A simple analysis of Twitch after Covid-19 pandemic

Francesco Baldi

2023-07-29

Table of contents

Chapter 1

Introduction

Twitch is a video live streaming established in 2011 and acquired by Amazon in 2014 for $1 billion. Initially designed for content like gaming and eSports, over time it has also expanded into non-gaming content like IRL (In-Real-Life), where creators share their activities with their viewers, Just Chatting, where streamers interact directly with their viewers on some topics, and so many other categories.

Personally, I frequently used Twitch during pandemic years like so many other people and that's why in this analysis I want to explore how Twitch has performed after its boom during the Covid-19 pandemic and identify the most relevant categories streamed from 2016 to 2023.

1.1 Data

The data I will analyze are two datasets: `Twitch_game_data` with observations per month about top games or categories, and `Twitch_global_data` with observations per month about general statistics from January 2016 to March 2023. These datasets are taken from the Kaggle's user Ran.Kirsh.

1.2 Libraries

```
library(tidyverse)
library(magrittr)
library(lubridate)
library(scales)
```

1.3 `Twitch_global_data` dataset

```
df <- read.csv("Twitch_global_data.csv")
head(df)
```

```
  year Month Hours_watched Avg_viewers Peak_viewers Streams Avg_channels
1 2016     1     480241904      646355      1275257 7701675        20076
2 2016     2     441859897      635769      1308032 7038520        20427
3 2016     3     490669308      660389      1591551 7390957        20271
```

```
4 2016     4     377975447      525696      1775120 6869719        16791
5 2016     5     449836631      605432      1438962 7535519        19394
6 2016     6     446429345      620903      1755888 6663363        18818
  Games_streamed Viewer_ratio
1          12149        29.08
2          12134        28.98
3          12234        28.92
4          12282        28.80
5          12424        28.85
6          12374        28.76
```

```
glimpse(df)
```

```
Rows: 87
Columns: 9
$ year           <int> 2016, 2016, 2016, 2016, 2016, 2016, 2016, 2016, 2016, 2~
$ Month          <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 1, 2, 3, 4, 5, 6~
$ Hours_watched  <dbl> 480241904, 441859897, 490669308, 377975447, 449836631, ~
$ Avg_viewers    <int> 646355, 635769, 660389, 525696, 605432, 620903, 600715,~
$ Peak_viewers   <int> 1275257, 1308032, 1591551, 1775120, 1438962, 1755888, 1~
$ Streams        <int> 7701675, 7038520, 7390957, 6869719, 7535519, 6663363, 7~
$ Avg_channels   <int> 20076, 20427, 20271, 16791, 19394, 18818, 18030, 16592,~
$ Games_streamed <int> 12149, 12134, 12234, 12282, 12424, 12374, 12961, 13693,~
$ Viewer_ratio   <dbl> 29.08, 28.98, 28.92, 28.80, 28.85, 28.76, 28.62, 28.47,~
```

