Twitch 2023 - Top 900 Streamers - Statistical Analysis

The Twitch 2023 Top 900 Streamers is both a prestigious and coveted rank among content creators everywhere because it means that you are among the best and most successful live streamers on the planet. This is thanks to the fact that Twitch is the worlds largest online live-streaming service and was acquired by Amazon about 10 years ago because Bezos saw the potential it had. Today millions of content creators log on to Twitch to share their art, music, gaming skills, start businesses, among others. With such a diverse array of creators it will be both interesting and in my opinion enlightening to see what really sets these creators apart from the millions of others with the same passion. Is it a massive following? Does that following mean the most followed will be the most watched? What about the most average viewers? It would make sense that the most followed pages would command the largest and most consistent slice of the viewing pie, right? And finally the biggest question of them all, how do you get into the Top 900? We know that Twitch (as a live streaming service) uses watch time as their determinant as that is the only goal of the platform, to keep users watching content. With that in mind we will look at the channels with the highest average watch time per live-stream, and watch time overall (recorded and live content) to determine if there is any correlation between the variables. These are the questions we will be asking going forward, and hopefully answering in the first half of the analysis. In the second half we will be looking at a random sample of 100 streamers from the Top 900, to see if we can make any accurate predictions about the language make up of the next Top 900. So without further delay, let's dive right in.

We begin with the raw data set of n=900; let's set up a simple scatter plot to answer our first question, "Do the most FOLLOWED pages have the MOST WATCH time overall? While the plot may be definitely scattered, we can faintly make out the weakest positive correlation possible. Some of the most followed channels are indeed the most watched, but it is also very clear that many of the most watched channels are absolutely no where close to the most followed channels either. What's going on here? Let's see if we cant uncover this mystery by answering some more questions. What about the most average viewers? The most followed pages may not have the most watch time overall, but that's just because all their watch time is live, right? Let's find out, enter scatter plot #2, comparing follower count to average watch time. These sure are some weak, positive correlations, but they are positive.

Again we do see many of the channels that command a larger following rising to the top of the average viewing, but at the same time we also see many examples of where that simply isn't the case as well. How can so many channels with such a smaller following command a greater average viewing than their super famous counter parts? Ok, ok lets take stock of what we've learned so far. While there is definitely a correlation between your channels follower count and its watch time and average viewer per stream, that correlation is insanely weak! A wonderful example of where correlation is absolutely not a factor in causation, because of the sheer examples that defy those bounds and soar to the top of view time both average and overall. But let's see about our big question, do these channels that have command a high average view time or a high watch time, coincide with each other? Enter scatter plot #3; and now there are even more questions. What we see is an absolutely positive correlation, but not in the way we think. It would seem almost like there's a threshold where after so many

average viewers are obtained, you've exhausted your watch time. This makes sense when we think of the pages that have essentially their entire fan base watching them at one given time, but then is it just the most followed pages that command that immense watch time? We already saw that to be false through the first scatter plot, the most followed pages are not all the most watched. And if the most followed pages aren't the most watched, and the pages with the highest average viewers aren't the most watched, how you do become the most watched? Is it stream time? That would make sense at face value, more time on air should equate to more content and in turn more viewership.

Let's test that theory with another, you guessed it, scatter plot! Enter scatter plot #4; well that was not what I was expecting at all. In my opinion it would seem there is no correlation at all. With many of the channels that stream the longest being the least watched. I suppose that eliminates that theory and reveals that it truly must be the content of the channel that commands viewing. Whatever that may be is the driving factor behind the increased retention rate and exposure that those channels are seeing. And that's what twitch is after. Attention is the true digital currency and it with the current data set we have it would be impossible to tell what kind of content will get you on the list because much of the content itself will be undefined. A personality is hard to quantify in terms and especially one people will gravitate towards through a digital medium like their phones. But that seems to be the determinant of watch time and thus the entrance into the TOP 900. Do people like your content enough to tell their friends about it, to sit down and watch a recorded episode because they connect to what you are presenting, playing or have created as a new activity.

