

DATA ANALYTICS FOR E-GAMING

PROJECT REPORT



Table of Contents

1. Introduction:	2
1.1 About PUBG:	2
2. Abstract.....	3
3. Description of the Data Used:.....	3
3.1 Understanding the Data:	3
3.2 Cleansing of the dataset:	3
3.3 Exploring the dataset:	4
4. SCATTER PLOT MATRIX FOR SOLO, DUO and SQUAD.....	5
4.1 SOLO.....	5
4.2 DUO.....	6
4.3 SQUAD.....	7
5. Now to Answer the Question.....	10
5.1 Win Ratio vs Kill death Ratio:	10
5.2 Average Survival time vs Win ratio:	11
5.3 Rounds Played vs Win Ratio:.....	13
6. Conclusion.....	14
6.1 To summarize.....	15

1. Introduction:

Esports is the electronic sports which is industry built off video games. We can play both video games or professional games like football, Cricket and so on. Even these tournaments are loved by the teens and young adults to compete against other skilled players in whatever games they prefer.

1.1 About PUBG:

Player Unknown's Battlegrounds (PUBG) is published by PUBG Corporation a subsidiary of Bluehole, a South Korean video game company on 23rd March 2017, it is the multiplayer online battle royale game. The game is based on previous mods that were created by Brendan "PlayerUnknown" Greene for other games, inspired by the 2000 Japanese film Battle Royale, and expanded into a standalone game under Greene's creative direction. In the game, up to one hundred players parachute onto an island and scavenge for weapons and equipment to defeat others while stopping destroying themselves. Over time, the game's map's accessible secure area declines in size, guiding remaining players to push encounters to tighter areas. The round is won by the last remaining player or squad. Battle Royale games have seen a meteoric rise in popularity over the last few years and Player Unknown's Battlegrounds (PUBG) is still going big, the game that arguably kicked off the craze along with Fortnite's the mobile version of the game is free to play and is a huge hit with children and young adults, with more than 50 million active players every day worldwide.

2. Abstract

The dataset has the collection of statistics of each player in the Player Unknown Battleground (PUBG) game which is one amongst the battle royale games. The given dataset contains various records about the player's win ratio, kill death ratio, rounds played, time survived, and the details of all other statistics over a period in which the player played the game. To begin with I tried cleaning the dataset using various data cleansing techniques. After completely having a glance at the dataset I tried to answer the following question – *“What aspect in the dataset determines the win ratio of the player and how can he improve his win percentage?”*. Data cleansing is the basic and important step in data analysis, so after completing this process I decided to filter the dataset based on the field which are needed to draw a solution to the question. Since, the relationship among the data in the dataset is simple to identify. I considered “Kill ratio”, “Win ratio”, “average time survived” and “Round played” Data to answer the question on how to improve the Win ratio of the player. Using the above data, I came to conclusion on how to increase the win ratio of the player in each game type such as Solo (Single player), Duo (Multiplayer), Squad (4 play er). Using various plotting libraries in python I was able to visualize the results and conclude.

3. Description of the Data Used:

3.1 Understanding the Data:

Dataset used is collected by Kaggle via the PUBG_Player_Statistics.csv file. The dataset contains player data for 87900 players who have played the game in different game types. The files contain 152 columns and 87898 rows of data. PUBG game contains three types of game play's such as Solo (Single player), Duo (Multiplayer), Squad (4 player). Our dataset contains columns such as “solo_KillDeathRatio”, “solo_WinRatio”, “solo_AvgSurvivalTime” and “Solo_RoundsPlayed” and so on for each game plays.

After analysis the data, I am trying to answer the following question - *“What aspect in the dataset determines the win ratio of the player and how can he improve his win percentage?”*.

For this idea I have filtered the columns which will possibly provide better results.

3.2 Cleansing of the dataset:

- Not all the columns given the dataset will be useful to build. Having unnecessary data will lead to improper result and decrease the efficiency of the result. So, we dropped the unwanted column from the given data set.

- I went through the complete data set before loading the data set into our environment.
- I then filtered the dataset and create a new dataset with the columns which are required to determine the win ratio of the player
- Checking the d-types of each column and checking of null values and treating them accordingly
- Dropping the duplicated values to achieve improved results and replacing the values as required.

