

Milestone Report Capstone 1 PUBG

1. Problem statement: Why it's a useful question to answer and for whom (get this from your proposal)
2. Description of the dataset, how you obtained, cleaned, and wrangled it (get this from your data wrangling report)
3. Initial findings from exploratory analysis (get this from your data story and inferential statistics reports)
 1. Summary of findings
 2. Visuals and statistics to support findings

Problem Statement

Players Unknown Battleground became the top game of 2017 worldwide. The battle royale genre of video games can be very competitive, and only one player / team can be the winner. There are many different approaches to trying to win a match, the bottom line comes to caution vs aggressiveness. So what correlations from other variables can we find, that first place consistently has. These answers can help create strategy guides for PUBG, and potentially other battle royale games, which have dominated the gaming industry in the past two years.

Dataset

I obtained the dataset from Kaggle. The dataset didn't have any missing entries. Fortunately due to the nature of getting the data from a video game, having a missing value is not likely to ever happen. The dataset came in 5 csv files each around 2 gb. The columns included player distance walked, player damage, team placement, distance driven, survival time, and party size.

To clean the data I narrowed it down to the useful columns. I removed undesirable rows such as players that did not survive for a minute, or didn't travel a distance of more than a few meters. I

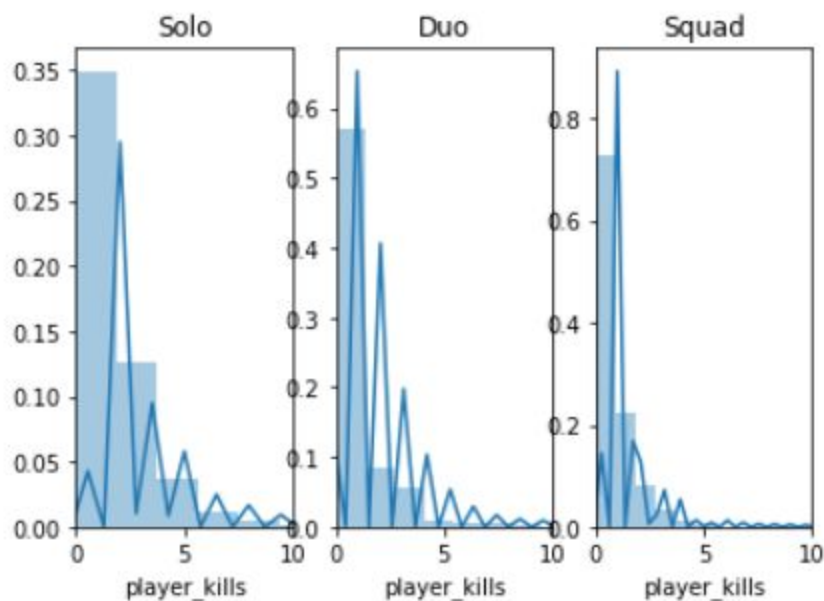
divided the player survival time column by 60 to convert it into minutes, making any graphs I conduct into a more discrete integer.

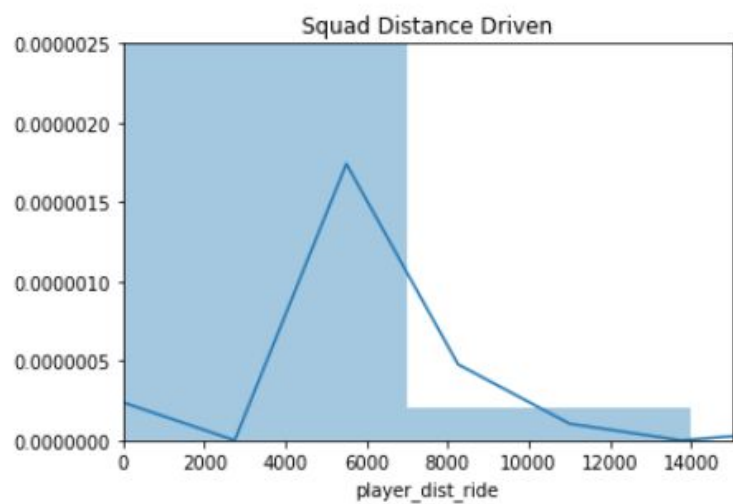
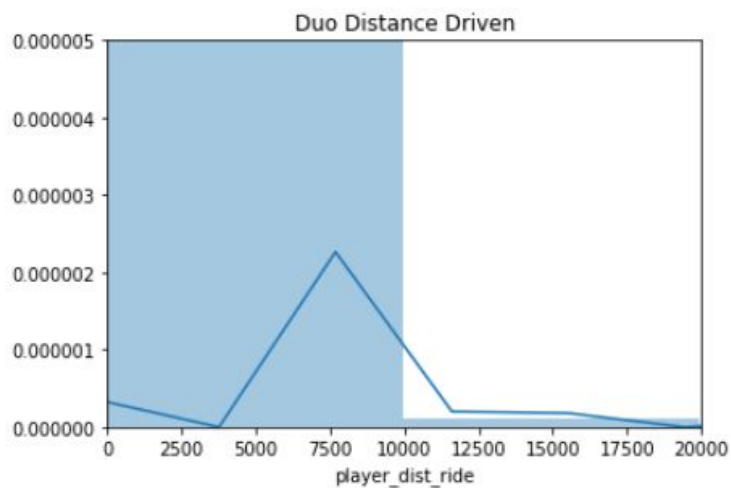
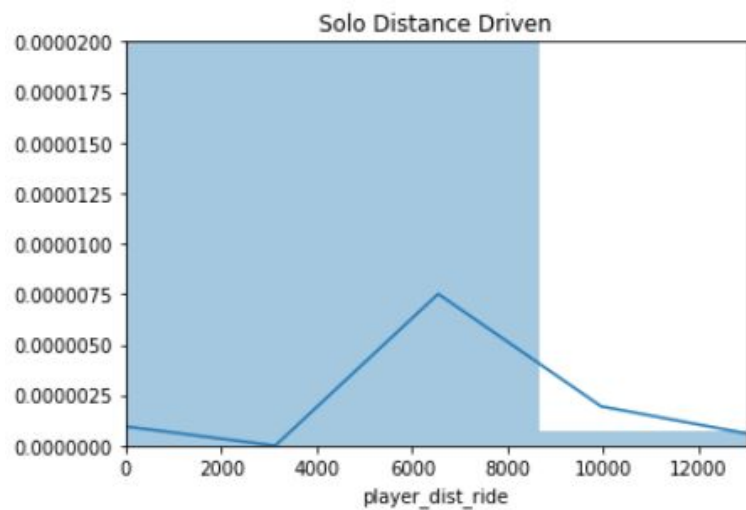
In my latest exploratory data analysis, I came to the conclusion that the remainder of my analysis should narrow down the data to only the top half of teams in a given match. The reason being from experience playing the game, I know that the first 50 players are eliminated within only a few minutes of a match, and the reason is simply due to instant fighting. Therefore data from these players won't be very insightful for the questions I am looking to seek, namely game strategy, rather than players being better at aiming / reaction time.

