Ryan Berger
SI 330
Final Project Report

An Analysis of Monthly Ranked Twitch Games

and Associated Esports Earnings

Motivation

Video games have been an integral part of my life so far. Playing them, watching them on live streams, and being a part of their communities have always been an essential part of who I am. They have shaped many of my relationships with both my family and friends. As such, I have utilized Twitch, an online video game streaming platform currently owned by Amazon, to become a part of many video game communities while watching content creators and professionals play the games I enjoy. In addition to being a part of these communities, I have also thoroughly enjoyed the competitive sides of video games and the level of expertise that can come with many of these games. So, for my project, I wanted to compare a database composed of the top 200 games on Twitch every month from 2016 to 2021 with esports data including the earnings of different games and their associated genres over time. The goal I sought to achieve with this project is to potentially see trends in games over time and see the earnings that correspond to them. Some of the different questions I asked during my exploration of these datasets include: How do earnings and ranks of some of my favorite games correlate? Do the earnings of games correlate with the Twitch rank they possess during those same months, as well as overall? Which games have the highest overall rank and earnings? Do specific genres fare differently in terms of esports earnings and Twitch rank? Overall, my goal was to compare and explore these datasets together to make inferences about esports and data related to popular games on Twitch.

Data Sources

Dataset 1: 'Monthly Top 200 Games on Twitch 2016-2021'

- Location: kaggle.com
  (https://www.kaggle.com/datasets/rankirsh/evolution-of-top-games-on-twitch)
- Format Used/Returned: CSV
- Access Method: pd.read_csv (pandas library)
- Size: 1.43MB
- Time Periods: January 2016 to December 2021
- Important Variables: The variables that I would be focusing on in this dataset were in the columns 'Rank', 'Month,' and 'Year'. The 'Rank' column is what I use for most of my analyses. Rank correlates to the number of concurrent viewers on a stream in any specific category or game. The more concurrent viewers, the higher the rank of that category. The 'Month' and 'Year' columns would be used to merge with the historical esports dataset to compare monthly data.

Dataset 2: 'Esports Earnings 1998 - 2021'

- Location: kaggle.com (https://www.kaggle.com/datasets/rankirsh/esports-earnings)
- Format Used/Returned: CSV
- Access method: pd.read_csv (pandas library)
- Size: 357.52 kB (Two datasets)
  - *Historical Esports Data*: Monthly esports data for games since January 1998
  - *General Esports Data*: Overall esports data for each game
- Time Periods: January 1996 to December 2021
- Important Variables: The 'Tournaments' and 'TotalTournaments' variables are used to manipulate the data and determine irrelevant values in each of the historical and general esports datasets respectively. The 'Earnings,' 'Total Earnings,' 'Date,' and 'Genre' variables are used in my analysis to answer my overarching questions and accomplish my analysis goals.

<u>Data Manipulation Methods</u>:

<u>Twitch Dataset</u>

To make my analysis, I would need to merge the historical esports dataset with the Twitch dataset to make inferences about competitive games over time. This would involve merging these files on the 'Date' column to make said inferences. In its original state, the Twitch historical dataset had two columns for the date: 'Month' and 'Year.' I used a function to change this into DateTime format under a new 'Date' column, and then removed the extraneous 'Month' and 'Year' columns afterward. In terms of missing or incorrect data, there was nothing other than one null value listed for a game, which I promptly removed. The rest of my manipulation methods were in preparation for my analysis. I created a data frame to determine the average rank on Twitch. For this, I did a groupby on the rank of Twitch games. This data frame would be used in my analysis when comparing the general statistics of games' esports data.

<u>Esports Datasets</u>

In terms of manipulating the esports datasets, I first took the historical dataset and converted it into DateTime format. Although this dataset was not as intensive to convert to this format, it would be necessary to merge it with the Twitch dataset for my analysis. Since the Twitch dataset only has values from 2016 to 2021, and the historical esports dataset has values starting from 1996, I used the DateTime format to select only those rows from 2016 onwards. Both datasets had the same ending date, December 2021, so this did not need to be changed. Something interesting to note about missing data was that there are some rows where the value for the column 'Players' was 0, but the number of tournaments and the prize money values are still there. Many of these games were widely popular during their peak, and the listed tournaments during those dates did have players in them. So, I decided to keep them in, as the 'Players' column's values were not essential in my analysis. However, there were many rows in

the general esports dataset that had values of 0 for all columns, indicating that these games had

no esports presence or that the data was incomplete. So, I removed these rows from the general

dataset.