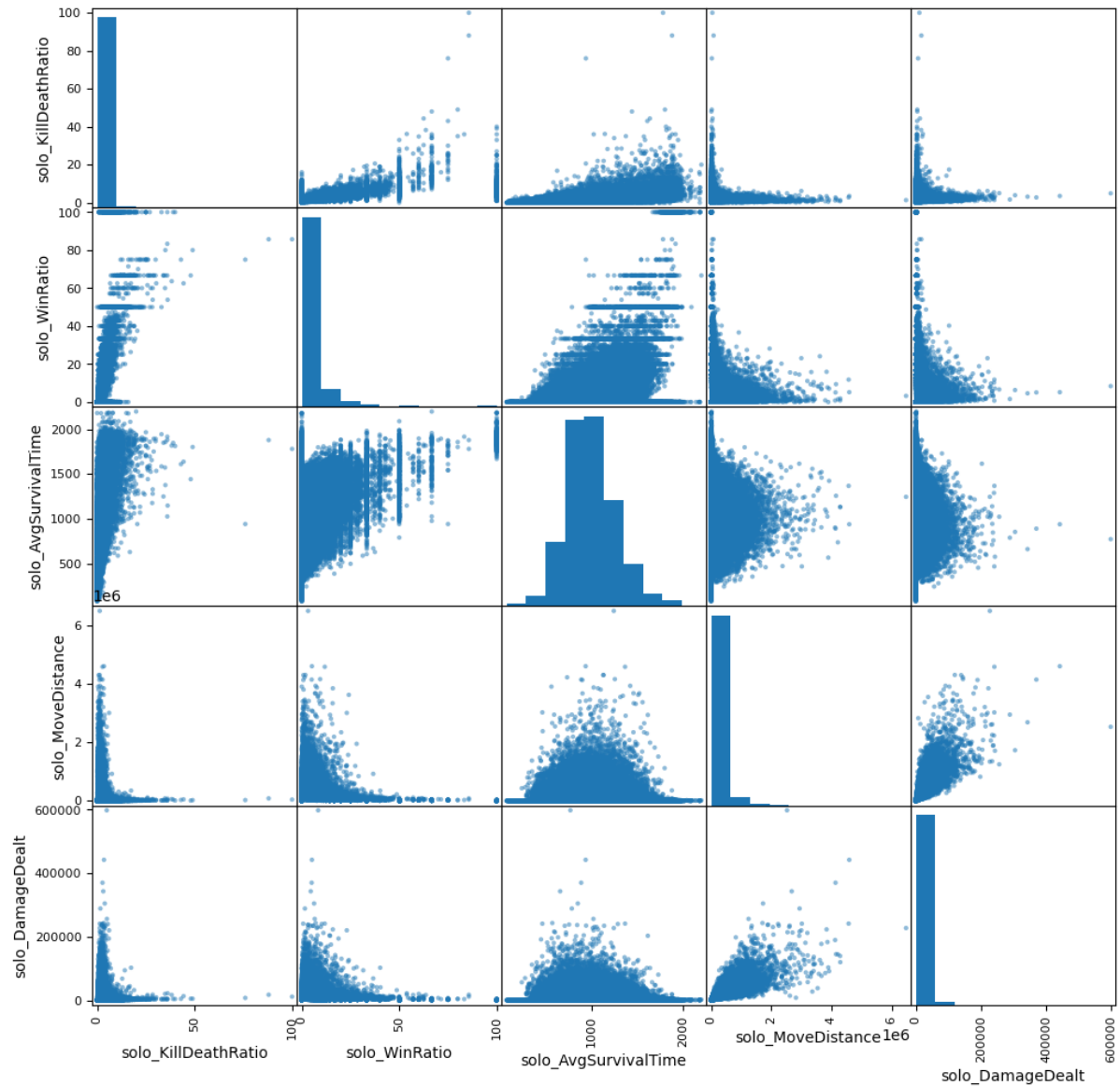
3.3 Exploring the dataset:

- After taking the required columns and cleansing the data, I decided to compare the Win ratio of each player with other categories in the game such as the Survival time, Rounds played, and the players kill death ratio. This helped us to provide user with a solution on how to increase their win ratio.
- We have taken the above-mentioned columns because in the overall game perspective these are the stats which define the overall performance of the player.
- The stats such as win ratio, time survived, rounds played and kill death ratio will enable me to provide a solution in such a way that it will benefit the new player and provide insights to the existing players in the game.
- I tried to keep the solution simple to provide better understanding and a better solution.

