

Twitch Top 1000 Streamer Analysis

Kim yeon jae (2019311895)

■ Dataset name: Top Streamers on Twitch

■ Brief description about the dataset

Twitch is the largest game streaming platform in the world. The data set I chose in this assignment is the information of the top 1,000 people who streamed on this Twitch platform. The data contains 11 pieces of information about streamers, such as the number of viewers, active viewers, and followers.

■ WHAT: Data Abstraction

- my dataset size - **1000 rows, 11 attributes, 78KB** / my dataset type - **.csv file**
- the type and semantics of each attribute

attribute	type	semantics
Rank	Int	The streamer rank of twitch platform
Channel	String	Streamer's channel name
Watch time(minutes)	Int	Total watch time for the channel
Stream time(minutes)	Int	Total stream time for the channel
Peak viewers	Int	Highest number of viewers during streaming
Average viewers	Int	Average number of viewers during streaming
Followers	Int	Total number of followers on the channel
Followers gained	Int	Number of new followers in the channel
Views gained	Int	Number of new viewers on the channel
Partnered	Boolean	Whether the channel has partnered with the Twitch platform (True/False)
Mature	Boolean	Whether the channel is only available to adults (True/False)
Language	String	What language does the channel speak

- nominal or ordinal attribute

Nominal value: **Partnered** attribute (Whether this streamer have partnership with twitch platform), **Mature** attribute (Whether this streaming channel is restricted above 18 years old), **Language** attribute (Represent what language does the streamer use).

Ordinal value: **Rank** attribute (Rank attribute represent streamers rank in twitch platform standard with watch time, follower number ect.) This attribute has descending order which means higher rank have small integer value.

- quantitative attribute

Interval value: There are no interval value, because each attribute's data is not evenly spaced, only have ratio value such as time data.

Ratio value: **Watch time** attribute (Show the total watch time, and this attribute follows ascending order), **Peak view** attribute (Represent highest view of streaming channel, and this attribute follows ascending order), **Average view** attribute (Represent average view of streaming channel, and this attribute follows ascending order), **Follower** attribute (Represent total number of follower of specific streaming channel, and this attribute follows ascending order), **follower gained** attribute, **Views gained** attribute (These attributes represent number of new arrivals, and this attribute follows ascending order).

- Separate the attributes into keys and values.

Key: **Channel, Rank**

Value: **Watch time, Stream time, Peak view, Average view, Follower, follower gained, Views gained, Partnered, Mature, Language**

● WHY: Task Abstraction

- **Which languages are actively used on Twitch platforms?**

[Analysis(discover) – Extremes] The purpose of this task is to find out which language speakers use the Twitch platform a lot. In order to proceed with the study, streamers were grouped according to the language used, and the average number of followers was calculated for each group. Through this, it is possible to determine which language speakers use Twitch a lot in the order of having a large average number of followers.

- **Which channel has more people coming in?**

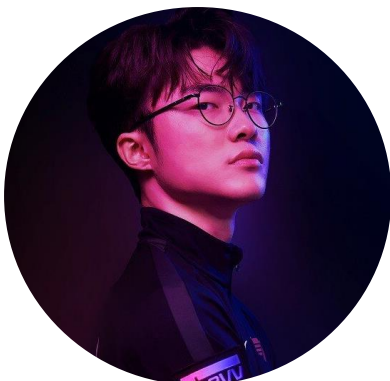
[Analysis(discover) – Distribution] The purpose of this task is to check which channels attract a lot of people. First of all, if you have a large number of gained followers compared to the number of followers, you can see that the number of people who have recently become interested in the channel has increased. Second, when the number of average viewers is large compared to the number of followers, it can be seen that there are many people who watch the channel even if they have not followed the channel. These two comparisons show which channels attract a lot of people.

- **Whether age-restricted broadcasting affects the growth of the channel.**

[Analysis(discover) – Correlation] The purpose of this task is to understand the correlation between channel growth and age-limited broadcasting. In the previous analysis, it was confirmed that there are many people who flow into the channel when the number of followers is higher than the number of followers and the number of viewers is higher than the number of followers. This also means that the channel is likely to grow due to the influx of population. Therefore, we divide the streamers into groups that broadcast age-limited and those that do not, and then analyze how the streamers in each group show in terms of the number of followers, the number of incoming followers, and the number of viewers. Through this, it is possible to determine whether the restriction of the viewing age affects the channel growth.