Generally, my source code flows in the order in which I answer my questions. I began by

cleaning and manipulating the data to prepare it to be merged and analyzed. Once my data was

ready for analysis, I merged the historical esports dataset and Twitch dataset, which both

matched in their date columns. I removed any null values from this merged data frame, as there

are likely many games and Twitch categories that do not have an esports presence. Next, I

answer my first question regarding some of my favorite games and the relationship between their

ranks and earnings. After that, I do an analysis of monthly ranks and earnings for all games in the

merged data frame and make another analysis accounting for outliers in the data. Then, I created

a second merged data frame of general esports earnings and the average rank of these games,

which was determined in a previous data manipulation step. Again, I removed the null values

from this data frame of those games and categories without esports data associated with them.

With this merged data frame, I could analyze the total earnings and average rank of every game,

as well as an analysis of these variables when compared to esports game genres.

Analysis and Visualization

Top Ranked Categories on Twitch

For the first part of my analysis, I wanted to determine the top categories on the Twitch platform.

I did this through two different methods. First, I found only those categories in the dataset that

have been ranked in the top 10 and then counted the number of times that they appeared there. I

sorted this by the largest values, and then converted it to a data frame so that I could name the column 'top10_count.' The aforementioned data frame is on the left. The Twitch categories ranked in the top 10 through this method are League of Legends (LoL), Counter-Strike: Global Offensive (CSGO), Dota 2, Fortnite, World of

Game	top10_count
League of Legends	72
Counter-Strike: Global Offensive	72
Dota 2	63
Fortnite	51
World of Warcraft	50
Grand Theft Auto V	46
Hearthstone	45
Overwatch	39
Just Chatting	39
Minecraft	33

Game	avg_rank
League of Legends	1.79
Just Chatting	2.77
Fortnite	4.67
Counter-Strike: Global Offensive	5.65
E3 2016	6.00
Dota 2	6.04
VALORANT	6.81
E3 2017	7.00
E3 2018	7.00
IRL	8.36
Hearthstone	8.99
Call of Duty: Warzone	9.21
World of Warcraft	9.56

Warcraft (WoW), Grand Theft Auto V (GTA V), Hearthstone, Overwatch, Just Chatting, and Minecraft. The other method was to determine the average rank of all categories, which I did with a groupby function to find the mean of each game's ranks. Similarly, I converted this to a data frame and named this column 'avg_rank.' This data frame is on the right. The annual Electronic Entertainment Expo, also known as E3, is a convention that lasts for 3 days over the beginning of the summer and has a large following during that time. I included an extra three categories in the average rank table to account for this event during these years. The Twitch categories ranked in the top 10 on average, other than E3, were: LoL, Just Chatting, Fortnite, CSGO, Dota 2, VALORANT, 'IRL', Hearthstone, Call of Duty: Warzone, and WoW. What was interesting to note was that although many of the categories were found similarly in both tables, several categories were only found in one of the tables. Categories such as 'Just Chatting' and 'IRL' are newer in terms of popularity and have lower counts of being in the top 10, but still perform extremely well on the platform as seen by their average ranks. Games that have a large number of top 10 counts, yet are not in the top 10 highest average ranked categories, such as

Overwatch, GTA V, and Minecraft, have a large gaming presence but are not found in the table

of the top 10 categories with the highest average rank. likely do not have as high of an average

rank because of lull periods in the games where they are not updated and much of the fanbase

loses interest.

### *My Favorite Competitive Games' Twitch Ranks and Associated Earnings*

After creating a merged historical data frame of each

game and their monthly Twitch rank and earnings, I

wanted to do an analysis specifically about some of my

favorite competitive games. I started by making a list of

three of these games: Hearthstone, Super Smash Bros.

