Analysis for PUBG Match statistics

Anonymous Author(s)

ABSTRACT

The development of games has been rapid in recent years. One such game is Player Unknown's Battleground(PUBG). With gaming in its all-time high and the recent increase in the number of tournaments, player analysis has never been more important. This analysis provides an insight into how users perform in this game and what kind of weapons or features attracts them to gross millions of matches.Predictive analysis using a decision tree was performed to predict if the player comes in the top 10 in the match. The accuracy of the decision tree is more than 90%. More analysis of the weapons showed that Assault Rifles are being used by most players and sniper weapons are used the least. A visual representation of the kills using hotspots showed the most active places on the entire map.

KEYWORDS

PUBG, Decision Tree, Heat-maps

1 MOTIVATION

The conception of simple games and simulations as a part of academic research can be traced back to the birth of digital gaming we know today. Multiplayer Online Battle Arena has become one of the most popular genres of video games played by people worldwide. Given the nature of these games, it is possible to extract huge amounts of data. There are a variety of video games that when listed will form a huge database. But, for this work, we pick one of the recent masterpieces of mobile/PC gaming: Player Unknown's Battleground also known as PUBG [3]. It is one of the best selling games in 2018 and 2019. The developers have provided an API [1] which can be used to get the data. We were motivated by this dataset as it is one of the first Battle Royale games to come out [14] and we had a decent understanding of the data due to our interest in the game.

In the last few years, esports and online competitive gaming are on the rise[15]. With the prize money of these tournaments being as high as a few million [7], the importance of understanding and getting better at these games has never been higher. In this project, we would like to gain a few insights into player preferences and general gameplay. We would like to perform analysis like most common weapons used for getting kills, whether more kills come in close-quarter combat or over a long distance. We could like to create a heat-map where we can map out high intensive areas of the map.

This paper is organized as follows: section 2 discusses the goals for the project. Section 3 presents our overall approach to the project, discusses the ER model in detail, and gives an overview of the implementation. Section 4 talks about the analysis done on the dataset. Section 5 talks about other work in the field of PUBG. Section 6 contains the details on the work completed till now and the future work to be done and section 8 discusses the legal and ethical considerations when analyzing your data.

2 PROJECT GOALS

Our goal for this project is to help pro-players to make some tactical decisions, which can help them win the game. Our analysis primarily focuses on picking out patterns in the existing data which were successful in winning the game. From the data we have at hand, we want to plot out heat maps of the kills, This would most likely point out areas with high activity of players, and players can make an informed decision on where to land. We would run a classification algorithm, to analyze the play styles of the players with the end up in the top 10 of each match. We could also look into the usage analysis of the weapons, to see if the weapon is overpowered and needs balancing so as to make the game more enticing to play.

3 OVERALL APPROACH

We used the PUBG Match Deaths and Statistics dataset [8]. The dataset contains statistics from over 72000 matches, which would be over 70M kills. The data was extracted from a game tracker website [10] which uses the PUBG API.

The dataset contains two different types of files, Aggregate match stats, and Kills stats. Aggregate match stats contains aggregate statistics of a player in the match and metadata of the match itself. The kills stats file records every death in the 720k matches and stores data regarding each death.

The project is primarily divided into two components:

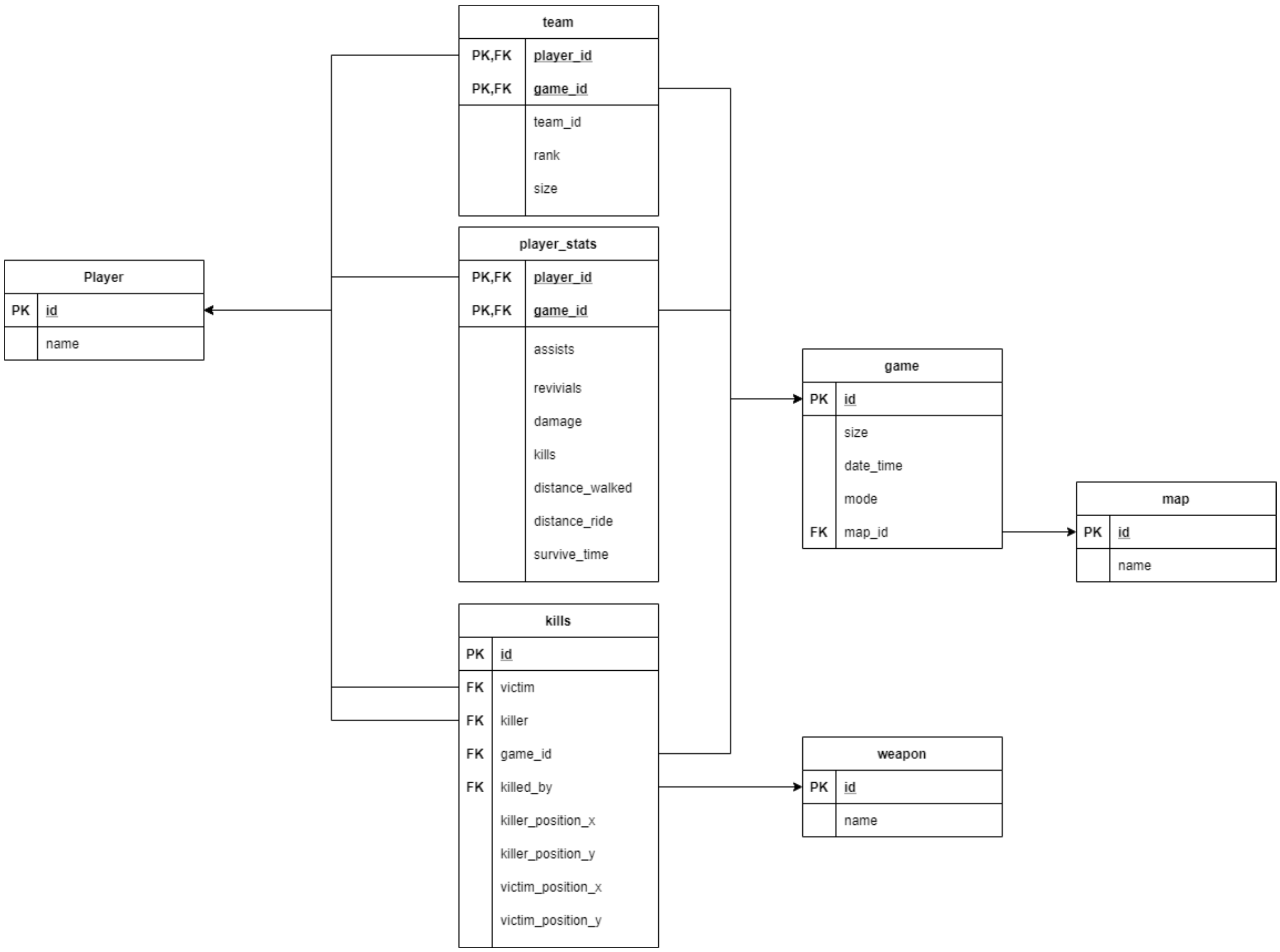
(1) Data management
(2) Data Analysis

The Data Management component was implemented using mySQL and Java using JDBC [16]. We had initially created an ER model, shown in figure 1. The dataset was divided into 7 tables.

(1) player: contains the details on the player. Currently contains just the username
(2) map: contains the details on the map. Currently contains just the name
(3) weapon: contains the details on the weapon. Currently contains just the name
(4) kills: contains the details of a kill, such a weapon used, victim, killer and their coordinates.
(5) game: contains the metadata of a game
(6) player_stats: contains the aggregate stats of each player in a match. Stats include assists, kills, travel distance and survive time
(7) team: contains the information of a team, such as size of team and team rank.

We had created the java class "CreateTables" to create the tables according to the ER model we have designed. The java class "LoadData" contains two functions, one each for loading the data from the Aggregate Match Stats and Kill Stats. Both functions work similarly. They go through the CSV file, split the data, and insert the data into the correct table. Two hashmaps were created to store the player name/id and game name/id while loading the Aggregate

Figure 1: ER model for PUBG statistics.



date	game size	match id	match mode	party size	player assists	player dbno	player dist ride	player dist walk	player dmg	player kills	player name	player survive time	team id	team placement
2017-10-31T02:41:53+0000	95	212	tpp	1	0	0	0	37.919838	20	0	m3xdave	106.351	100000	88
2017-10-31T02:41:53+0000	95	212	tpp	1	0	0	292.205048	3457.03613	387	3	addyHere	1690.021	100006	8

Table 1: Sample elements of Aggregate Data

Match Stats. This was done, to avoid querying the database for the player id when inserting data from the Kill Stats.