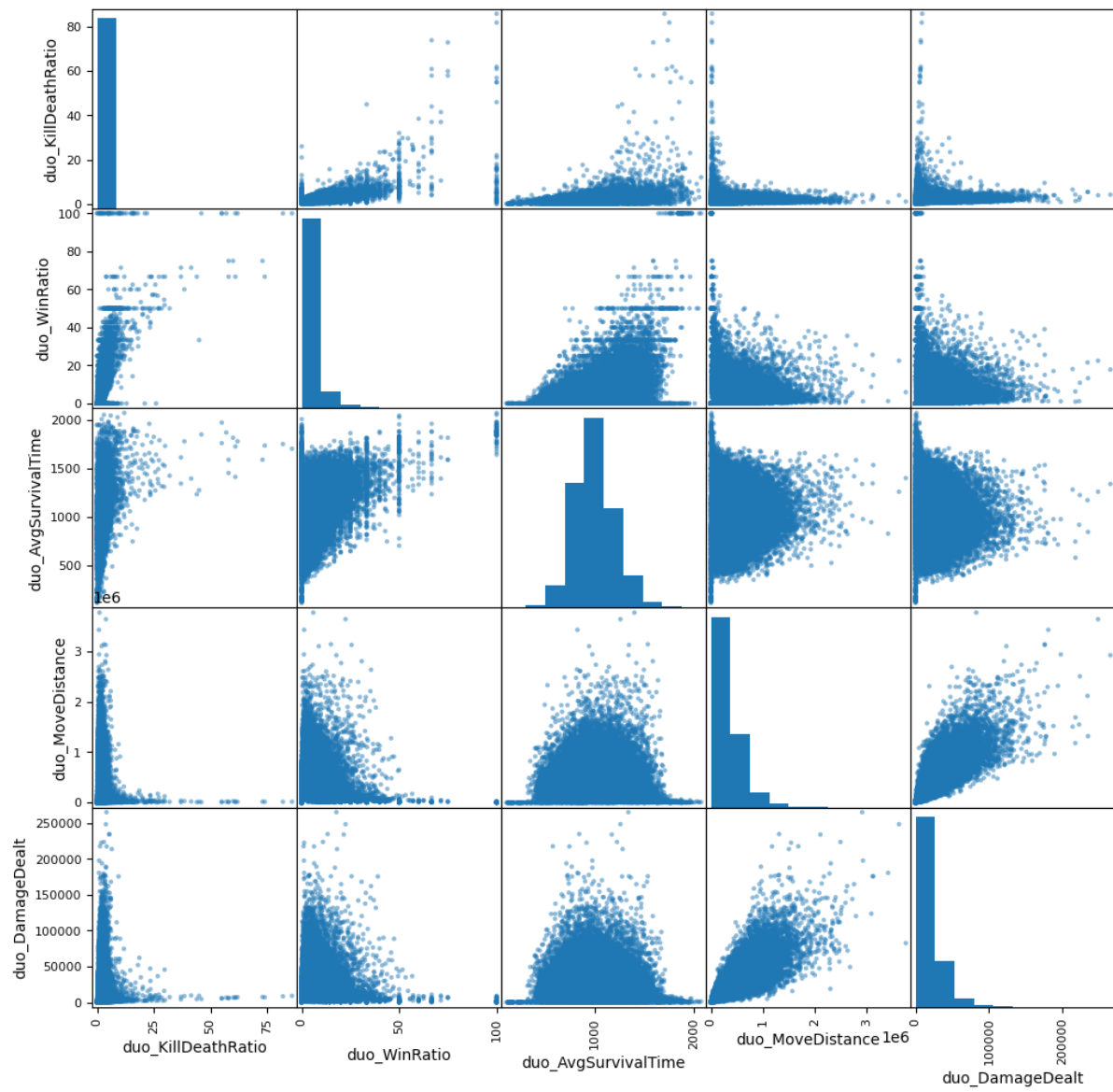
4. SCATTER PLOT MATRIX FOR SOLO, DUO and SQUAD

Using this matrix to determine which of the following have a higher correlation.

