

## Project 1 | 13th November 2020

---

For any doubts regarding the assignment, ask questions in the [Data Science 101 B1](#) Group in the Community.

Submit Assignments by **16th November 2020 11:59 PM.**

Assignment Submit Form : <https://letsupgrade.in/ds101submission>

**Submit assignments in Appropriate Dropdowns.**

---



So, where are we droppin' boys and girls?

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.

The team at PUBG has made official game data available for the public to explore and scavenge outside of "The Blue Circle." This competition is not an official or affiliated PUBG site - Kaggle collected data made possible through the PUBG Developer API.

You are given over 65,000 games' worth of anonymized player data, split into training and testing sets, and asked to predict final placement from final in-game stats and initial player ratings.

What's the best strategy to win in PUBG? Should you sit in one spot and hide your way into victory, or do you need to be the top shot? Let's let the data do the talking!

Dataset's columns Description:

groupId - Integer ID to identify a group within a match. If the same group of players plays in different matches, they will have a different groupId each time.

matchId - Integer ID to identify match. There are no matches that are in both the training and testing set.

assists - Number of enemy players this player damaged that were killed by teammates.

boosts - Number of boost items used.

damageDealt - Total damage dealt. Note: Self inflicted damage is subtracted.

DBNOs - Number of enemy players knocked.

headshotKills - Number of enemy players killed with headshots.

heals - Number of healing items used.

killPlace - Ranking in match of number of enemy players killed.

killPoints - Kills-based external ranking of player. (Think of this as an Elo ranking where only kills matter.)

kills - Number of enemy players killed.

killStreaks - Max number of enemy players killed in a short amount of time.

longestKill - Longest distance between player and player killed at time of death. This may be misleading, as downing a - player and driving away may lead to a large longestKill stat.

maxPlace - Worst placement we have data for in the match. This may not match with numGroups, as sometimes the data skips over placements.

numGroups - Number of groups we have data for in the match.

revives - Number of times this player revived teammates.

rideDistance - Total distance traveled in vehicles measured in meters.

roadKills - Number of kills while in a vehicle.

swimDistance - Total distance traveled by swimming measured in meters.

teamKills - Number of times this player killed a teammate.

vehicleDestroys - Number of vehicles destroyed.

walkDistance - Total distance traveled on foot measured in meters.

weaponsAcquired - Number of weapons picked up.

winPoints - Win-based external ranking of player. (Think of this as an Elo ranking where only winning matters.)

winPlacePerc - The target of prediction. This is a percentile winning placement, where 1 corresponds to 1st place, and 0 corresponds to last place in the match. It is calculated off of maxPlace, not numGroups, so it is possible to have missing chunks in a match.

Tasks to be performed:

1. Read the dataset.
2. Check the datatype of all the columns.
3. Find the summary of all the numerical columns and write your findings about it.
4. The average person kills how many players?
5. 99% of people have how many kills?
6. The most kills ever recorded are how much?
7. Print all the columns of the dataframe.
8. Comment on distribution of the match's duration. Use seaborn.
9. Comment on distribution of the walk distance. Use seaborn.
10. Plot distribution of the match's duration vs walk distance one below the other.
11. Plot distribution of the match's duration vs walk distance side by side.
12. Pairplot the dataframe. Comment on kills vs damage dealt, Comment on maxPlace vs numGroups.
13. How many unique values are there in 'matchType' and what are their counts?
14. Plot a barplot of 'matchType' vs 'killPoints'. Write your inferences.
15. Plot a barplot of 'matchType' vs 'weaponsAcquired'. Write your inferences.
16. Find the Categorical columns.
17. Plot a boxplot of 'matchType' vs 'winPlacePerc'. Write your inferences.
18. Plot a boxplot of 'matchType' vs 'matchDuration'. Write your inferences.

19. Change the orientation of the above plot to horizontal.
20. Add a new column called 'KILL' which contains the sum of following columns viz. headshotKills, teamKills, roadKills.
21. Round off column 'winPlacePerc' to 2 decimals.
22. Take a sample of size 50 from the column damageDealt for 100 times and calculate its mean. Plot it on a histogram and comment on its distribution.

## FAQs

---

**Q. How to upload a jupyter notebook as a part of an assignment?**

- A. 1. Click "File" option in notebook
- 2. Go to "Download As" -> "Notebook(.ipynb)"
- 3. Upload the downloaded .ipynb file to github and share the link in google form.

**Q. When do I submit the Project and how?**

- A. The assignments for the week should be submitted by 16th November, Monday 11:59 PM IST.  
You can use Jupyter Notebook or python files or even Google Colab to Solve your Assignments

**Q. Where do I get class links for next session?**

- A. All sessions will be Live on our Youtube Channel. It will be also updated in the Community Group in the pinned post.

**Q. I have some doubt, whom do I ask?**

- A. Post your Queries on the community, someone will help you out.

**Q. Sir don't have anaconda so how can I solve the project?**

- A. Use Google Colab : [Click me](#)

**Q. Can we submit multiple .py or .ipynb project solution files for each question separately?**

- A. Solve all assignments for a day in a single notebook. Make sure you are submitting a single file.