A few rows of the Aggregate match stats do not have the player name information. As the player id is a primary key for many tables in our ER model and we can not get that information for these rows, we decided to remove these entries as part of the data cleaning process. Another issue we had was the empty killer name field. There are a couple of ways where a player can die without being killed by another player, like dying after receiving fall damage or dying due to receiving damage from the blue-zone. As in these situations, there is no killer, the field was left empty in the dataset. As there is no killer information for these rows, we have skipped these tuples for our analysis.

The second component, Data Analysis, was done primarily using R. We used the 'RMySQL' module to fetch the data from the MySQL database. For visualizing the hot zones of the map, we used the 'plotly' library. The classification analysis was done using the 'rpart'

killed by	killer name	killer placement	killer position x	killer position y	map	match id	time	victim name	victim placement	victim position x	victim position y
Kar98k	Chewwe	35	375243.2	422101.2	ERANGEL	212	638	KingCobraGut	52	381097.3	448397.5
AKM	jiuxijiuxi	56	627858.1	379821.1	ERANGEL	212	207	Thecrossbows	72	628253.9	378554.9

Table 2: Sample elements of Kills Data

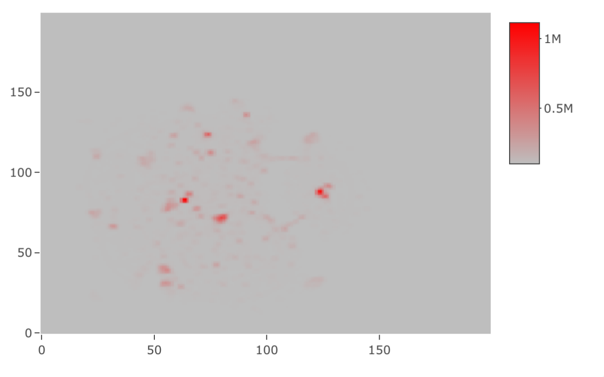
module. For more details regarding the data analysis, refer section 4.

4 ANALYSIS

We have performed the following analysis:

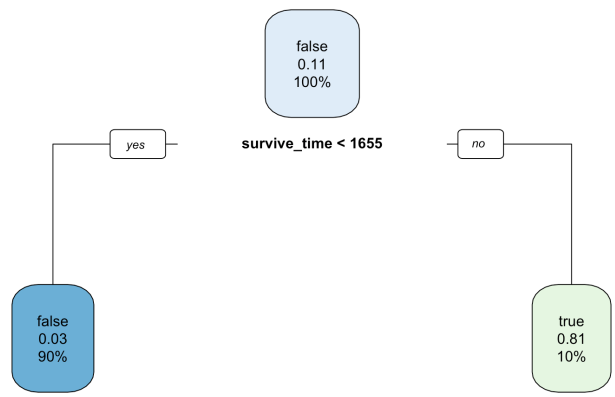
(1) Heat map of kills: We used the 'plotly' library from R to visualize the killer positions on the map. We extract the killer positions from the kills table. As the map coordinates range from 0 to 1000000, to get a more meaningful representation of the data we made the grid size of 200 and 200. We scale the x and y co-ordinates to fit the grid size. From the figure we can see that there are couple of spots where most kills are happening. Players could use this information to avoid those places. This is not as useful as it seems as each match lasts more than 30 minutes and the no information regarding time is being depicted in the analysis.

Figure 2: HeatMap for Erangel



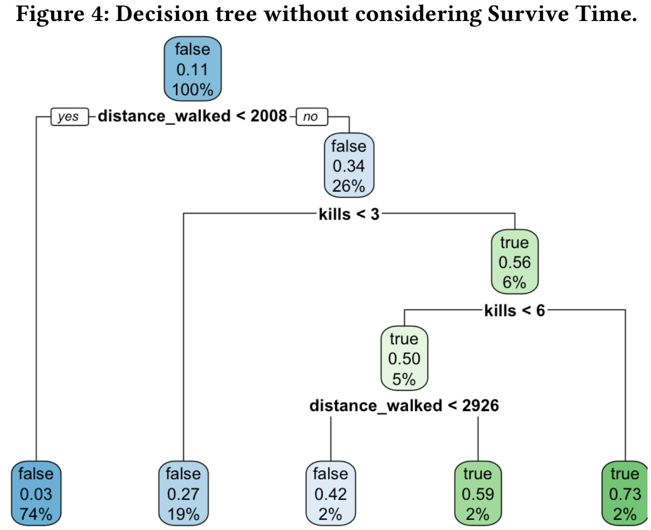
(2) Classification by Decision Tree: We used the 'rpart' library for the decision tree. Our aim was to use the player stats of a match and check which ones were affecting a players ability to win. For the sake of consistency, instead of just considering first position, we consider it a positive result if the player ends up in the top 10. Initially we had used damage, kills, distance_ride, distance_walked and survive_time for the decision tree, which proved to be a mistake. From Figure 2, we can see that the decision tree just considered the survival time of the player to make a prediction, which on second though seems obvious that a player with longer survive time would be ranked higher.

