League of Legends Artificial Intelligence Coaching

Kadin McWilliams

kadinmcwilliams@gmail.com

2/3/2024

Introduction

This data set is challenger League of Legends games spanning 10,957 rows and 138 columns. This data set contains key player statistics that measure player performances in games. I created this data set by creating a data gathering python script that pulled solo queue ranked games from the official league of legends server. The data set can be found on my GitHub profile under DS_Final_Project_KCM. I chose this data set due to my extensive knowledge and experience with league of legends. I figured my long-time experience with the game would allow me to make more meaningful insights into the data set that I am researching. Key data findings such as potential important performance factors and player performance trends can be found inside of the report.

Data Set Description

The dataset consists of 10,957 rows and 138 columns, containing statistics about players' game performance. The data includes a variety of data types, with the majority being integer values (86 columns), followed by a smaller number of boolean (10 columns), object (41 columns), and float (1 column) data types. There are a few missing values in the dataset. The Summoner Name column is missing 972 values, which are not critical for understanding the gameplay or performance of the players. This field can be disregarded when focusing on the analysis. Additionally, the Team Position column is missing 2 values, but these are easily inferable from other columns that specify the player's lane. Thus, the missing data does not pose a significant issue for the overall analysis. In terms of data quality, the dataset is relatively clean with a minimal amount of missing data, making it suitable for analyzing trends, player behavior, and performance metrics across multiple games and variables. A full data dictionary is provided to understand each column. One row represents the statistics of a player in one match. To access the dictionary go to this link.

Data Set Summary Statistics

Looking at individual player performance statistics, there were a few noteworthy findings in the data set. Looking at the kills, players average around 5.6 kills per game, with a relatively wide spread of values (standard deviation of 4.29). In comparison, deaths are slightly lower, with an average of 4.87, and a smaller variation (standard deviation of 2.95). Assists, on the other hand, average at 8.2, but show the highest degree of variation, with a standard deviation of 5.91, this could be to because multiple people can be credited with an assist for one kill. The average game duration is 26.5 minutes, though the median game length is slightly longer at 27.18 minutes, with a standard deviation of 6.87 minutes. This indicates that while game times are fairly consistent, there is some variation. When examining turret takedowns, players typically destroy about 2 turrets per game, but the variation (standard deviation of 2.15) suggests that some matches have much higher or lower numbers. Similarly, gold earned averages at 10,741, with a relatively high standard deviation of 3,669, showing that gold income can vary significantly across players and matches.

Looking at team objectives, baron kills occur on average, 0.31 times per game, while dragon kills are more frequent, with an average of 1.79 per game. These averages set a realistic goal for game performances. From the player's perspective, baron kills, and dragon kills occur much less frequently: 0.07 and 0.37 times on average, respectively. This could be because mainly jungle players smite objectives.

