**Players Unknown’s Battlegrounds**

Oluwatofunmi Alabi, Noelle Cespedes, Sai Kiran Gontyala, Yoshitha Gattu and Uchechukwu Okafor

Department of Information Systems, California State University

Los Angeles

**Abstract:** Video games are a vast source of information for data extraction due to its digital nature. PUBG is a popular game known as “PlayerUnknown's Battlegrounds.” The dataset applied has provide over 720,000 competitive battles. PUBG is a first/third-person shooter battle royale style game that matches over 90 players on a large island where multiple teams/players fight to the death until only one party remains. The following project demonstrates the use of Hadoop, MapReduce, and Hive to interpret and analyze big data. We will apply the knowledge learned during the lectures, extensive research and development of HiveQL in order to generate data and visualize it on Power BI, Tableau and 3D maps.

Data information:

URL: https://www.kaggle.com/skihikingkevin/pubg-match-deaths

Dataset size: 8GB

Cluster version: Oracle Cloud

No of nodes: 6

Memory size: 180 GB

CPU Speed: 2.195 GHz

**1. Introduction**

Based on the list of data provided by our instructor, we did some research which lead to exclusively decided data which would be used for this project. We then manipulate and filter the datasets via the following step:

* PUBG game analysis; data size is 8GB.
* Cleaning down the information to have a detailed comparison between the kills and attacks.
* From each dataset, we then sorted out the interests of the players, to see which locations are preferred, the survival time, and also to better understand the reasons for the players death.
* The tools used were HiveQL, Putty, Oracle Cloud, Microsoft Excel 3D maps and Power BI.

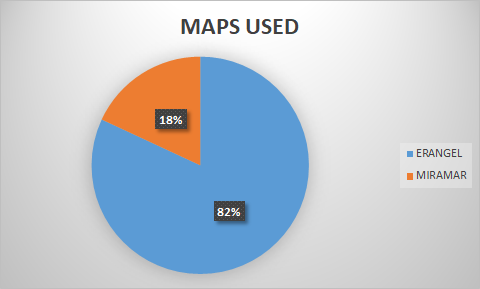
**2. Manipulating datasets**

* 1. **Tools and data processing**
* We extracted our PUBG game data from the corresponding website in a .csv format: PUBG Match Deaths and Statistics from Kaggle.com.



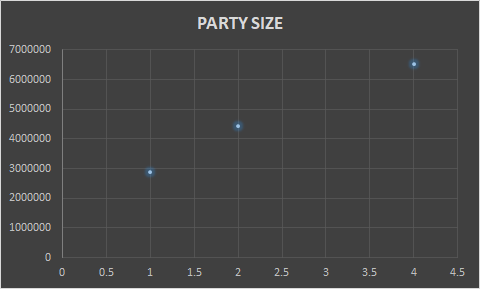
**Figure 1: Data Processing**

* We used basic commands to connect to the data source such as “pscp,” “mkdir,” and then uploaded our data to HDFS and used Hive queries to create “external tables” with the .csv data.
* Then, we made Hive queries to select the desired data from the external table and filtered out any null or unimportant data. (cleaned the data)
* We created hive queries to analyse the data
* Lastly, we used Power BI, Tableau and 3D maps to reproduce the selected data in the form of information by generating the appropriate graphs, maps and chat.



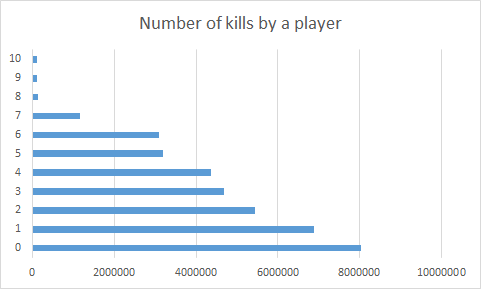
**Figure 2: Map Locations Players Are Interested In**

Figure 2 displays the map locations that players are active in. By analyzing the above pie chart, one can see the game has 2 locations. Of which, 82% of player activity taking place in ERANGEL. With 18% of players in MIRAMAR.



