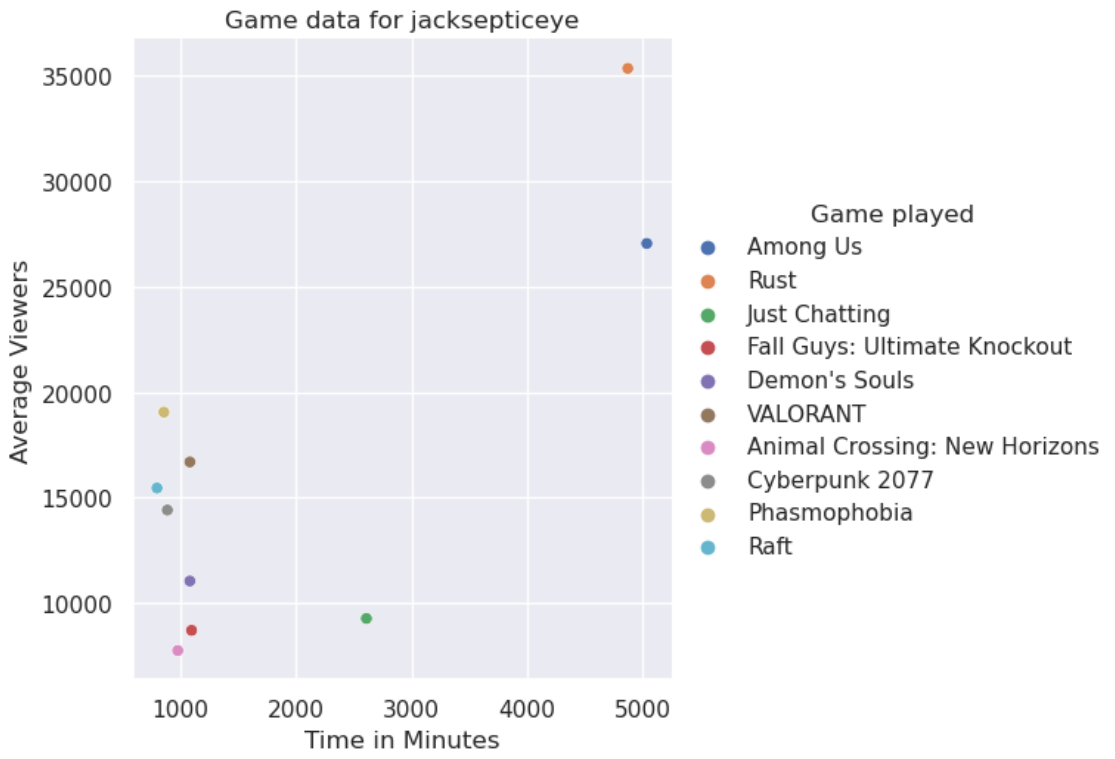

tommyinnit
Peak viewers
Avg viewers
Viewers
0
100000
200000
300000
400000
500000
600000
February 2019
July 2019
August 2019
September 2019
October 2019
November 2019
December 2019
January 2020
February 2020
March 2020
April 2020
May 2020
June 2020
July 2020
August 2020
September 2020
October 2020
November 2020
December 2020
January 2021
February 2021