For example, the channel with the most average viewers out of the entire TOP 900, is @kingsleague. Not a personality streamer, not a professional gamer, not even a famous celebrity, but a 7-on-7 soccer league created by Gerard Pique (his fame definitely helped, but not as much as one would think since its following is still small compared to the exposure he got at FC Barcelona, for instance I know OF him as a player but didn't know he created this until now!). In my opinion its the focus that the channel represents that commands such insane viewer loyalty, with a little over 4% of their followers consistently tuning in (Avg. Viewers 132,356/ Followers 3,043,007). While this seems like a miniscule percentage, the top 10 most followed channels each only claim; 0.3% (#1 Auronplay), 0.4% (#2 ibai), 0.2% (Rubius), 0.4% (xQc) and 0.2% (TheGrefg). The immense following that these channels have does reflect in their average viewership being higher than most, but not all. To make it to the top of the 900 doesn't require the biggest following, it requires the biggest audience. While your following may be large, they may not be watching. KaiCenat was the 2023 most watched streamer and he is no where near the most followed, nor did he stream the most either. There can be no other explanation than that people simply loved what he streamed when he was on the air, and they told their friends who may have tuned into his channel to watch, but didn't necessarily follow. So I guess what it really takes to get into the TOP 900 is word of mouth and a little bit of luck because if watch time is the determinant, and following, stream time, and average viewers doesn't get you there, lady luck and a great marketing team will.

Now that we have established what it takes as a creator to reach the TOP 900, lets see if we can predict that makeup of languages for the next listing in 2024. We see the language breakdown from our n=900 initial data set and can already see that English and Spanish make up almost 55% of the list at 43.78% and 11%. But how confident can we be moving forward that these percentages will be the same? Let's find out. Now we can refer to the second data set attached, a random sample of 100 streamers from the TOP 900. In our random sample language pie chart we see that there are 47/100 (47%) English speaking channels. Lets construct a 95% confidence interval and see what our upper limits and lower limits will be for the 2024 list based on this random sample. First lets conduct a hypothesis test.

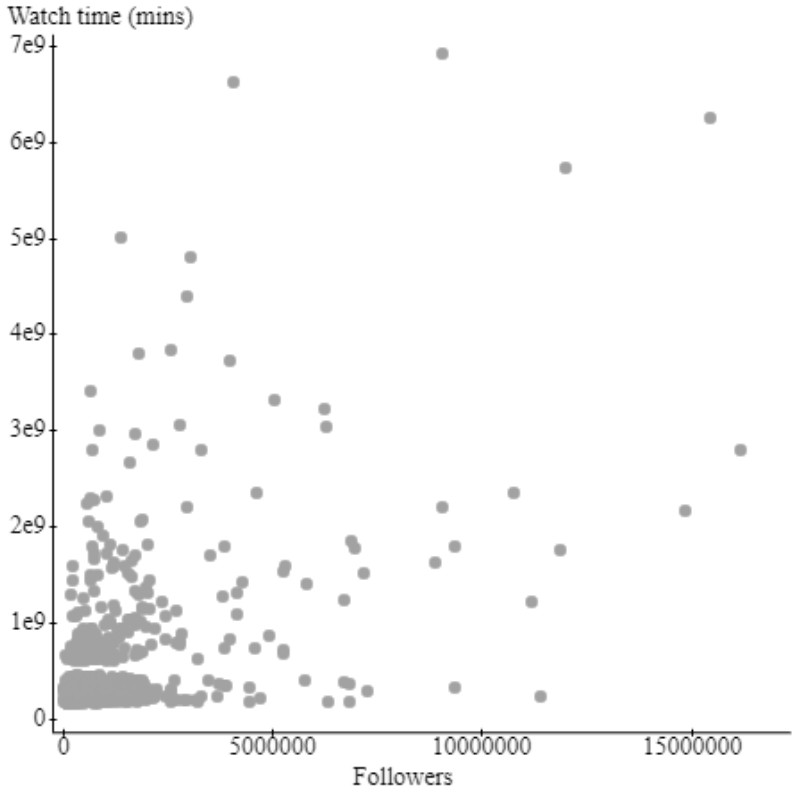
The null hypothesis states that the future proportion of English speaking streamers will equal 47%.