Figure 3: Decision tree considering Survive Time



We then tried the decision tree without the survive time. From Figure 3 we can see that distance_walked and the number of kills of the player were given importance. We can see that the decision tree predicts that if a player walks more than 2000m and kills 6 other players, he is almost guaranteed a top 10 position. In its current state, the decision tree doesn't provide a whole lot of value, but with more datapoints and more types of data this can be more useful.

Figure 4: Decision tree without considering Survive Time.



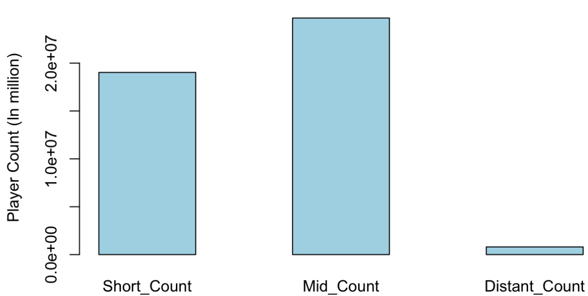
(3) Kills per weapon: We grouped the kills by each weapon and get the count of kills per weapon. This is done to the most used weapons, to understand the player preferences and relative effectiveness of the weapons. From figure 4 we can see that Assault Riffles like M416, M14A4, SCAR-L, AKM are used the most. The sniper Rifle weapon class is one the least used, as it has only one entry. This could be because the skill level required to use a sniper rifle is much higher.

Figure 5: Kills per weapon

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

(4) Types of combat: We grouped the kills at different ranges, and depict them in a bar plot. We find that the most kills happen at the medium range followed by short range and then long distance. This corresponds to the the above analysis where we saw AR weapons were most prominently used(AR weapons are most effective at medium range). We also see that long range kills are less, which follows the above analysis of sniper rifles being used less.

Figure 6: Types of Combat



5 RELATED WORK

There are many websites dedicated for understanding and distribution of knowledge regarding PUBG weapon statistics. OP.GG[10] is one such website. The website has stats like damage, reload speeds, rate of fire, avg. kill distance etc for each weapon. PUBG Lookup[12] is a website that can be used to get stats for any player as long as we have the player username. There are more websites[4, 6, 11, 13] that are dedicated to weapon and vehicle spawn heatmaps, PUBG leaderboards and other information related to PUBG.

There are many research papers on the psychosocial effects that PUBG had on people. Mamun and Griffiths [9] talk about incidents that have occurred in India. They mention cases of incidents ranging from a student failing an exam to death. D'Souza. et al. [5] talk about PUBG addiction and develop a testing methodology to test the addiction. Xu et al. [17] talk about the loyalty of the players of the game PUBG. They take into consideration gender, game design, and other aspects to see the trust and loyalty of different players. They have some interesting findings, that the women have a stronger influence than men on the game.

Though there are alot of literature on PUBG, most focus on psychosocial effects that the game has on the players. To our knowledge, this is the only paper talking about strategies and weapon effectiveness in the game.

6 CURRENT STATUS & FUTURE WORK

We have completed both the Data management and the Data Analysis components for this project. We have inserted the data into the database using Java and JDBC. We have implemented R analytics and extracted meaningful information from the analysis.

The dataset used by us mostly comprises of players in lower-tier rankings. A similar analysis can be done on a different dataset to get a more accurate state of the game at the higher tiers. If there was data available related to landing of the players, that could be used to produce a heatmap which is more useful as all players go in weaponless and it would be safer to land where there are no people.