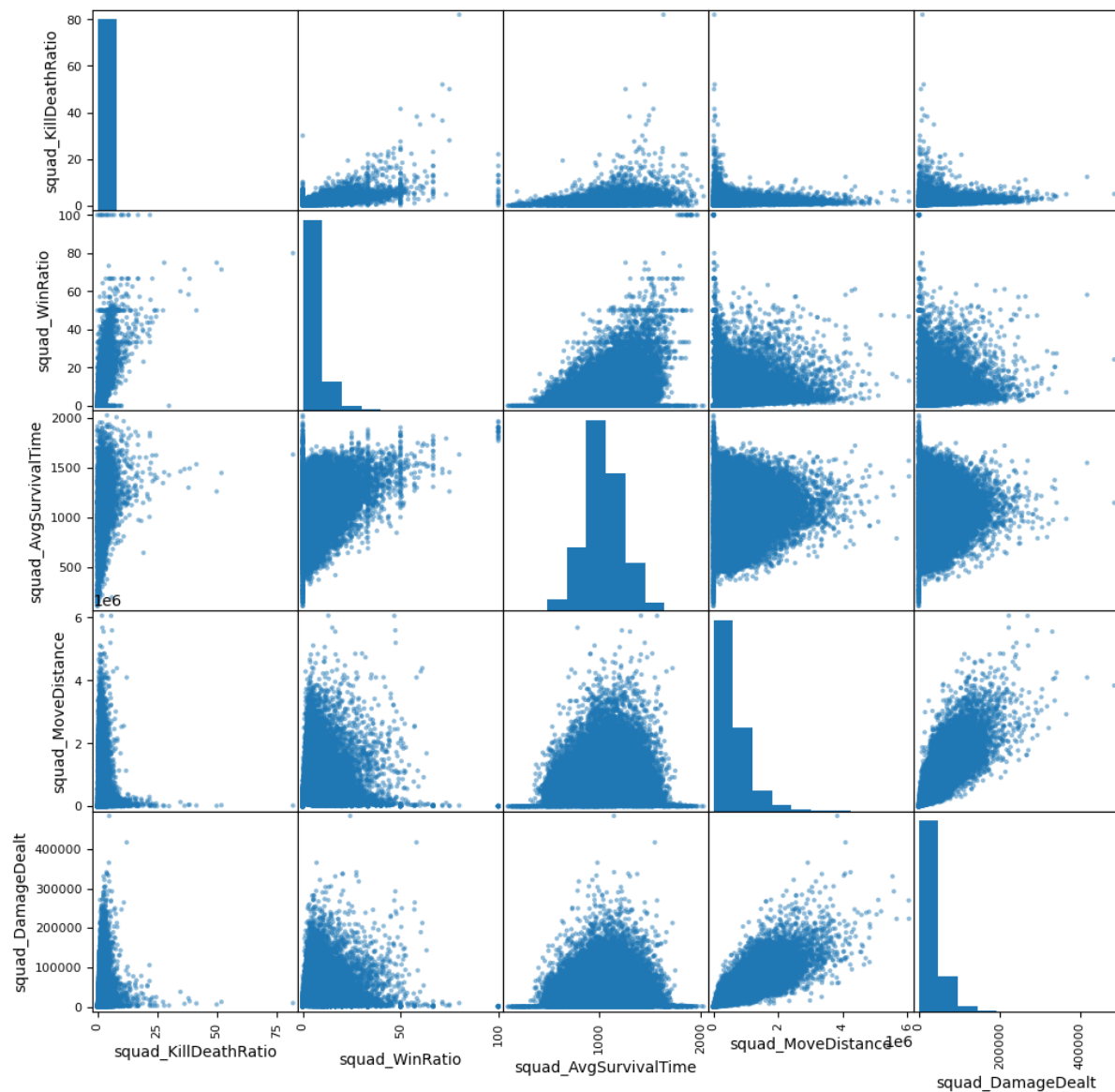
4.1 SOLO



4.2 DUO



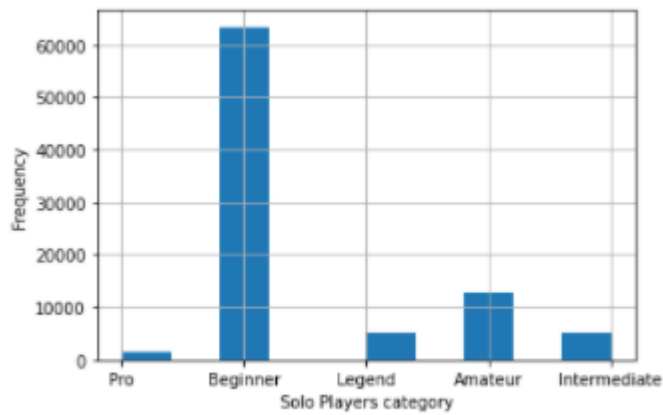
4.3 SQUAD



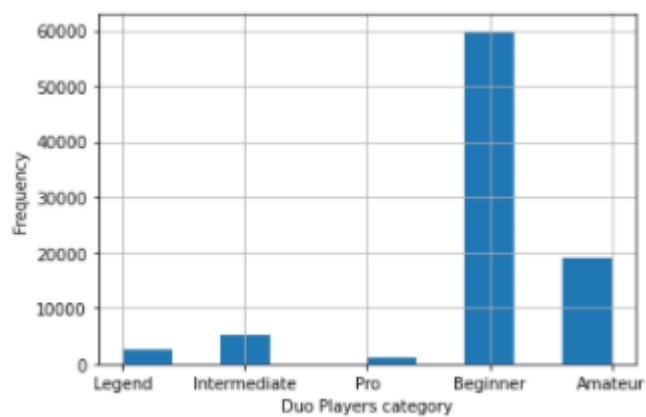
After Determining the relation between the data in the dataset, we selected the required columns which are necessary to answer the question appropriately.