Ultimate, and Overwatch. Each of these games is from a

different genre of gaming. Hearthstone is an online

collectible card game, Super Smash Bros. Ultimate is a

fighting game, and Overwatch is a first-person shooter
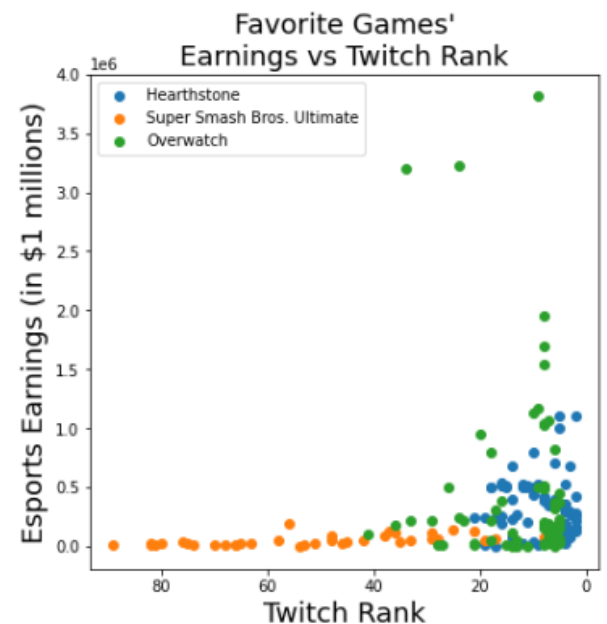
game. Next, I created a for loop to go through this list, find the game's data in the merged

historical data frame, and then plot each point onto a scatter plot. The top 3 results for each game

were printed as well to determine potential outliers. I created a legend with distinct colors for

each game to distinguish between them, and inverted the x-axis to indicate that lower values for

rank are preferable. The labeled graph can be found to the right. Some interesting points to note

are the Overwatch outliers on the top of the graph, which can be attested to the 'Overwatch

League,' a well-known competitive Overwatch league with high prize pools which was very

popular on Twitch. Other than these outliers, both Activision-Blizzard titles, Overwatch and

Hearthstone, consistently stay top ranked out of all Twitch categories regardless of esports

earnings each month. These games could additionally have smaller tournaments occurring during some months when esports earnings are low. Despite having very low esports earnings, the fifth edition in a series of Nintendo fighting games released in 2018, known as Super Smash Bros. Ultimate, has been at a wide variety of Twitch ranks. The competitive scene has a very dedicated audience, but very little competitive support from Nintendo, which is why esports earnings are consistently low.