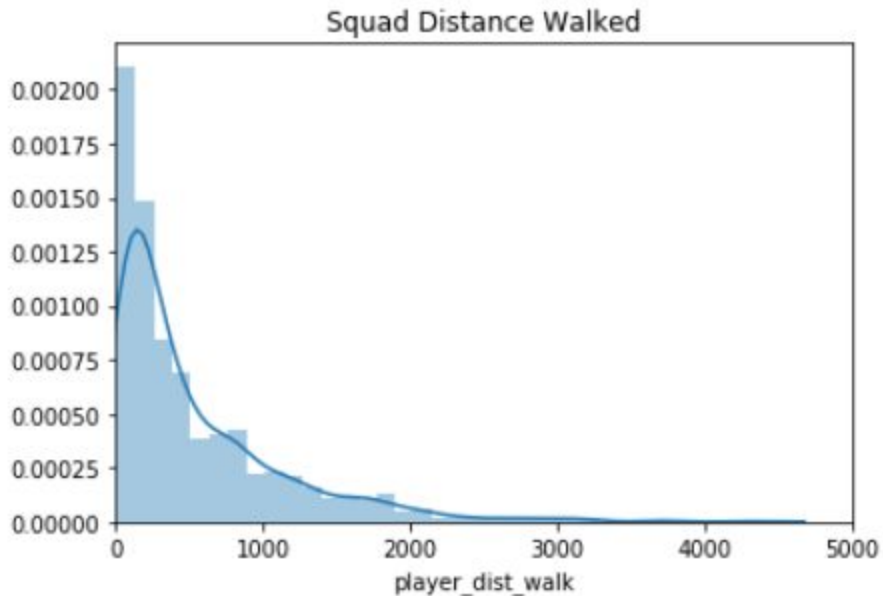
Most importantly I divided the data by game modes. The game can be played in solo's, duo's, or in a squad (up to 4 players on a team). Not mixing this data is essential. For the solo matches of top half of teams, there are 7,201,035 observations, and close to 2 million for duo and squads.

Inferential Statistics

First I wanted to examine if the independent variables of interest were normally distributed.







They all appear to be normally distributed to me, but they skew to the right heavily for distance walked and player kills variables.

I utilized the correlation method to view the correlations of each of the variables. I found which independent variables had a high correlation to the dependent variables. I did not examine independent variables that had high correlations with other independent variables that I already selected. I came up with the following observations.

Solo

1. Player Kills & Team Placement : $-.39$
2. Player Distance Walked & Team Placement : $-.09$
3. Distance driven & team placement: $-.3$

Duo

1. Player Kills & Team Placement : $-.39$
2. Player Distance Walked & Team Placement : $-.22$
3. Distance driven & Team placement: $-.52$

Squad

1. Player Kills & Team Placement : $-.144$
2. Player Distance Walked & Team Placement : $-.11$
3. Distance driven & Team placement: $-.31$

I also took from the complete dataset for solo matches all the players that won first place, took bootstrap samples 1000 times, and got the confidence intervals for the means. I did the same for

second place. First place consistently has a mean kill of about 7, with a confidence interval of [6.94742282, 6.99468351], while second place had a consistent mean of about 3.6, and a confidence interval of [3.58724136, 3.61285905].