PUBG Finish Placement Prediction

Made by-
Ritu Agrawal
Ankit Yadav
Preetam Jain

1. Introduction

Battle Royale-style video games have taken the world by storm. 100 players are dropped onto an island empty-handed and must explore, scavenge, and eliminate other players until only one is left standing, all while the play zone continues to shrink.

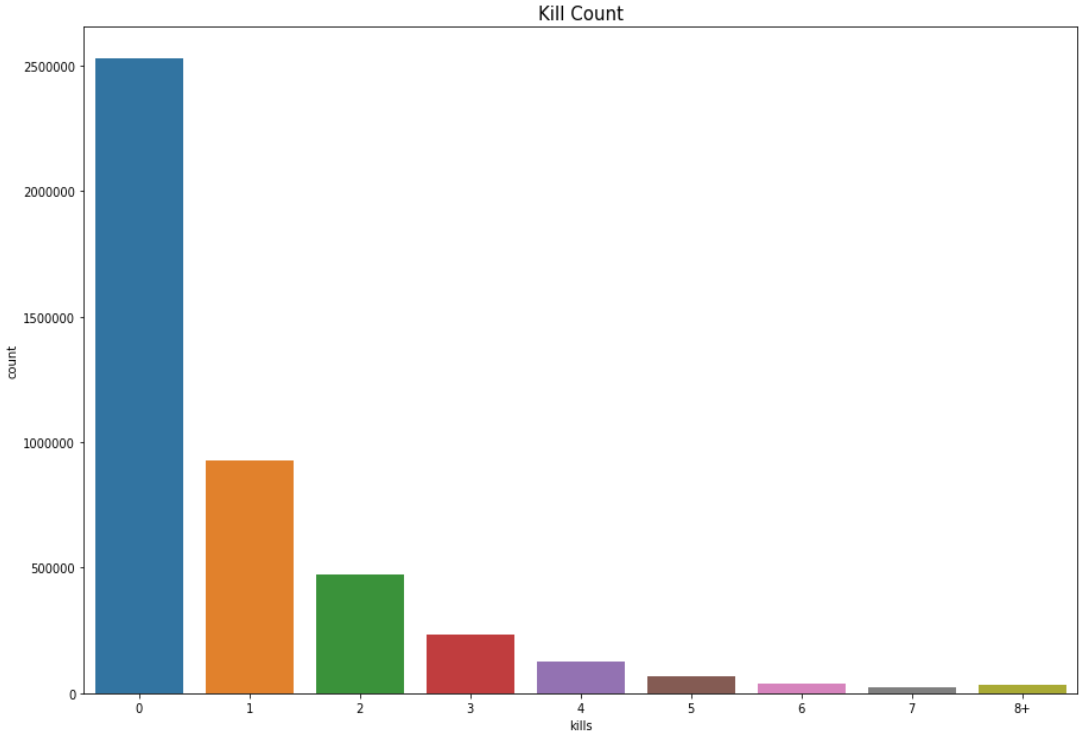
PlayerUnknown's BattleGrounds (PUBG) has enjoyed massive popularity. With over 50 million copies sold, it's the fifth best selling game of all time, and has millions of active monthly players.

Thus, this is a very interesting problem. We are given a train as well test data (csv) with around 4.5 million rows in it. In this we are trying to predict the finish placement of a winner in a scale of 0 to 1, the higher the number, the more the winning likeliness.

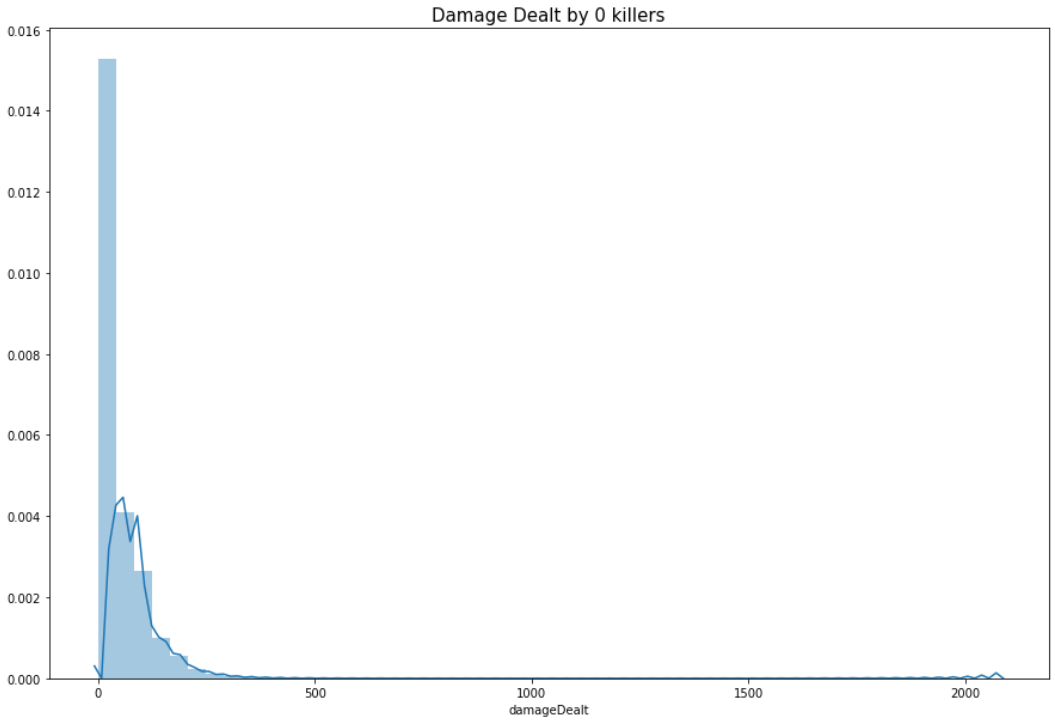
2.Exploratory Data Analysis

It is one of the most important steps in finding the insights from the data. In this, we tried to find the following-