1.4 Twitch_game_data dataset

```
df_games <- read.csv("Twitch_game_data.csv")
head(df_games)
```

```
  Rank                            Game Month Year Hours_watched Hours_streamed
1    1                League of Legends     1 2016      94377226        1362044
2    2 Counter-Strike: Global Offensive     1 2016      47832863         830105
3    3                          Dota 2     1 2016      45185893         433397
4    4                      Hearthstone     1 2016      39936159         235903
5    5      Call of Duty: Black Ops III     1 2016      16153057        1151578
6    6                        Minecraft     1 2016      10231056         490002
  Peak_viewers Peak_channels Streamers Avg_viewers Avg_channels
1       530270          2903    129172      127021         1833
2       372654          2197    120849       64378         1117
3       315083          1100     44074       60815          583
4       131357           517     36170       53749          317
5        71639          3620    214054       21740         1549
6        64432          1538     88820       13769          659
  Avg_viewer_ratio
1            69.29
2            57.62
3           104.26
4           169.29
5            14.03
6            20.88
```

```r
glimpse(df_games)
```

```
Rows: 17,400
Columns: 12
$ Rank             <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16~
$ Game             <chr> "League of Legends", "Counter-Strike: Global Offensiv~
$ Month            <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ Year             <int> 2016, 2016, 2016, 2016, 2016, 2016, 2016, 2016, 2016,~
$ Hours_watched    <int> 94377226, 47832863, 45185893, 39936159, 16153057, 102~
$ Hours_streamed   <int> 1362044, 830105, 433397, 235903, 1151578, 490002, 342~
$ Peak_viewers     <int> 530270, 372654, 315083, 131357, 71639, 64432, 46130, ~
$ Peak_channels    <int> 2903, 2197, 1100, 517, 3620, 1538, 1180, 460, 148, 75~
$ Streamers        <int> 129172, 120849, 44074, 36170, 214054, 88820, 33375, 2~
$ Avg_viewers      <int> 127021, 64378, 60815, 53749, 21740, 13769, 11805, 106~
$ Avg_channels     <int> 1833, 1117, 583, 317, 1549, 659, 461, 276, 71, 274, 9~
$ Avg_viewer_ratio <dbl> 69.29, 57.62, 104.26, 169.29, 14.03, 20.88, 25.57, 38~
```

Chapter 2

Data Analysis #1: How Twitch has performed after Covid-19 pandemic?

Now let's have a look into `Twitch_global_data` to observe how `Hours_watched`, `Avg_viewers` and `Avg_channels` have evolved through the years and, particularly after the Covid-19 pandemic.

```
df <- df |>
  mutate(Date = as.Date(paste("01", Month, year, sep = "-"),
                        format ="%d-%m-%Y")) |>
  mutate(year = as.factor(year))
```

2.1 Hours watched

Figure 2.1 shows how many hours were watched on Twitch by viewers. As we know, at the beginning of 2020 there was a significant increase in hours watched, going from approximately 1 billion hours watched to approximately 1.8 billion hours watched per month due to the first lockdowns. Then, with following lockdowns, hours watched reached its peak in May 2021 with approximately 2.8 billion hours watched. After this peak, we can observe a costant decrease of hours watched over the months.

```
df |>
  ggplot(aes(Date, Hours_watched/1E6)) +
  geom_col(aes(fill = year)) +
  theme_minimal() +
  scale_fill_brewer(palette = "Spectral") +
  scale_x_date(date_breaks = "1 years", date_labels = "%Y") +
  labs(x = "", y = "")
```

