



Players Statistical Analysis

PLAYERUNKNOWN'S BATTLEGROUNDS

By
Gaurav Kshirsagar
Ajinkya Ingle
Shrawan Sapre

Team



Shrawan Sapre



Ajinkya Ingle



Gaurav Kshirsagar

What is PUBG?

- ▶ PlayerUnknown's Battlegrounds is a multiplayer online battle royale game developed and published by PUBG Corporation, a subsidiary of publisher Bluehole.
- ▶ Up to one hundred players parachute onto an island and scavenge for weapons and equipment to kill others while avoiding getting killed themselves.
- ▶ The available safe area of the game's map decreases in size over time, directing surviving players into tighter areas to force encounters. The last player or team standing wins the round.
- ▶ There are three modes of play: solo, duo and squad, with the latter letting you team up with three other players for a team of four. Each is a little different with various pros and cons, the big pro of team play being that have people to support you, revive you and bring an element of co-operative tactics to the game.

Data

```
> summary(d)
player_name      tracker_id      solo_KillDeathRatio      solo_winRatio      solo_TimesSurvived      solo_RoundsPlayed
Length:87898    Min. : 0.00      Min. : 0.000000      Min. : 0.000000      Min. : 76.10      Min. : 1.00000
Class :character 1st Qu.: 51925.25 1st Qu.: 1.030000      1st Qu.: 0.000000      1st Qu.: 17897.05 1st Qu.: 17.00000
Mode :character  Median :112705.50 Median : 1.470000      Median : 2.000000      Median : 45173.33 Median : 48.00000
                  Mean :110794.36 Mean : 1.865855      Mean : 5.017506      Mean : 69715.71  Mean : 79.27534
                  3rd Qu.:164879.75 3rd Qu.: 2.140000      3rd Qu.: 5.560000      3rd Qu.: 92462.65 3rd Qu.: 105.00000
                  Max. :222490.00 Max. :100.000000      Max. :100.000000      Max. :1521200.13 Max. :1681.00000

solo_wins      solo_winTop10Ratio      solo_Top10s      solo_Top10Ratio      solo_Losses      solo_Rating
Min. : 0.000000 Min. :0.0000000      Min. : 0.00000      Min. : 0.00000      Min. : 0.00000      Min. : 984.580
1st Qu.: 0.000000 1st Qu.:0.0000000      1st Qu.: 4.00000      1st Qu.: 10.80000      1st Qu.: 16.00000      1st Qu.:1477.233
Median : 1.000000 Median :0.1100000      Median : 8.00000      Median : 16.70000      Median : 46.00000      Median :1754.210
Mean : 2.080093 Mean :0.1792507      Mean : 12.10628      Mean : 20.66733      Mean : 77.19525      Mean :1768.627
3rd Qu.: 3.000000 3rd Qu.:0.2500000      3rd Qu.: 15.00000      3rd Qu.: 25.90000      3rd Qu.: 102.00000      3rd Qu.:2041.138
Max. :106.000000 Max. :1.0000000      Max. :386.00000      Max. :100.00000      Max. :1673.00000      Max. :2967.090

solo_BestRating      solo_DamagePg      solo_HeadshotKillsPg      solo_HealsPg      solo_KillsPg      solo_MoveDistancePg      solo_RevivesPg
Min. :1200.000      Min. : 0.0000      Min. : 0.0000000      Min. : 0.000000      Min. : 0.000000      Min. : 0.000      Min. :0
1st Qu.:1467.050      1st Qu.: 126.5500      1st Qu.: 0.1800000      1st Qu.: 0.930000      1st Qu.: 1.010000      1st Qu.: 2049.835      1st Qu.:0
Median :1759.740      Median : 169.9100      Median : 0.2900000      Median : 1.290000      Median : 1.430000      Median : 2624.765      Median :0
Mean :1769.416      Mean : 194.2561      Mean : 0.3769248      Mean : 1.415902      Mean : 1.682336      Mean : 2829.960      Mean :0
3rd Qu.:2053.285      3rd Qu.: 231.5200      3rd Qu.: 0.4600000      3rd Qu.: 1.730000      3rd Qu.: 2.000000      3rd Qu.: 3385.938      3rd Qu.:0
Max. :2963.690      Max. :2029.4200      Max. :13.7500000      Max. :24.000000      Max. :20.000000      Max. :26737.340      Max. :0
```

Data

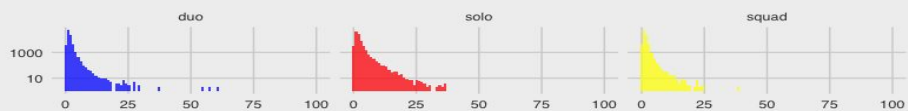
- ▶ Player statistics of the top PUBG players as tracked by <https://pubgtracker.com/>.
- ▶ All statistics were gathered using aggregate region filters (all regions) and feature labels are subdivided by server type: solo, duo, and squad.
- ▶ 87,898 players with 150 numerical game-play features per player (+2 for player name and PUBG Tracker ID).

