Introduction

Practice makes perfect! Or at least that’s what people say. In our project, we sought to discover if playing more games competitively would correlate to a better performance output. We took data from “speedrun.com” and analyzed the times of players, and compared them, to see if factors such as gaming platform, run category, or game affect these participants. The tasks ahead of us were to first find a way to scrape the data from “speedrun.com”’s API for a sample set of 10 different games, clean the data, and create visuals that displayed trendlines based on the data. We needed to do all this to grasp a better understanding of the story that the data showcases.

Data Description

Our dataset is stored onto csv files and takes data from the Speedrun.com API. In the files, we store attributes for Run, User, and Platform ID. Additionally, we store the user’s name for each run that they submit. We also scrape and separate the total playtime into hours, minutes, and seconds. The platform ID is converted to a platform name which is a plain text representation. Since runs can also be verified or rejected, we have an attribute to represent this. A direct link to the API page for a run is also stored. Lastly, the date for each run is saved for reference.

Methods Section

The data is pulled directly from the Speedrun.com API. The API stores its data in a JSON element, which acts in a hybrid form combining dictionaries and lists. All data is stored in the same format regardless of which game we look at, allowing us to programmatically filter our data from theirs. First, we generate a filename based on the game’s title. The scraper then builds a list of strings used to populate a row in the table, writing the previously described data in the order in which they appear. This has been specified and written in the first row designated as column titles. Lastly, the API only returns 20 runs per page so we must increase the offset specified in the HTML query (i.e. ?offset=20) to cycle through all available data.

Analysis Section

During the analysis, we wanted to try and uncover a few things: 1. We wanted to see if a specific player improved over many runs. 2. We wanted to see how much each player improved from their speedrun data alone. 3. We wanted to see if playing on different platforms for a game was beneficial.

To see if a specific player improved over many runs, we took the top five fastest runners from each game and plotted their speed run performance over time. For all players that we plotted, they either improved over time, or they only had a single point of data.

For the second point, we were able to graph every unique player that has submitted a run and determine the improvement from their first speedrun entry to their fastest time. Almost all players across all games and categories that we tested for were able to out-perform their first times.

When we plotted the usage over different platforms, typically there is only one platform that is commonly used among the runners. Once an optimal platform is found, it is used commonly by the community to preserve consistency within the runs. The more consistent a runner is, the more quality runs they can produce in hopes of a world record time.

Results

We noticed from the graphs we made that over time, the time that it took to complete the games for the particular categories got better (lower time). For example when we graphed the world progression and the progression for the top 5 speedrunners for a particular category in each game, the graphs would have a negative correlation. This indicates that as people played the game more and more as time passes, their times would improve for that particular category.

We also made graphs to analyze other details such as the improvement of each speedrunner’s runs between the first and last run and for displaying the number of runs for all the speedrunners combined by platform and per year. When looking at each individual person’s progression, there are a good amount of speedrunners whose runs have improved between their first and last runs, but it’s surprising that those who made the most improvements didn’t rank high in the leaderboard, at least for the games we looked at.

When analyzing runs by platform, we see that it varies greatly by game. For example, in Spelunky, a lot of players on the leaderboard played on the PC. Wii Sports had a split between the Wii and Wii U. A lot of the Mario Bros. games were played on the NES. This may be the result of the original release for that game that we see a greater distribution of games played on the original platform. When analyzing runs per year, we see that the graphs are skewed left to indicate that more runs were submitted now in 2020 for these games compared to when they were first released. This shows how much speedrunning has gotten popular over time, especially now.

Conclusions

We definitely see in our results that as time goes on and as people play the game more and more, their times for particular games improve. As explained in the results section with the progressions, we see that the times get shorter and shorter as people play more and time goes on. This is likely due to the discoveries of skips, glitches and other time saves that people find out and use in order to improve their times. It also shows that practice makes perfect as the more these speedrunners play, the more they learn about these discoveries and strategies that they can develop in order to improve their times. Speedrunning has definitely become more and more popular now, and it’s become more competitive as people try to show off how quickly they’re able to complete these categories by putting their names on the leaderboard and comparing themselves to others.

Link to all files: <https://github.com/AndrewMacatangay/UIC-CS-418-Data-Scrapers/tree/main/WebScrapers>