Samuel Gadson

10/20/2021

Colab link: <https://colab.research.google.com/drive/1mnxPtDIdGTU28xsdEsnrG6z1F7b2bE-o#scrollTo=xH3zG9L9dFVm>

Dataset link: <https://www.kaggle.com/rankirsh/evolution-of-top-games-on-twitch>

My dive into the statistical side of Twitch

Twitch.tv is one of the world's most popular online streaming services to use to watch online media created by influencers live.Bringing in nearly thirty million different users daily to view content everyday according to Twitch. Making them ranked thirty-seven amongst overall internet engagement. Thus why I wanted to see what content streamed on Twitch did the best overall over the course of the last five year. My goal is to analyze a dataset off kaggle.com that contains different information based off the monthly top 200 games on the Twitch.tv platform and figure out the average monthly watch time, and use that data to see what game did the best overall over its lifetime on Twitch. As well as seeing what games placed within the top five the most as well as taking a deeper look into the top three games peak viewers.

I chose this dataset solely for the purpose that I have spent many hours myself watching streamers on Twitch as well as having played many of the games that have been boosted to high popularity such as the Call of Duty Franchise, Hearthstone, and League of Legends. With that said I don’t think my analysis will be too different from others. Since the dataset was fully complete before I started doing all of my filtering. The author of the data set Ran Kirsh did a similar but more in depth analysis of the dataset only using the global twitch dataset also put on kaggle by him. Another User used the same dataset to try and predict what games will be watched in the future.

Since this is my first big project for coding started by looking at examples of people’s data analysis such as the project demo. After I finished looking at examples I immediately hopped into setting up a function to determine a games placement in the .csv file. Since one of the columns was called rank it made it easy to count a games recurrence in the file by setting up a for loop and using an empty dictionary to hold all the information I needed. Therefore I called the function top\_count. From there I decided to setup a bar graph to show visually how many times a game showed up on the csv in general. League of Legends was the game that showed up the most on the csv a total of sixty-seven times. The next two were Dota 2 and Fortnite at forty-one times, Just Chatting a total of thirty-four times, and Finally Counter Strike: Global Offensive showed up a total of thirty-three. Thus giving us our top five games that have been on twitch since 2016, but more importantly our top three games.

Next I made histograms for each game in the top three based on their hours watched over the years. I can confidently say if I made any mistake in my data analysis that would make me doubt if it was a success it would be my histograms. However after alot of struggling with setting them up I removed the scientific annotation of 1e8, and 1e7 off the tick axis so the regular numbers would show. Then since Dota 2 didn’t do as well in overall hours watched as League of Legends and Fortnite I made an overlapping histogram showing the correlation between the two. It shows that overall League of Legends has done better more consistently over time then Fortnite even in Fortnite’s most popular time. Finally I set up a box and whisker plot base off the three games’ peak viewers. I figured a box and whisker plot would be the best since it is very useful when it comes to figuring out different quartiles and percentiles. With it I was indeed able to conclude that Dota 2 did much worse than the other two. Dota 2’s peak viewers at the most was 1,045,985 which while respectable is nowhere close to League’s 2,020,835 views. However in a strange turn of events Fortnite, which has less entries in the top 200 games and doesn’t do as well as League of Legends does in terms of hours watched per month. Fortnite has the highest peak viewer count at a staggering 2,331,987 viewers. At least three-hundred thousand more than league of legends has had!

Overall I would say that my analysis was a success. I have figured out that the content streamed on Twitch that did the best overall over the course of the last five years was League of Legends based on its Ranking in the dataset and it’s Hours watched. Fortnite is second or possibly possibly first if you go off peak viewers.

Sources:

https://www.businessofapps.com/data/twitch-statistics/