Data Preparation

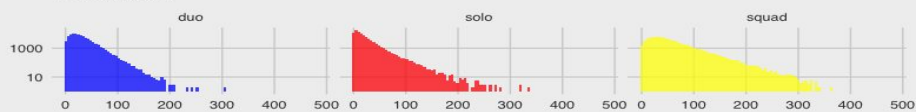
```
#Defining the features in group
performances_features<-c('KillDeathRatio','winRatio','TimeSurvived','RoundsPlayed','Wins',
                        'winTop10Ratio','Top10s','Top10Ratio','Losses','WinPoints')
ratings_features<-c('Rating','BestRating')
support_features<-c('Heals','Revives','Boosts','DamageDealt','DBNOs')
survival_features<-c('Days','LongestTimeSurvived','MostSurvivalTime','AvgSurvivalTime')
distance_features<-c('WalkDistance','RideDistance','MoveDistance','AvgWalkDistance','AvgRideDistance','LongestKill')
combat_features<-c('Kills','Assists','Suicides','TeamKills','HeadshotKills','HeadshotKillRatio','VehicleDestroys',
                  'Roadkills','DailyKills','weeklyKills','RoundMostKills','MaxKillStreaks','WeaponAcquired')
```

Exploratory Data Analysis

KillDeathRatio



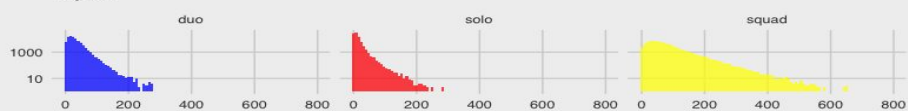
TimeSurvived



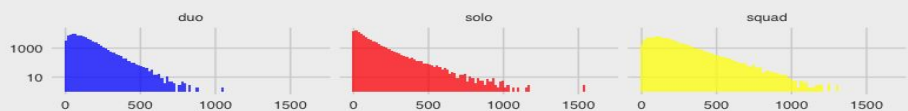
Wins



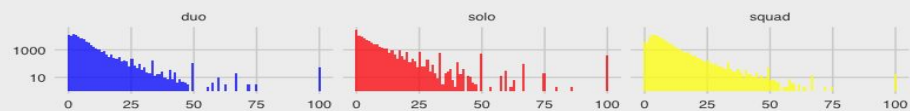
Top10s



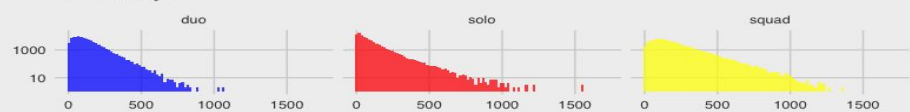
Losses



WinRatio



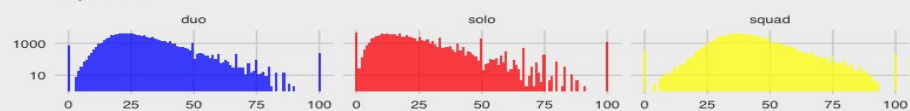
RoundsPlayed



WinTop10Ratio



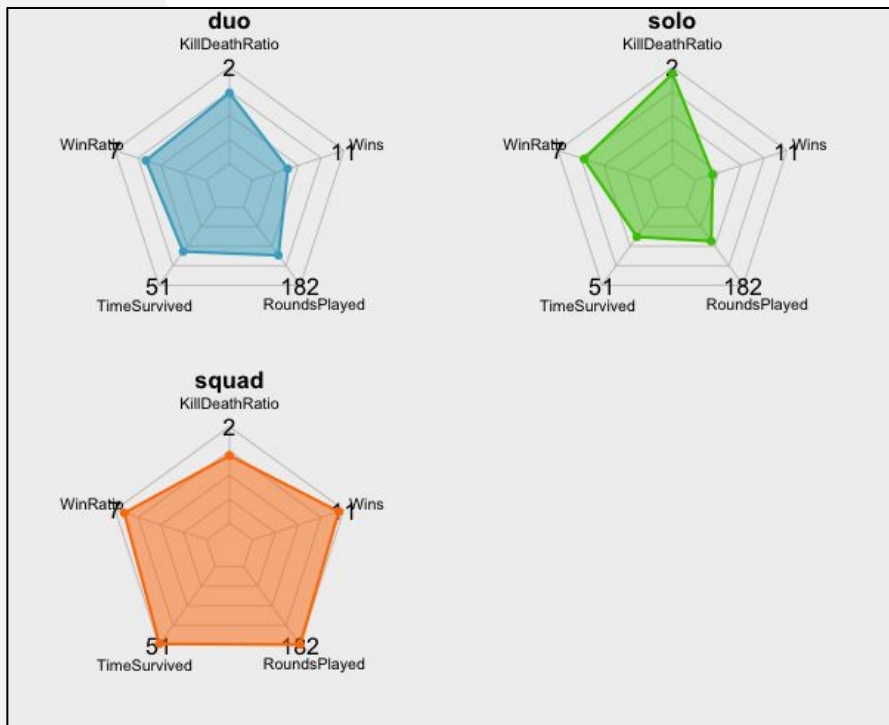
Top10Ratio



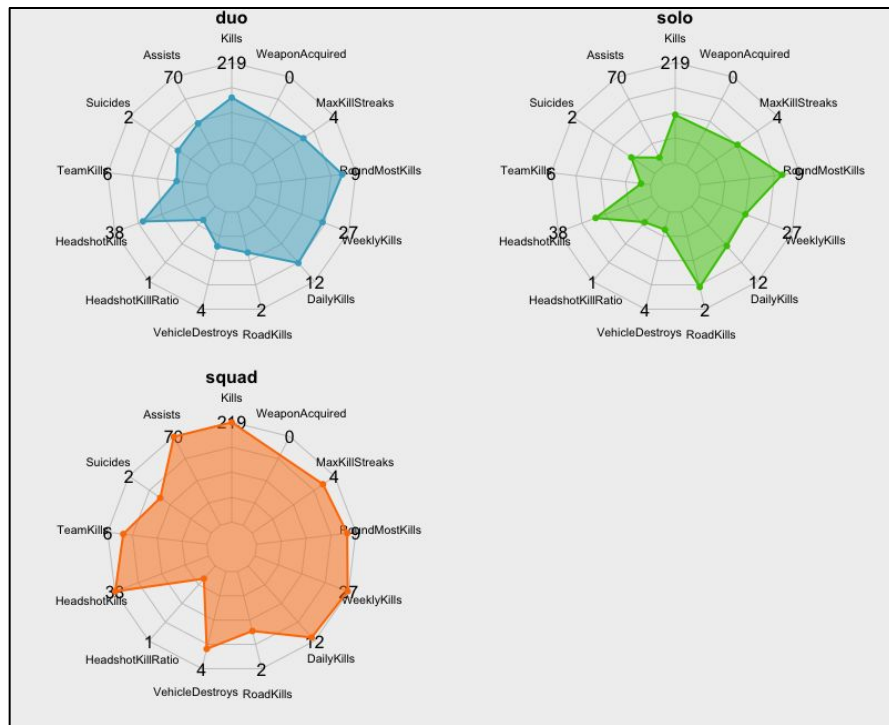
WinPoints



Exploratory Data Analysis

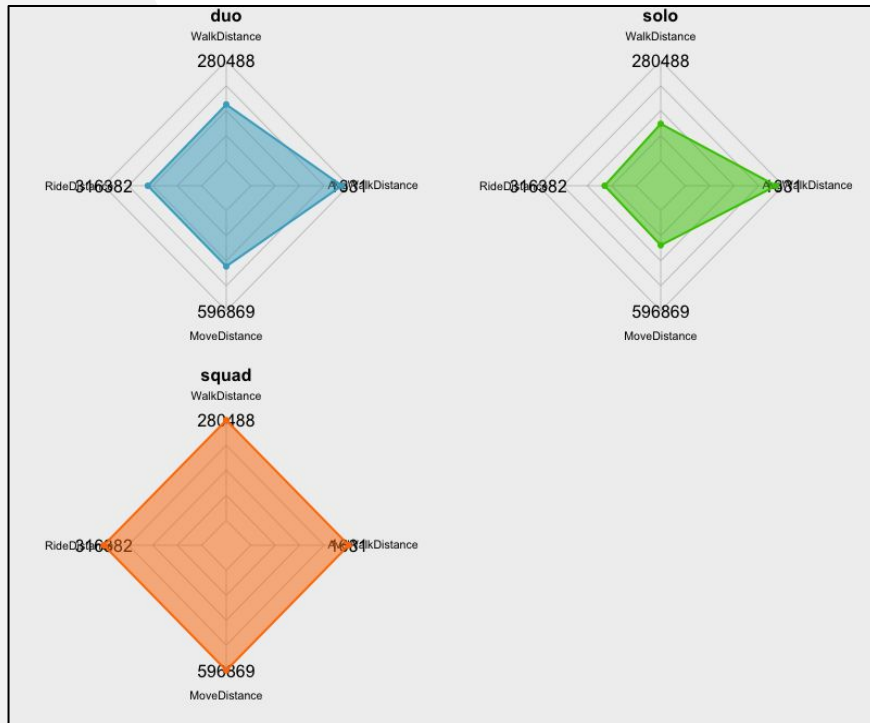


Performance features