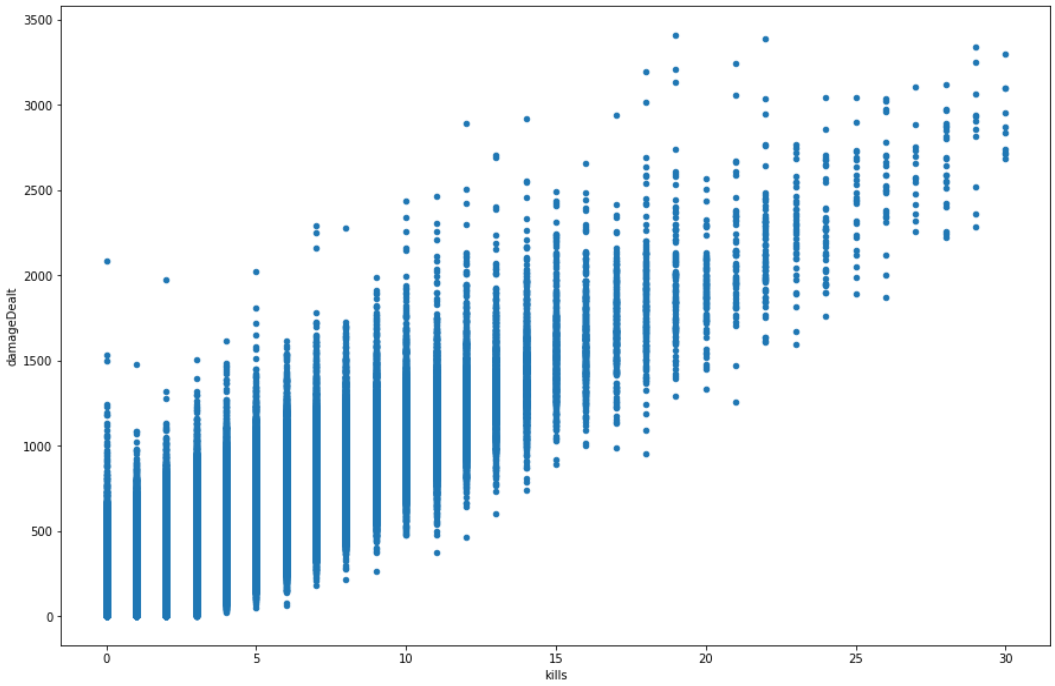
A. Analysis on the basis of kills-



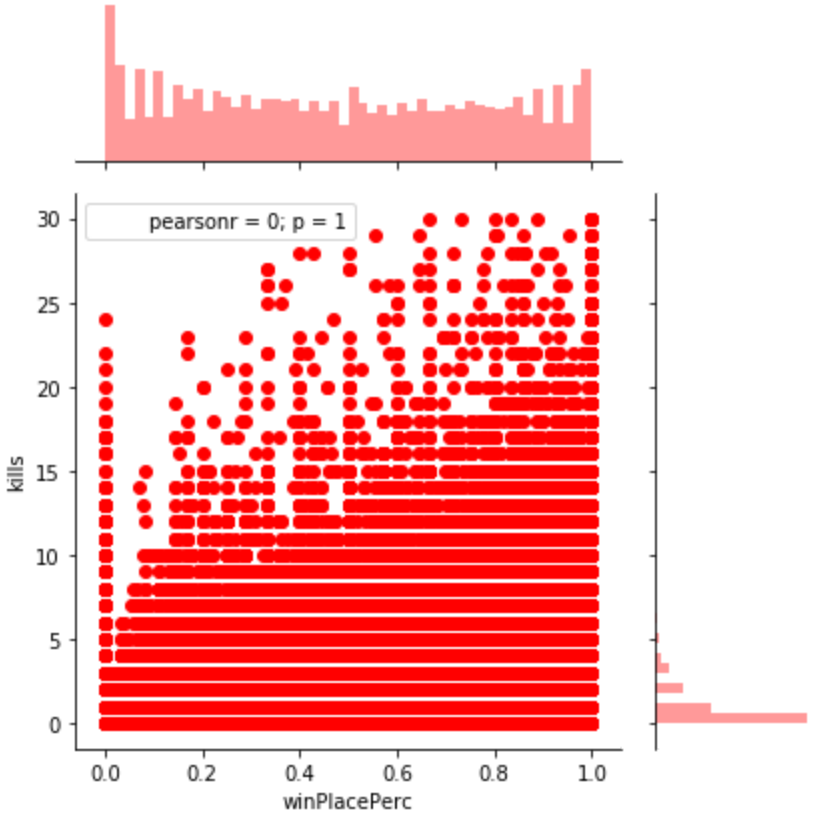
B. Damage dealt by 0 killers-



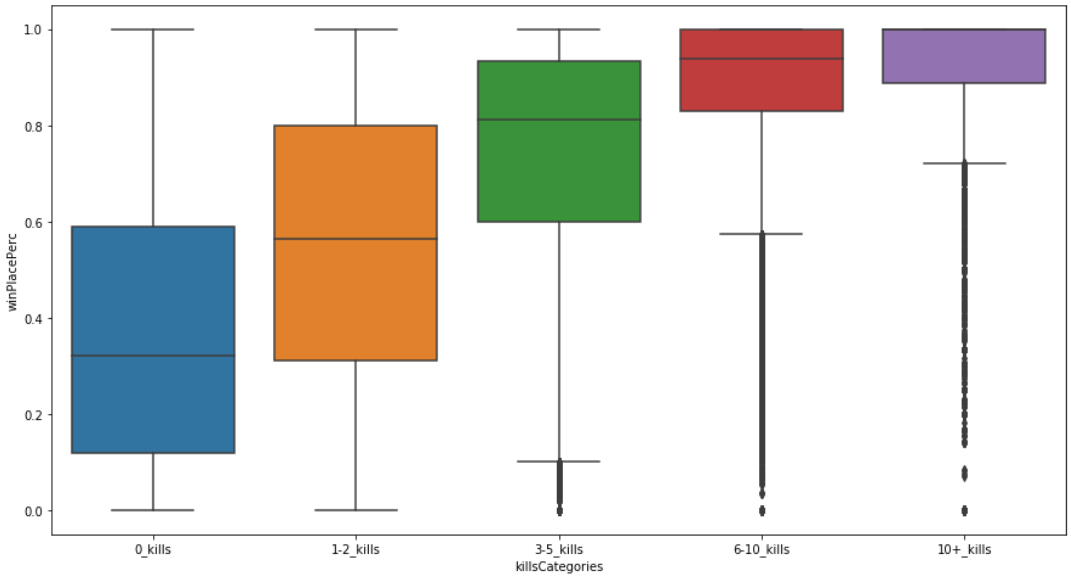
C. Kills vs. Damage Dealt plot-



