PUBG Battle Stats

Joshua Ramos
Tyler Moore

Description

- Our dataset takes player information in million of matches and records death times, placement, distance walked, and team information.
- Our data is separated into two different csv files. One with specific death information and the other with specific information about a player i.e. team name, player name, team placement.

Questions

- Does the amount of distance a player runs during the match affect their placement percentile?
- Does the more kills a player gets during a match increase their chance to win?
- What are the best places to land to increase a players chance to win?
- What team size gives you the best winning percentages?

Prior Work

- The owner (skihikingkevin) of the dataset created a heat map of deaths in the final circle
- Another contributor (chegerland) did a heat map of where most of the early deaths take place
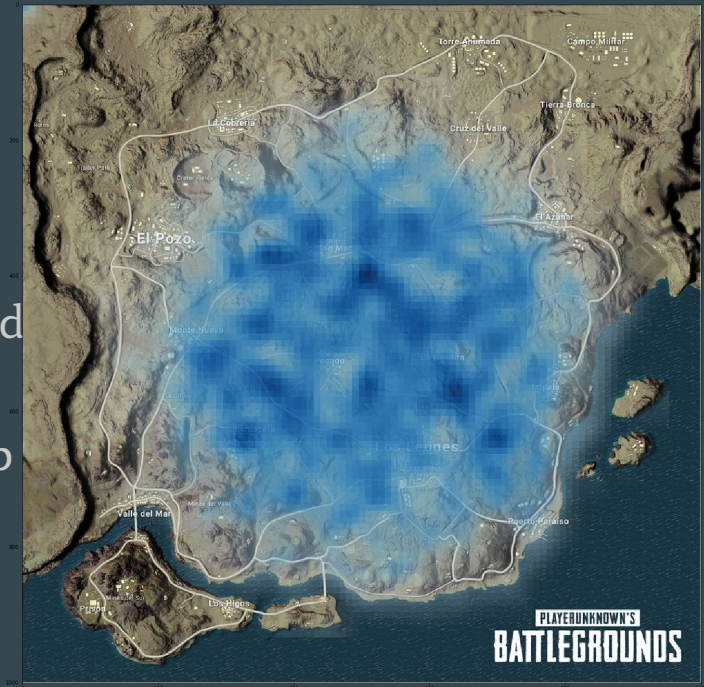
Datasets

- We have two datasets: one about player information and one about death information
- URL: https://www.kaggle.com/skihikingkevin/pubg-match-deaths/
- Kaggle is where we found the datasets
- Josh has downloaded it on his computer

Proposed Work

- Data Cleaning:
  - There are players who started the match but left early and so all of their data is blank
- Data Preprocessing:
  - We need a way to normalize our data so that it can be compared
    - Ex: If we are comparing kills from two different teams, but one team has 4 people and the other has 3 people, this would not be a fair comparison since one team has one more player
- Data Integration:
  - One of our datasets has information about individual players. We need to be able to create a new dataset that groups each player into the team they were on during the match

Tools Used

- Python
  - Numpy
  - matPlotLib
  - Pandas
- Excel

Evaluation

- We will evaluate our results by testing them in game. Then we'll compare our game stats before and after we have gained knowledge from our dataset.
- Another way to test is by conditional probability. Before we begin mining, we can use this dataset to see how many times a person wins. And then after we found a pattern, we can go back to the dataset and see if a player has a higher chance to win, given they meet the requirements
  - Our pattern will be correct if we get a higher win percentage.