__Total Esports Earnings and Historical Twitch Ranks Per Game__
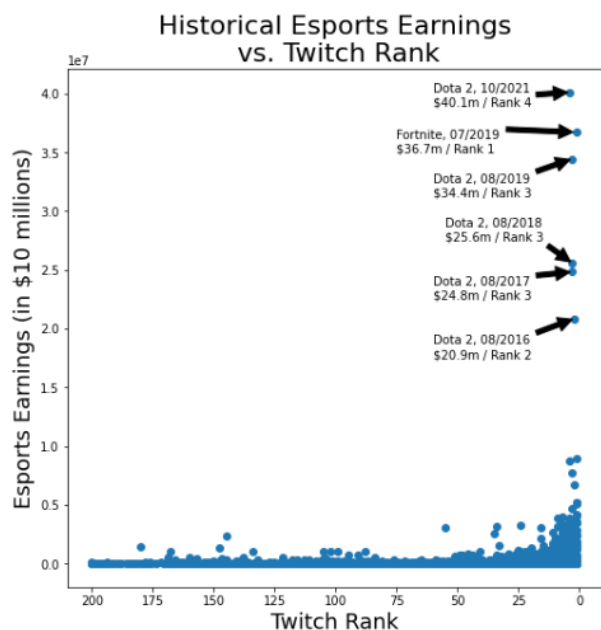
|	Game	Rank	Date	Earnings
13802	Dota 2	4	2021-10-01	40125696.75
8399	Fortnite	1	2019-07-01	36713555.90
8601	Dota 2	3	2019-08-01	34460762.14
6202	Dota 2	3	2018-08-01	25584943.71
3802	Dota 2	3	2017-08-01	24824809.87
...	...	...	...	...
8768	Yu-Gi-Oh! Duel Links	170	2019-08-01	20.00
2747	Yu-Gi-Oh! Duel Links	148	2017-02-01	12.97
6311	Clash Royale	112	2018-08-01	0.00
9350	Quake Champions	152	2019-11-01	0.00
6410	Call of Duty: Black Ops 4	11	2018-09-01	0.00
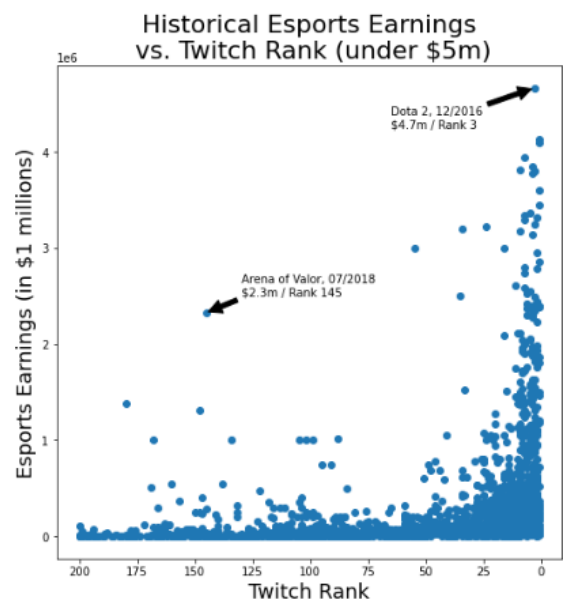
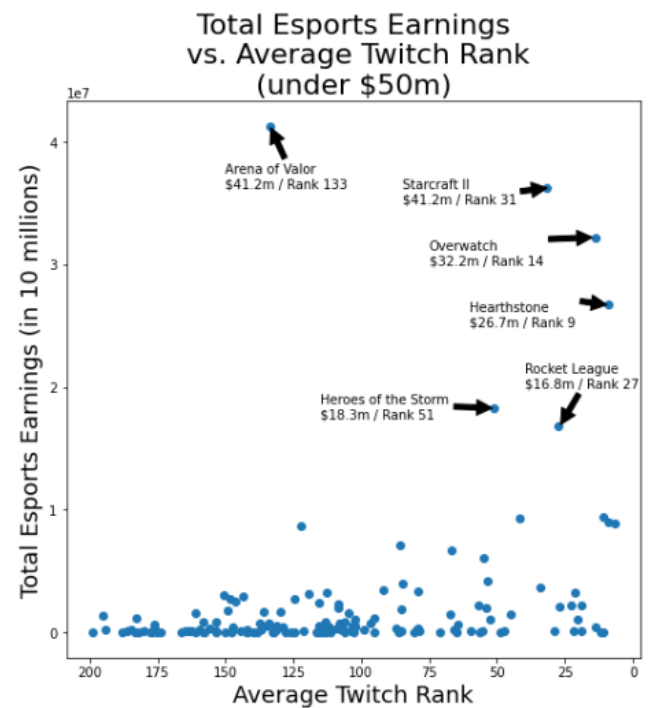|	Game	Rank	Date	Earnings
2202	Dota 2	3	2016-12-01	4667978.88
6200	Fortnite	1	2018-08-01	4132307.15
7999	Fortnite	1	2019-05-01	4095331.48
13605	Counter-Strike: Global Offensive	7	2021-09-01	3941868.94
14002	Counter-Strike: Global Offensive	4	2021-11-01	3853484.14
...	...	...	...	...
8768	Yu-Gi-Oh! Duel Links	170	2019-08-01	20.00
2747	Yu-Gi-Oh! Duel Links	148	2017-02-01	12.97
6311	Clash Royale	112	2018-08-01	0.00
6410	Call of Duty: Black Ops 4	11	2018-09-01	0.00
9350	Quake Champions	152	2019-11-01	0.00

With the merged historical esports earnings and Twitch rank data frame, I was able to answer my question regarding the potential correlation between Twitch games' ranks and their associated esports earnings. By comparing Twitch rank with esports earnings through this data frame using a scatter plot, I found a significant correlation between the Twitch rank of games and esports earnings during specific months. Those games which had larger prize pools and overall monthly earnings typically saw higher viewership on Twitch. Some of the games that earned the most were Dota 2, Fortnite, and League of Legends. These three games almost single-handedly make up the top 20 of the highest-earning and ranking games in the data frame. Dota 2 is a Multiplayer Online Battle Arena (MOBA) game in which two teams of players compete against each other on different battlefields. Counter-Strike: Global Offensive (CSGO) is an online team-based first-person shooter game in which players work together to kill the enemy team and capture or defend objectives depending on the team's tasks. Fortnite is a battle royale game, where players are put on a map and must explore their surroundings to gather equipment and fight other players. These games have significant competitive and live streaming followings and immense player bases. Dota 2 tops this list with their tournament known as the 'International 10,' which occurred this past October and had a record-breaking prize pool of over $40 million. To view and interpret the data on a greater scale, I made a second table with those outlier values removed, including only those games with historical earnings under $5 million. This value for earnings is over the 99th percentile for this data frame. Although the majority of the top plotted values are Dota 2, one of the most interesting outliers for this plot is a game called Arena of Valor, as I had never heard of this game before. This Chinese MOBA game has a point on the graph from July of 2018 which has $2.3 million in earnings and is ranked 145 on Twitch. This game appears again in my analysis, and it is interesting to see as an outlier because of its increasing popularity

in esports despite its low ranking on the Twitch platform. The game likely has a significant following abroad that has not made itself as well known domestically in the United States, which is why its rank on the American live streaming platform is so low in comparison to its esports earnings.