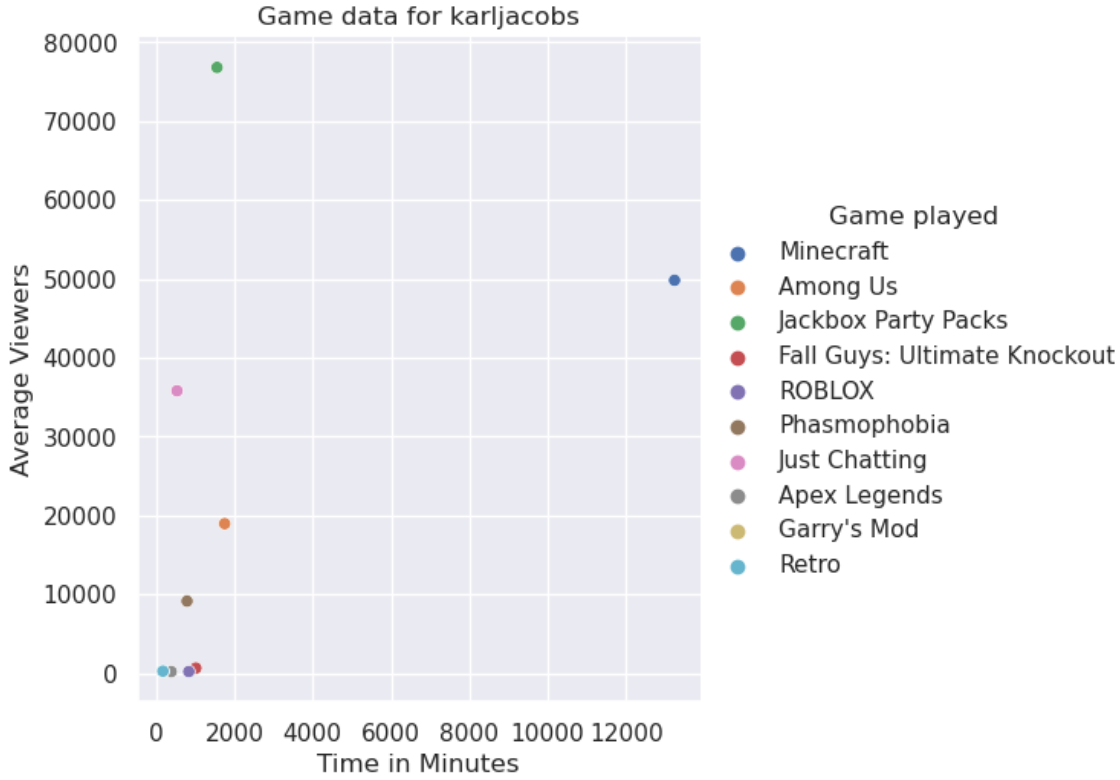


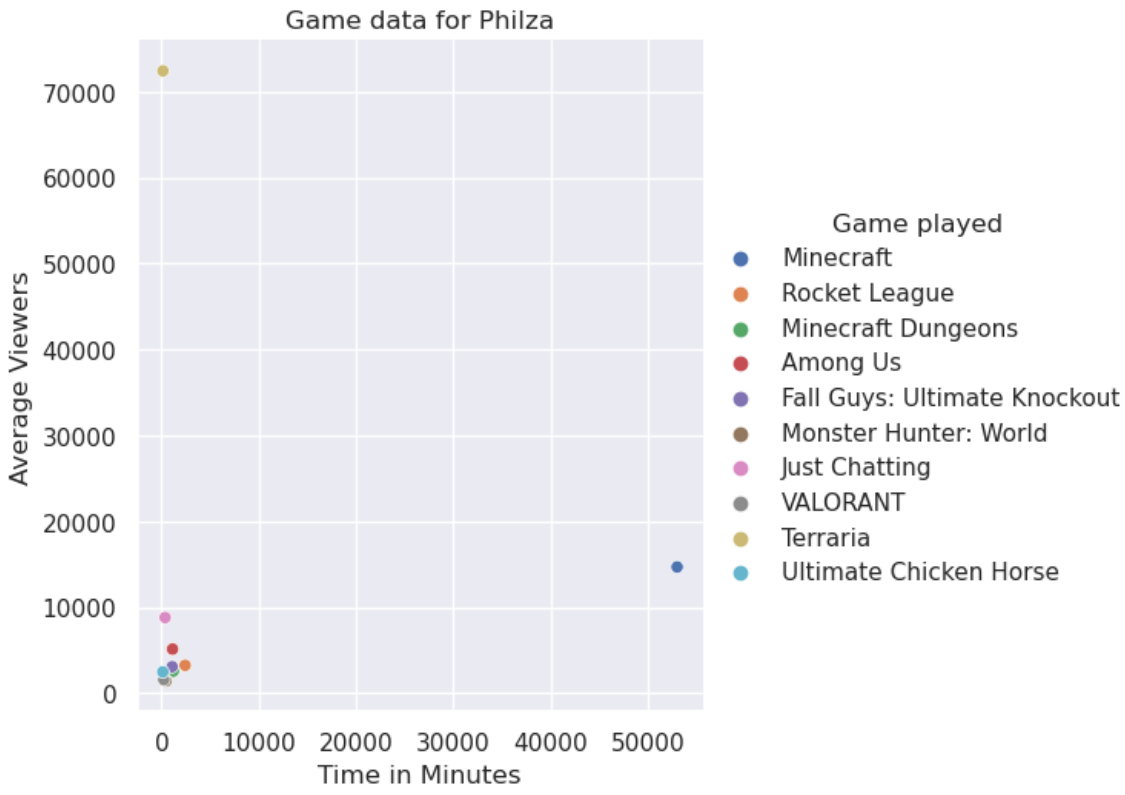


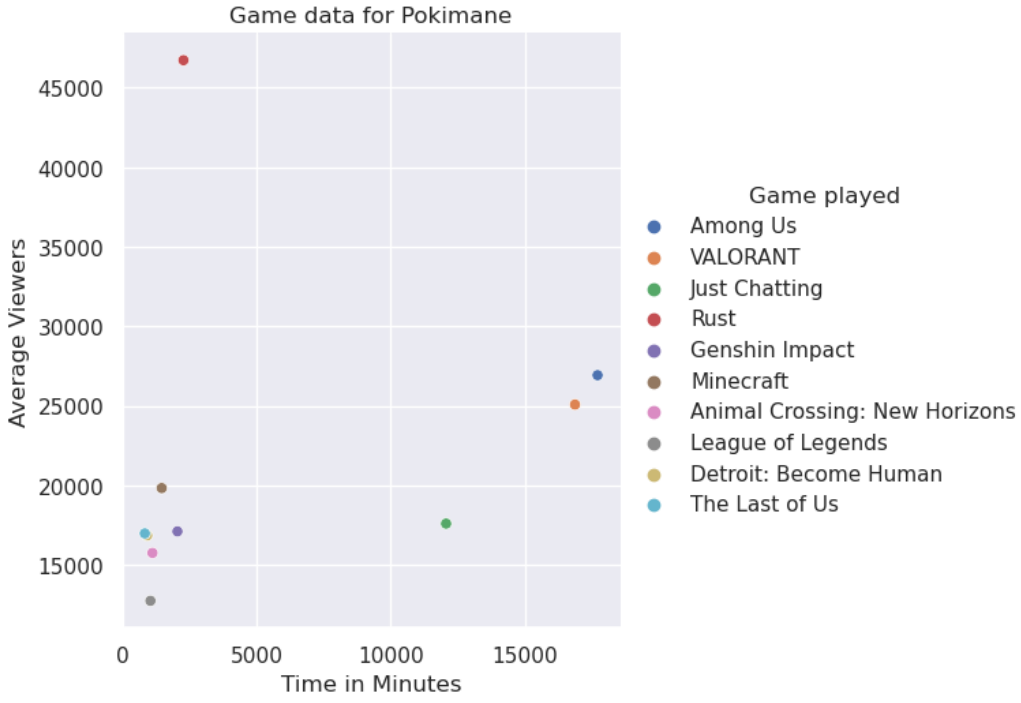
We plotted the average and peak viewers each month for each streamer we had data for, as well as for the entirety of the Twitch platform from 2019 to now. The Twitch data was interesting because before quarantine it had fairly stable average viewer counts, sitting at around 1.2 million a month. However, since the beginning of quarantine in March 2020, there has not been a single month with fewer than 2 million average viewers. The peaks are also greater than any peak before March 2020. This trend can also be seen in the individual channel graphs, with every channel having impressive growth since the beginning of quarantine. Most channels showed growth resembling an exponential trend, and 7 out of 10 channels had their highest peaks between December 2020 - January 2021.