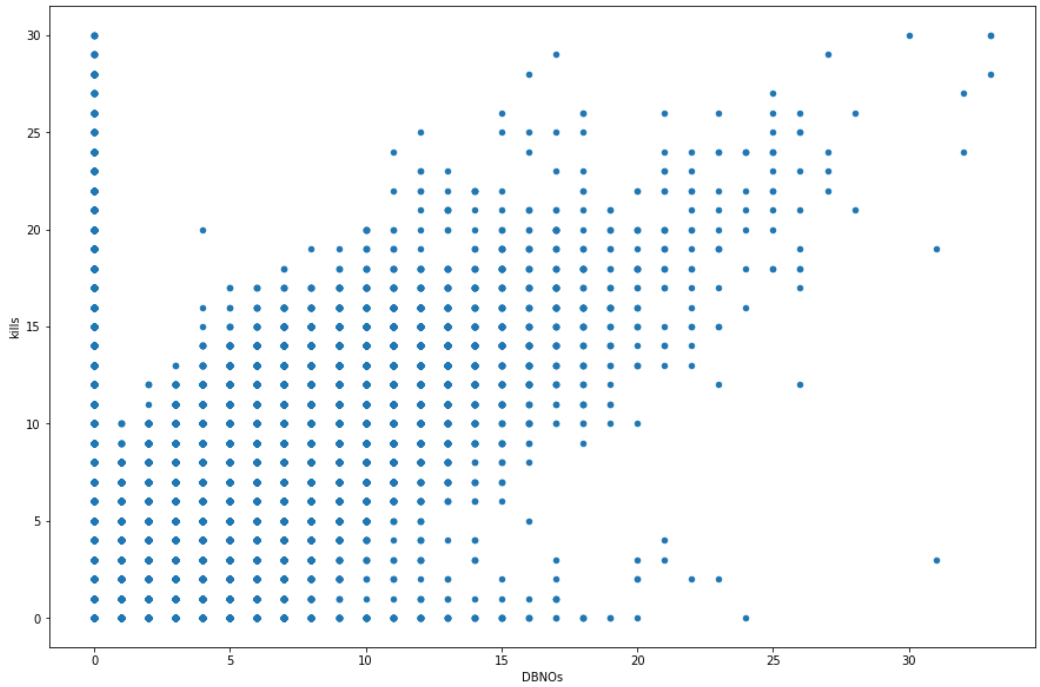
D. Plotting win placement percentage vs kills



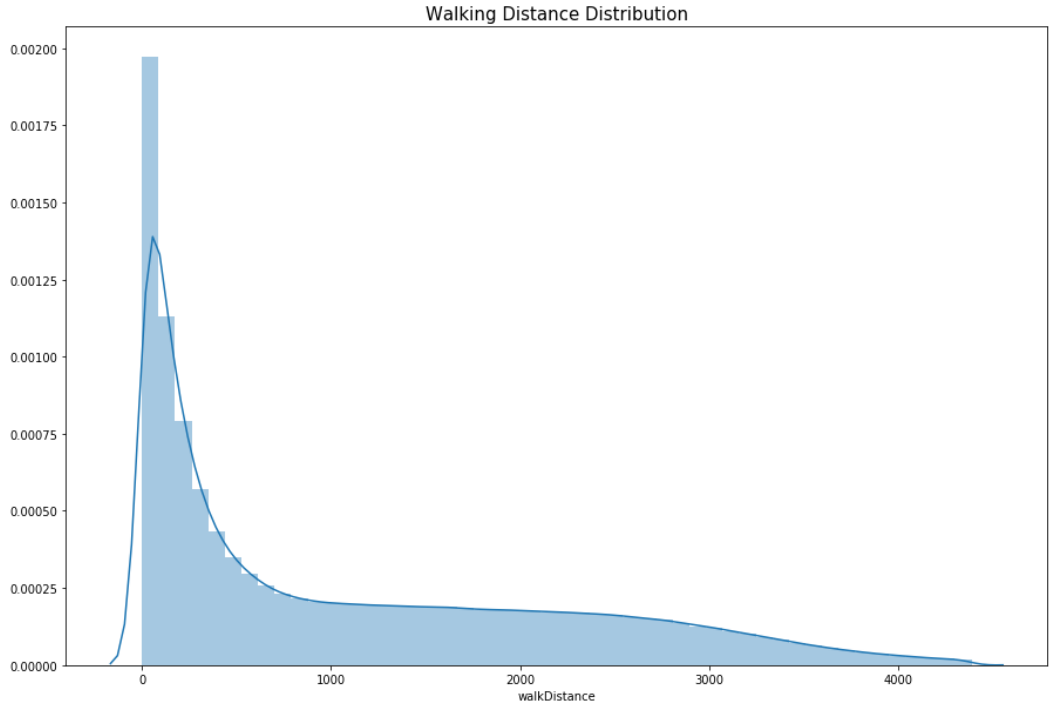
E. Kills vs. Winning-



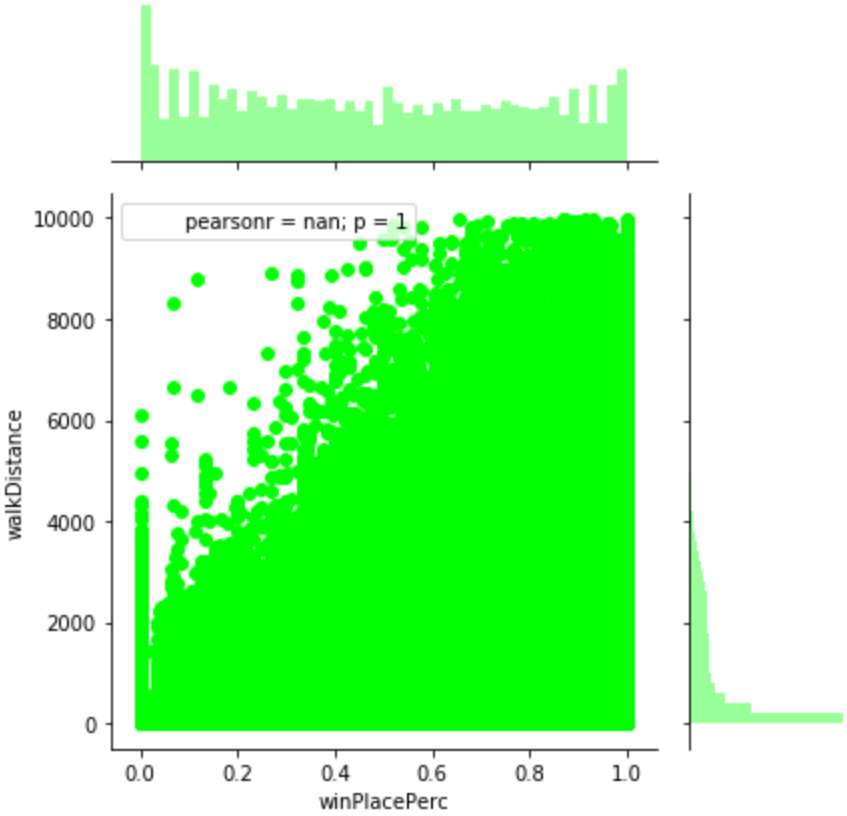
F. Plotting the relation between DBNO and kills



