Analyzing Human Decision Making in the Virtual World

Scott Zhao miz107@ucsd.edu
Yuchun Hsieh yuh378@ucsd.edu
Qihan zhang qiz237@ucsd.edu
Gautam Nain gnain@ucsd.edu

1. Introduction

Video games are a rich area for data extraction due to its digital nature. Notable examples such as the complex EVE Online economy, World of Warcraft corrupted blood incident and even Grand Theft Auto self-driving cars tells us that fiction is closer to reality than we really think. Data scientists can gain insight on the logic and decision-making that the players face when put in hypothetical and virtual scenarios.

PUBG is another such popular survival competitive video game where hundreds of players every match dive into a sandbox battlefield to fight with each other after looting randomly generated weapons and try to survive till the end. There are tens of thousands of matches played on the official server every day and a lot of data is recorded for each match. Examining these match data would yield interesting discoveries in human behavior and psychology under different circumstances while playing the game. We focus on the players personality and their behavior. For example, What are player preferences for a place to start the match, how people tend to hide more as fewer people are present in the game, difference in weapon choice for different fighting distance, average distance of fight for players with different skill level, etc.

2. Related Work

There are a few projects online that analyze similar datasets to provide advice to improve gaming experience, and analysis of certain action / behaviour's relationship to winning rate. Examples includes heatmap for location for last people surviving, relationship between weapon of choice and numbers of kills, between number of kills and surviving time, etc. In this project, more focus will be put onto analysis on human's psychological behaviour under different circumstances.

3. Dataset

- Match Deaths and Statistics: This Kaggle Dataset provides over 720,000 competitive matches from the popular game PlayerUnknown's Battlegrounds. The data was extracted from pubg.op.gg, a game tracker website. This dataset includes killers positions, placement as well as victims. By analyzing the interaction between killers and victims, we can infer their game playing process about how the players decision making affects the end results.
- First Person Squad Matches: This dataset includes all player death record in 31k matches, including every killer and victim s information (rank, position, weapon used). This dataset was inspired by the dataset mentioned above.

4. Project Breakdown

This Project can be broken down into three main components as mentioned below:

4.1. Pre-Processing Data

Before doing data analysis, our first goal is to import the data to python and pre-process it, so that it can be used to visualized for data analysis.

4.2. Data Visualization Techniques

The following strategies (including, but not limited to) will be used to help process and visualize the dataset:

1. From the map (provided in the dataset), set up axis and classify different positions on the map. (e.g. Indoor, town, forests, bridge, etc), and match the location in the dataset to these environmental types
2. Use bar chart to compare different choices and popularity
3. Use heatmap on top of the field map to visualize location choice and popularity
4. Use pie chart to visualize composition (e.g. by which weapon / what distance range are players killed)

4.3. Final Analysis

After getting all the data in the desired visual format, we need to utilize the visual information to make insights into the human decision making process and understand human behavior in different scenarios while interacting with the virtual world.

5. Project Timeline

- 7th November: Survey related work to analyze how deep can we dive into the data-set to get more information about PUBG Players.
- 14th November: Import Data to Python and Pre-Process it to desired format.
- 21st November: Get Insights from the data using various visualization techniques mentioned above.
- 28th November: Display the Insights in an interactive manner.
- 5th December: Prepare Final Report.

6. Division of Labor

Each one of us will contribute equally towards the different topics mentioned above. But to finish this project efficiently, we will create sub topics for each of the main topics and divide the work equally. we will get to know more about the sub topics as the project progresses.

7. Tools Used

To implement this Project, we will be coding in Python and using the existing `scikit`, `numpy`, `pandas` and `matplotlib` libraries.