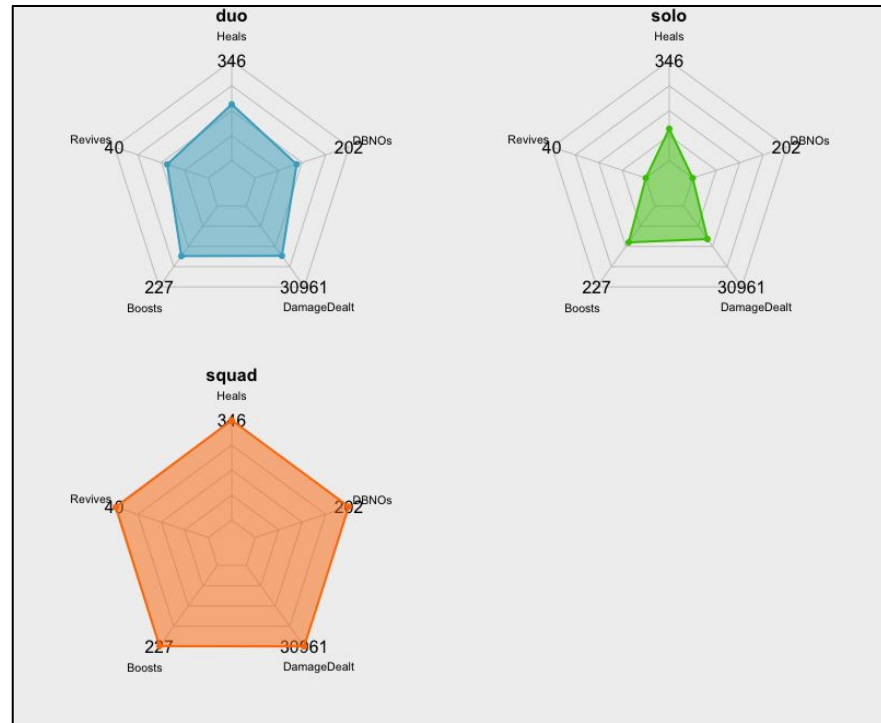


Combat features

Exploratory Data Analysis



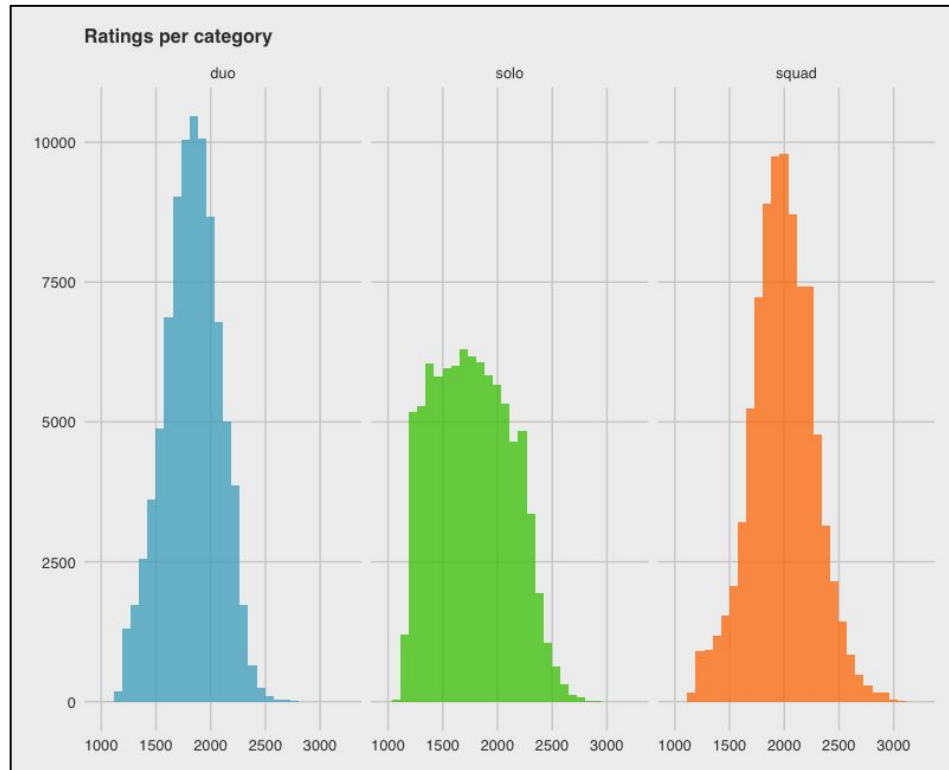
Distance features



Support Features

Exploratory Data Analysis

- ▶ The best rating distribution is broader for solo players compared to duo and squad players.
- ▶ Playing in a group (as a duo or squad) will result in a global score/rating
- ▶ Skill of a team is ~ the average of each players skills.

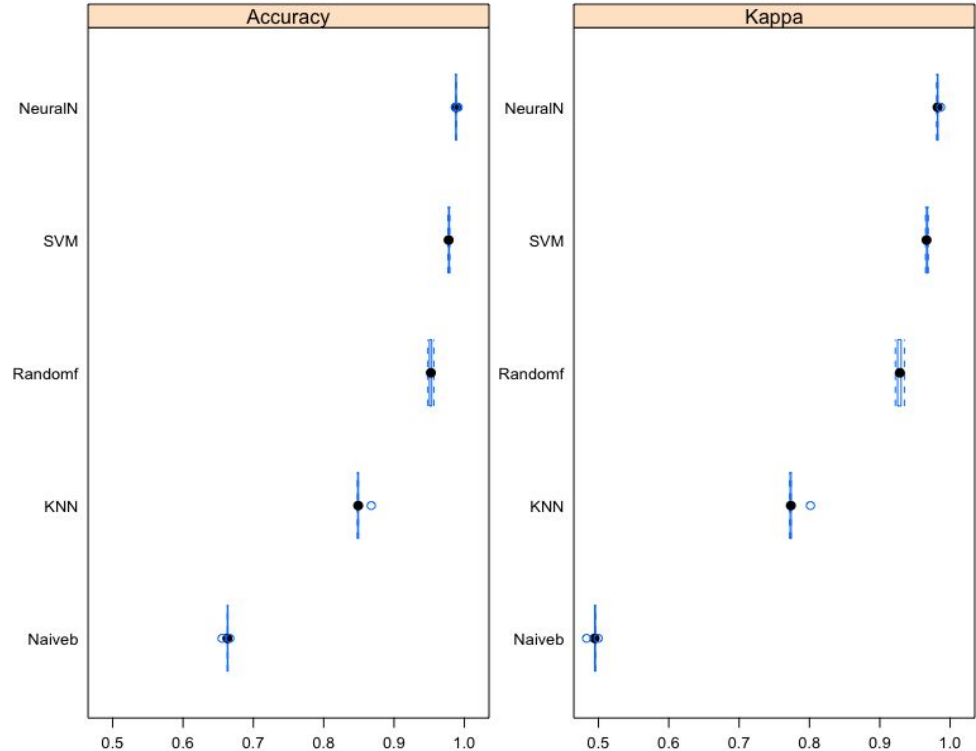


Machine Algorithms used

- ▶ K- Nearest Neighbors (k-NN)
- ▶ Naive Bayes
- ▶ Support Vector Machine (SVM)
- ▶ Artificial Neural Network (ANN)
- ▶ Random Forest
- ▶ K-means Clustering
- ▶ Hierarchical Clustering