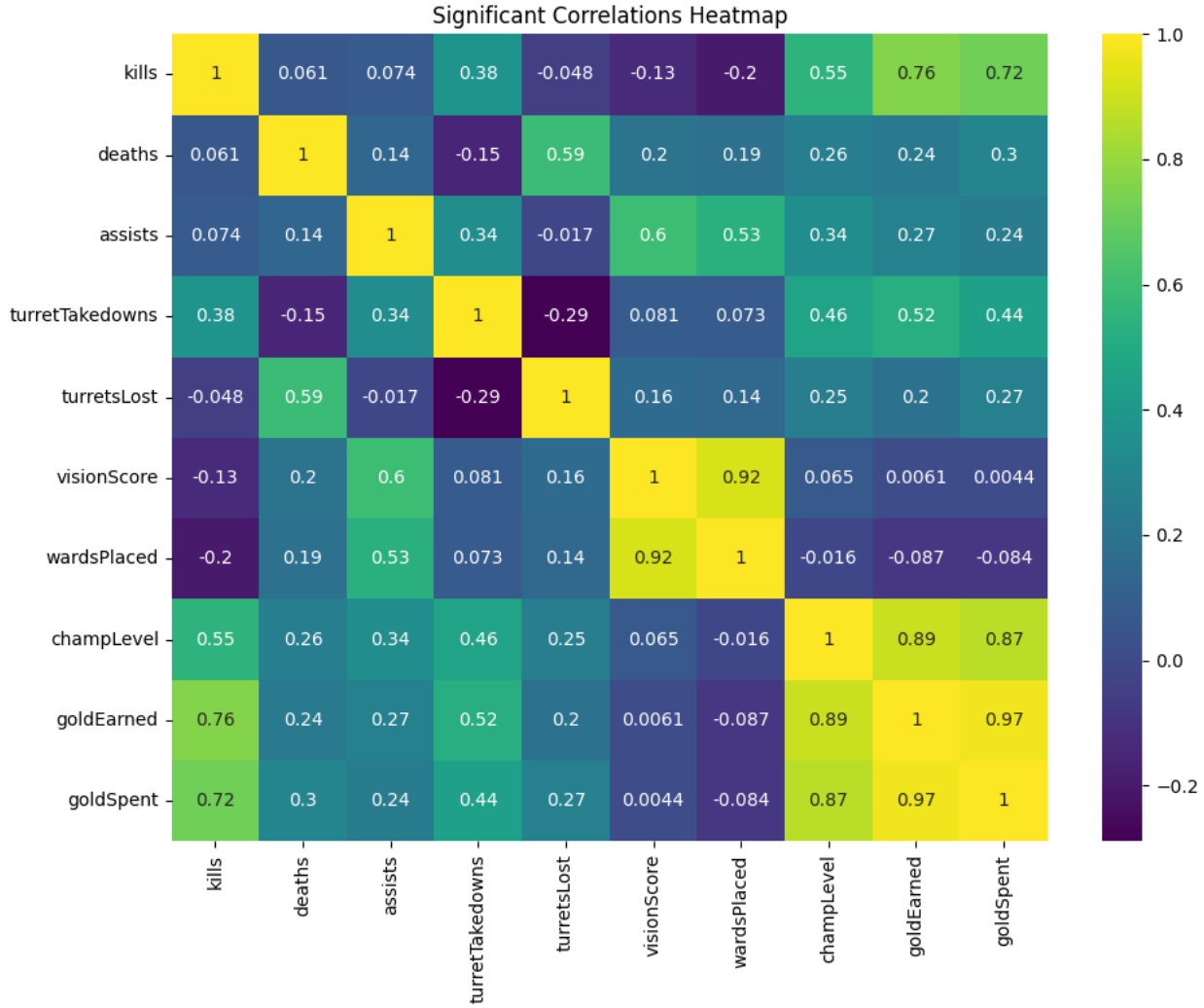
Several interesting correlations emerged from the data. There is a moderate correlation between turrets lost and deaths, as well as between time spent dead and percentage time dead, reflecting how players' performance in taking or losing turrets may be influenced by their ability to stay alive in a game. Additionally, assists show moderate correlations with vision score and wards placed, which is consistent with the role of supports in the game who focus on vision control and leaving the kills for other players, mainly getting assists. Furthermore, player kills and turret takedowns also show a moderate positive correlation, implying that players that are successful in securing kills are likely to also take down more turrets. A particularly strong correlation exists between gold earned and gold spent, especially in relation to kills. This suggests that players who earn more gold are also likely to spend it effectively, and their gold usage is tightly linked to their kill performance.

Interestingly, there are no significant outliers in categories such as CS, ward placements, dragon kills, and other major objectives like rift herald kills or turret losses, which are key performance indicators. However, neutral minions killed, and dragon kills show larger numbers of outliers, possibly indicating extreme performances in certain games. A potential explanation for these numbers could be the jungle player's role in securing these objectives and taking that credit for securing them. One surprising finding is the number of baron kills, which have

