Best Strategies for PUBG

Alexander Pearson and Mariette Souppe

November 15, 2018

Introduction

How the game works: In PlayerUnknown's Battleground (PUBG) is a shooting game where up to one-hundred players are put into a battlefield with the goal of being the last one standing on a given map. In addition of trying to stay alive and not being shot by an opponent another challenge of the game is that over time the surface area of the initial map shrinks. A player must always be in the new map or else they automatically die. In the battlefield, there are empty buildings and ghost towns that contain weapons and equipment to help the player achieve the goals of the game to be come the last person or team standing.

Motivation: PUBG has become a growing popular game and since there has been a big craze about the game our motivation behind this analysis is to predict the best strategy to win the game.

Main Question: Can we use clustering to identify different strategies for playing PUBG, and then regression to determine which strategies are the most successful?

The Data

The dataset we choose is from the Kaggle competition, "PUBG Finish Placement" (**link here**). The data consists of the PGaPs (Post Game Player Statistics) from 47,965 games. In total we have 4,446,966 observations from the PGaPS. Some of the statistics consist distance traveled by foot, vehicle, or by swimming to the type of match was played such a solo, duo, or group match, and which perspective the game was played in (first person perspective or first and third person perspective) just to name a few.

The figure below shows sample of a basic exploratory data analysis showing the different group types versus the numbers post game player statistics for the two different perspectives the game can be played in for our dataset.
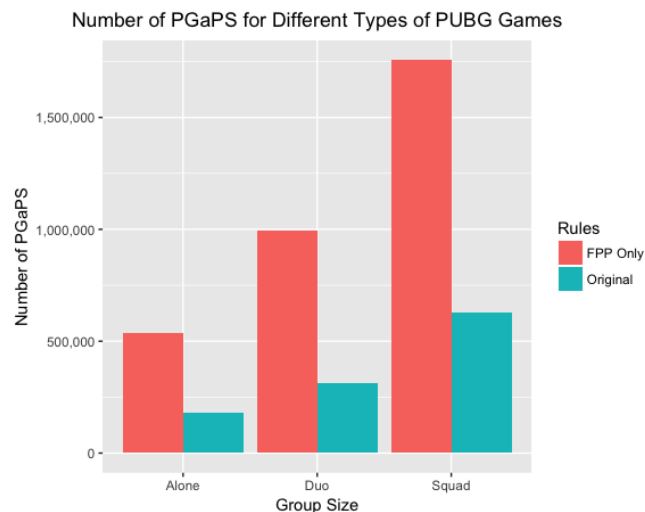
Tools and Models

Below are the tools and models that we had in mind for our dataset.

Tools:

- RStudio (ggplot2, reshape2, tidyr, scales)

Models:

- Clustering (k-means)
- Regression
    - Linear: Run Linear Regression models (with transformed response) on subsets of the data determined by type of game and cluster.
    - Logistic: Define a win place percentage greater than or equal to 0.9 as a success and less than or equal to 0.9 as a failure.

Potential Challenges

With this data set we are predicting that there will be a couple of potential challenges.

1. Inconsistency between games can make finding clusters that reflect difference in strategy difficult.

    (a) Every game of PUBG has a different number of teams competing we have this information.

    (b) Different PUBG games may be played on different maps (we don't have this information but given the differences in duration, I suspect they are played on different maps).

2. Creating a good regression model that does not overfit the data or give us numbers that lie outside of the range of possible outcomes (i.e. the unit interval).

The challenge that is finding good clusters that do not just cluster by number of teams in the because we want our cluster to be based on movement. We are predicting that there might be a dependency with the amount of movement and the type of match that was played so we want to condition on number of players. Another potential challenge we are predicting is when reducing the teams. This means that we take the team data team and flatten it down to single players for the regression model.

Prior Analysis

As mentioned previously, our dataset was pulled from the Kaggle competition "PUBG Finish Placement". The prior analysis for this dataset are based on machine learning such as neural networks and basic explanatory data analysis.

Presentation Feedback

1. Look at the clustering methods that are being used for this competition to get an idea of the types of methods that would work well for this dataset size.

2. Look into Bayesian regression, since the dataset size is pretty massive (Ask Raquel for more information).

3. Be careful about the question that is being asked because we can easily go into too many paths and end up not answering the inital question.

4. In final report, remember to add in types of methods were used and explain why one method worked and why the method other did not work.