● WHO: Creating Personas

1. **Min-joon Kim** (21, college student, Korea)



Goal and Needs: Min-joon Kim, a college student, is not good at games. but he's trying to start broadcasting on the Twitch platform. He is a fan of Twitch LoL streamer Faker and wants to become famous streamers by using him as role models.

Pain Points: Unable to determine which language to broadcast. Thinking about partnership with Twitch.

2. **Oliver Miller** (48, advertiser, USA)

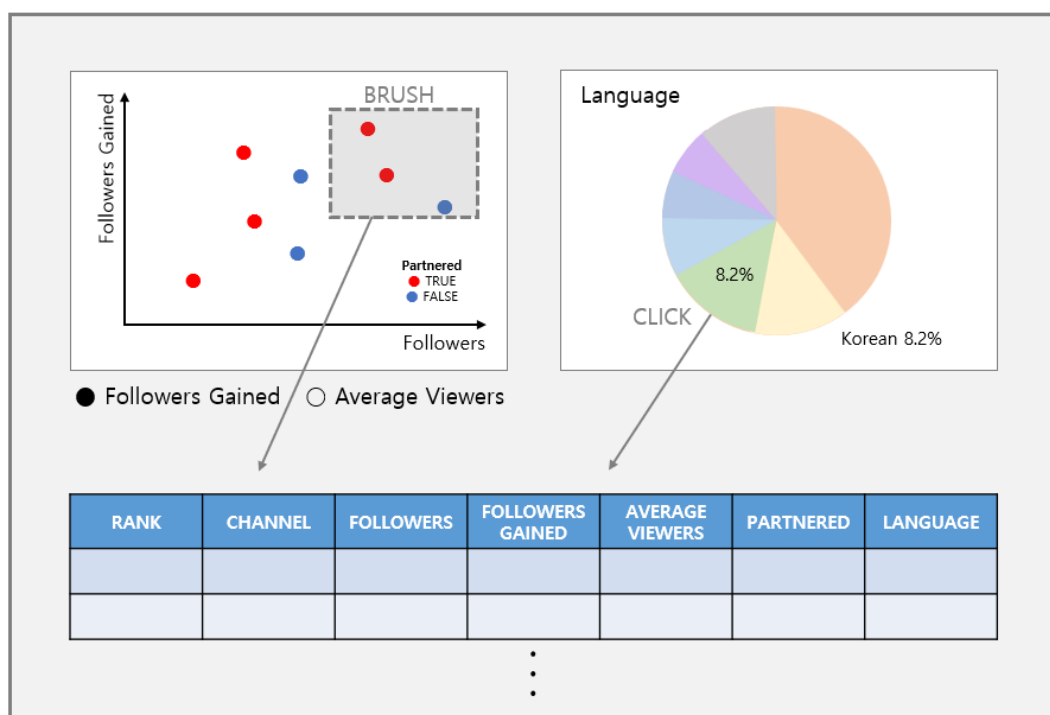


Goal and Needs: Olive Miller, the advertiser of the U.S. chicken franchise company, wants to maximize the advertising effect by suggesting 'mukbang' advertisements to famous streamers who broadcast on the Twitch platform.

Pain Points: Can't decide which streamers should be responsible for advertising.

Difference between personas: The primary persona is interested in how the channel can grow because he will be broadcasting on the Twitch platform. On the other hand, the secondary persona is interested in how to maximize the effectiveness of the advertisement, so it is necessary to find the best streamer by considering the number of subscribers and viewers of the already successful Twitch streamers.

• **HOW: Vis Idiom Design**



- The scatterplot on the left was selected to compare the number of gained followers and the number of viewers compared to the number of followers. Each element is expressed in a different color depending on whether it has partnered with the Twitch platform. Through this, it is possible to compare how the partnership affects the growth of the channel and the number of viewers for each channel.
- The pie chart on the right was selected to express how many streamers it has in total according to each language. Through this, it is possible to check which language speakers use the Twitch platform a lot.
- The table below was selected to output information on both selected elements in the above two graphs.
 - ⇒ Interaction: In the scatter plot on the left, you can select attributes to enter the y-axis from the choices below the graph. Certain elements can also be selected through drawing. The pie chart on the right may select a corresponding language by clicking the bar. The elements selected in the two graphs are printed in detail in the table below.