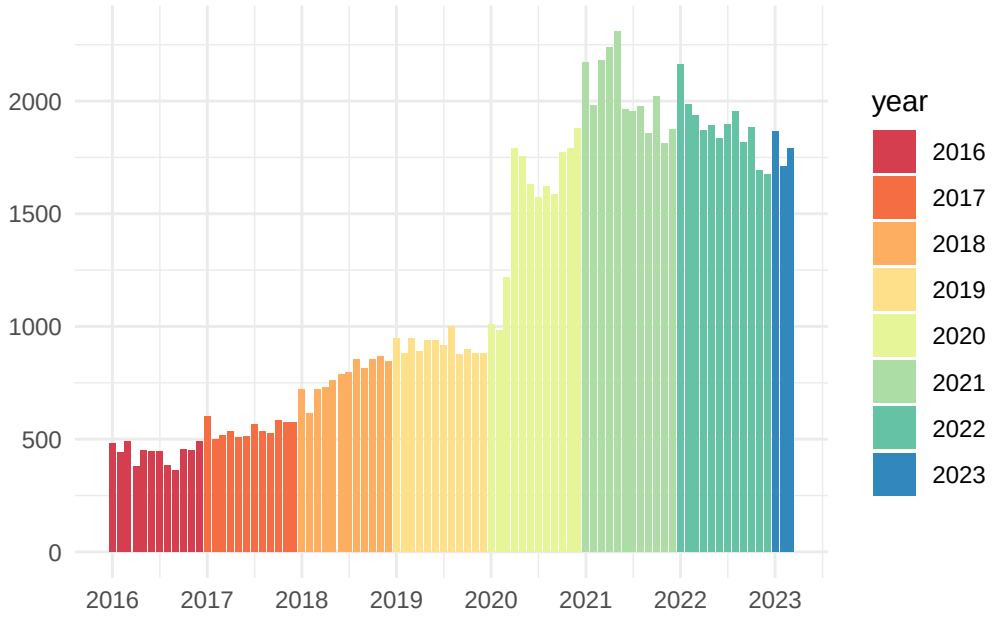

Figure 2.1: A bar plot of hours watched on Twitch from 2016 to 2023 in millions of viewers.

2.2 Average viewers

As we can observe, Figure 2.2 follows the same pattern as Figure 2.1. At the beginning of 2020 there was an increase, going from approximately 1.4 million average viewers to approximately 2.5 million average viewers. It then reached its peak to approximately 3.1 million average viewers in April 2021. After this point, we can observe the similar decrease, probably due to the same reasons.

```
df |>
  ggplot(aes(Date, Avg_viewers/1E6)) +
  geom_col(aes(fill = year)) +
  theme_minimal() +
  scale_x_date(date_breaks = "1 years", date_labels = "%Y") +
  scale_fill_brewer(palette = "Spectral") +
  labs(x = "", y = "")
```

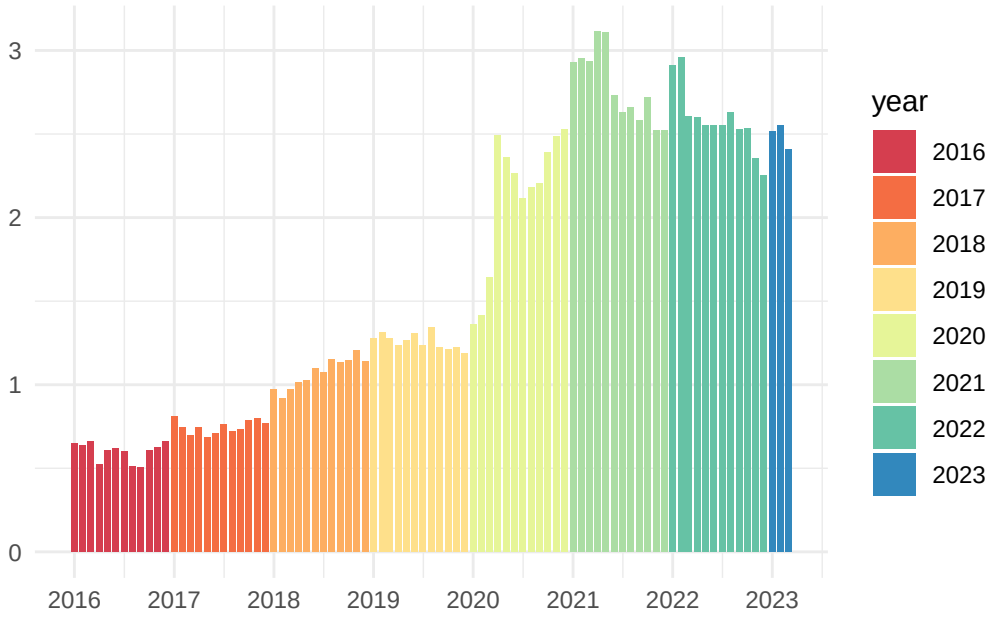

Figure 2.2: A bar plot of average viewers on Twitch from 2016 to 2023 in millions of viewers.