Before Going into providing the solution I tried to rate the players based on the win ratio.

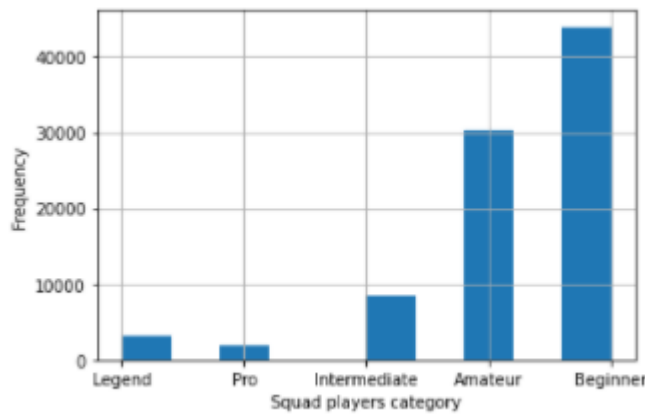
I Categorized players into levels such as Beginner, Amateur, intermediate, Pro and Legend for each game type such as solo, Duo and Squad.



SOLO Players Categories



DUO Players Categories:



SQUAD Players Categories:

I have also got the count of players based on their level in each game type.

	Solo_Categories	Duo_Categories	Squad_Categories	Players_count
0	Amateur	Amateur	Amateur	6252
1	Amateur	Amateur	Beginner	2694
2	Amateur	Amateur	Intermediate	2436
3	Amateur	Amateur	Legend	603
4	Amateur	Amateur	Pro	414
...
120	Pro	Pro	Amateur	39
121	Pro	Pro	Beginner	21
122	Pro	Pro	Intermediate	66
123	Pro	Pro	Legend	66
124	Pro	Pro	Pro	33

I have done to get an overall summary of the dataset and to give an overview of the player statistics in the game by categorizing them based on the level of play.

These level which I have categorized players into is based on their Win ratio. So, for example if a player has a win ratio less than 5, I have considered those players as beginner and so on for the remaining levels.

5. Now to Answer the Question

– “What aspect in the dataset determines the win ratio of the player and how can he improve his win percentage?”

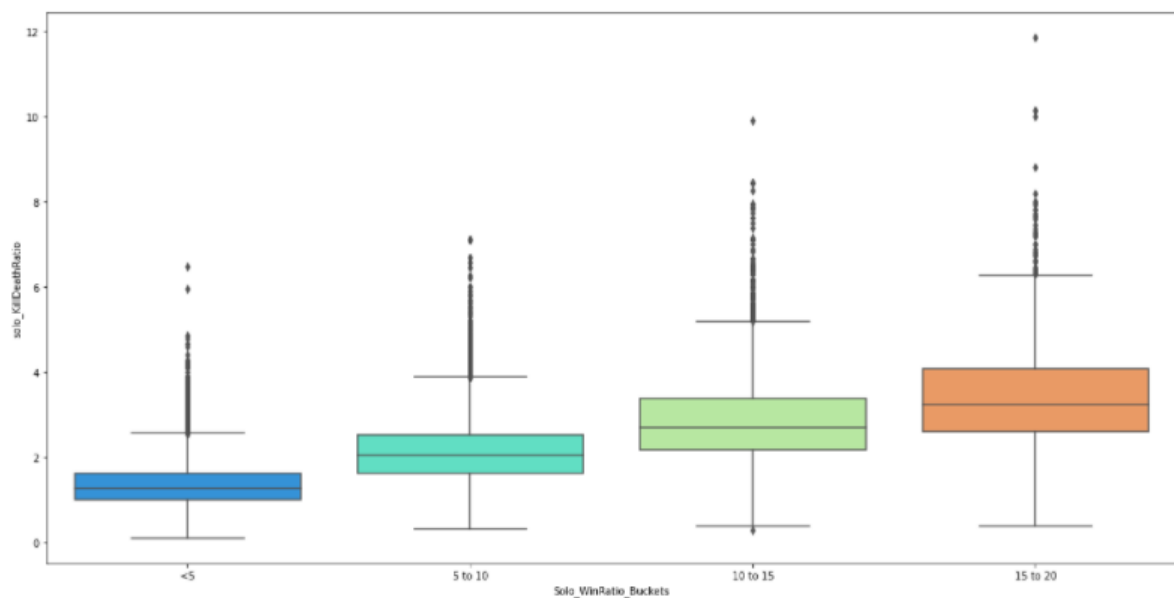
I have done the comparison of win ratio with other columns. Which I will explain one by one.

