Analysis for PUBG Match statistics

Suhas Cholleti[1], Santosh Kumar Nunna[1]

[1] Rochester Institute of Technology, Rochester, NY

Motivation

Online Gaming and esports tournaments are on the rise and prize money for these tournaments in reaching millions. Any competitive edge in tournaments like these could determine the result of the tournament. With these in mind, it has never been more important to understand and getting better at the game of PUBG.

Project Goals

To help Pro players tactical decision by:

- **Performing predictive analysis using classification to predict the top 10 players in each match using the player match statistics**
- **Produce heatmaps using the kill statistics to determine areas with high levels of activity**
- **Analyze weapon usage by players to see if there is any preference regarding weapons**

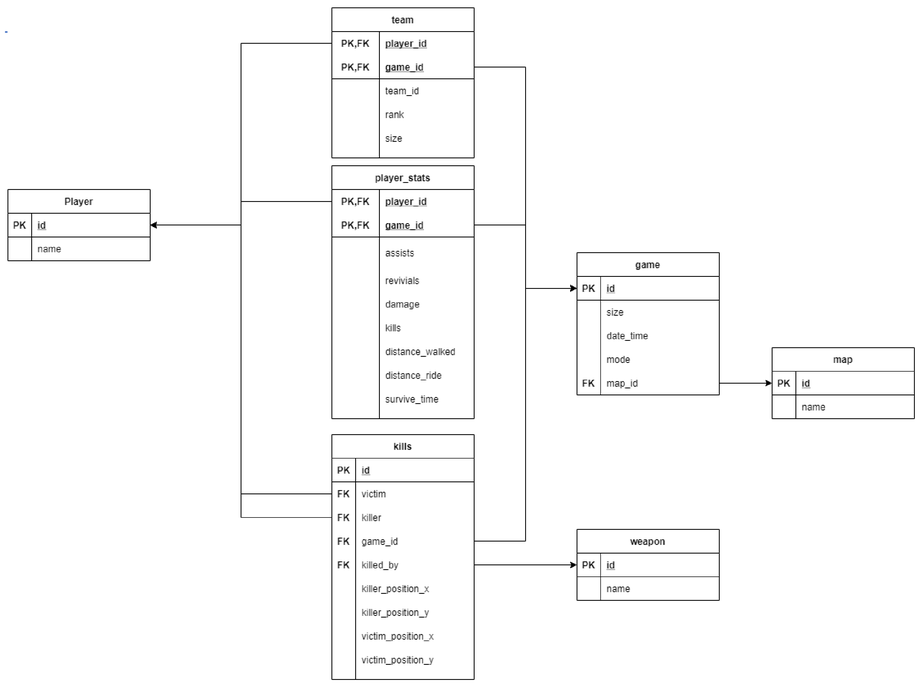
Solution approach

We used the PUBG Match Deaths and Statistics dataset from Kaggle. The ER model given below was used and data was added to MySQL database using JDBC and java.

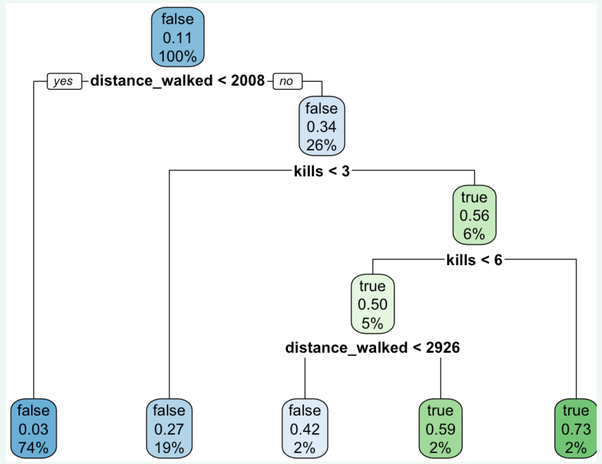
Data is cleaned while insertion by skipping over the tuples with erroneous data. A few tuples do not have player name. Such tuples were removed as part of the Data cleaning process.

R was used for the Data analytics part of the project. 'RMySQL' module is used to access data from the SQL database. 'rpart' module was used for decision tree classification. The heatmap was created using the 'plot_ly' module.