2.3 Average channels

Figure 2.3 follows a similar pattern compared to the previous Figure 2.1 and Figure 2.2. At the beginning of 2020 there was a growth from approximately 53 thousand average channels to approximately 100 thousand channels, and reached its peak to approximately 123 thousand average channels in February 2021. After its peak, we can observe a similar decrease in the following months.

```
df |>
  ggplot(aes(Date, Avg_channels/1E3)) +
  geom_col(aes(fill = year)) +
  theme_minimal() +
  scale_fill_brewer(palette = "Spectral") +
  scale_x_date(date_breaks = "1 years", date_labels = "%Y") +
  labs(x = "", y = "")
```

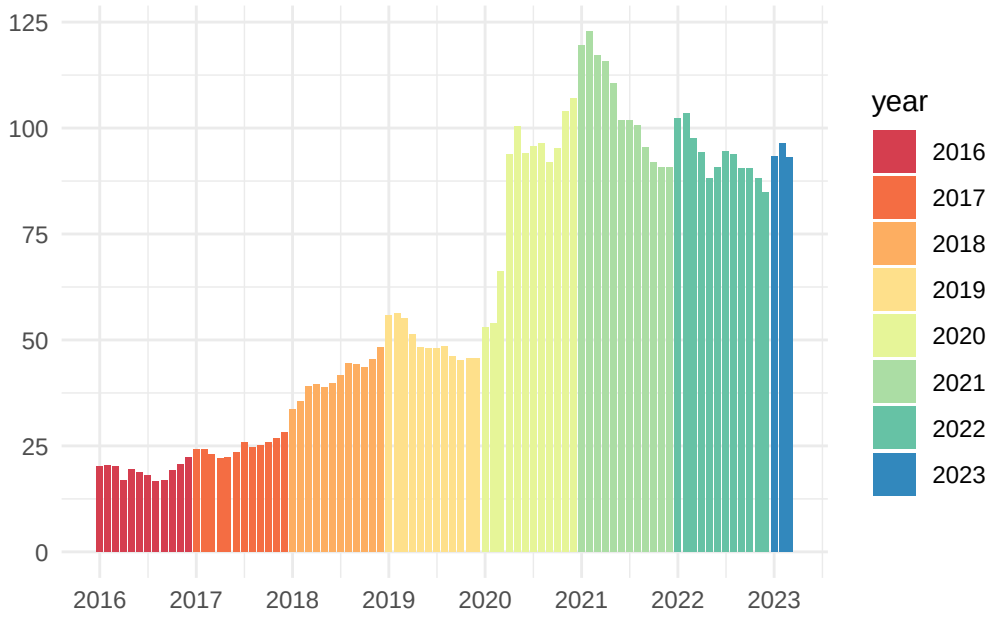

Figure 2.3: A bar plot of average channels on Twitch from 2016 to 2023 in thousand of viewers

2.4 Summary

In this chapter, we observed how Twitch has performed after its spike from 2020-2021 and how `Hours_watched`, `Avg_viewers` and `Avg_channels` have constantly decrease over the months after it reached its peak. This can be explained by a reduction in lockdowns and a return to daily life, resulting in a decline in casual viewers' interest in Twitch. Furthermore, another possible reason could be that Twitch has reduced its revenue share between streamers and the platform, resulting in some top streamers leaving along with their audience.

Chapter 3

Data Analysis #2: Which are the most relevant Twitch's categories?

In this 2nd part of Twitch's data analysis, the goal is to explore `Twitch_game_data` and identify which are the most relevant categories.

