In-Depth Statistical Analysis and Survivability of PlayerUnknown's Battlegrounds

Extended Abstract
Group: 02

Phi Trang
B.S. Computer Science, University of Colorado Boulder
phi.trang@colorado.edu

Jacob Tran
B.S. Computer Science, University of Colorado Boulder
jacob.tran@colorado.edu

Neil Nguyen
B.S. Computer Science, University of Colorado Boulder
neil.nguyen@colorado.edu

ABSTRACT

PlayerUnknown's Battlegrounds is a battle royale game-mode that will put up against 99 other players on a remote island for a winner-takes-all showdown. Players enter the island and try to locate weapons and supplies in a tactical battlefield that forces players into a shrinking playzone.

CS CONCEPTS

• **Data mining, aggregation, plotting, correlation, data analytics**

Figure 1: The official banner for PlayerUnknown's Battlegrounds

1 Problem Statement

A complex problem that players run into before every game is where to eject from the plane and which site to land. Our focus in this project will be to optimize player choices within the early, middle, and late game. Early game decisions such as areas to drop will greatly increase a player's ability to survive towards mid game. By analyzing the areas where players die the most, we can see areas to avoid, allowing players a higher probability of making it past the 40 players left mark. We will look at distance traveled to see if travelling more often will give a player a higher chance of surviving. This can affect player behavior during the middle and end of the game. If travelling decreases a players chance of surviving, we can draw the conclusion that "camping" can be an efficient method to win the game. We can also do analysis of weapons to help players choose an arsenal that best fits their needs. We will aggregate weapon and death data to determine which weapons are most efficient at long range. We can also see which weapons have the most kills at the end of the game, so that players can determine what weapon they should have to win the battle royale.

2 Literature Survey

● "PUBG Analytics: Final Circle"
https://www.kaggle.com/skihikingkevin/final-circle-heatmap
This data mining project plotted a heat map of the frequency of where the final circle appears.

- "PUBG Data Analysis"
https://www.kaggle.com/chegerland/pubg-data-analysis

This project explored where most people jump out in PUBG. Findings indicate that most players jump near military base, school, prison, and the northern cities.

- "PUBG 50 Million Death Map"
https://public.tableau.com/profile/stewart.uyeda#!/vizhome/pubg50mildeathmap/pubg50mildeathmap
This project explored where the deaths occurred during the game. They plotted a interactive Tableau heat map that steps through game time and illustrate the locations of deaths.

3 Proposed Work

First we will search for data containing null entries where there should be an entry and remove such instances. Then, we will split entries between the two different maps, because the two different maps have two different sets of weapons, and the location data is map specific. We can look at a lot different events, and so we can separate the data into the appropriate forms.

Our work will involve data manipulation in order to separate the large data set into relevant subsets. We will calculate the effectiveness of weapons using statistical analysis methods discussed in class. We will use players' location data to find the "hot" areas of the map and to find the distance between the players.

For drawing a conclusion on weapon effectiveness, we will look at entries from solo matches, in order to remove uncertainty in the data. This is because the game allows players to get back up after getting "knocked out" in games with teammates, and the cause of death will not reflect what knocked them down in the first place. Thus, we will be uncertain of a certain

weapon's actual effectiveness in the game. Therefore, we will only look for games in which the players do not have duplicate placements when data mining for weapon effectiveness.

We will use values such as killer_postion_x, killer_position_y, and victim_position_x, victim_position_y in order to create heat maps of "hot" areas on each of the maps. We can also use these values as a method of gauging the effective range for each weapon type. We will publish heatmaps and player locations overlaid over the map of the play area.

Our work will use Jupyter Notebook in conjunction with the Pandas and Numpy modules to reveal such things about the matches in Player Unknown's Battlegrounds.

4 Dataset

The original dataset:
https://www.kaggle.com/skihikingkevin/pubg-match-deaths/data

This csv / xlsx document contains the first 1 million entries from the Kaggle PUBG death dataset:
https://github.com/jatr6994/csci4502/blob/master/PUBG%20Death%201million.xlsx

The dataset contains information regarding:
match_id: nominal string
map: ordinal string
time: interval integer
killed_by: nominal string
killer_name: nominal string
killer_placement: ordinal int
killer_position_x: interval Int From 0 to 800000
killer_position_y: interval Int
victim_name: nominal string
victim_placement: ordinal int
victim_position_x: interval int
victim_position_y: interval int

5 Evaluation Methods

- Finding the best weapons for use early game
- Average range of kills with each gun
- Heat maps of kills vs deaths?
- Average heatmap of casualties for each map
- Common weapons for players in top 10

6 Tools

We will be using:
- Anaconda/Python
  - High-level programming language focused on statistical analysis.
- Jupyter Notebook
  - Open-source web application that allows us to create and share live code, equations and visualization of data.
- Tableau
  - Will be used to also visualize our data. Easy to use drag and drop to create visualizations.
- Matlab Plot
  - Allow us to give visuals of our data.
- Pandas
  - Used to read in our dataset and store it into a dataframe. Allows for easy data cleaning and processing.
- Numpy
  - Fundamental python library used for scientific computing. It contains tools that are useful for algebra and arrays.
- Scipy
  - Python-based software for mathematics, science and engineering.

- Other Python libraries as needed

7 Milestones

- Clean data, remove null and missing values. Sort by map and do initial testing for average values.
  - March 14, 2018
- Analyze deaths according to position dropped. Analyze survival rates depending on position dropped.
  - March 21, 2018
- Get data on weapons. Statistical analysis on most efficient weapons, distance of bullets travelled, most popular guns.
  - March 28, 2018
- TBA

8 Summary of Peer Review Session

Differentiate ourselves from the other two PUBG groups as well as communicate with them to help each other mine data. Set a clear problem statement and how to go about finding correlations within the data.

ACKNOWLEDGMENTS

REFERENCES

TBA