

Valorant Masters 2021 & Player Earnings Analyzation for Linear Regression Prediction

Dataset Introduction:

Valorant Masters is an international competition running through the year to determine the greatest team of players in the world. Competitors from the United States to South Korea face off online and in person through the brackets for a chance to perform in the grand finals held in a new location every year. For the purposes of this research, I focused on the 2021 Valorant Masters Tournament held in Berlin. Sixteen Teams qualified, and within our dataset found on kaggle (links to sets below), we had the stats of ten teams and their players. Obviously not the perfect spread, but the data was more than useful to look at insights and trends within the top players in the scene.

Initial exploration of the data concluded that a good amount of cleaning would be needed before any calculations, comparisons, or correlations could be performed. Agent composition is arguably the most important foundation to base your team's game plans and executions when competing. Our initial dataset had player's agents grouped together for all games essentially making any separation based on agent order useless. Before interacting with this data it would need to be properly organized by separating each agent into its own row and assigning it to the order in which they were played.

Dataset Cleaning:

First, a separate data frame was made for our agents row from the original, then using regex each agent was split upon the comma and separated into a new column(1-3). Some matches never made it to their third game, so the empty rows were filled with nulls so as to not disrupt data visualization in the future. Finally, a simple merge of the new columns and removal of the old allowed a much clearer look into the match history of players.

Here, we may now begin a dive into the visualization and trends of the data. Our main data frame was accompanied by two adjoining data frames dictating wins, losses, picks, and bans of the maps played in the tournament. Unfortunately, the data was not immediately useful to any hypotheses towards the gameplay of our teams, but by stacking the picks and bans of the maps we were able to see favored maps overall, favored map by team, the sides they performed best on, and maps that were likely to be avoided. If we had more data on how each team and players performed throughout the game we might be able to source information about why each chose their composition and gameplans based on that. With the data I had however I would need to look for a target variable to base the correlation and output of our player's stats.

Data Merging:

Our second main data set was also found on kaggle (linked below), containing tournament placements and earning estimations of professional players. As luck would have it, a large portion of the dataset matched with the competitors of the 2021 Valorant Masters dataset. I wanted to merge the two, but I did not want to ruin the integrity of either data set. By copying

our earnings dataset, I could filter through this version by matching the player names from each set and only keeping ones that appeared in both, creating a new dataframe from the original and only keeping the players involved in the tournament along with their stats. Now, we were almost ready to run our data through a linear regression model.

Machine Learning Preparation:

Before splitting our data into training and testing, we still had a few columns not fit to be run through. Most of the columns that had come from the second data frame had been stored as strings, including any numeric ones. A quick check into the data types and changing them back into integers made them accessible for our future work. Now, the data had been neatly organized, but it was still not ready for our model. Agents, teams, and countries were all categorical objects and not usable within our fit. Dummies for each variable would need to be made in order for the model to accept the inputs. After replacing our columns with the new dummies, the data was finally acceptable for linear regression input. I had no intention of blindly inputting data and hoping that it worked though, I wanted to see the correlation of my variables individually and how much influence they had on the final product. Understanding what made the players great was just as important as how much they could profit off that skill.

Dashboarding and Data Visualization:

Through Tableau, I manufactured a dashboard for complete story compatibility to my intention and purpose of the project. I wanted to make sure any participants of this data had a complete understanding of where that data came from, how it impacted the tournament, and the sheer diversity of our teams. Each came from all parts of the globe with different play styles, games mechanics, agents, and more. Highlighting these factors in an easy to use dashboard grants users accessibility to the work and how it might stack up globally. Back in python, I also implemented a heat map and pair plot, easily visible methods of finding correlation. Ensuring the dashboard was not too crowded, I wanted only the most important variables to appear.

The dashboard can be found linked to this summary. It includes individual player stats like their KDA, placements, agents, country of origin, links to tournament gameplay and more. The dashboard comes with an instructional video as well, but hopefully the annotations and interactivity are more than enough for a complete understanding of the data's process and placings.

Linear Regression Implementation:

Finally, it has come to the implementation of our Multiple Linear Regression Model. I wanted to create a prediction model to label an estimated earnings of professional players where our second (earnings) data frame had fallen short. With our newly merged and manipulated data frame, we could easily import our train-test split method to prepare our data for fitting into the model. I chose to use an 80/20 train test split, and after separating our dependent and independent variables, ran our X & Y values through. Importing from `sklearn.linear_model`, I initialized the regression model and fit our data in.

Our training model did very well scoring nearly perfect, but unfortunately the testing data fell flat on its face only predicting correctly just under 40% of our data. Perhaps the amount of

dummy variables, or low correlation variables might have thrown off our model and created an overfitting problem. If I were to continue this project I might run a Lasso/Diamond method to only keep the best variables, or even group in clusters to find players that are the most similar and range their earnings.

Summary:

The fitted model was now in a usable state to predict any earnings for unknown players. From our masters data set, I filtered through once again, but this time keeping players that lacked an earnings column. Our new data frame now contained all of our players lacking an earnings column. Not forgetting our dummy variables I once again changed our categorical variables, and they were ready for prediction. All I have to do from here is convert my data frame into a dictionary, and finally call the index of whichever player I want to estimate, drop their name, and save to a converted list. A complete dictionary is added near the bottom of the workbook, and the player Effys was run through the model to predict their earnings for proof of concept. A continuation of this project might be running each player through our model iteratively, and adding their earnings into a new complete data frame for future use of other developers.

The purpose of this project was a deeper understanding into the professional scene of valorant, what made up the factors of a high performing player, and to use this data to create a model for earning predictions. All portfolio project steps are connected to this summary for replicability and viewing. Hopefully, through this analysis users have a clearer understanding of the top players and teams, and how individuals might impact the revenue flow of a team's enterprise.

Bibliography:

<https://www.kaggle.com/datasets/jaykumar1607/valorant-masters-2021-player-map-data?resource=download>

<https://www.kaggle.com/datasets/evangower/valorant-esports-top-earnings/code?select=Teams.csv>