## *Total Esports Earnings and Average Twitch Rank Per Game*
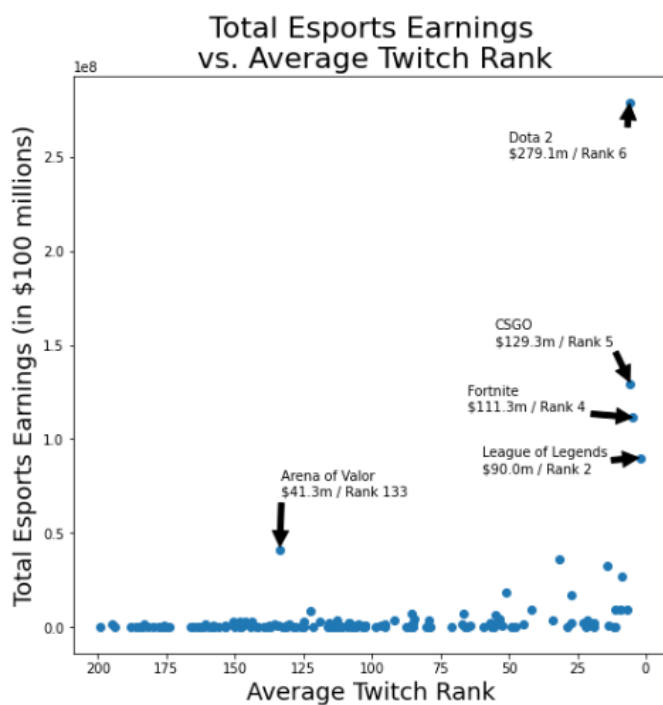
|	Game	avg_rank	ReleaseDate	Genre	TotalEarnings
5	Dota 2	6.04	2013.00	Multiplayer Online Battle Arena	279153000.65
3	Counter-Strike: Global Offensive	5.65	2012.00	First-Person Shooter	129352068.91
2	Fortnite	4.67	2017.00	Battle Royale	111339316.05
0	League of Legends	1.79	2009.00	Multiplayer Online Battle Arena	90041272.98
930	Arena of Valor	133.60	2015.00	Multiplayer Online Battle Arena	41290753.03
43	StarCraft II	31.71	2010.00	Strategy	36249504.04
18	Overwatch	13.97	2016.00	First-Person Shooter	32213162.28
10	Hearthstone	8.99	2014.00	Collectible Card Game	26689371.78
99	Heroes of the Storm	50.94	2015.00	Multiplayer Online Battle Arena	18280006.65
35	Rocket League	27.28	2015.00	Sports	16787959.34

	Game	avg_rank	ReleaseDate	Genre	TotalEarnings
930	Arena of Valor	133.60	2015.00	Multiplayer Online Battle Arena	41290753.03
43	StarCraft II	31.71	2010.00	Strategy	36249504.04
18	Overwatch	13.97	2016.00	First-Person Shooter	32213162.28
10	Hearthstone	8.99	2014.00	Collectible Card Game	26689371.78
99	Heroes of the Storm	50.94	2015.00	Multiplayer Online Battle Arena	18280006.65
35	Rocket League	27.28	2015.00	Sports	16787959.34
14	Apex Legends	10.94	2019.00	Battle Royale	9437038.69
65	SMITE	41.82	2014.00	Multiplayer Online Battle Arena	9250112.75
11	Call of Duty: Warzone	9.21	2020.00	First-Person Shooter	8996609.13
6	VALORANT	6.81	2020.00	First-Person Shooter	8859667.50