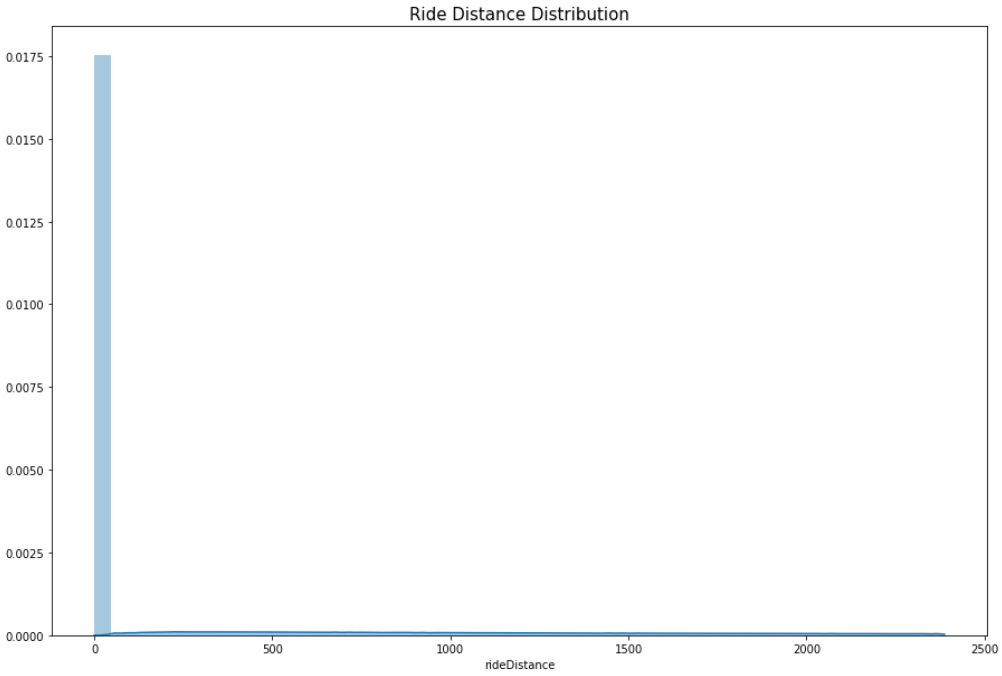
G. Analysis on the basis of distance run-