Due to the lack of data, we have not performed any analysis for the health supplies and weapon attachments like silencers, scope, stock, etc. Due to this, we have no information regarding the configuration of the weapon used. We would like to see the weapon configurations being analyzed in the future. Similarly, we have no information regarding health supplies and vehicles. A similar analysis can be performed with each of them.

7 LESSONS LEARNED

From the technological stand point, we have learned using JDBC on java and analysis using R. We have gotten ourselves familiar with a lot of libraries in R.

Our biggest take away from the project is that we need to work on the initial design. Due to the kind of dataset we had selected, coming up with a predictive analysis for PUBG was difficult. As so many of the decisions in the game are done in real time, coming up with a meaningful predictive analysis was difficult. We had initially faced difficulties with the Java classes as there was a huge performance hit while performing for the whole dataset. This was good learning experience.

8 LEGAL AND ETHICAL CONSIDERATIONS

PUBG has made an API publicly available to be able to get the stats of each match and player. They even allow a developer to monetize the PUBG data by analyzing it as long as they stick to the guidelines [10] [11]. 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. PUBG is a game that people are addicted[5] to and ruin their lives over it. 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. The paper uses data that was collected in the American server, and the games that were present in the dataset were with players that ranked low. Due to this, even if the information provided by the paper is genuine, its true mostly for those players as the player meta and preferences change as they get better at the game.

The API was created with privacy in mind [2]. The data doesn't have any private information of the user. The only information related to the player would be the username. The only way to know the user's personal information is if the player uses it in the username, which is considered a bad practice.

The analysis from the project can be used to understand the tactics of a player. So if the right analysis carried out on certain users, we could understand their tactics which could lead to them losing in tournaments. If the data is analysed wrong and we reach wrong conclusion, it could lead to the person using the analysis lose in the tournaments.

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

[2] PUBG CORPORATION. 2019. PUBG Developer API FAQ:Can I use the API to get a Steam ID from an IGN or vice versa? https://developer.pubg.com/faq [Online; accessed 09-July-2020].

[3] PUBG CORPORATION. 2020. Playerunknown's Battlegrounds. https://www.pubg.com/ [Online; accessed 12-June-2020].

[4] Mitch Dempsey. 2019. PUBG INTERACTIVE MAP. https://pubgmap.io/erangel.html?/v2/15/4ncpcd/BLeG [Online; accessed 09-July-2020].

[5] Deeksha.S D'Souza.L, Manish.S. 2019. Development and Validation of PUBG Addiction Test (PAT). *International Journal of Indian Psychology* (2019), 562–574. https://doi.org/10.25215/0701.063

[6] fellow gamers. 2019. PUBG INTERACTIVE MAP. https://pubgg.com/maps/erangel [Online; accessed 09-July-2020].

[7] Kevin Hitt. 2020. PUBG Corp. Details Events and Prize Money Distribution for the PUBG Global Series. https://esportsobserver.com/pubg-global-series-2020/ [Online; accessed 01-Aug-2020].

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

[9] Griffiths M.D. Mamun, M.A. 2019. The Psychosocial Impact of Extreme Gaming on Indian PUBG Gamers: the Case of PUBG. *International Journal of Mental Health and Addiction* (2019). https://link.springer.com/article/10.1007/s11469-019-00102-4

[10] OP.GG. 2020. OP.GG Analytics Inspire Better Gaming. https://pubg.op.gg/ [Online; accessed 09-July-2020].

[11] pubganalyzer. 2020. PUBGANALYZER. https://www.pubganalyzer.com/

[12] pubglookup. 2020. Detailed PUBG Stats. https://pubglookup.com/ [Online; accessed 29-July-2020].

[13] PUBGStats. 2020. PUBG Stats Leaderboards. https://pubgstats.com/

[14] GAMINGSTREET STAFF. 2019. How battle royale games exploded in popularity. https://www.gamingstreet.com/how-battle-royale-became-popular/#:~:text=as%20creative%20director%20of%20the,by%20the%20end%20of%202017.

[15] Mark Stock. 2020. The Rise of eSports and Online Competitive Gaming. https://www.themanual.com/culture/rise-of-esports-online-competitive-gaming [Online; accessed 01-Aug-2020].

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

[17] Zhuqing Xu, Ming Xiang, and Lijun Pang. 2018. Study on the influence factors of college students' loyalty in PUBG game. *IOP Conference Series: Materials Science and Engineering* 439 (nov 2018), 032059. https://doi.org/10.1088/1757-899x/439/3/032059