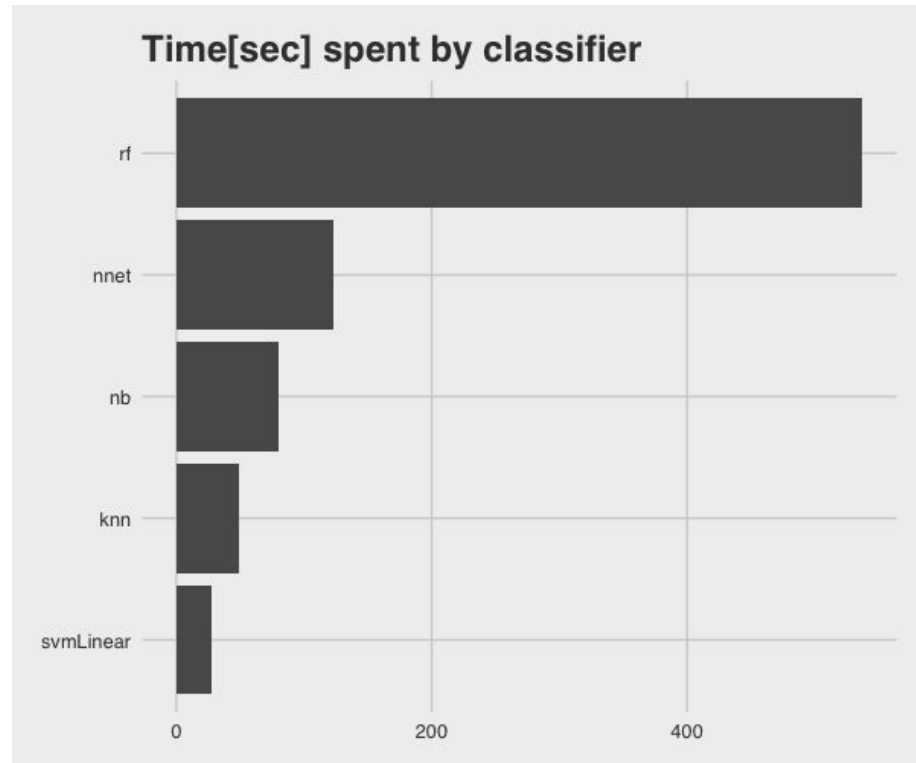
Accuracy of Classifiers

- ▶ Accuracy of Artificial Neural Network algorithm is the highest (98.7%).
- ▶ SVM was the close second.
- ▶ Naive Bayes algorithm has the highest error rate (33%).
- ▶ Kappa values gives us more reliable results as they are based on Expected Accuracy and not the Observed Accuracy.



Time taken by Classifiers

- ▶ Most time efficient Classifiers:
 1. SVM Classifier
 2. KNN
 3. Naive Bayes
 4. Neural Network
 5. Random Forest
- ▶ Since Random Forest have time complexity of $O(M(mn \log n))$ it takes the maximum amount of time.
- ▶ While working on large datasets SVM might be a better option as it is comparatively more accurate along with being time-efficient.



Confusion Matrix for

- ▶ We got an accuracy of 98.76 % with ANN
- ▶ That gets us an error rate of 0.2 %.
- ▶ The kappa value was 98.14 %.

```
> confusionMatrix(test_pred, all_test$cat)
```

Confusion Matrix and Statistics

	Reference		
Prediction	solo	duo	squad
solo	6978	34	9
duo	20	6857	86
squad	2	109	6905

Overall Statistics

Accuracy : 0.9876
95% CI : (0.986, 0.9891)
No Information Rate : 0.3333
P-Value [Acc > NIR] : < 2e-16

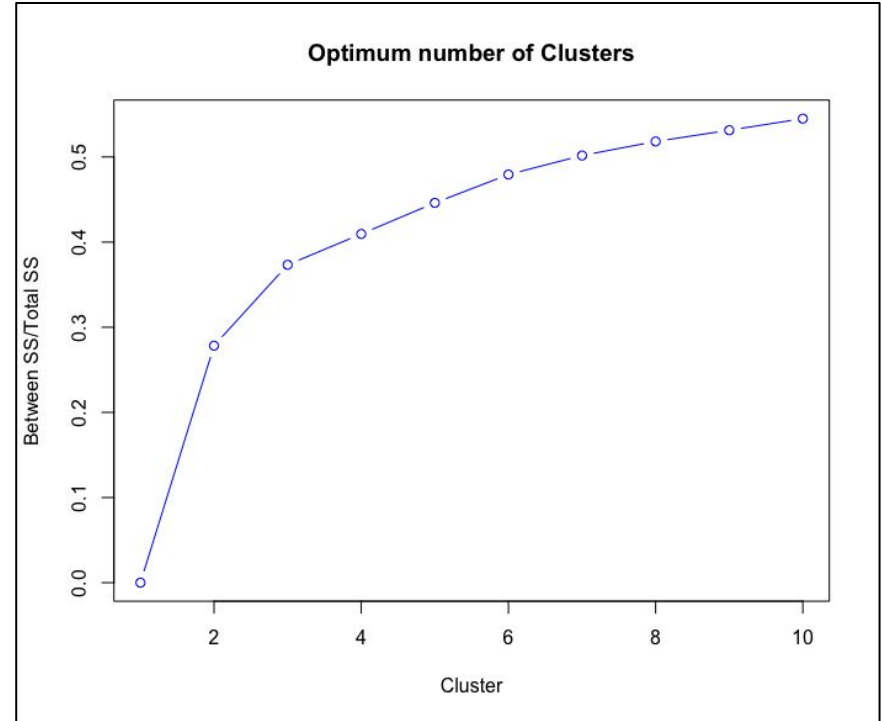
Kappa : 0.9814
McNemar's Test P-Value : 0.01288