The Alternative hypothesis states that it will not be equal to 47%.

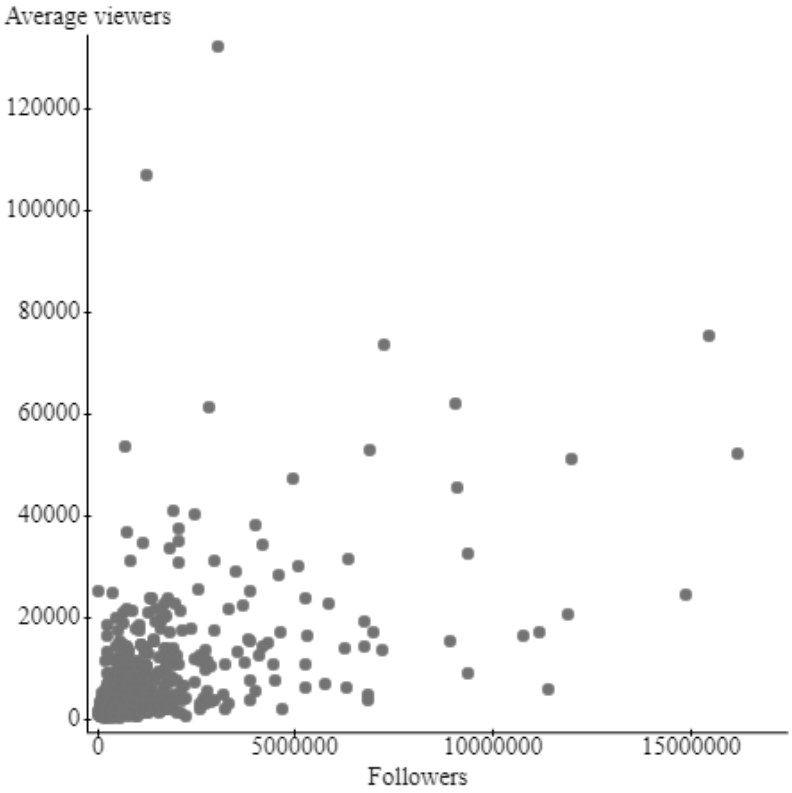
With a p-value of 1, and a statistically significant threshold of .5 we cannot reject that null hypothesis and must accept the fact that it is entirely possible that the future breakdown of English speakers in the next TOP 900 will accrue for 47% of the list. But now lets make a prediction with 95% certainty as to what the percentage will really be. After calculations we can be certain that there is a 95% chance that the proportion of English speakers in the TOP 900 will be between 31.2178% and 56.7821% with a standard error of 4.9%.

Thanks for joining me on this quick statistical analysis of the TOP 900 streamers on twitch, while we may be able to predict the language these creators will be speaking in the TOP 900, a deeper dive into the content of the most successful channels both in average watch time and overall watch time is necessary to determine what gets you into the TOP 900, because it definitely isn't just being able to speak English, that's only half the battle!

