Capstone Project 1

For my capstone project, I would like to use the PlayerUnknown's Battlegrounds (PUBG) dataset found on Kaggle. PUBG is one of the top-selling games on Steam, and at its peak had around 3.2 million players. This has been one of the most popular games on Steam for a long time, and a personal favorite of mine having played close to 400 hours.

- **What is the problem you want to solve?**

Quoting from Kaggle's Playground Competition Website, the underlying problem is to use the available anonymized data to "*predict final placement from final in-game stats and initial player ratings.*" Once I model the final placement (expressed as a percentage) as a function of the players' performance stats, I can attempt to analyze the way in which these stats influence the final placement, hoping that such analysis will reveal strategies that lead to better and worse placements in the game. For instance, the competition's Website mentions the following questions: "*What's the best strategy to win in PUBG? Should you sit in one spot and hide your way into victory, or do you need to be the top shot?*"

- **Who is your client and why do they care about this problem?**

The client in this case is the PUBG Corporation, mainly the developers of the game. This will potentially help the developers of PUBG create a better game if it seems like players are spending too much time camping and not enough time looking for kills, which the community, including me, finds it to be a boring play style. The developers could find ways to incentivize

movement in the game and finding other players to kill, making the game a more fun experience overall.

- **What data are you using? How will you acquire the data?**

The data for this project comes from a competition on Kaggle. Can be downloaded off the site and is already split up into training and testing data.

- **Briefly outline how to solve the problem**

In PUBG, there are two different perspective styles: first person and third person, as well as three different game modes: solos, duos, squads. I plan on only analyzing solos data, since that does not have the extra variable of people in teams and having a friendly player with them as they fight. Solos data could possibly be extrapolated up to duos and squads, as the playstyle that is found most effective could translate to the other game modes. For perspectives, I plan on breaking those up and analyzing them separately, as the playstyle that is most effective in the third person perspective could be different than the playstyle most effective in the first person. Based on all this data, I plan on using regression algorithms to express final placement as a function of relevant variables in the dataset--and potentially others that I might engineer.

- **What are the deliverables?**

The way I want to deliver this is through a presentation. Presentations and public speaking are not my strong suit and doing so would give me more practice in that area and will allow me to be more comfortable speaking in front of people.