3.1 Most watched categories

I filtered the monthly top 5 games and then counted how many times each game appeared in top 5 ranks. After this, I create a bar plot using the new variable `Count_Months_Top5` to identify the most watched categories.

```
df_games |>
  filter(between(Rank,1 ,5)) |>
  group_by(Game) %>%
  summarize(Count_Months_Top5 = n()) |>
  arrange(desc(Count_Months_Top5)) |>
  ggplot(aes(Count_Months_Top5, reorder(Game, Count_Months_Top5))) +
  geom_col(fill = "#6441a5") +
  theme_minimal() +
  labs(x = "", y = "") +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75, 90))
```

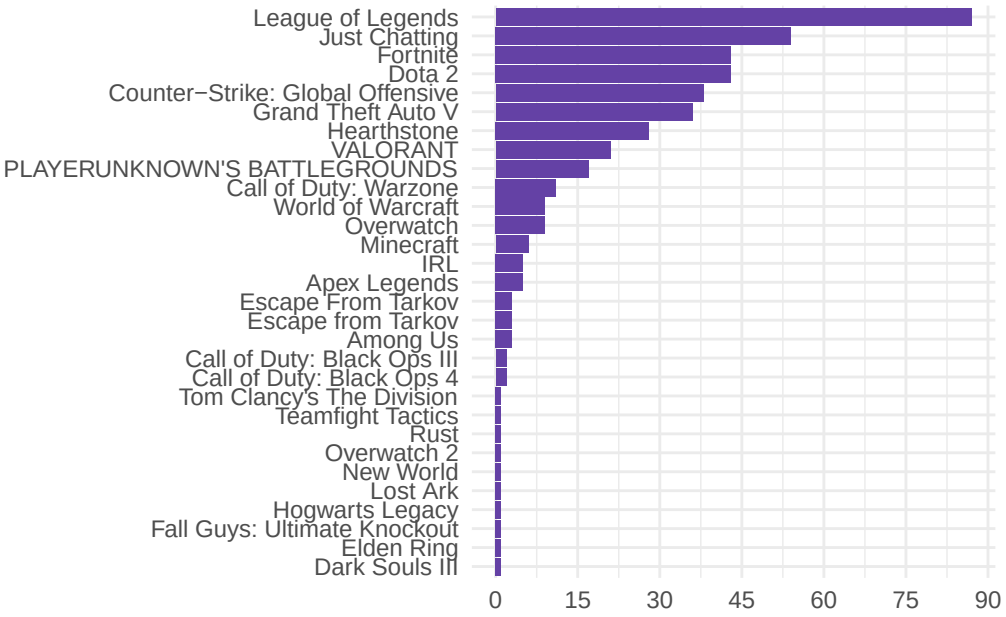

Figure 3.1: A bar plot of months in most watched categories from 2016-2023 - Top 5

Figure 3.1 shows that the most watched category is League of Legend. One notable observation is that all the games in the high position on the bar chart are free-to-play, except for Grand Theft Auto V. This can be explain by free-to-play games' replayability which is often consistent over the time.

The only non-games categories are Just Chatting in 2nd position and IRL in lower position, which confirms that Twitch is not only a live streaming platform for gaming, but also a platform to directly interact with viewers.

```
df_games |>
  filter(Year == 2023) |>
  filter(between(Rank,1 ,5)) |>
  group_by(Game) %>%
  summarize(Count_Months_Top5 = n()) |>
  arrange(desc(Count_Months_Top5)) |>
  ggplot(aes(Count_Months_Top5, reorder(Game, Count_Months_Top5))) +
  geom_col(fill = "#6441a5") +
  theme_minimal() +
  labs(x = "", y = "")
```