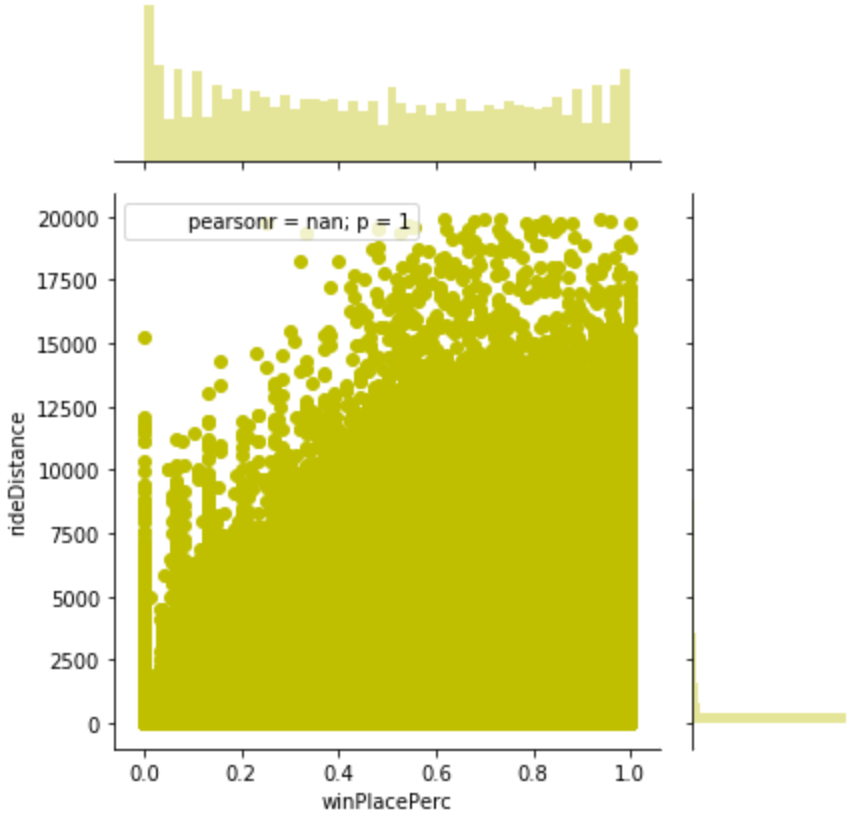


H. Relation between walk distance and winner placement

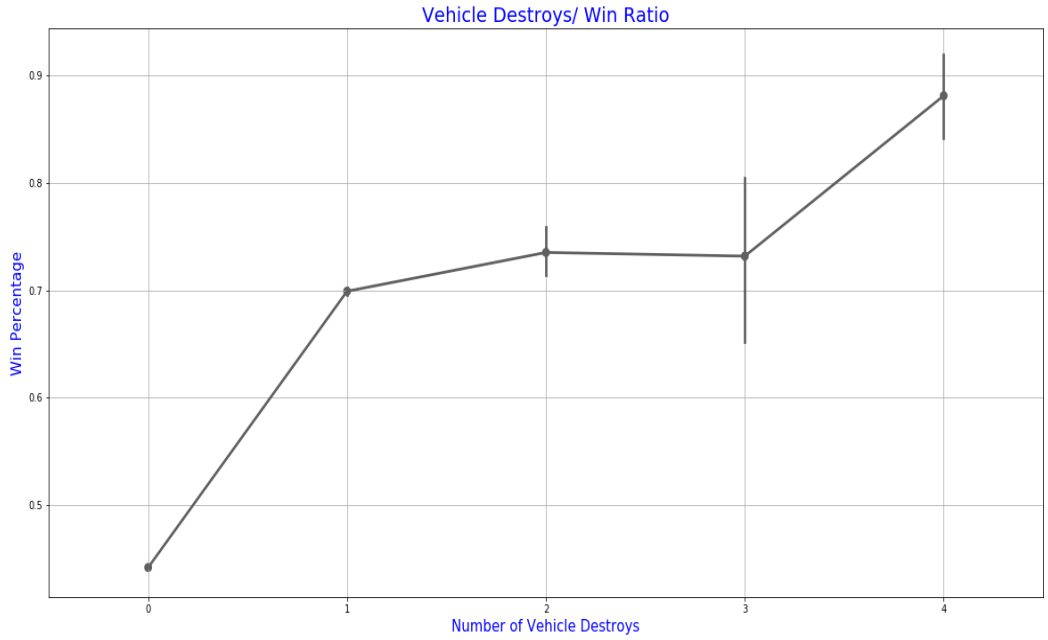


I. Analysis on the basis of Driving Vehicles

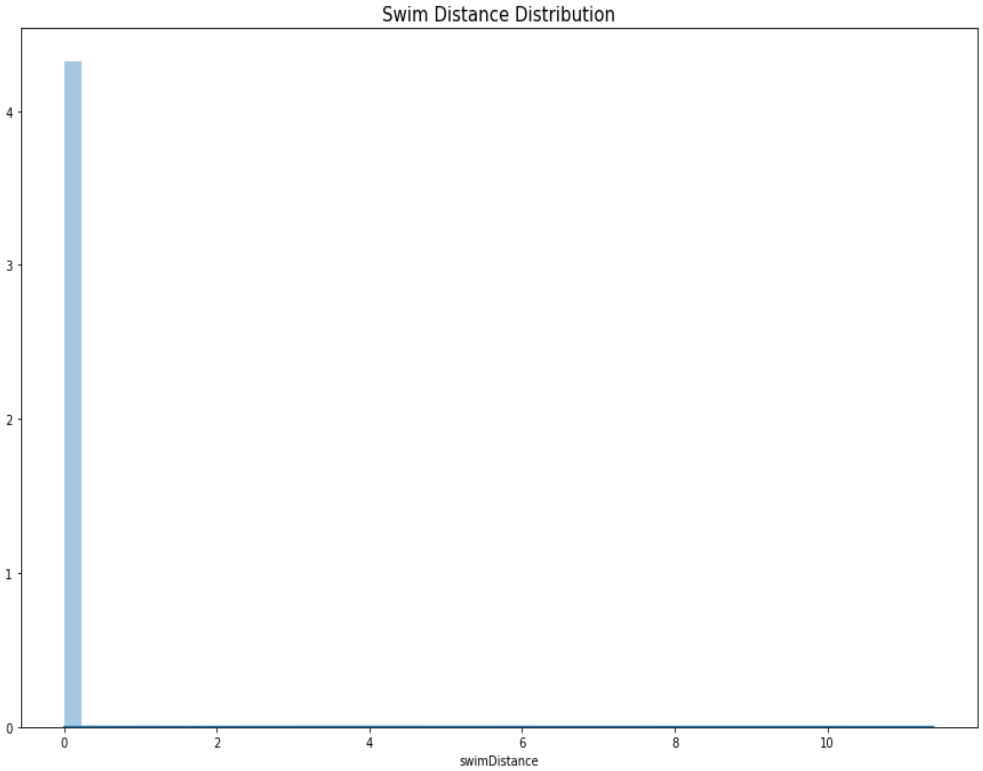




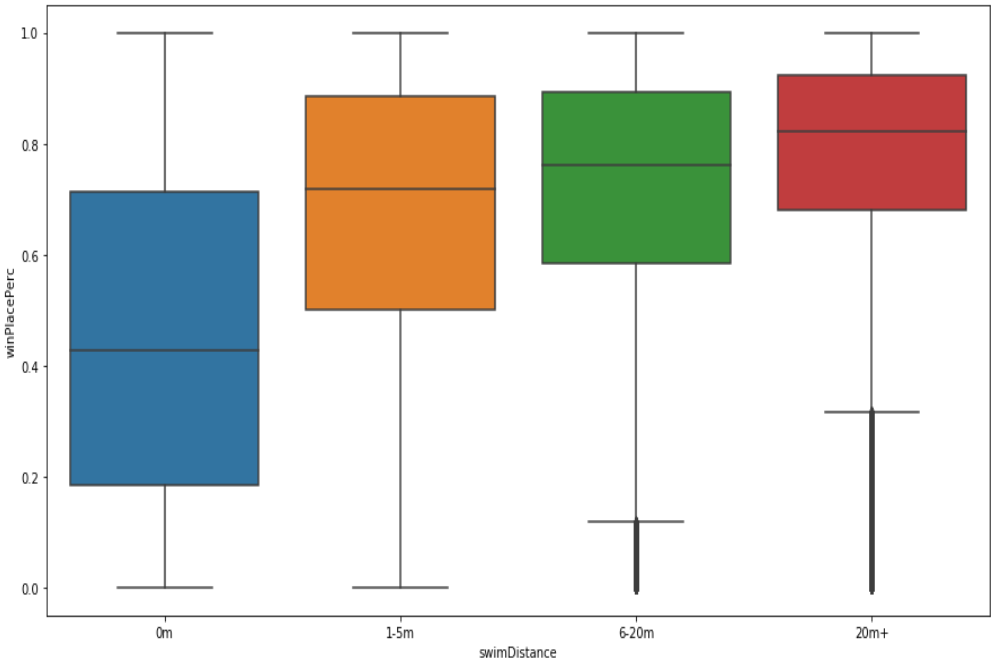
J.



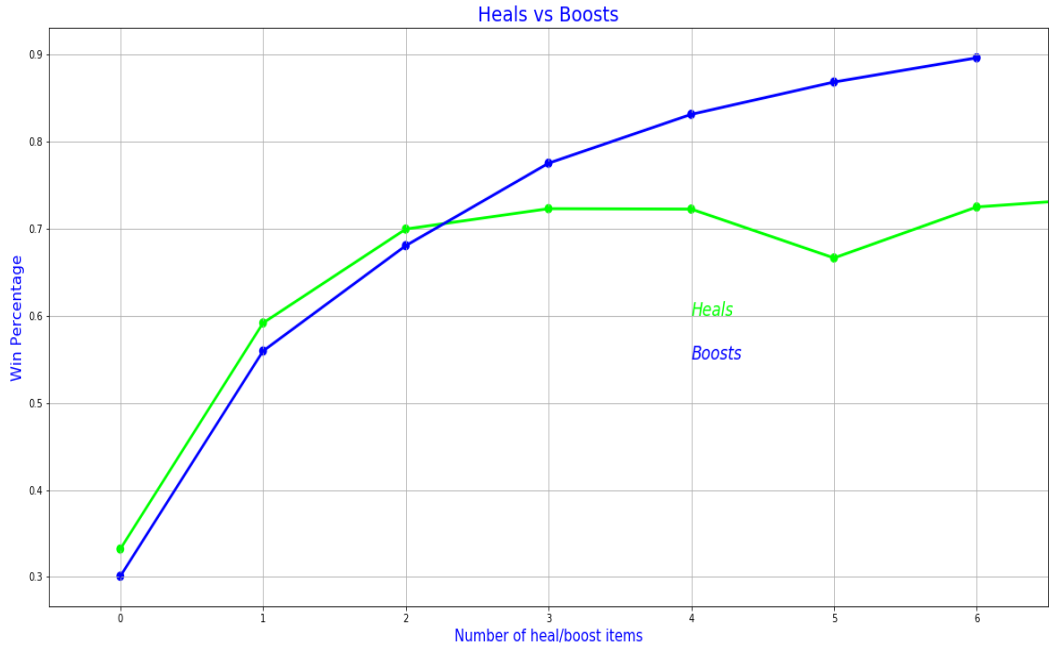
K. Analysis on the basis of swim distance