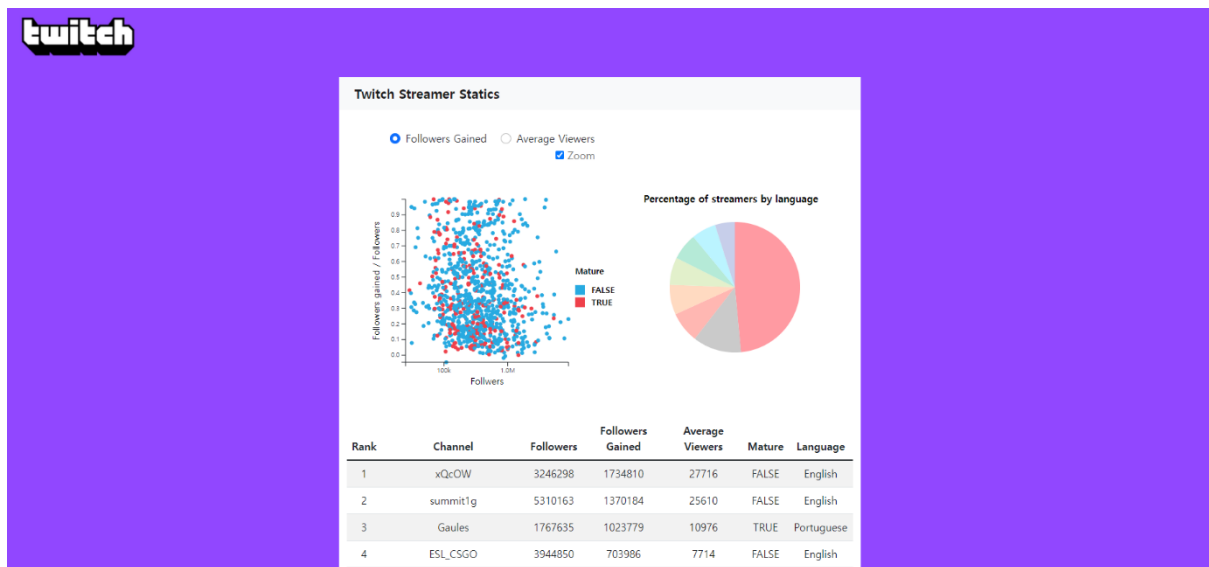
● Case Study

- Primary persona is thinking about which language to broadcast in, and is wondering which language the Top 1000 streamers use a lot. Therefore, he learned from pie chart that the most frequently used language by top 1000 streamers is English, followed by Korean. Based on this, he decided to conduct broadcasts using English and Korean.
- Primary persona is wondering if age restriction on broadcasting affects the growth of channels. Therefore, he analyzed the distribution shown in the scatter plot, and confirmed that age-restricted channels were evenly distributed in all channels that grew rapidly or slowly. Therefore, it was confirmed that the age limit was not related to the growth of the channel. However, it was confirmed that among the channels with a large number of viewers compared to the number of subscribers, there were relatively few age-limit channels. Therefore, the analysis results showed that it would be a little more advantageous not to limit the age.
- Secondary persona is considering which channel it would be efficient to leave ads on. Therefore, he was able to check the list of channels in the table at the bottom by dragging channels that were rapidly growing in the scatter plot, or had a large number of viewers compared to the number of subscribers.