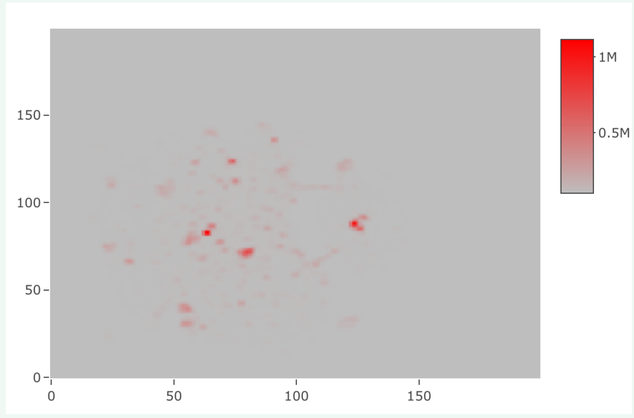
ER Model



Decision Tree Classification: Prediction weather a player finishes in the top 10 is predicted using this classification. Distance walked, distance ridden, kills and damage were given as input for the decision tree, but only distance walked, and kills were used.



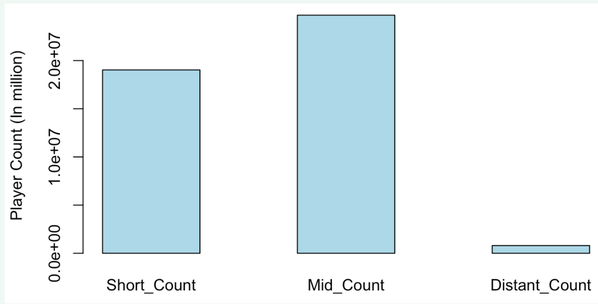
Heatmap: This heatmap was created using killer's x and y co-ordinates on the Erangel map. This map clearly shows a few points on the map where encounters are more common. Such areas can be avoided by players to survive for longer.



Kills at Different Weapons: Kills from each weapon are grouped together. Top 10 weapons are shown below. Assault Refiles are the most used weapon class(M416, M16A4, SCAR-L, AKM) and sniper weapons are one of the least used weapon class.

Weapon_Name	Weapon_Kills
Down and Out	4202259
M416	3036067
SCAR-L	2471793
M16A4	2452245
AKM	2261315
UMP9	1459782
S1897	1128271
Mini 14	814039
Punch	700295
Kar98k	687194

Kills at Different Ranges : The bar plot depicts the kills at different ranges. Medium range is the most common combat range and long range is least common combat range. This is not surprising after the previous result as sniper rifles are most effective at the long range and it is evident that players do not prefer to combat at that range. Similarly Assault riffles are the most used weapon class and it is depicted here as AR's are most effective at medium range.



Legal and Ethical Considerations

PUBG has made an API publicly available to be able to get the stats of each match and player. So accessing the data and analyzing the data for the project is completely legal.

The problem with this project is that it doesn't contribute anything to the society. This paper gives ways to get better at a game, which in all honesty the player is better off not playing, for his/her own well being.

As the data used for this project is relatively old and uses data from low tier players, the analysis might not be suitable for pro players. This could lead to a player using this data losing in the tournament.

Conclusions

In this project we proposed an ER model to store and retrieve this data. We provided an analysis of different weapons and areas of high activity on each map.

We also analyzed the type of combat that players prefer, up close or long distance. We found that the Assault riffle is the most used weapon class. This could mean that the weapons in the class are overpowered and needs some balancing, or it could also mean they are more readily available and easy to use. The heatmap clearly showed areas on the map which have high activity. This information can be used by pro-players to help them make better decisions on the weapons they choose, areas they want to land etc. Players with an aggressive play style would like to land in areas with high activity and vice versa. This information can also be used to balance the weapon damage.

References

[1] KP. 2018. PUBG Match Deaths and Statistics.https://www.kaggle.com/skihikingkevin/pubg-match-deaths

[2]PUBG CORPORATION. 2019. Playerunknown's Battlegrounds Developer API.https://developer.pubg.com/?locale=en [Online; accessed 12-June-2020].

[3]Y. H. Gu Weiming Wu, Xiaolong Pan. 2011. Study and Optimization Based onMySQL Storage Engine.At SIGMOD(January 2011), 185 – 189.

[4]Zhuqing Xu, Ming Xiang, and Lijun Pang. 2018. Study on the influence factors ofcollege students' loyalty in PUBG game.IOP Conference Series: Materials Scienceand Engineering439 (nov 2018), 032059. https://doi.org/10.1088/1757-899x/439/3/0320595