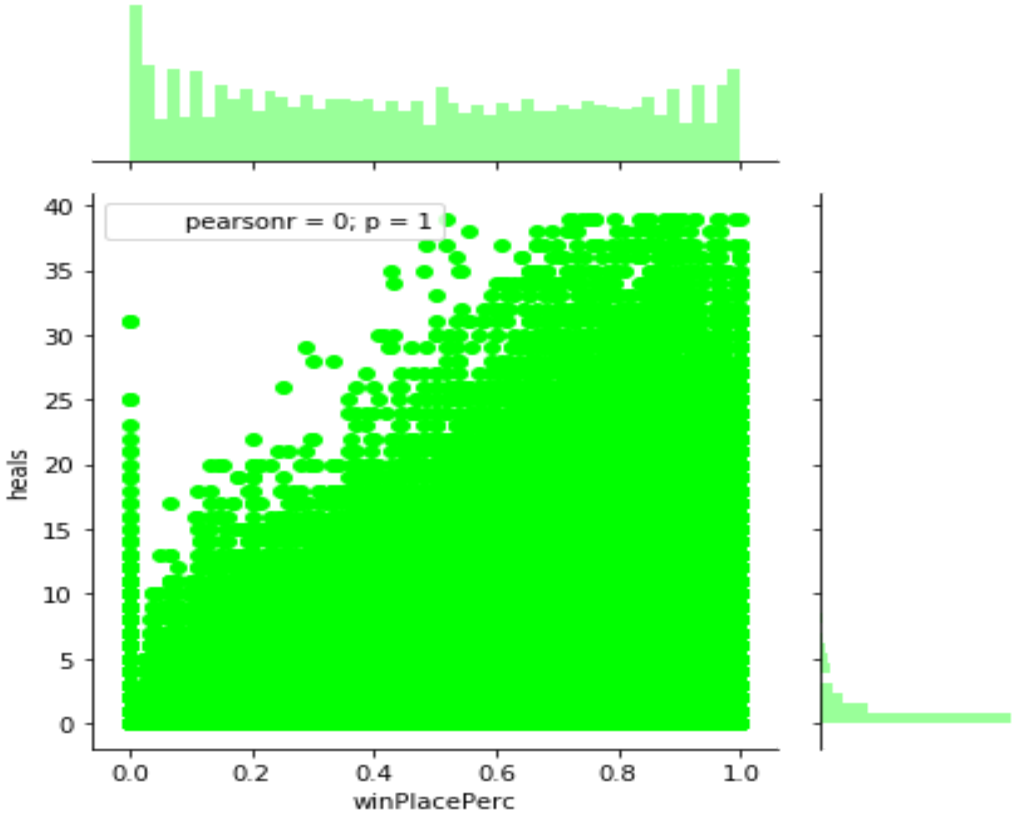
L.



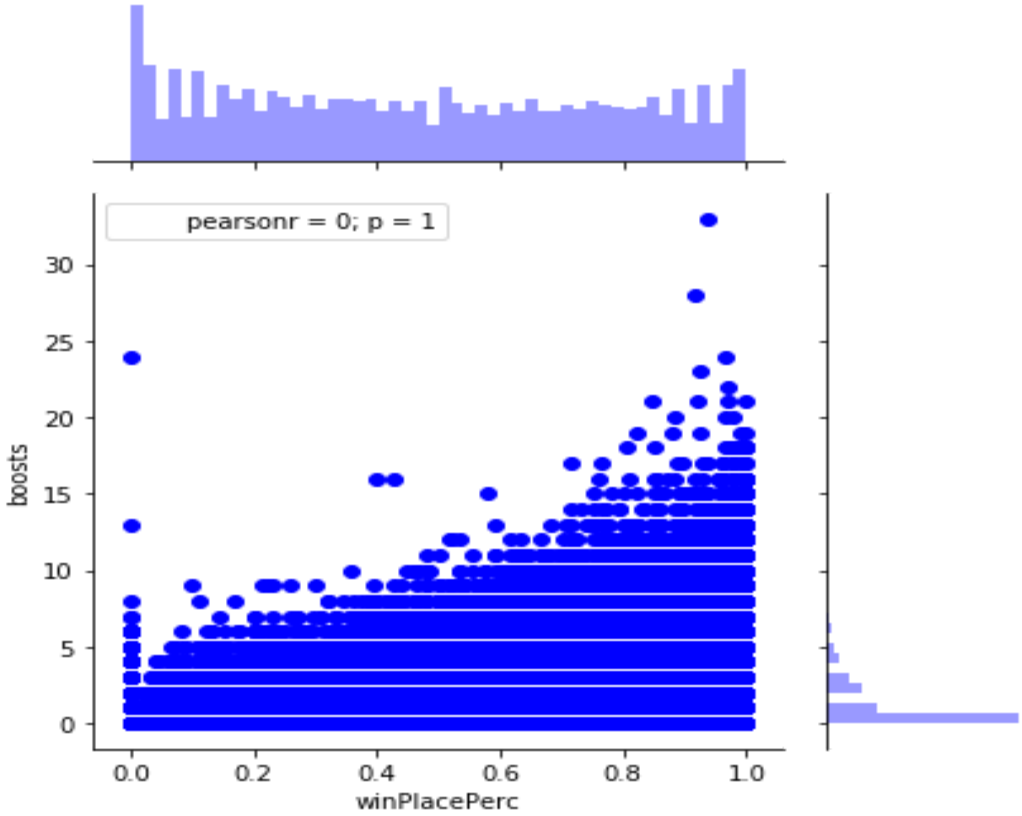
M. Using heal & boost vs winning percentage


