

STAT 486 Report

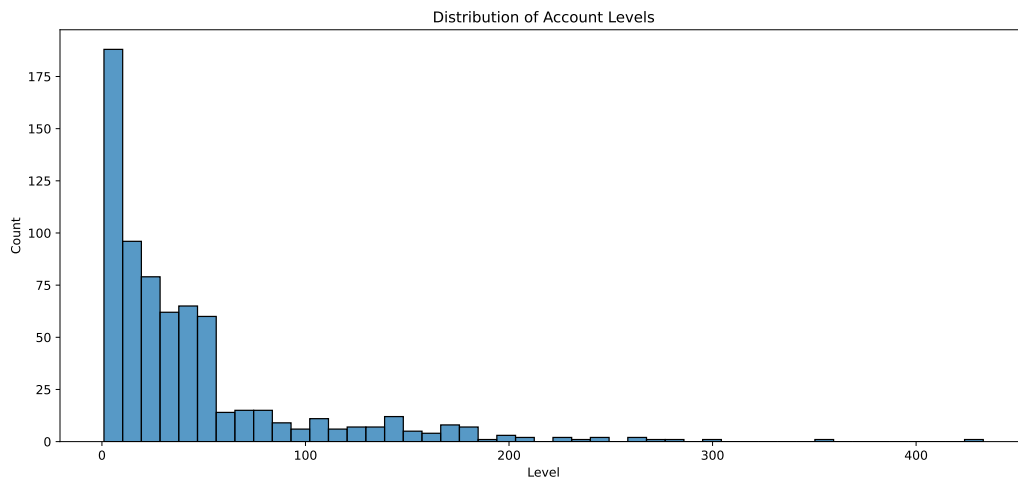
Justin Ross

2024-04-15

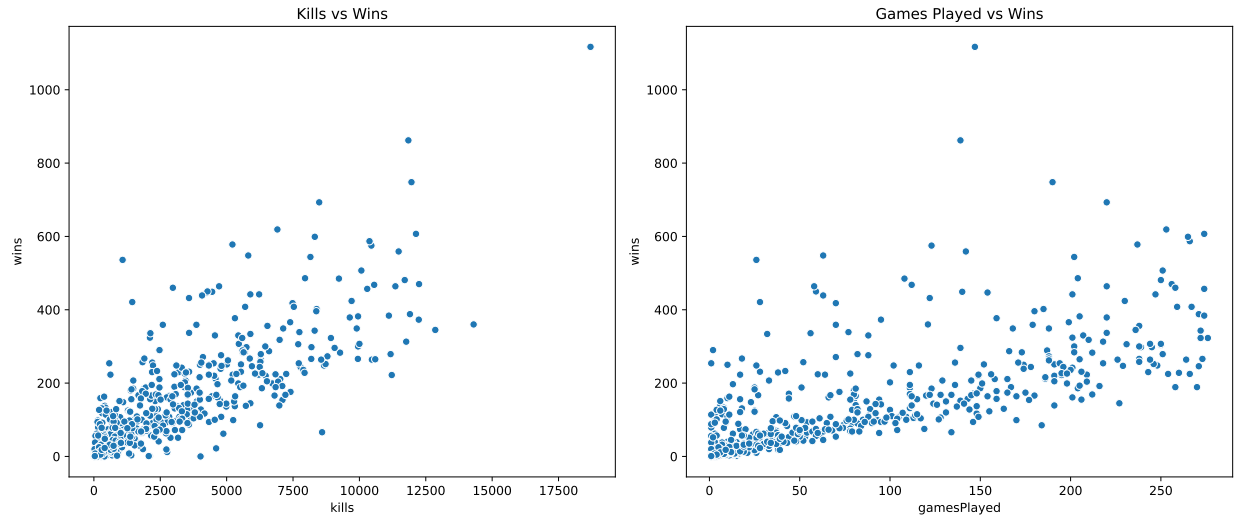
Introduction One of life's most desperately desired achievements is a win in Call of Duty. As such, our goal in this analysis is to determine the factors that contribute to success in online multiplayer matches. The data for our analysis was collected from Kaggle, and contains a number of variables relating to player performance.

Exploratory Data Analysis Our dataset originally included data on 1558 unique Call of Duty players. However, many of these players had not played any games and as such they were removed. We also removed significantly large outliers. As a result, our dataset now consists of 694 unique players.