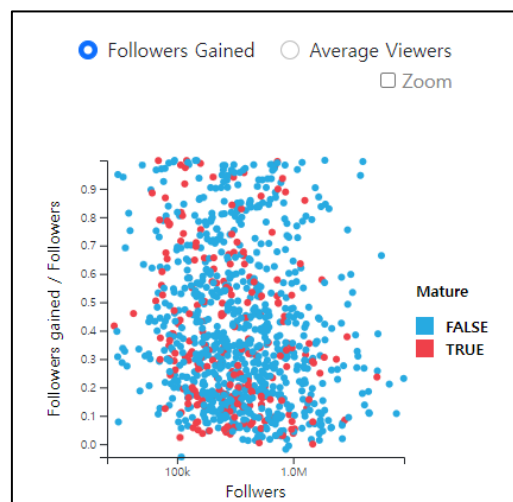
- Web Page Link

<https://yeonjaekim99.github.io/Information-Visualization-Project/index.html>

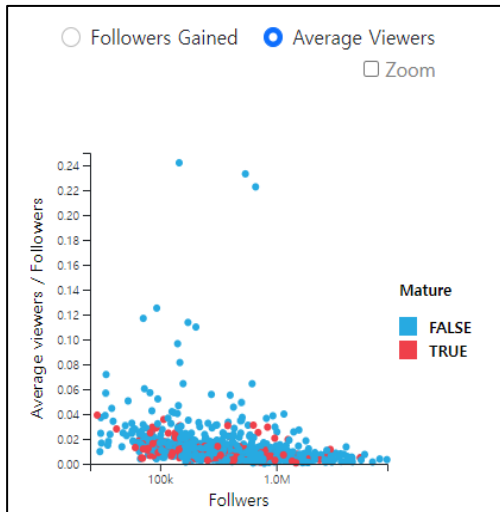
- Screen Shot



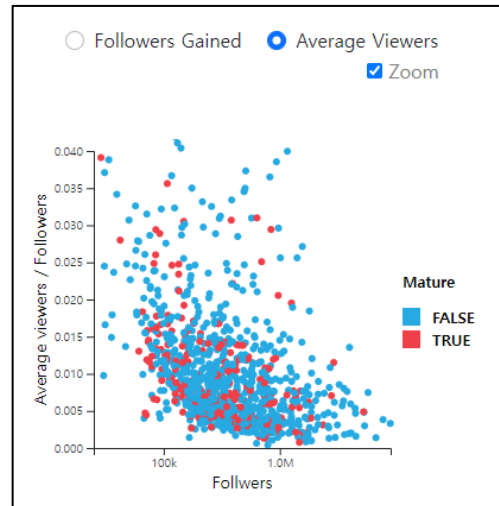
<Full screen shot>



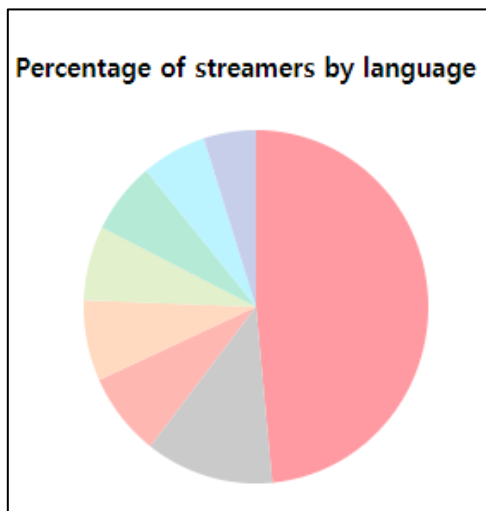
<Followers gained scatter plot>



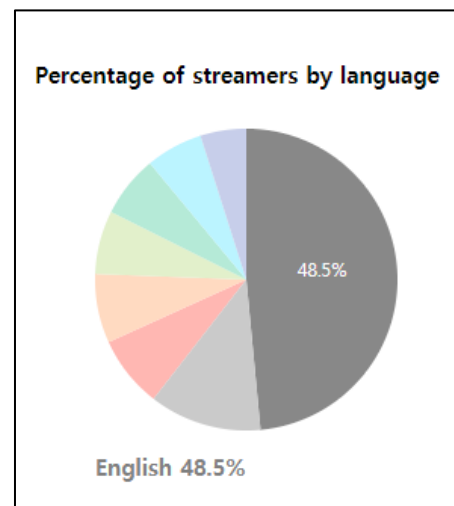
<Average Viewers scatter plot>



<Average Viewers scatter plot(Zoom)>



<Language pie chart>



<Language pie chart (mouse over)>

Rank	Channel	Followers	Followers Gained	Average Viewers	Mature	Language
1	xQcOW	3246298	1734810	27716	FALSE	English
2	summit1g	5310163	1370184	25610	FALSE	English
3	Gaules	1767635	1023779	10976	TRUE	Portuguese
4	ESL_CSGO	3944850	703986	7714	FALSE	English
5	Tfue	8938903	2068424	29602	FALSE	English
6	Asmongold	1563438	554201	42414	FALSE	English

<Data Table>