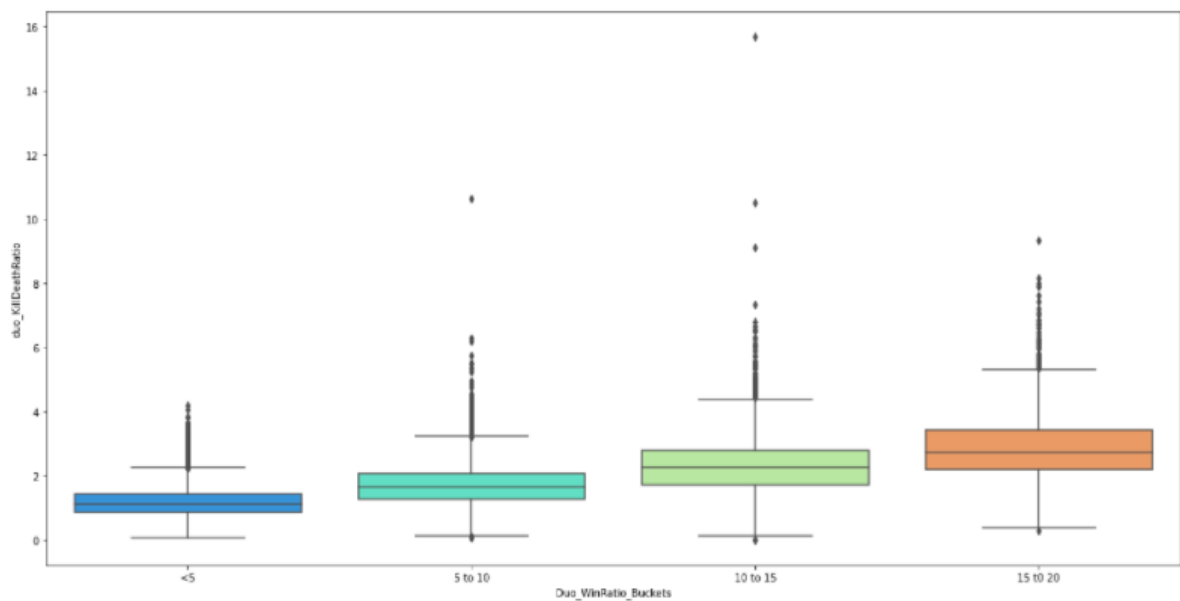
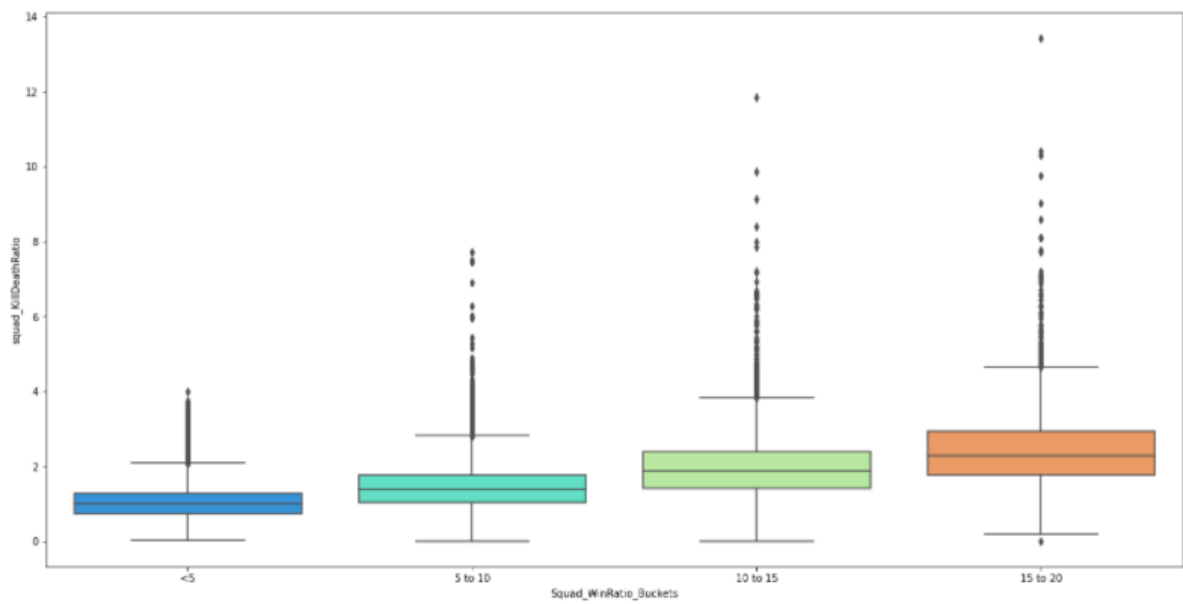
5.1 Win Ratio vs Kill death Ratio:

So, by comparing these two columns of data, I am trying to see whether the win ratio of a player increase when the player has more kill death ratio.

I have compared it for each game type – Solo, Duo and Squad.

```
import seaborn as sns
fig_dims = (20,10)
fig, ax = plt.subplots(figsize = fig_dims)
#sns.boxplot(x='solo_KillDeathRatio', y='solo_WinRatio', data =Solo,palette = 'rainbow')
sns.boxplot(x = "Solo_WinRatio_Buckets", y = "solo_KillDeathRatio", data=Solo, palette = 'rainbow')
plt.show() # to show graph
```





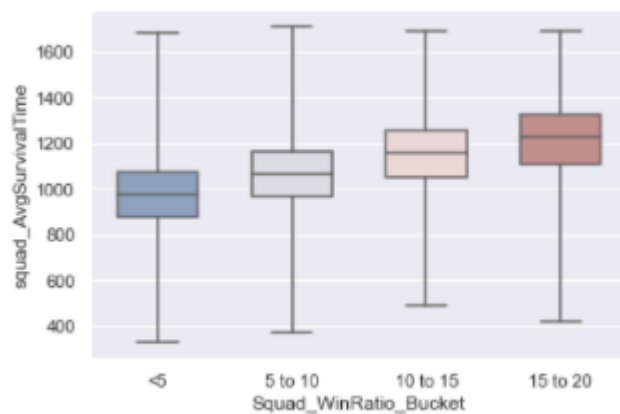
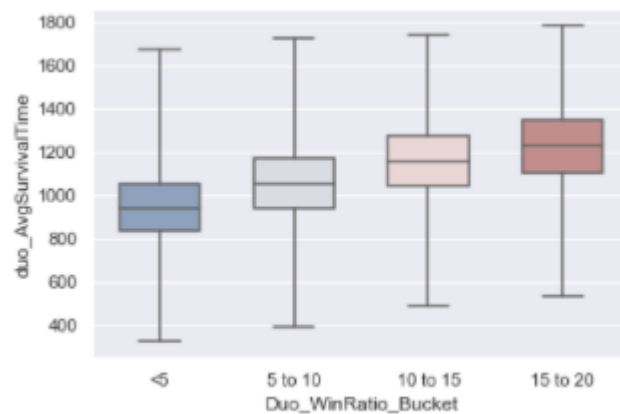
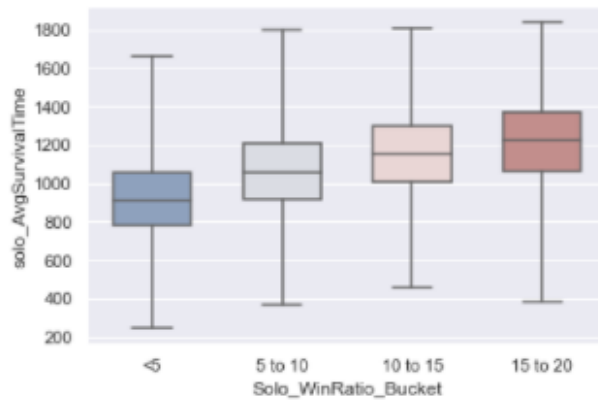
5.2 Average Survival time vs Win ratio:

Average survival time of a player is the average of the time which he/she survived in each game which they have played. So, using this data I am trying to determine whether the players who survived more time in the game have the possibility of winning the games and increasing their win ratio.