The next part of my analysis utilizes the merged Twitch and general esports data frame, which takes the average rank for each game and compares it to the game's total earnings and genre. With this data, I utilize scatter plots to analyze the potential correlation between average Twitch rank and overall esports earnings. Overall, I did see a correlation between these two variables, but not as strong as that of the historical data. There are only four games that have total esports earnings larger than $50 million. These include Dota 2, CSGO, Fortnite, and League of Legends (LoL). In addition to the first three games mentioned previously, LoL is a MOBA game similar to Dota 2 with a very big Twitch and esports following. To account for these outliers, I created a second scatter plot with only those games with under $50 million in total esports earnings, which can be found on the right. Both of these scatter plots had their x-axis inverted to show that lower rankings are more desired. The previously mentioned game, 'Arena of Valor,' holds a spot as the fifth highest-earning esports game, despite its low Twitch rank. What is surprising to me about this data is that, after Arena of Valor, the next four games are all under the same IP. StarCraft II, Overwatch, Hearthstone, and Heroes of the Storm are all Activision-Blizzard titles. These games are all from different genres, and despite Overwatch and Heroes of the Storm not seeing any big updates to their competitive play in a long time, they still have a strong standing in the top-earning esports and are ranked Twitch games of all time. A challenge with this analysis was determining where to cut off values for the second scatter plot to analyze further data and outliers.

Total Earnings by Esports Genre

Total Earnings by Esports Genre
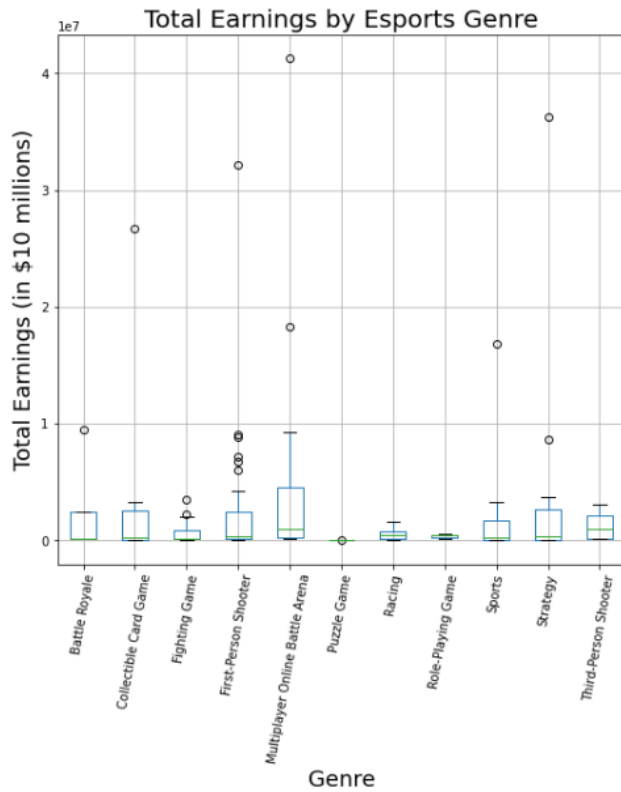
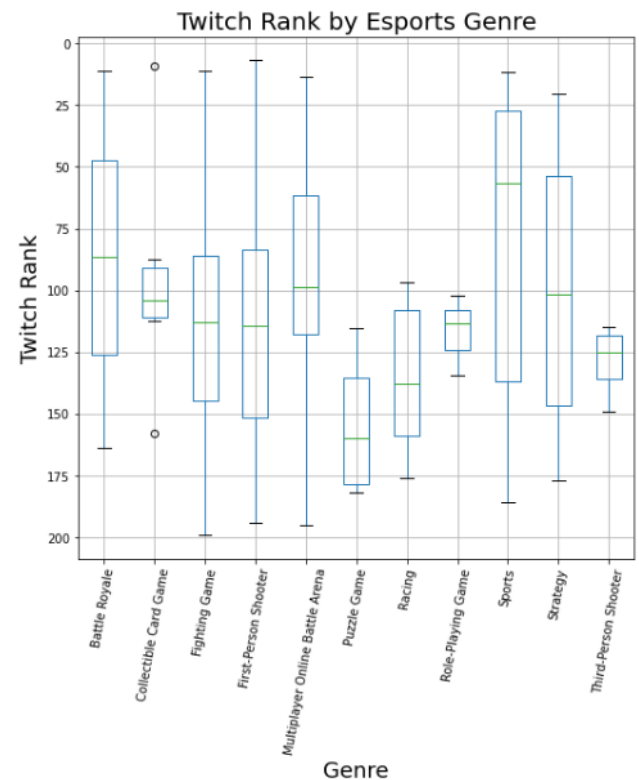| Genre	count	mean	std	min	25%	50%	75%	max
Multiplayer Online Battle Arena	12.00	6313912.57	12278888.48	45950.00	240543.36	910854.59	4501558.11	41290753.03
Collectible Card Game	6.00	5046429.44	10676520.57	4100.00	15936.68	184463.88	2488428.08	26689371.78
Strategy	20.00	3135657.80	8073682.26	220.00	31433.32	288746.84	2598675.72	36249504.04
Battle Royale	4.00	2426759.67	4673612.07	50000.00	87500.00	110000.00	2449259.67	9437038.69
First-Person Shooter	40.00	2414631.50	5455031.78	1390.00	116670.48	337537.24	2460242.12	32213162.28
Sports	21.00	1534987.36	3631992.36	61.30	27243.84	216000.00	1681120.00	16787959.34
Third-Person Shooter	4.00	1272571.25	1444999.11	50000.00	140000.00	976500.00	2109071.25	3087285.00
Fighting Game	25.00	680655.86	946419.22	2004.59	50741.58	123611.42	837447.16	3497427.71
Racing	12.00	481646.63	466101.95	1754.44	136955.64	403164.19	750813.00	1594178.42
Role-Playing Game	3.00	313203.67	226722.98	54250.00	231802.76	409355.52	442680.51	476005.50
Puzzle Game	4.00	11917.21	20442.33	185.00	1205.00	2483.92	13196.13	42515.99