Statistics by Class:

	Class: solo	Class: duo	Class: squad
Sensitivity	0.9969	0.9796	0.9864
Specificity	0.9969	0.9924	0.9921
Pos Pred Value	0.9939	0.9848	0.9842
Neg Pred Value	0.9984	0.9898	0.9932
Prevalence	0.3333	0.3333	0.3333
Detection Rate	0.3323	0.3265	0.3288
Detection Prevalence	0.3343	0.3316	0.3341
Balanced Accuracy	0.9969	0.9860	0.9892

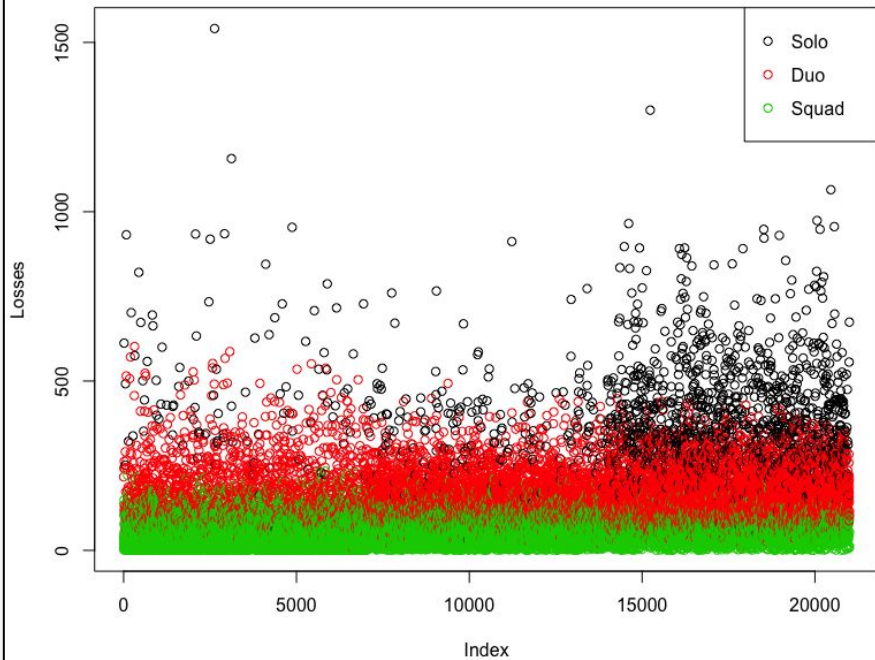
K-Means Clustering

- ▶ BetweenSS/TotalSS ratio basically demonstrates the integrity of the clustering the k-means has found.
- ▶ Ideally you want a clustering that has the properties of internal cohesion and external separation
- ▶ It is observed that, there's little deviation after 3 clusters.