I have compared it for each game type – Solo, Duo and Squad.

```
sns.boxplot(x="Solo_WinRatio_Bucket", y="solo_AvgSurvivalTime", data=Solo_TimeandWin,
            whis=[0, 100], width=.6, palette="vlag")
```

```
<AxesSubplot:xlabel='Solo_WinRatio_Bucket', ylabel='solo_AvgSurvivalTime'>
```



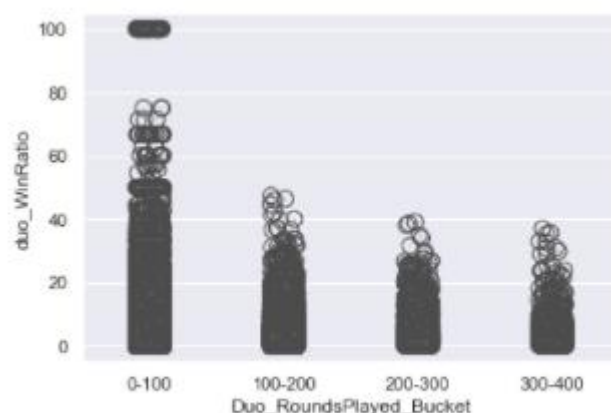
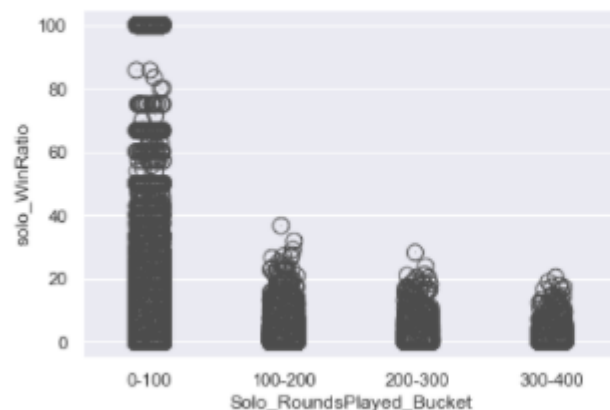
5.3 Rounds Played vs Win Ratio:

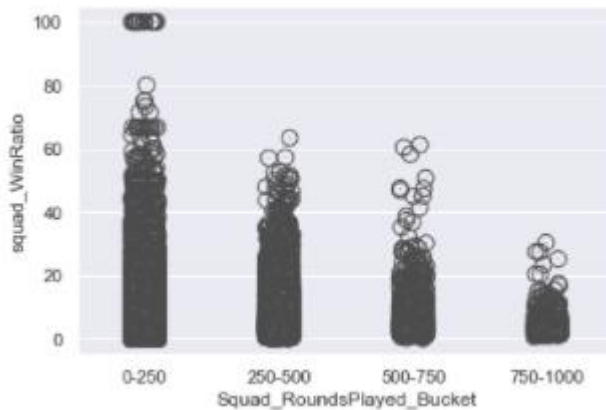
I have taken rounds played to determine whether the win ratio varies based on the number of games played. This was based on my experience when playing the game. As the level of player increases the number of games, he played come into factor more than we expect. The players who have played more games have more experience, but will this help them increase their win ratio?

I have compared it for each game type – Solo, Duo and Squad.

```
sns.stripplot(x="Solo_RoundsPlayed_Bucket", y="solo_WinRatio", data=Solo_RP,
              size=1, color=".5", linewidth=10)
```

```
<AxesSubplot:xlabel='Solo_RoundsPlayed_Bucket', ylabel='solo_WinRatio'>
```





6. Conclusion

So, to conclude my solution – I would say that to increase the player's win ratio there are many factors to consider. I have taken the rounds played, kill ratio of the player and, Average time survived by the player as these are the major stats in this Battle Royale Game.

When I compared the Win ratio with the players Kill death ratio, I concluded that Win ratio increases when players get more kills in the game.

After the above analysis, I then correlated the players win Ratio with Average survival time in the game. During this analysis I came to know that the more time the player survives in the game, the more chances are there for winning the game.

The above analysis was giving the results as expected. So, to answer my question more efficiently. I Compared the Win Ratio with the round played to check whether the Players win ratio increases when he/she plays more game. This analysis led to an interesting result, result was different compared to previous analysis because as the player plays more and more game the win ratio of the player is decreasing. Which I found as an interesting finding.

6.1 To summarize:

Question - *“What aspect in the dataset determines the win ratio of the player and how can he improve his win percentage?”*

Answer – *“The players Win ratio decrease as he plays more and more game as the level of difficulty in the game increases. So, my suggestion to the player will be to win as many as games possible within 200 games. As the probability of win between these games is more when compared to 250 or more game. My above analysis summarizes this.”*