Total Earnings by Esports Genre (under $10m)
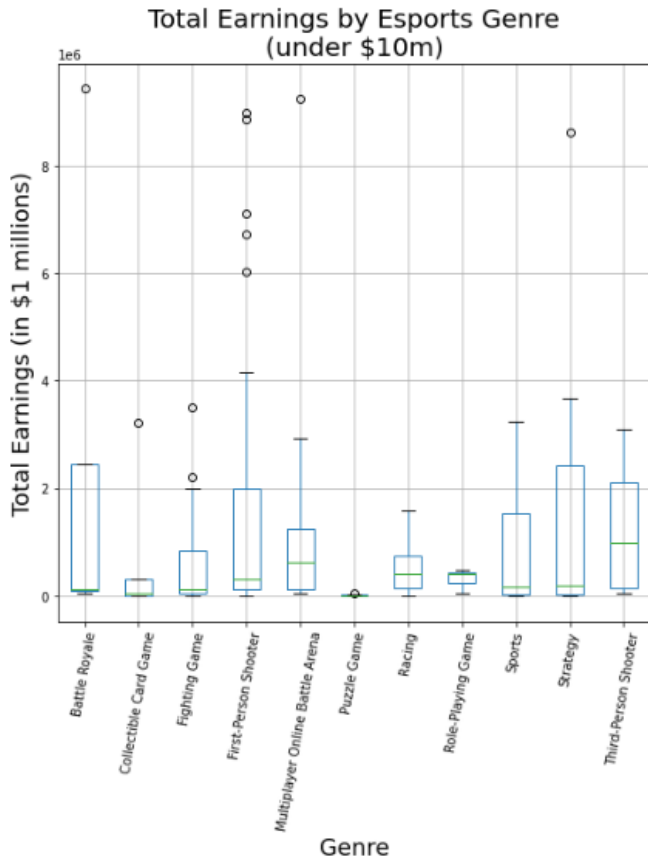
Genre	count	mean	std	min	25%	50%	75%	max
Battle Royale	4.00	2426759.67	4673612.07	50000.00	87500.00	110000.00	2449259.67	9437038.69
First-Person Shooter	39.00	1650566.61	2564007.18	1390.00	115980.32	299281.00	1993410.04	8996609.13
Multiplayer Online Battle Arena	10.00	1619619.11	2819602.80	45950.00	123630.09	613480.17	1243072.00	9250112.75
Strategy	19.00	1392823.79	2163636.48	220.00	30955.55	191181.81	2427226.40	8627165.61
Third-Person Shooter	4.00	1272571.25	1444999.11	50000.00	140000.00	976500.00	2109071.25	3087285.00
Sports	20.00	772338.76	1014151.39	61.30	26074.55	170500.00	1532239.47	3242642.34
Collectible Card Game	5.00	717840.97	1400010.18	4100.00	5000.00	48746.72	320181.03	3211177.09
Fighting Game	25.00	680655.86	946419.22	2004.59	50741.58	123611.42	837447.16	3497427.71
Racing	12.00	481646.63	466101.95	1754.44	136955.64	403164.19	750813.00	1594178.42
Role-Playing Game	3.00	313203.67	226722.98	54250.00	231802.76	409355.52	442680.51	476005.50
Puzzle Game	4.00	11917.21	20442.33	185.00	1205.00	2483.92	13196.13	42515.99