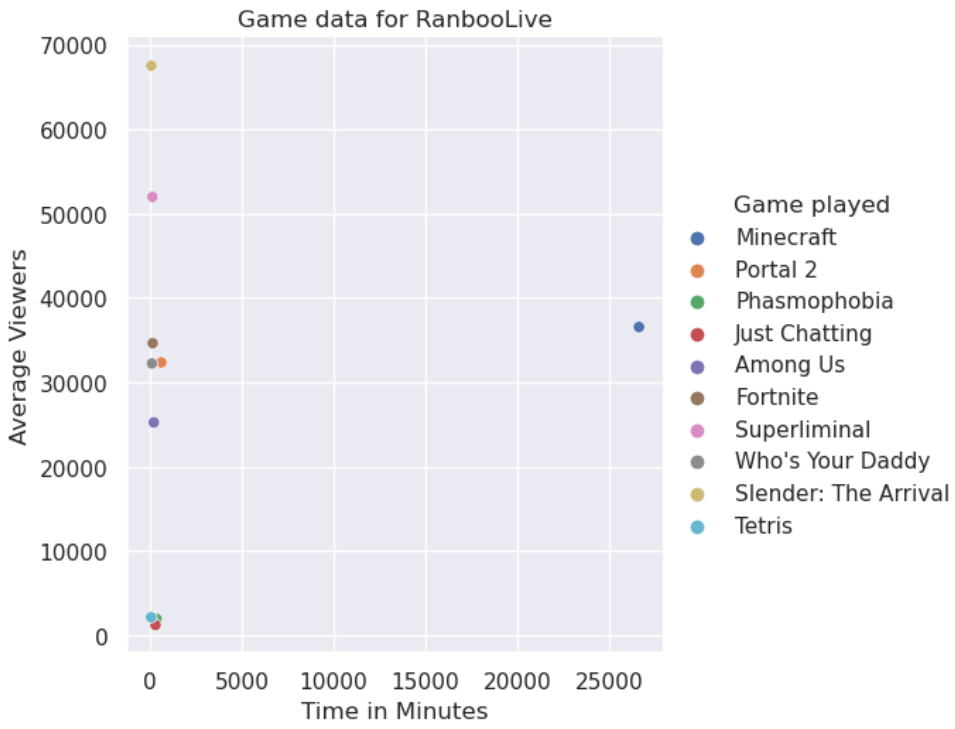
Stream Data based on Games (Research Question 3)