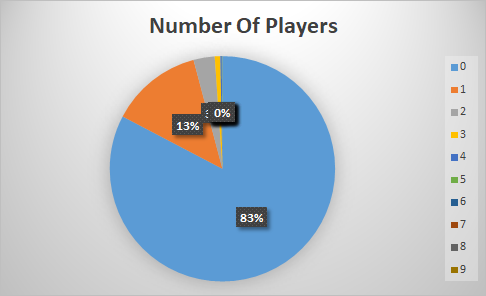
**Figure 3: Party Size Players Are Most Interested In**

Figure 3 shows the party sizes which players use. There are 3 different party sizes: a squad of 4, a duo of 2 and a single player. The most common size party played is the SQUAD [of approx. 6.5M,] composed of 4 players. The single player is the least common party size [of approx. 2.9M.] With the median party size of 2 or the Duo [at approx. 4.4M.]



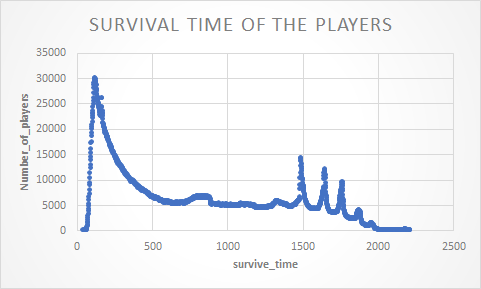
**Figure 4: Number Of Kills By A Player**

Figure 4 shows the number of kills by a player. From the above graph, 8,043,195 players die without killing any enemies. 6,881,553 players have killed only 1 player before being killed.



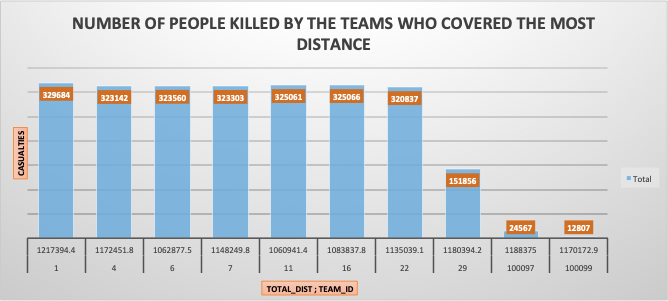
**Figure 5: Players Who Scored Assist Points**

Figure 5 offers the information pertaining to kill credits. Prior to the most recent update, if a teammate shot and finished off an enemy that you had “Downed,” they would get the credit for the kill and the victim gets nothing. The latest update changes the kill credit and has created an “assist” credit. Therefore, if a player downs an enemy with another teammate finishing them off, the kill-stealer will be credited with an assist. The kill credit goes on to the player who originally downed them.



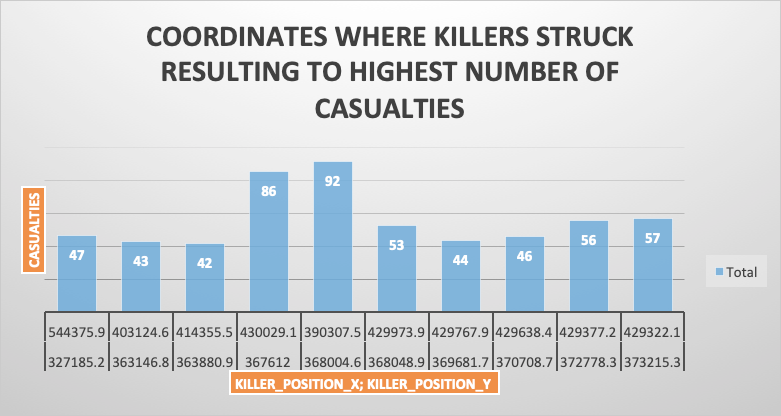
**Figure 6: Survival Time Of The Players**

Figure 6 demonstrates the survival time of the players. Resulting in the finding of over 30K players dying within 125 seconds into the game. Others survive longer into the limited game time.



**Figure 7: Number Of People Killed By Team With Highest Covered Distance**

Figure 7 iterates the number of people killed by the team with the highest covered distance. The analysis illustrates the Teams with the highest distance covered correlating with the highest number of casualties.



**Figure 8: Highest Number Of Casualties Coordinates With Highest Number Of Casualties**

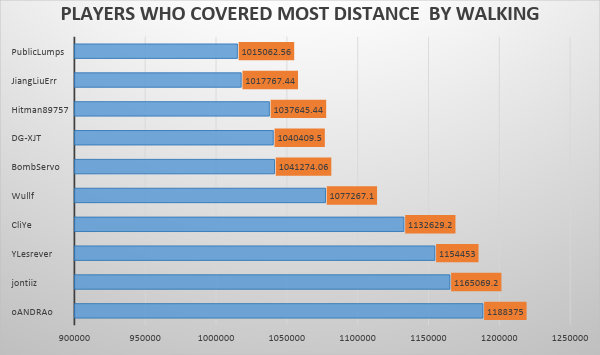
Figure 8 gives the coordinates where killers struck other players, resulting in the highest number of casualties. The exact coordinates where killers  struck of which resulted in 92 as the highest number of casualties at 390,307.5, 368,004.6 (x , y).