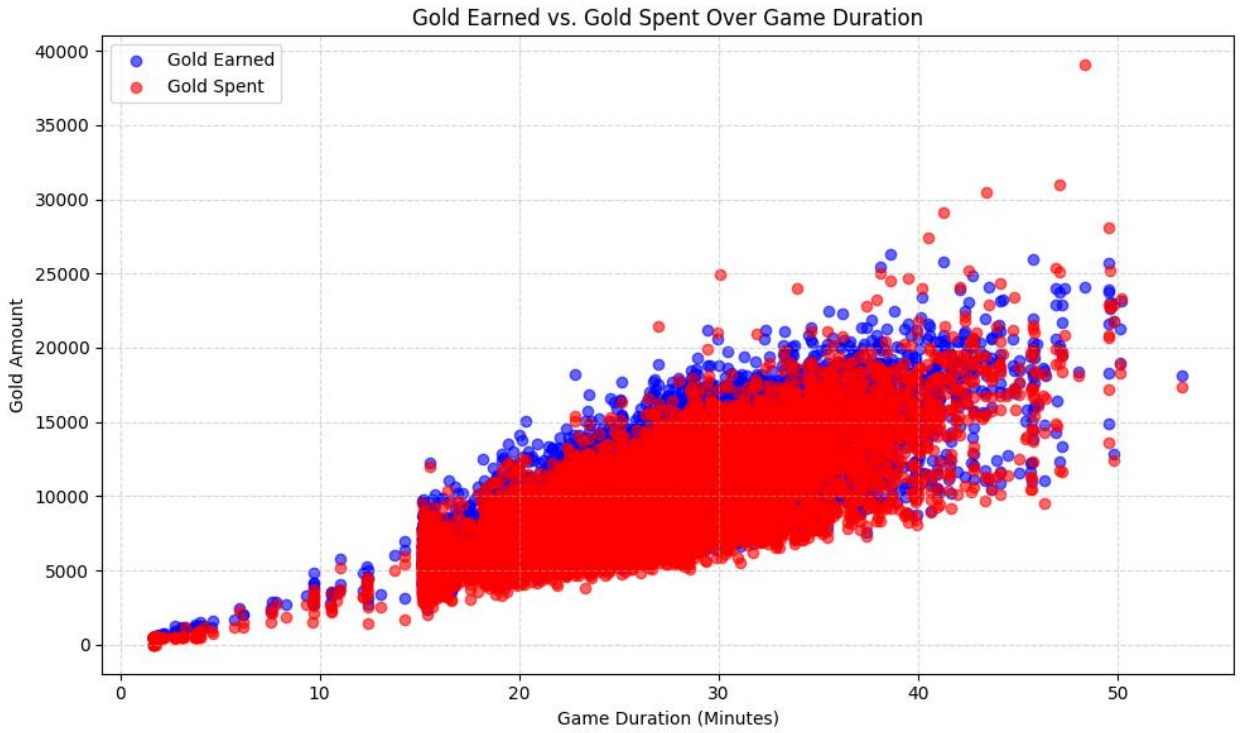
around 600 outliers, much higher than expected for individual performance, particularly since these kills are often a team effort. When it comes to pings, the enemy missing ping showed fewer outliers compared to vision pings, likely because vision pings are more commonly used for strategic purposes during gameplay. Other ping types, such as alert pings, displayed larger outliers.

Overall, the data reveals notable trends and correlations that reflect the dynamics of team strategy, player performance, and the impact of objectives like turrets and dragons. The presence of outliers in certain categories, such as gold spent and turret takedowns, points to exceptional gameplay moments that may deserve closer analysis. The relatively low number of missing values and the consistency in certain categories make this dataset well-suited for further in-depth analysis and decision-making regarding gameplay strategies.