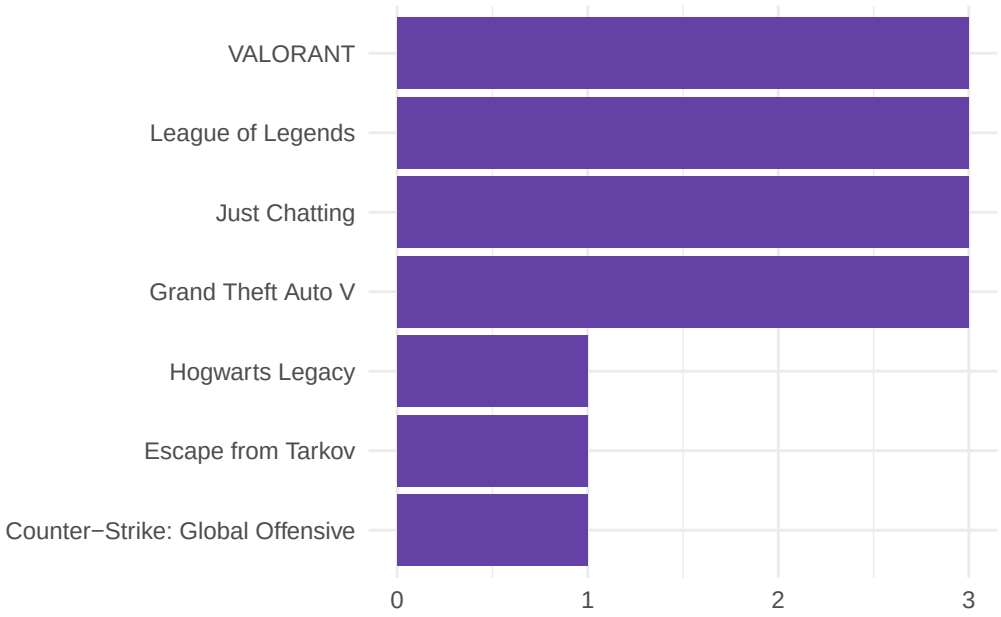

Figure 3.2: A bar plot of months in most watched categories from January 2023-March 2023 - Top 5

Here is Figure 3.2 which shows the most watched categories from January 2023 to March 2023. Once again, I can highlight that the most watched categorie are free-to-play games like VALORANT and League of Legends, as well as Grand Theft Auto V, a game realese in 2013 with ongoing popularity, and Just Chatting.

3.2 Hours watched by categories

```
df_games |>
  group_by(Game) |>
  summarize(Tot_Hours_watched = sum(Hours_watched)) |>
  arrange(desc(Tot_Hours_watched)) |>
  slice(1:10) |>
  ggplot(aes(Tot_Hours_watched/1E6, reorder(Game, Tot_Hours_watched))) +
  geom_col(fill = "#6441a5") +
  theme_minimal() +
  labs(x = "", y = "") +
  scale_x_continuous(labels = comma)
```

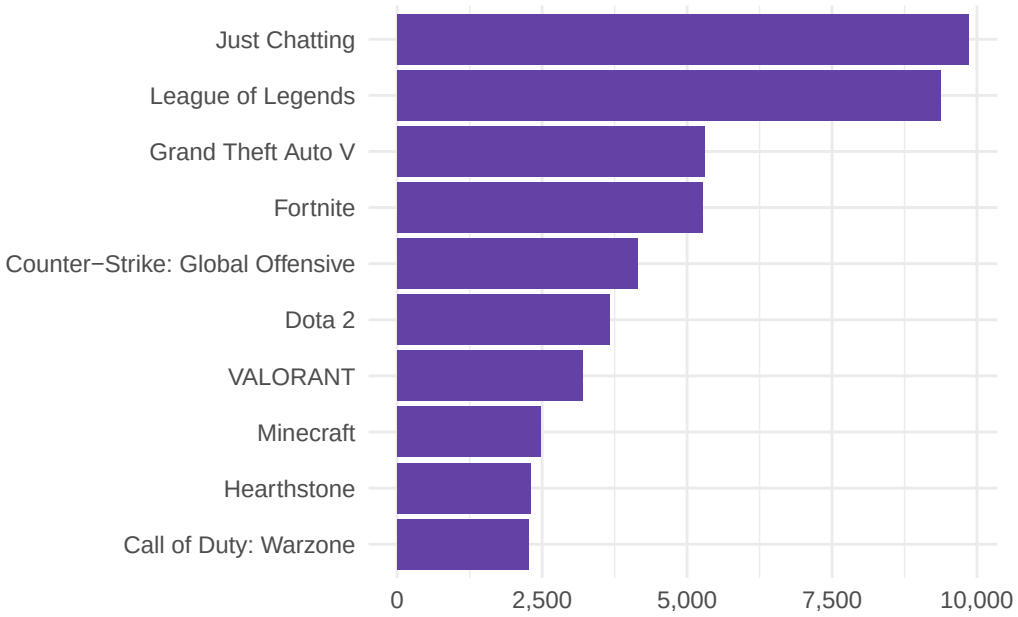

Figure 3.3: A bar plot of Hours watched by categories from 2016-2023 in millions of hours