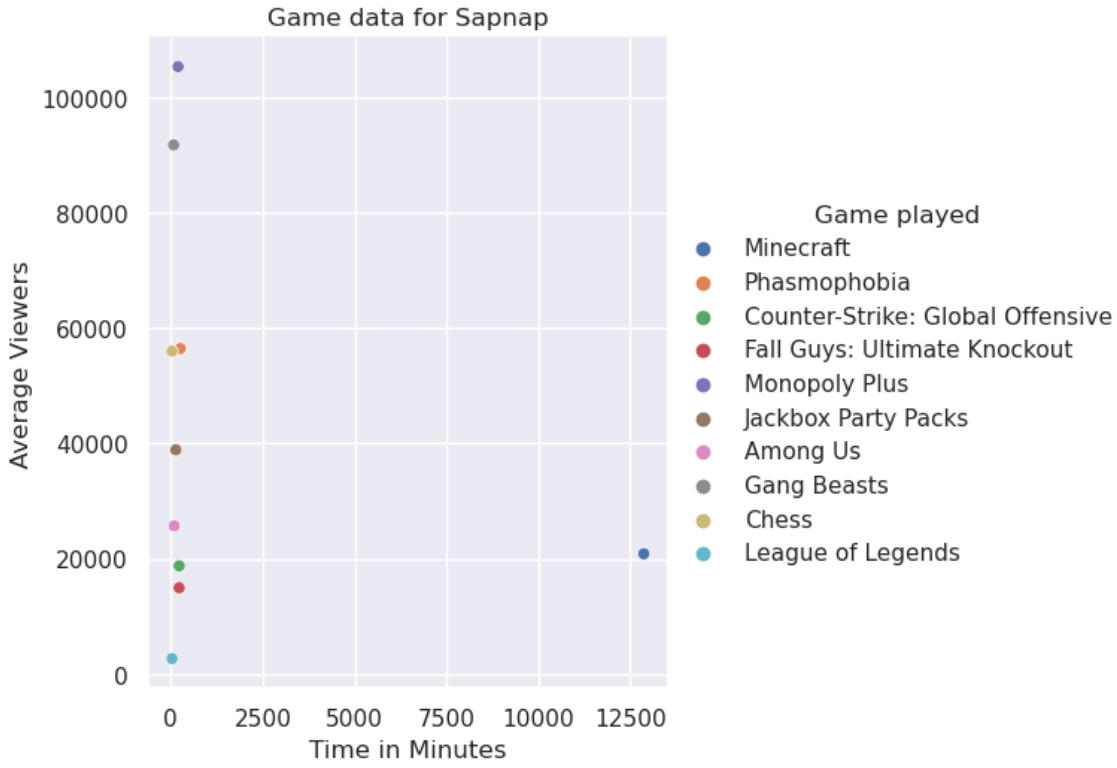

Game data for jacksepticeye
Average Viewers
35000
30000
25000
20000
15000
10000
1000
2000
3000
4000
5000
Time in Minutes
Game played
Among Us
Rust
Just Chatting
Fall Guys: Ultimate Knockout
Demon's Souls
VALORANT
Animal Crossing: New Horizons
Cyberpunk 2077
Phasmophobia
Raft

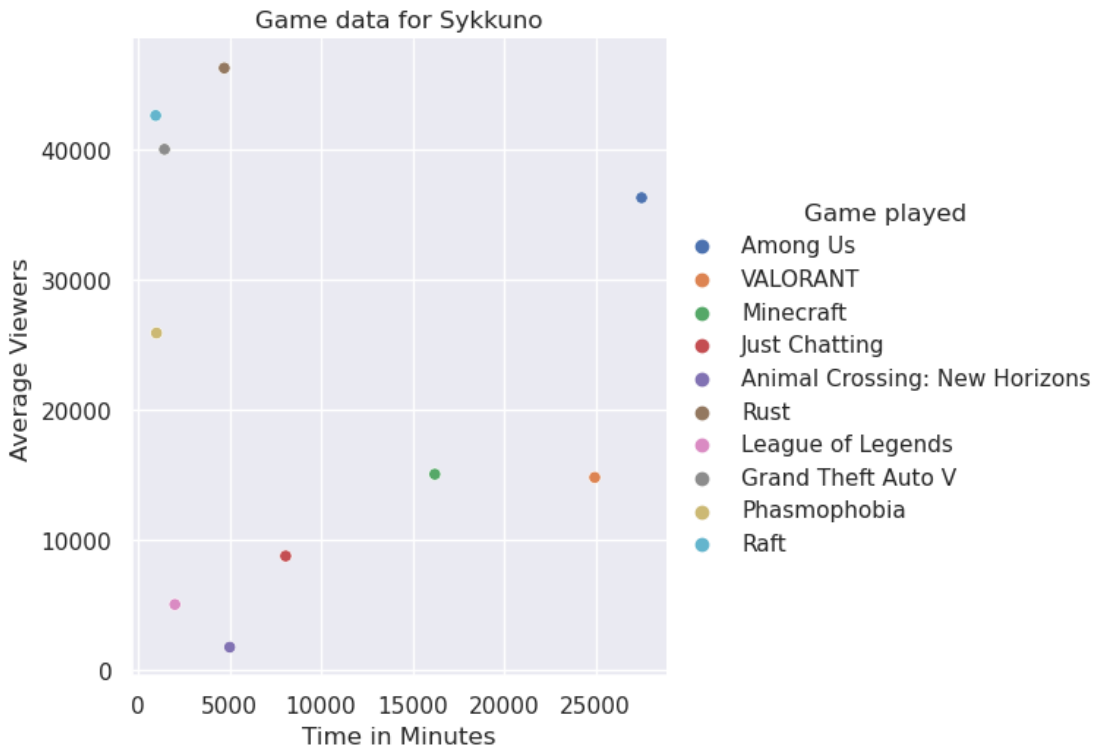

Game data for karljacobs
Average Viewers
80000
70000
60000
50000
40000
30000
20000
10000
0
0
2000
4000
6000
8000
10000
12000
Time in Minutes
Game played
Minecraft
Among Us
Jackbox Party Packs
Fall Guys: Ultimate Knockout
ROBLOX
Phasmophobia
Just Chatting
Apex Legends
Garry's Mod
Retro

