

Joe Lollo

LIS 572

Professor Melanie Walsh

31 October 2022

## Final Project Proposal

**Title:** [“Cheers, Love! The Analyst’s Here!”](#): Data Analysis of the *Overwatch 2* Competitive Scene, One Month In.

(the link just leads to the working title’s reference)

**Name:** Joe Lollo

**Introduction:** One of my favorite video games of all time is *Overwatch*, which I’ve recently dedicated a lot of my gaming time to, and I thought that it would be interesting, and timely, as *Overwatch 2* was just released in September 2022, to perform a data analysis of the game’s community through analyzing player performance and competitive matches, building narratives from that data, and making predictions for what will happen in future games.

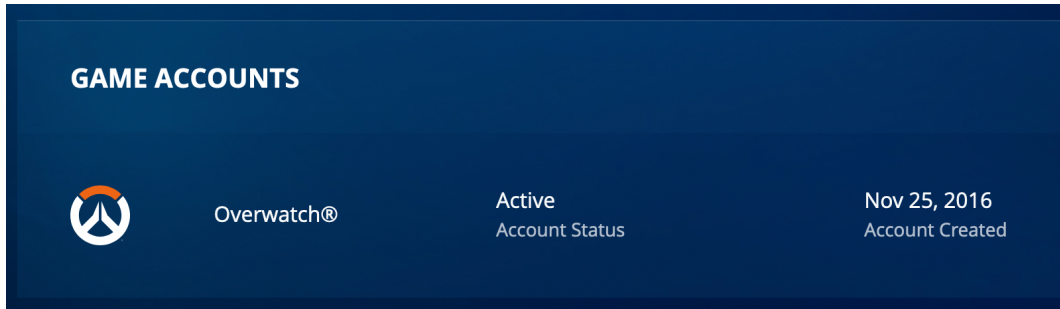
Although the dataset I picked (which I will go into more depth on later) tracks a very small amount of player data, as the game has only been out for a little over a month, I found many interesting potential correlations and connections within the *Overwatch 2* data that I wanted to explore, because they could help me gain a better understanding of why *Overwatch* players play the way that they do and pick the characters they choose:

- Is there a correlation between a character’s gender and their pick rate among players?
- Do the new characters introduced in *Overwatch 2* have a higher pick rate now because they are new?
- Do characters with a higher win rate also have a higher Player of the Game (POTG) rate? (POTG is a special award given to players who play particularly well in each match).

- Are there generally higher or lower pick rates for specific types of characters? Is there a correlation between type and pick rate?

**Related Work:** There has not been much data analysis on *Overwatch*, especially not in the subject I am looking to explore through this project, but this sort of computational analysis is very common within the interdisciplinary field known as game studies, which combines questions of how games generate meaning with how humans interact with them, blending modes of inquiry from all academic disciplines. [This project on “Fighting Game Data Analytics”](#) by Medium writer TyE explores many similar questions to the ones that guide my data analysis of the *Overwatch* community, as the author explores win rates and loss rates with specific characters in different arcade fighting games, namely from the *Street Fighter* series. A similar project from *Towards Data Science*, [“The Science to Super Smash Success”](#) by Aditi Mahabal, which analyzes all character matchups using information theory, machine learning, and data analysis using Python to predict win rates. As shown in the two examples I chose to share, the majority of work in this area revolves around fighting games, while *Overwatch* is a MOBA (multiplayer online battle arena)-shooting game hybrid which does not have a community that is super-dedicated to tracking, analyzing, and making meaning from data, meaning that my work on this project will be rather innovative. Something I think is very interesting to address is how I am aiming to create narratives from the numerical data, and this article by Melissa Somerdin titled [“The Game Debate.”](#) from the University of South Carolina’s academic journal *The Oswald Review*, talks about how games that don’t follow a linear narrative, such as *Overwatch*, can still have “stories” built out of the player experience and the emotions built out of it. Design is, functionally, a form of play, Somerdin’s piece argues, and the way that game designers create the players’ experience can give way to a wealth of individualized stories and meanings that depend on the player’s successes and failures throughout their experience in the game, and the affects and emotional responses tied to them.

On a more practical note, I have a strong connection to this data:



As this screenshot from my battle.net account shows, I have been an *Overwatch* player since around the first game's release, and have been a dedicated member of the fan community since around that time too: I [regularly contribute to the \*Overwatch\* wiki](#) and am involved in the game's subreddit, and have met many interesting people just by being involved in the fandom. There is actually a very good chance that, since my time playing *Overwatch 2* started, I am actually *a part of the data myself*. Having a tangible connection with the data I am analyzing will make my analysis much stronger, because I will know some more intricate connections that others will not and my passion for the game and status as a member of the game's community of players gives me more authority to discuss the subject than the distance that many pre-2020s scholars have kept when analyzing games. There has been a strong positive trend lately in academics who are also gamers analyzing video games using methods from their disciplines, and this is a sort of discourse that I see myself becoming a part of someday. Whether that's through public and open-access work or through a future graduate degree and academic employment, only time will tell.