11

The merged total esports and average ranked Twitch games data frame also includes the genres of those listed games. Using this data, I can answer my last research question and interpret the earnings by genre as well as the Twitch rank of those games by their respective genre. The only limit of these interpretations is that this only includes those games that are in the esports datasets. This would not include any games without an esports presence or at least without any values listed in the dataset. I created a boxplot with this data using average Twitch rank as the 'column' value and using the Genres as the 'by' value. From this data, we can see that MOBA games, First-Person Shooters, and Battle Royale games hold many of the outliers for this data, which were also seen in both of my previous analyses. Most of these genres have relatively low means with large outliers except for Puzzle Games, which have both low mean and max values. I created a second graph only including those games with under $10 million in total earnings to try and make further inferences from this data. Many of these genres' box plots grow in shape with the exclusion of these outliers, such as First-Person Shooters. There appears to be one outlier for this genre, with most of its other outlier values below $10 million in esports earnings. These values are more clearly seen in the boxplot under $10 million. However, the shape of the Collectable Card Games genre's boxplot retracts significantly. This indicates that Hearthstone, among other popular card games, is a significant leader in the genre's gaming industry and holds a lot of influence in this space. Overall, it is interesting to see the influence that many of the larger games have over their associated genre's esports earnings and how these genres compare in numbers when those outliers are removed.

Twitch Rank by Esports Genre

The last part of my analysis took the merged general data frame and compared Twitch ranks of each esports genre. Again, this analysis does not include all Twitch games under those genres, only those games with mentioned general esports earnings in the dataset that I utilized. I created a boxplot to analyze these values and inverted the y-axis containing Twitch ranks to again indicate that lower ranks are better in terms of my analysis. Many of these genres have maximum values ranked in the top 25 categories on the platform, except for Puzzle, Racing, Role-Playing, and Third-Person Shooter games. Collectible Card Games have one outlier among the top-ranked categories, which is Hearthstone. It is fascinating to see how influential this game continues to be in my analyses overall. The lowest mean value for Twitch rank is in the genre of Battle Royale games, and the highest value is in Puzzle Games. Fighting Games have possibly the

Genre	count	mean	std	min	25%	50%	75%	max
Battle Royale	4.00	86.96	66.25	10.94	47.36	86.45	126.05	164.00
Sports	21.00	87.00	63.52	11.75	27.07	56.50	137.00	186.00
Multiplayer Online Battle Arena	12.00	94.97	49.44	13.76	61.49	98.87	117.96	195.00
Collectible Card Game	6.00	95.90	48.79	8.99	91.12	104.33	110.95	157.91
Strategy	20.00	99.85	51.99	20.23	53.53	101.81	146.51	177.00
Fighting Game	25.00	111.35	40.81	11.00	86.00	113.10	144.50	199.00
First-Person Shooter	40.00	112.74	50.91	6.81	83.37	114.49	151.38	194.00
Role-Playing Game	3.00	116.83	16.38	102.31	107.96	113.61	124.09	134.58
Third-Person Shooter	4.00	128.68	15.33	115.00	118.17	125.26	135.77	149.22
Racing	12.00	134.91	28.91	96.83	107.81	138.00	158.75	176.00
Puzzle Game	4.00	154.21	31.37	115.50	135.38	159.83	178.67	181.67

largest range in games, with its maximum value at a rank of 199 and its minimum at 11. This analysis truly shows the wide range of popularity of games and their associated genres. Although many games may not have as big of a following, they can still hold an esports presence for their respective audiences.

Parts of My Analysis That Did Not Work

Some parts of my project plans were too difficult to fit into my overall analysis. Initially, I had planned to use a third type of dataset including video game ESRB ratings and different characteristics of games under their associated ratings. Unfortunately, this dataset only contained video games on PS4 or Xbox 1 consoles, which would not incorporate well into my analysis, as many listed Twitch and esports games are computer-based. This would have otherwise been interesting to see what percentage of games were different ratings based on genre, earnings, Twitch rank, and other factors.