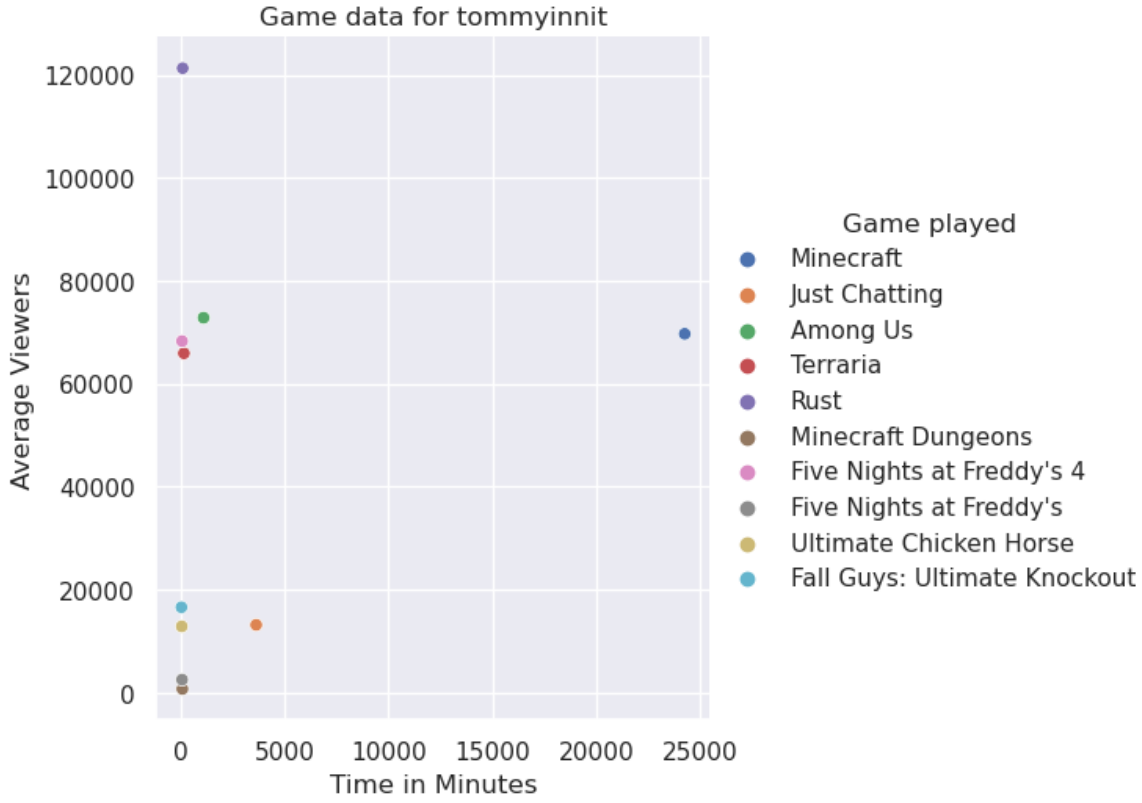

Game data for Philza
Average Viewers
Time in Minutes
0
10000
20000
30000
40000
50000
60000
70000
Game played
Minecraft
Rocket League
Minecraft Dungeons
Among Us
Fall Guys: Ultimate Knockout
Monster Hunter: World
Just Chatting
VALORANT
Terraria
Ultimate Chicken Horse