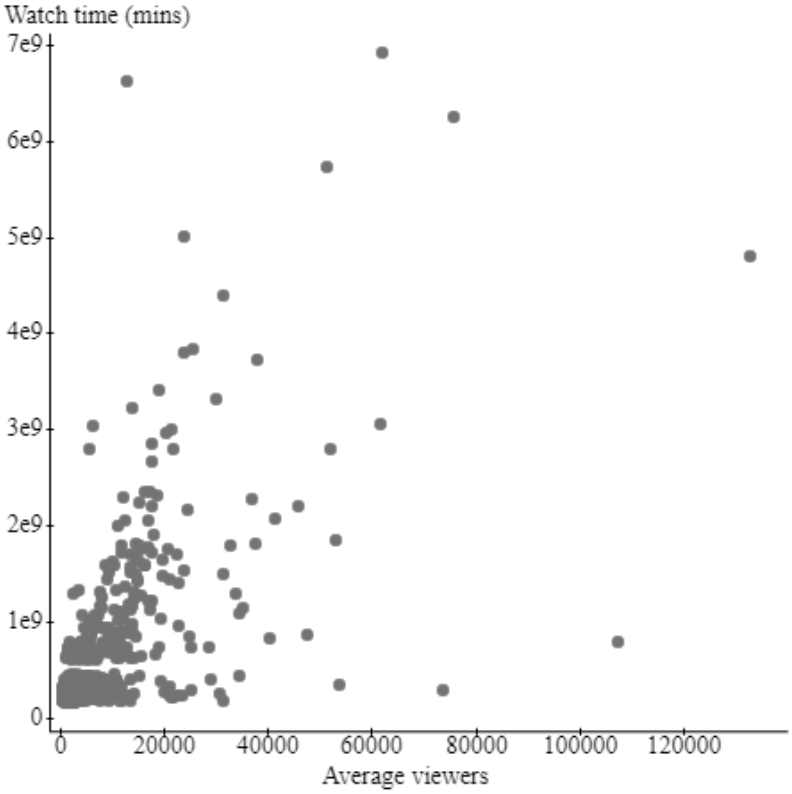
Result 1: Scatter Plot



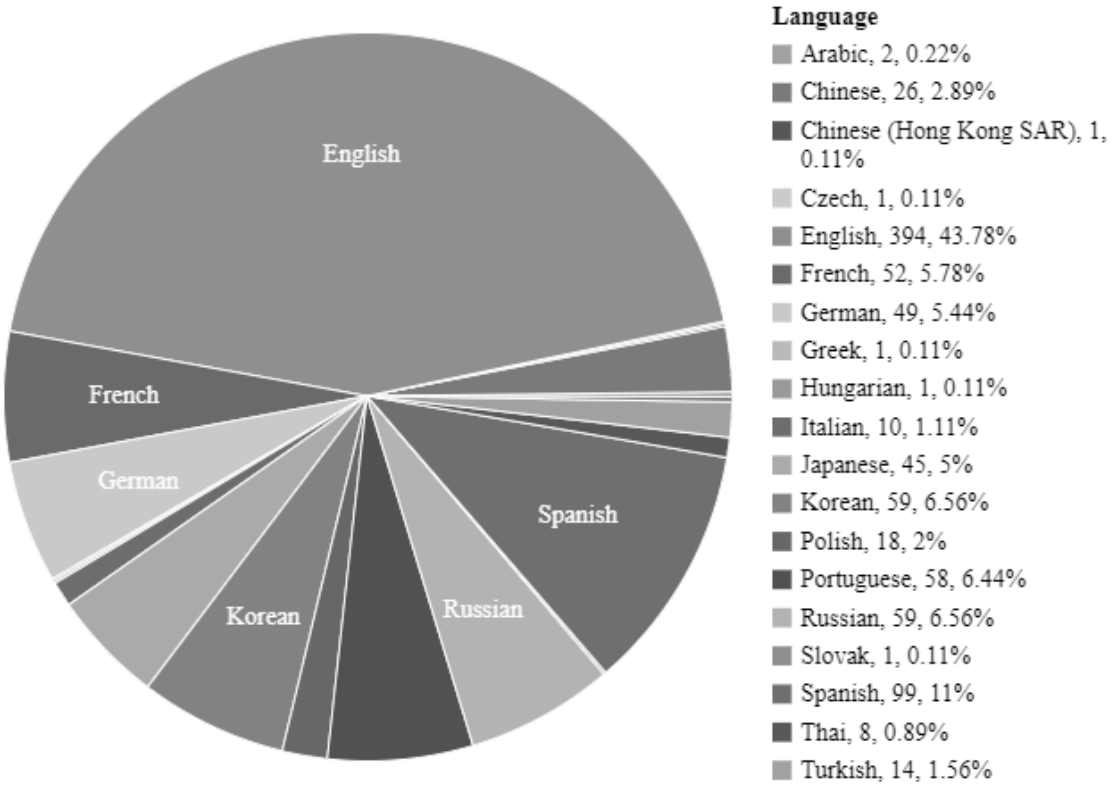
Result 2: Scatter Plot #2



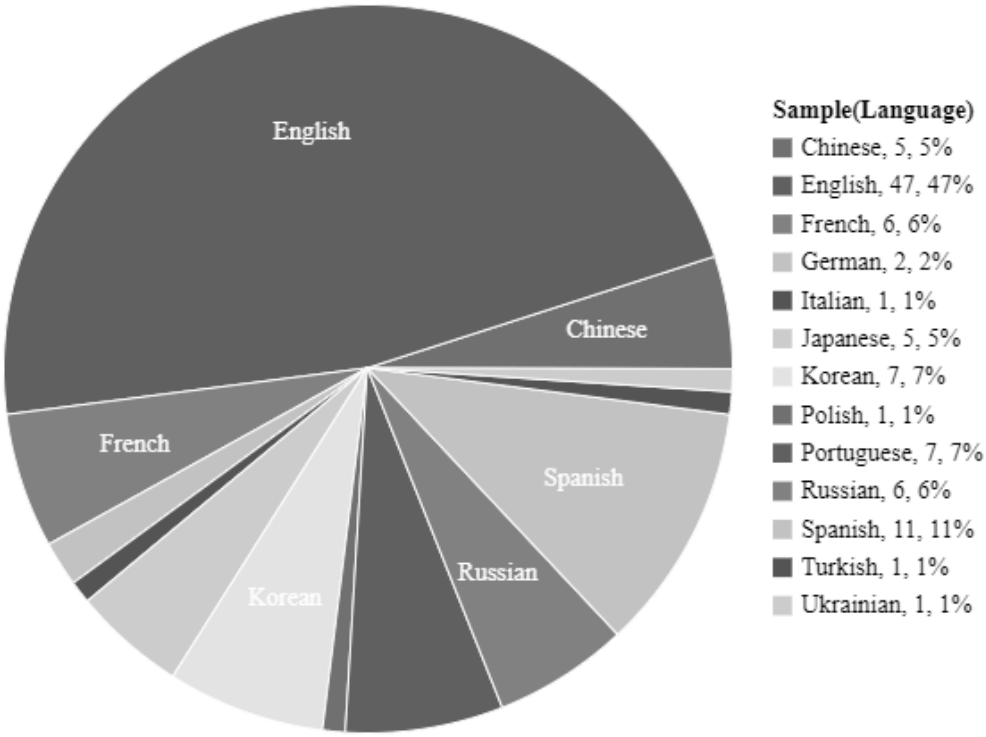
Result 3: Scatter Plot #3



Result 4: Language Pie



Result 5: Random Sample Language Pie



Result 6: One sample proportion hypothesis test

One sample proportion hypothesis test:

Outcomes in : Sample(Language)
Success : English
p : Proportion of successes
H_0 : p = 0.47
H_A : p ≠ 0.47

Hypothesis test results:

Variable	Count	Total	Sample Prop.	Std. Err.	Z-Stat	P-value
Sample(Language)	47	100	0.47	0.049909919	0	1

Result 7: One sample proportion confidence interval

One sample proportion confidence interval:

Outcomes in : Sample(Language)
Success : English
p : Proportion of successes
Method: Standard-Wald

95% confidence interval results:

Variable	Count	Total	Sample Prop.	Std. Err.	L. Limit	U. Limit
Sample(Language)	47	100	0.47	0.049909919	0.37217836	0.56782164

Data set 1: Twitch Streamer Data 2023

Data set 2: Sampled from Twitch Streamer Data 2023