Heals vs Boosts
Win Percentage
Number of heal/boost items
Heals
Boosts

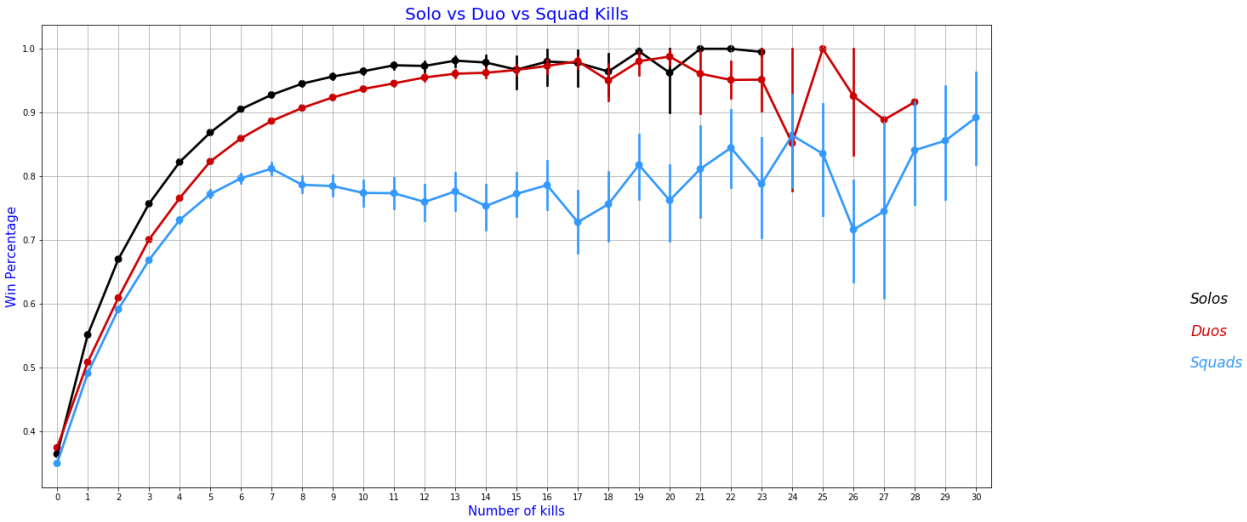
N.


pearsonr = 0; p = 1
heals
winPlacePerc

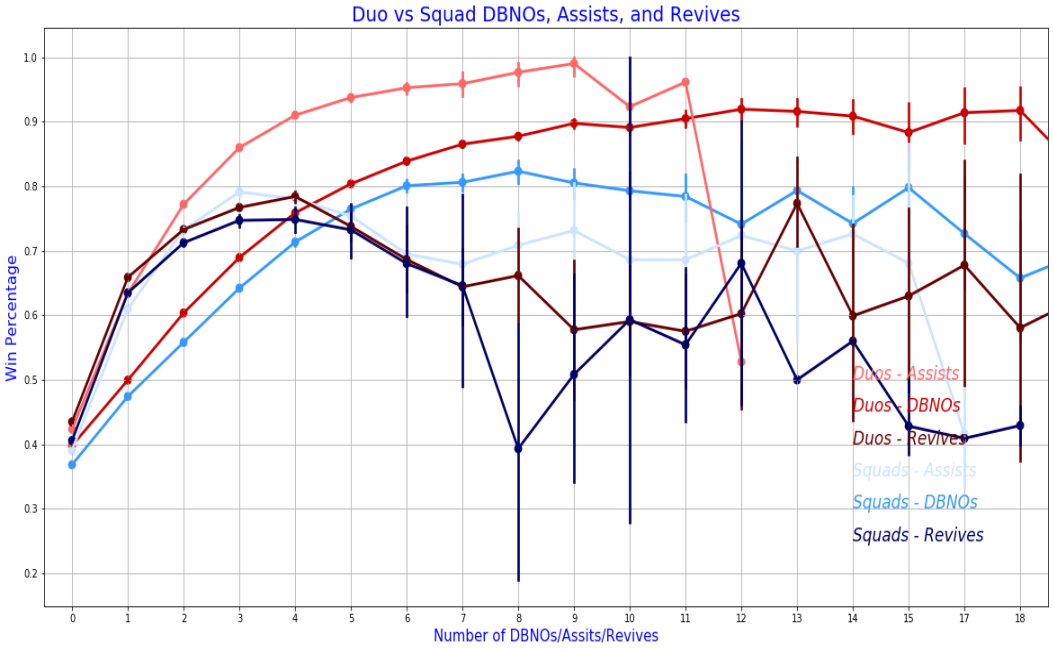
O.



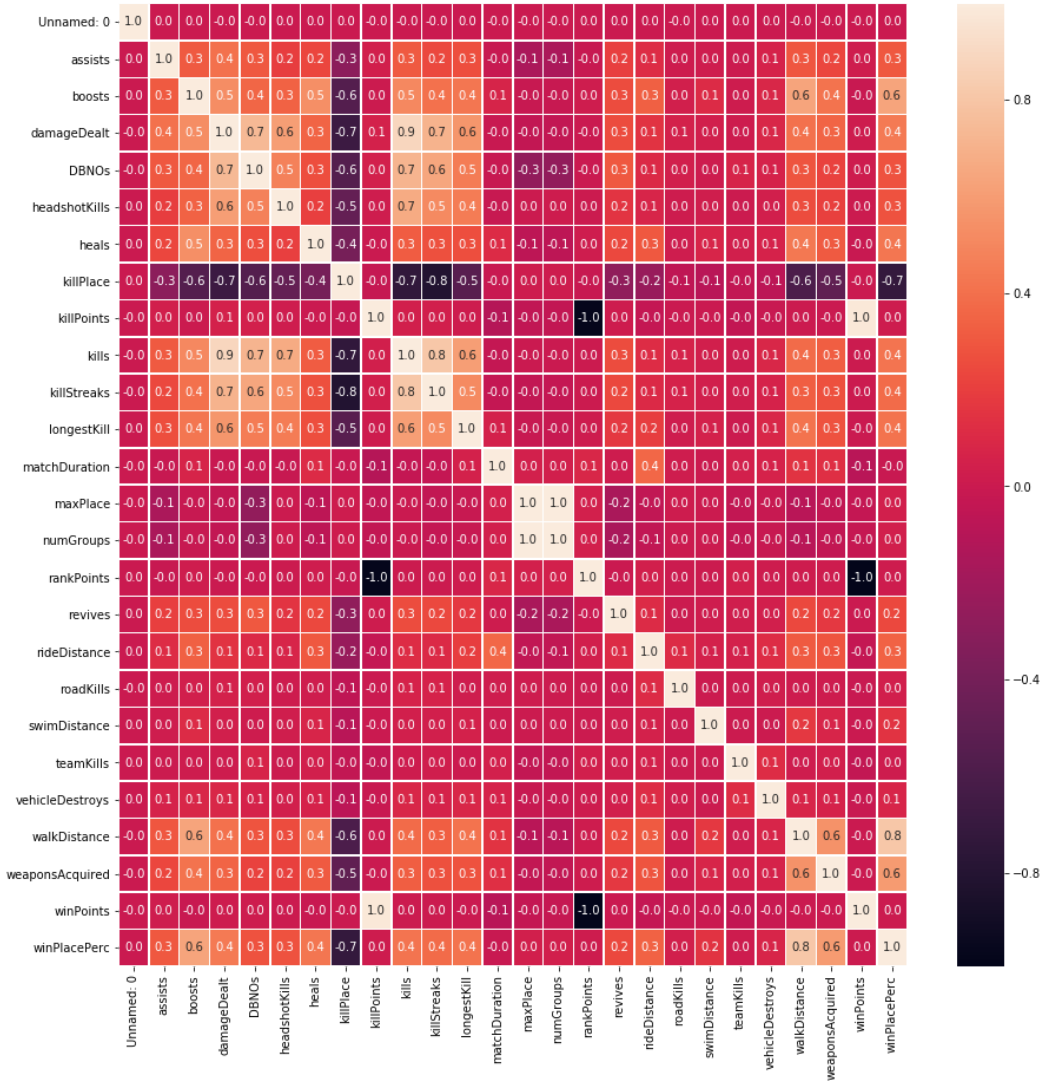
P. Analysis in different modes of play- Solos, Duos and Squads