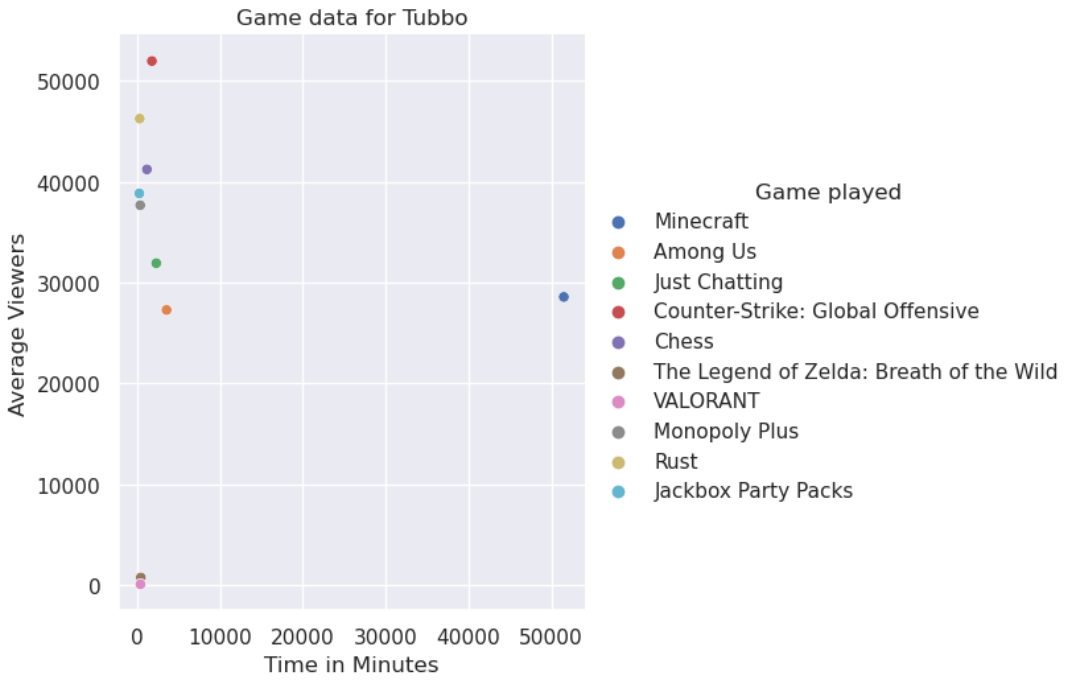

Game data for Pokimane
Average Viewers
Time in Minutes
0
5000
10000
15000
20000
25000
30000
35000
40000
45000
Game played
Among Us
VALORANT
Just Chatting
Rust
Genshin Impact
Minecraft
Animal Crossing: New Horizons
League of Legends
Detroit: Become Human
The Last of Us

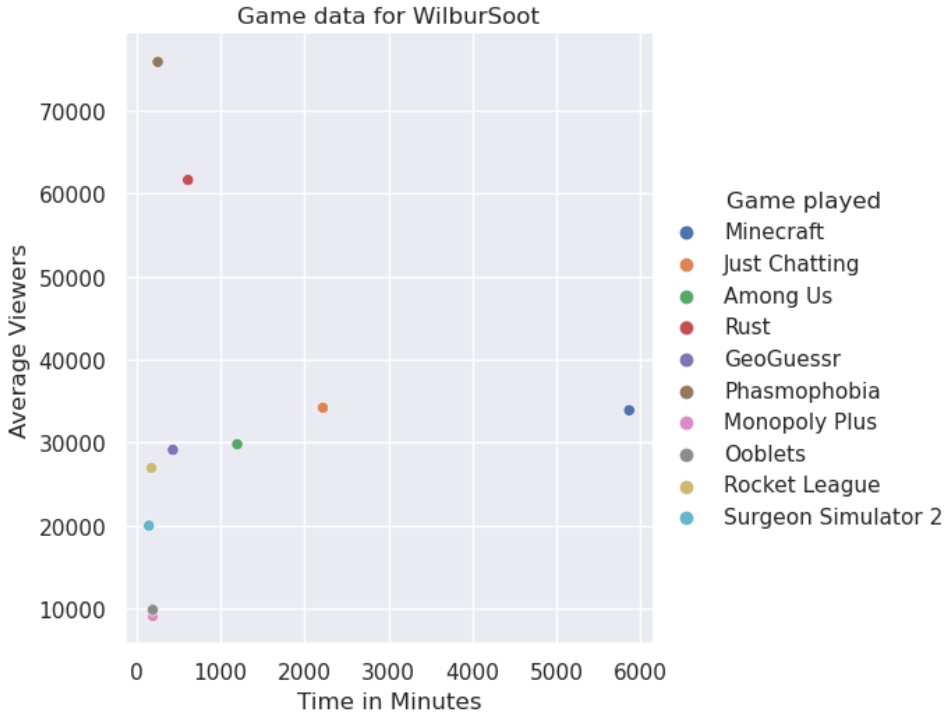

Game data for RanbooLive
Average Viewers
70000
60000
50000
40000
30000
20000
10000
0
0
5000
10000
15000
20000
25000
Time in Minutes
Game played
Minecraft
Portal 2
Phasmophobia
Just Chatting
Among Us
Fortnite
Superliminal
Who's Your Daddy
Slender: The Arrival
Tetris