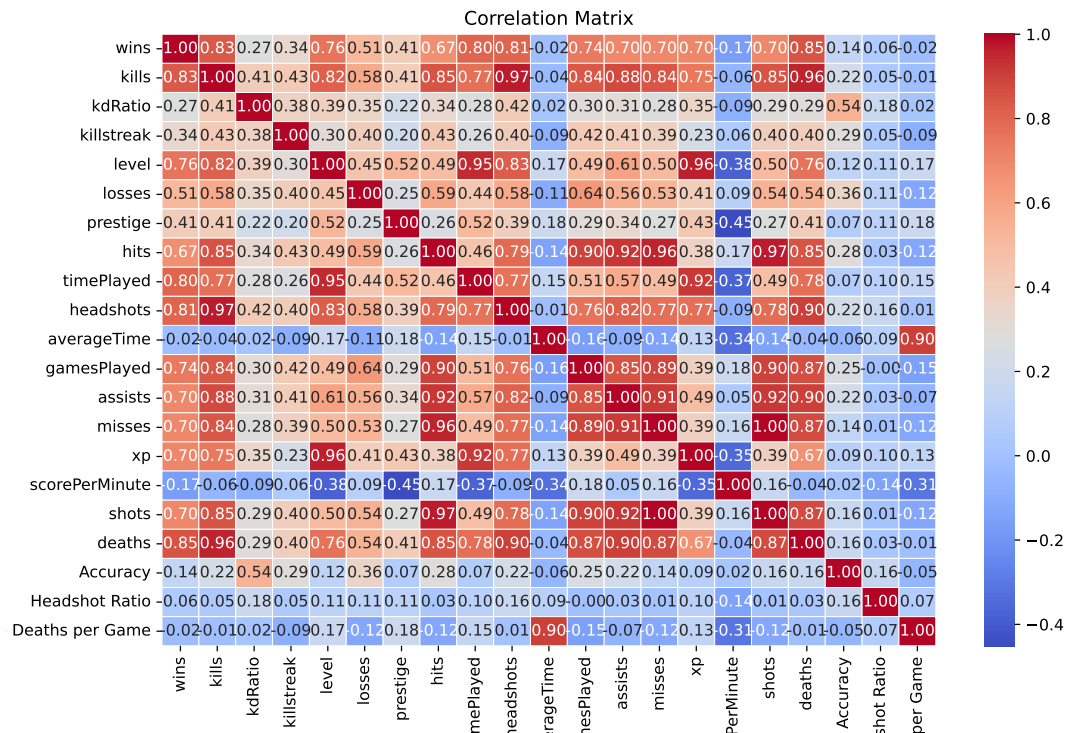
For the players in our analysis, the average number of online matches is 68, and the average number of total online wins is 116. The mean kill-death ratio is 0.838, suggesting that on average, players are killed more often than they kill their opponents. As the plot below demonstrates, most players have relatively young accounts with levels less than 50.



To better understand the relationships between the explanatory variables and the number of wins, we plotted some of the variables we thought would be most related to wins.



These plots demonstrate the high correlations between these variables and a player's number of wins. Kills has a correlation of 0.833 and games played has a correlation of 0.742. However, an issue we face is that there is correlation between some of our explanatory variables.



Due to the multicollinearity in the data, we have to be decisive in the model fitting process. We will have to be careful with linear models, as multicollinearity can cause a model to produce inaccurate model coefficients. To combat this, we will try a number of linear and non-linear models, as well as regularization and feature selection to create our best possible model.

Methods (depending on the structure of your project, only some of the following are applicable):

- Overview of feature engineering and how it improved your model (including any features from dimension reduction)
- Present a list or table of all models used. For each model, provide a very brief description (slightly longer if the model is not commonly known)

Model	Description	Hyperparameters
Linear Regression	This does Linear Regression and some cool other info goes here	hyperparameters
K-Nearest Neighbors	New predictions are the averaged target values of the k nearest neighbors in the training data	Number of neighbors, Weights
Lasso/Ridge	Linear models that penalize the size of the model coefficients (this helps combat multicollinearity)	α
SVM	Creates a margin around predicted values, trying to find an optimal balance between fitting the data and avoiding overfitting	C, Kernel, Gamma

Discussion on Model Selection: Without going into deep details, mention any patterns observed across models, such as:

- Were tree-based models consistently outperforming linear models?
- Did ensemble models show significant promise over single models?
- Talk about major pitfalls or challenges faced with certain models, e.g., overfitting with a deep neural network, or convergence issues with a certain algorithm. Briefly touch upon why certain models didn't make the cut. This doesn't have to be detailed but can include reasons like:
- Poor performance on validation data.
- Overfitting issues.
- Computationally too intensive for the marginal gain in accuracy.
- Difficulty in hyperparameter tuning.

Detailed Discussion on Best Model:

- Go in-depth into the model that performed the best
- Hyperparameter tuning
- Performance metrics
- SHAP and/or feature importance results including a few individual false positives/false negatives and true positives/true negatives (adapted to regression if applicable)
- Insights from cluster analysis or anomaly detection

Conclusion and Next Steps:

- Conclude by reinforcing the choice of the best model and its implications.
- Report or reinforce the answer/conclusion about your original problem statement
- Discuss any potential future steps or improvements that could be made, perhaps leveraging models or techniques not yet explored.