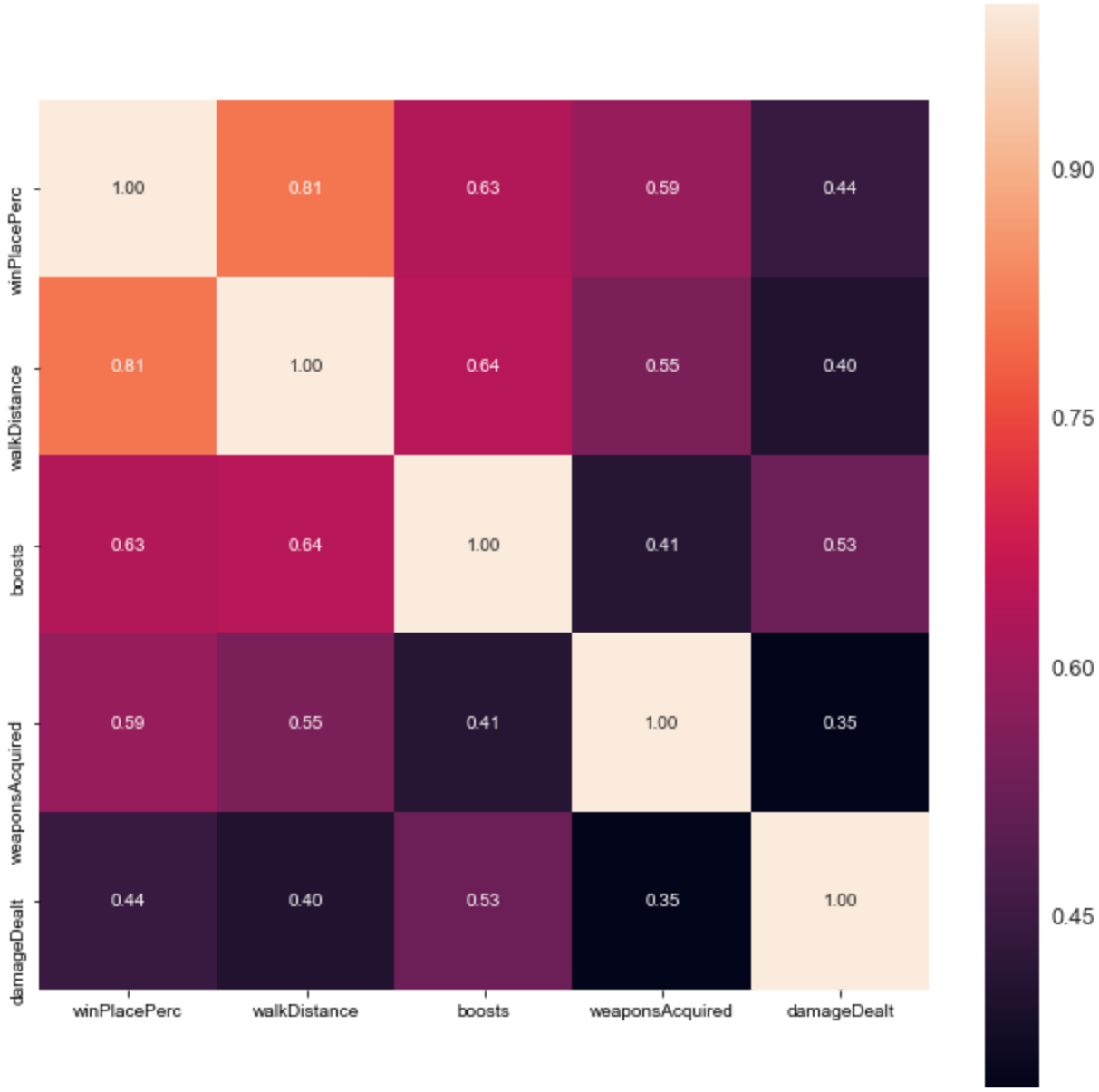


Q.



R. Finding correlation between various features-


Unnamed: 0
assists
boosts
damageDealt
DBNOs
headshotKills
heals
killPlace
killPoints
kills
killStreaks
longestKill
matchDuration
maxPlace
numGroups
rankPoints
revives
rideDistance
roadKills
swimDistance
teamKills
vehicleDestroys
walkDistance
weaponsAcquired
winPoints
winPlacePerc
0.8
0.4
0.0
−0.4
−0.8



3. Feature Engineering

It is one of the most important part in Machine Learning as it helps in finding out the best feature for prediction as well as polishing certain features. The various features that we have engineered are-

1. Normalising the columns based on the players joining the match.
2. Changing the datatype for reducing the memory usage.
3. Using SHAP to find the best features for Prediction and keeping only them for prediction.

4. Anomaly Detection

Here, we have found all the rows in the dataset that contains anomalous data. This is found by mapping the rules of the game with practicality.

5. Prediction with ML Algorithms

Here, we try to predict the winner's probability of winning on the scale of 0-1.
The ML algorithms used are-

a. Linear Regression-
b. Random Forest Regressor-
c. Multilayer Perceptron Regressor-
d. LightGBM

6. Luigi Pipeline-

This is for the automation of the complete Machine Learning process. This includes from cleaning the data, error handling, prediction and each and every process.

7. Flask-

We have created a Web application using Flask. It consists of the following-

1. The Exploratory Data Analysis
2. Taking user input and predicting the winner placement
3. Anomaly Detection Code

8. Containerize-

We created a docker image of the whole process uploaded it in the S3 Bucket.