The Figure 3.3 shows how many hours were watched on Twitch by categories. Just Chatting is the most watched category by hours watched from 2016 to 2023 which confirms what I said before about the fact that Twitch is used to engage with viewers and other streamers.

3.3 Average monthly streamers by categories

```
df_games |>
  group_by(Game) |>
  summarize(Avg_streamers = mean(Streamers)) |>
  arrange(desc(Avg_streamers)) |>
  slice(1:10) |>
  ggplot(aes(Avg_streamers, reorder(Game, Avg_streamers))) +
  geom_col(fill = "#6441a5") +
  theme_minimal() +
  labs(x = "", y = "") +
  scale_x_continuous(breaks = seq(0, 600000, by = 150000),
                     labels = comma)
```

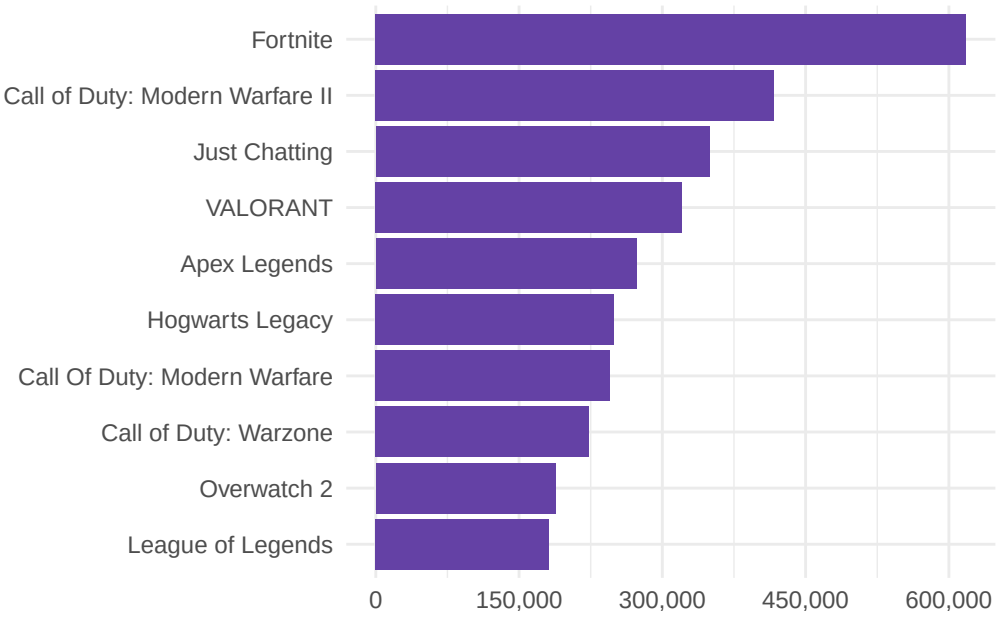

Figure 3.4: A bar plot of average monthly streamers by categories from 2016-2023

In conclusion, Figure 3.4 shows that Twitch's streamers prefer to stream Fortnite instead of Just Chatting or League of Legends which are the most watched categories.

3.4 Summary

In this final chapter, we observed which are the most watched categories, how many hours were watched for each category and which are the most streamed categories by streamers. Specifically, we observed that, generally, the most watched categories are games like free-to-play titles or GTA 5, as well ass Just Chatting and IRL which are categories focused on interact and engagement.