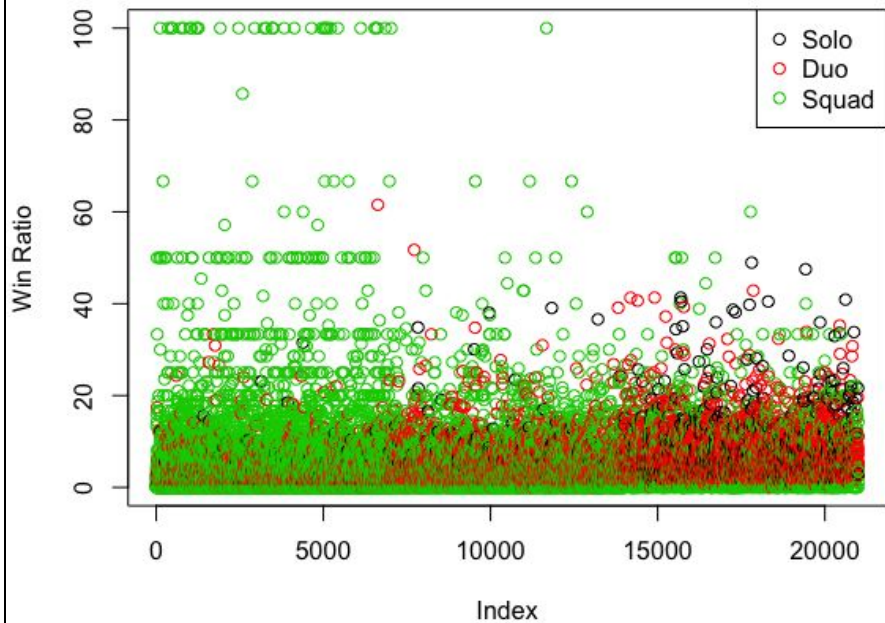


Clusters with Features

Losses of Solo, Duo and Squad

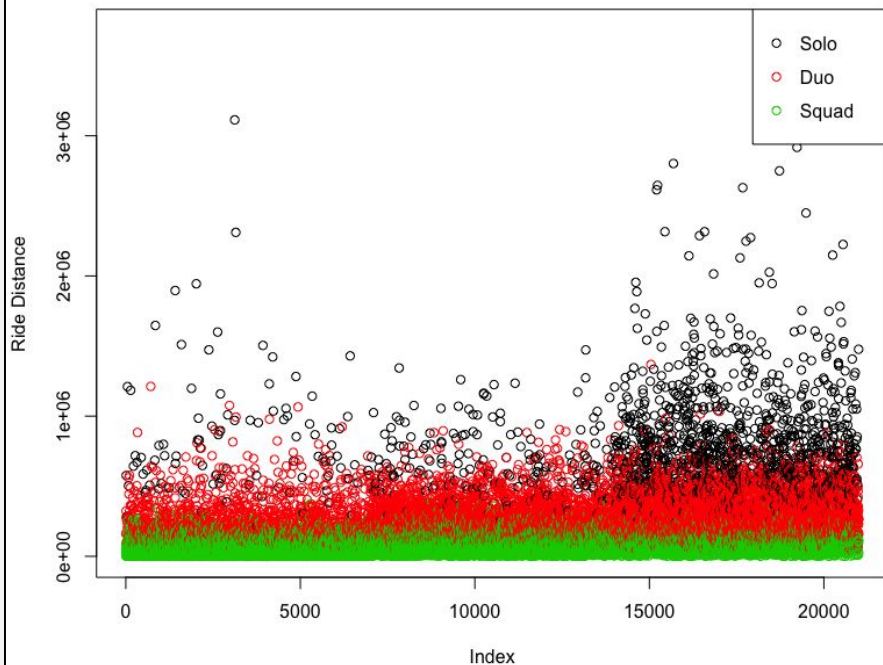


Win Ratios of Solo, Duo and Squad

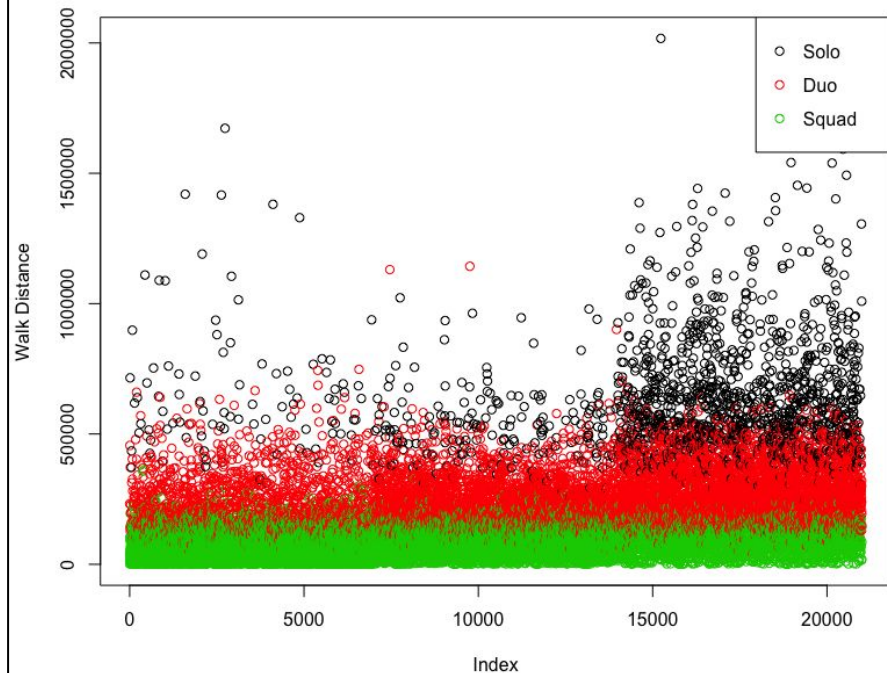


Clusters with Features

Ride Distance of Solo, Duo and Squad

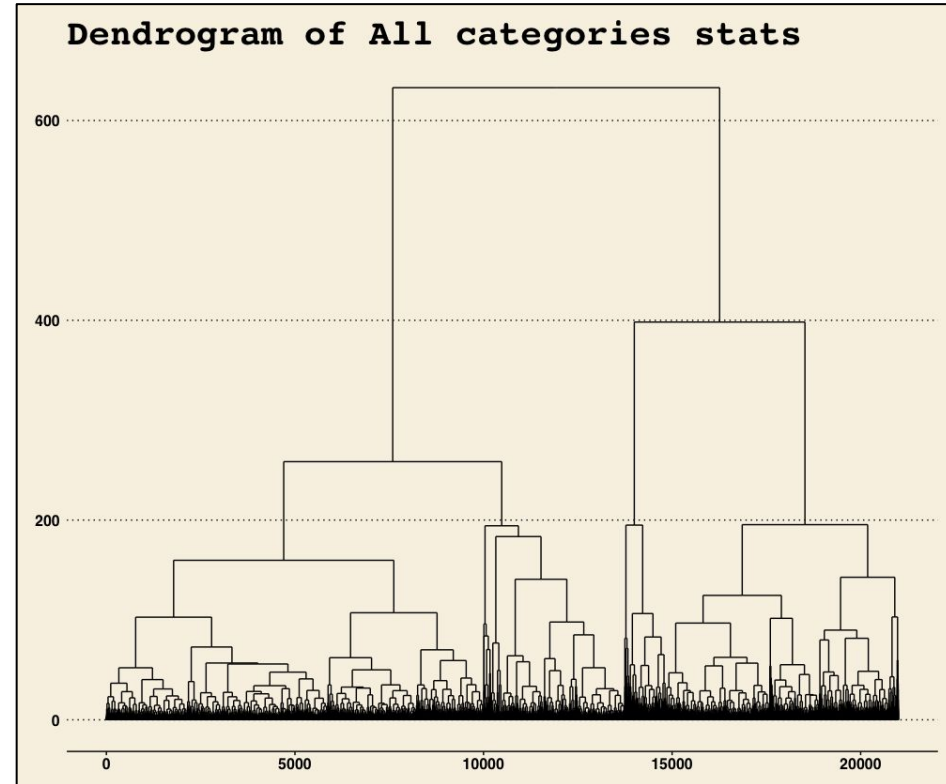


Walk Distance of Solo, Duo and Squad



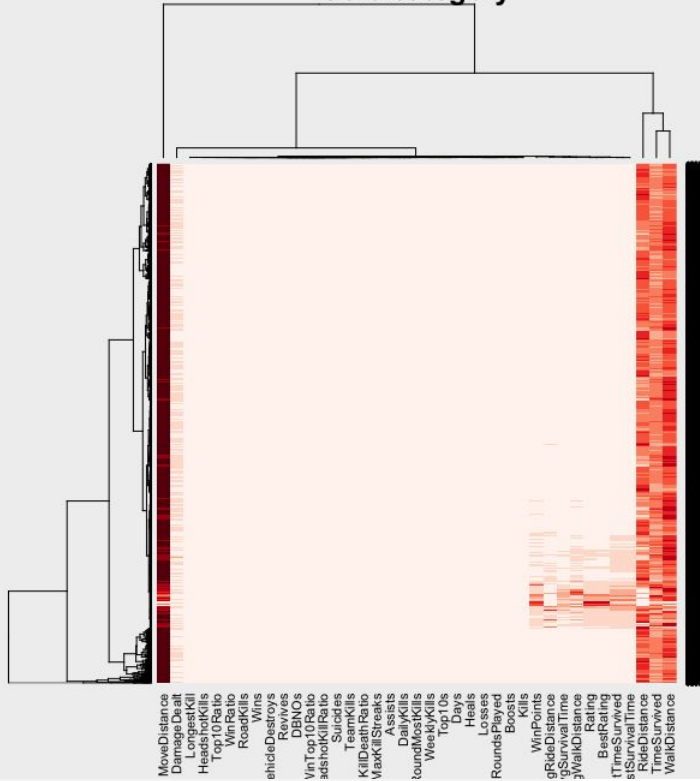
Hierarchical Clustering

- ▶ We used “ward.d2” method for Hierarchical Clustering.
- ▶ “ward.d2” uses agglomerative Hierarchical Clustering.
- ▶ This is a “bottom up” approach.
- ▶ Starts with each user as a clusters and merges them as it moves up

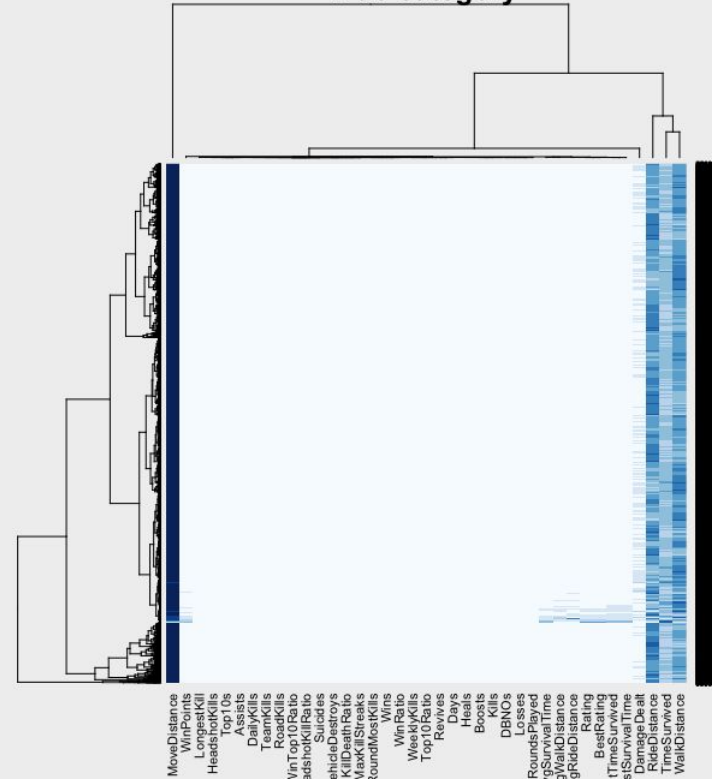


Heatmap

Solo category



Duo category



Heatmap Analysis

- ▶ We get a glimpse of the most used features in the game by carefully observing the heatmap.
- ▶ The game developers and designers can focus on the most used aspects of the game.
- ▶ The resources wasted in developing the least used features can be utilized to improve the existing most used features.
- ▶ The features like “Vehicle destroys” and “Road kills” can be overlooked since very few players are using such features.
- ▶ Different categories like solo, duo, and squad have different preferences that needs improvement.

Future Scope

- ▶ There are numerous business benefits for data analysis and exploration of this dataset.
- ▶ Bluehole can focus on the development of a **particular** mode of game-play: solo, duo or squad based on the predictive model's results.
- ▶ A potential customer can be grouped into any one of the modes and marketing the product using techniques like personalised social media advertising can be used to attract more customers.
- ▶ The dataset can be further analyzed to eventually get an idea of the player statistics that actually predict the probability of **winning the game** (can be dangerous in the wrong hands!)

“

Thank You!

”