**The Dataset:** I was able to download a publicly available dataset (in CSV format, although there is also a .db format) from [Overbuff](#), one of the top *Overwatch* fan sites and a hub for the competitive *Overwatch* community. The site has a main table (which I downloaded) featuring both qualitative categories, such as characters' genders, "type" (as there are four roles characters can fulfill in matches), and debut game, and quantitative data related to win rates, tie

rates, and pick rates, among others. There are also more detailed stats for each specific character (including their most used attacks and the stages they win the most in) on a page dedicated to them, but I knew that analyzing *those* datasets would take up way too much time. I thought that working with [the general dataset](#) accessible from the landing page would give me enough information to create readable, compelling visualizations. There are 35 playable characters, and the dataset has 35 rows (as expected) and 12 columns indicating both categorical and quantitative variables that fit with each character.

The columns are:

- The character's name (string, categorical)
- The character's gender (string, categorical)
- The character's role or "type" (string, categorical, out of 4)
- The game the character debuted in (string, only *Overwatch* or *Overwatch 2*)
- The character's relative difficulty, as indicated in-game (integer from 1-3)
- Four columns for pick rate, win rate, tie rate, and POTG (player of the game) rate (float/decimal)
  - Another interesting question I thought about just now is exploring average pick rates and win/tie rates for characters of different types.
- Three columns for average KPG (kills per game), DPG (damage per game), and HPG (healing per game), based on each character's abilities (integer)
  - This would also be interesting to explore, namely in correlation with character types, as I have noticed some generally perform more kills than others.

One interesting data point is the "difficulty" column, which indicates the character's relative difficulty (out of 3) to use and master compared to the others. I can see myself using this column for exploratory analysis, especially to test a correlation between difficulty and pick rate.

There is another interesting dimension with the project that leads to a question of power: in 2019, Overbuff and similar *Overwatch* fan sites were hit with a cease and desist letter from

Activision Blizzard, the game's developers, citing the "unfair" usage of their copyrighted content including the game's logo and various characters. Overbuff shut down for a while, but resurfaced on the Internet fairly recently when *Overwatch 2*, the game's sequel was released in September 2022. The new version is built on an API and hosted on the creator's GitHub page. This is interesting to note because it asks an important question about the positions of fans and their creations within Blizzard's power structure: are we allowed to freely publish and explore the game we love, or should we keep our work hidden or small-scale to accumulate the lowest risk possible in our enjoyment of the material? This is something that will guide me throughout my analysis, as I strongly believe that corporate meddling, especially in the gaming and film industries, ruins what fan material creators can do with intellectual property.

**Methods:** I would hope to explore the above questions, and any more I'd have, in a project that uses a mix of DPLYR and ggplot2 methods. I think I will definitely use the `group_by()` method, and summarize means after, to look at average pick rates for characters of different genders and types, because the disparity between player choice and character gender was a question I realized almost immediately when looking at the dataset. At this point, I am imagining making bar charts of characters with the highest and lowest pick rates and win rates, as well as the grouped bar chart of pick rate by gender and character type, and I also think that maybe a box plot or two related to win rate or POTG rate by type would be interesting to see if it is true that some characters are just "better" than others or not. I have a clear idea of how I will carry out the final product, as I already have a Medium account and have used it for many things in the past, and think it would be a good idea to share the findings and connections with interested parties on that platform.

**Implications:** I think that this project can add to the growing body of research being created by gamers who also have proficient research skills. Until between the late 2010s and the 2020s, game studies research was primarily done by scholars, mostly in anthropology, communication/media studies, or psychology, who did not play games that often, and only

wanted to apply their methods to the study of other cultural factors that were not really studied in the past. This changed when younger academics and graduate students began to notice the disconnect, as I did when I was in college, and now the field is almost entirely populated by gamers offering their critiques, in various ways, on the cultural phenomenon of gaming. I hope to use my academic background in both LIS and the humanities to continue influencing this trend of work created by people who meaningfully engage with games outside of academia, and I think this project is a good starting point as it will use a combination of skills from both spheres to communicate effectively, which could potentially inspire others, especially if I share it among both the *Overwatch* fan community and my academic peers.

**Challenges and Limitations:** As I said earlier, this dataset documents just one month of data, as *Overwatch 2* was released in September. The data only tracks win rates for the second game, and not the six years prior that the first game was released and widely played. As a result, there is a very limited scope of the playerbase, as many players who were active at the start of the first game's competitive scene are no longer playing and there is a higher rate of new players now that *Overwatch 2* became free to play. This limits my data analysis quite a bit because I cannot do any time-based analysis, which was something I was interested in tracking but knew I did not have the data for since Overbuff is the only site that really *has* publicly available *Overwatch* data. Aside from that, I do not see any limitations and see many possibilities within my dataset.

## Citations

Mahabal, Aditi. "The Science of Super Smash Success." *Towards Data Science*, 10 June 2020. Web. Accessed 31 October 2022.

Somerdin, Melissa. "The Game Debate: Video Games as Innovative Storytelling." *The Oswald Review* 18.1 (March 2016): 69-82. Web. Accessed 1 November 2022.

TyE. "Raging Downloads: Fighting Game Data Analytics." Independent blog, Medium, 15 April 2021. Web. Accessed 31 October 2022.