**Figure 9:** **Coordinates With Highest Number Of Casualties**

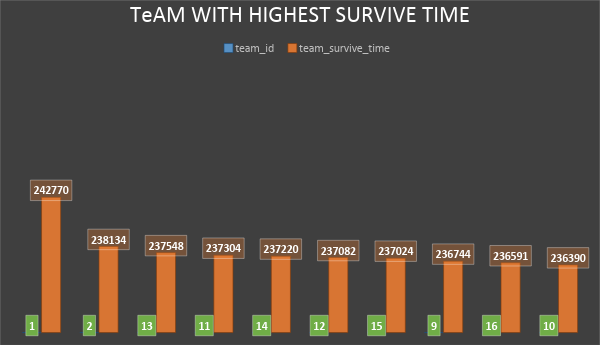
Figure 9 presents the coordinates where the highest number of casualties occured. With the exact points at which victims were struck and killed the most within the 2 location in the virtual map of the game.

`



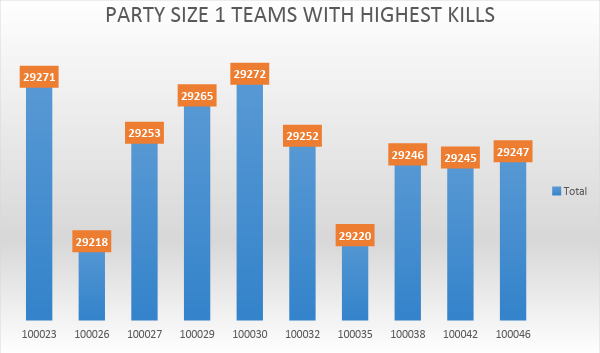
**Figure 10: Distance Covered On Foot**

Figure 10 exhibits the players who covered the most distance by walking. The highest distance covered on foot is approx. 1,188,375 miles by a single player named ‘Oandrao’.



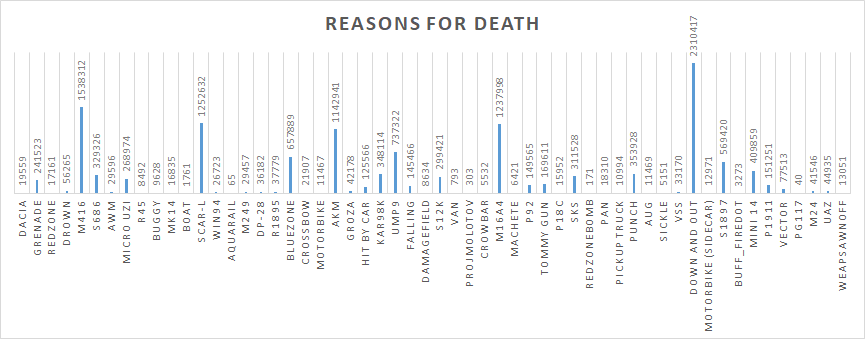
**Figure 11:** **Highest Survival Time**

The Figure 11 graph shows the team with the highest survival time. It can be observed that Team ID = 1 which the highest number of kills (242,770) is also the team with the highest survival time on the battle ground.



**Figure 12:** **Highest Kills-Single Player**

The Figure 12 graph shows the highest kills amongst single players. Player 100,030 has the highest number of kills (29,272.)



**Figure 13: Reasons for Death Of A Player**

Figure 13 defines the reasons (as in weapon used) for a players death and the frequency. The highest number of players are “Down and Out” by a kill with the “M416” weapon.

**3. Summary**

* Successful application of the many tools learned in class: HiveQL, Oracle Cloud, and Tableau to use and manipulate data.
* By creating and running individual queries we were able to discover specific information about the game.
* After cleaning and filtering data we further discovered that PUBG has a great following.

**4. Github URL**

<https://github.com/sgontya/hiveanalysisonpubgdeathrate>

**5.Reference**

https://www.kaggle.com/skihikingkevin/pubg-match-deaths

https://www.pubg.com/PLAYERUNKNOWNS\_BATTLEGROUNDS/

https://www.youtube.com/watch?v=6TOSV3-7hzU