Game data for Sapnap
Average Viewers
100000
80000
60000
40000
20000
0
0
2500
5000
7500
10000
12500
Time in Minutes
Game played
Minecraft
Phasmophobia
Counter-Strike: Global Offensive
Fall Guys: Ultimate Knockout
Monopoly Plus
Jackbox Party Packs
Among Us
Gang Beasts
Chess
League of Legends


Game data for Sykkuno
Average Viewers
Time in Minutes
0
10000
20000
30000
40000
0
5000
10000
15000
20000
25000
Game played
Among Us
VALORANT
Minecraft
Just Chatting
Animal Crossing: New Horizons
Rust
League of Legends
Grand Theft Auto V
Phasmophobia
Raft


Game data for tommyinnit
Average Viewers
Time in Minutes
0
20000
40000
60000
80000
100000
120000
0
5000
10000
15000
20000
25000
Game played
Minecraft
Just Chatting
Among Us
Terraria
Rust
Minecraft Dungeons
Five Nights at Freddy's 4
Five Nights at Freddy's
Ultimate Chicken Horse
Fall Guys: Ultimate Knockout





For a majority of the channels Minecraft was the longest played game by a large margin. For the other channels, Among Us and Valorant seemed to be the most played games. Either Minecraft or Among Us was the top played game for every single channel. Each channel's top game had a high number of minutes and a number of average viewers around the middle of the graph. Surprisingly, on almost every channel that played it, Rust had a very high number of average viewers and low time played, leading us to believe it was a game that became popular very quickly and attracted a lot of viewers, but ultimately died out quickly as well. It appears that

Minecraft and Among Us are the games with the most longevity, or at least for the year of 2020. These graphs do not take into account data from before March of 2020 so it is difficult to tell how long these games can last past that.

Challenge Goals

We met the web scraping and using a new library challenge goals. We originally planned on doing machine learning and messy data, however we realized the way we were combining CSV files was not significant enough to count as a challenge goal. We also struggled to implement a machine learning algorithm that clearly and accurately addressed any of our research questions. Most of our research questions were too vague to implement without making them more specific, which made machine learning pretty difficult. That is why we decided to focus on web scraping, which is how we got most of the data we used in our analysis. We had to scrape the average viewers and peak viewers on twitch per month over the last two years from SullyGnome in order to answer how twitch has grown since the pandemic began. In order to scrape, we had to learn how to use a new library, BeautifulSoup, which took many hours of research to learn, both through reading documentation and YouTube videos.

Work Plan Evaluation

Original plan:
Get the data: 3/5/21 (~2 hours)

- Web scrape necessary data to perform analysis.
- Validate all the data

Write functions/codes 3/7/21 (~7 hours):

- Write functions to calculate results in order to answer research questions.

Challenge goals 3/10/21 (~7 hours):

- Apply challenge goal technique to further analyze the data.

Validate results 3/12/21 (~2 hours):

- Check if the result is good to go.

Our estimate for getting the data was off because although it was easy to physically download the data, we realized that the data we originally had was not as useful for our research questions as we thought. We had to reevaluate the research questions and restructure them so that they were quantitative questions we could answer/visualize rather than broad statements that had vague constraints. This caused our writing functions to be longer than expected as well, because we had to basically come up with new functions that could answer tangible questions, with the new data we decided to use. The estimates turned out to be accurate based on the work we decided to keep, however there was a lot of extra work and code writing that came before that before we decided on what we wanted to use for our project.

Testing
We used the aggregate data to manually calculate growth rates for each channel. We also compared the peak data for each graph with the data in the csv to make sure it was accurate.

Collaboration
We had no outside help from other people. The online resources we used were YouTube videos on python web scraping using BeautifulSoup and library documentation websites for the assortment of libraries that we used.