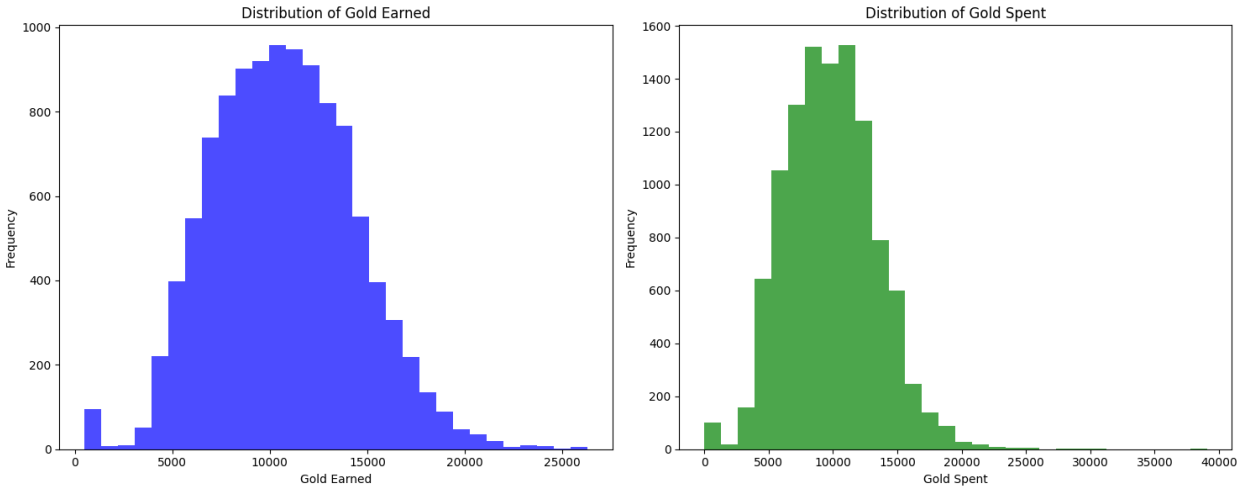
Data Set Graphical Exploration



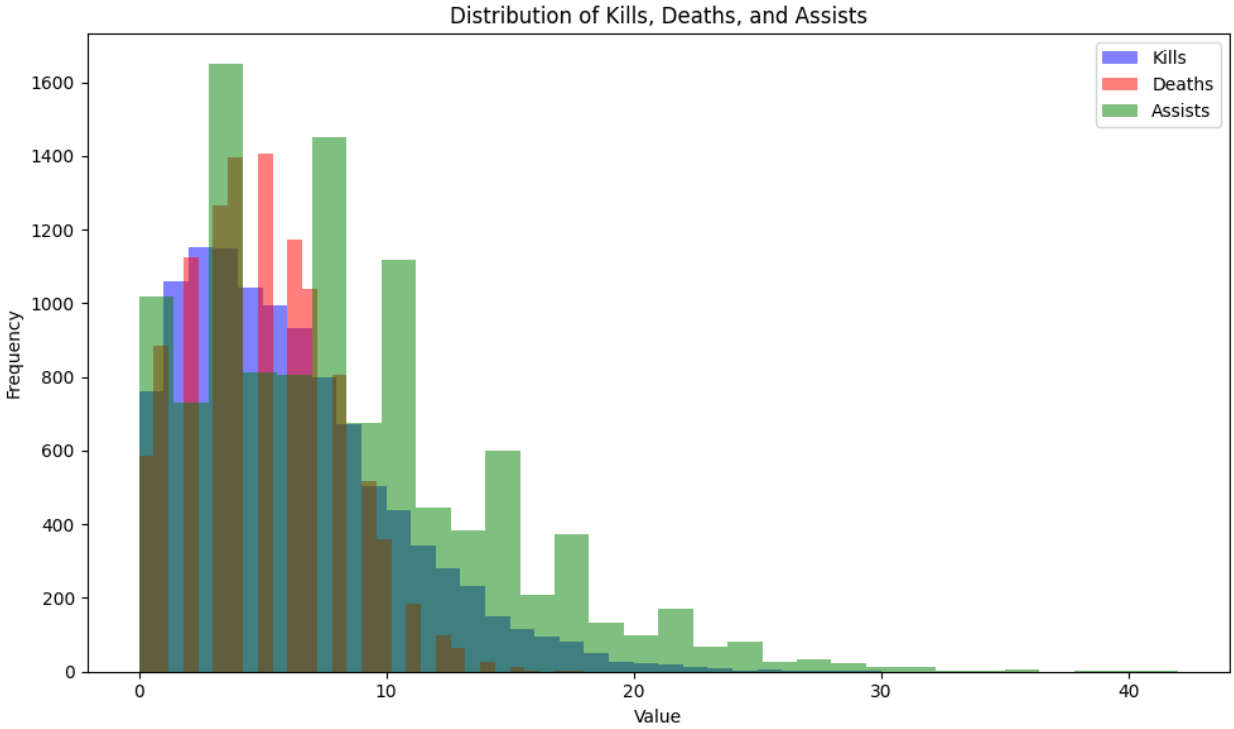
This heat map visually displays the trends and correlations mentioned earlier in the analysis. It shows how key performance metrics, such as kills, deaths, assists, gold earned, and gold spent, are related to team objectives like turret takedowns and dragon kills. Specifically, it highlights the moderate correlation between kills and turret takedowns, as well as between assists and vision score. The heat map makes it easier to see these relationships and provides a clear overview of the trends that affect player performance and the overall dynamics of the game.



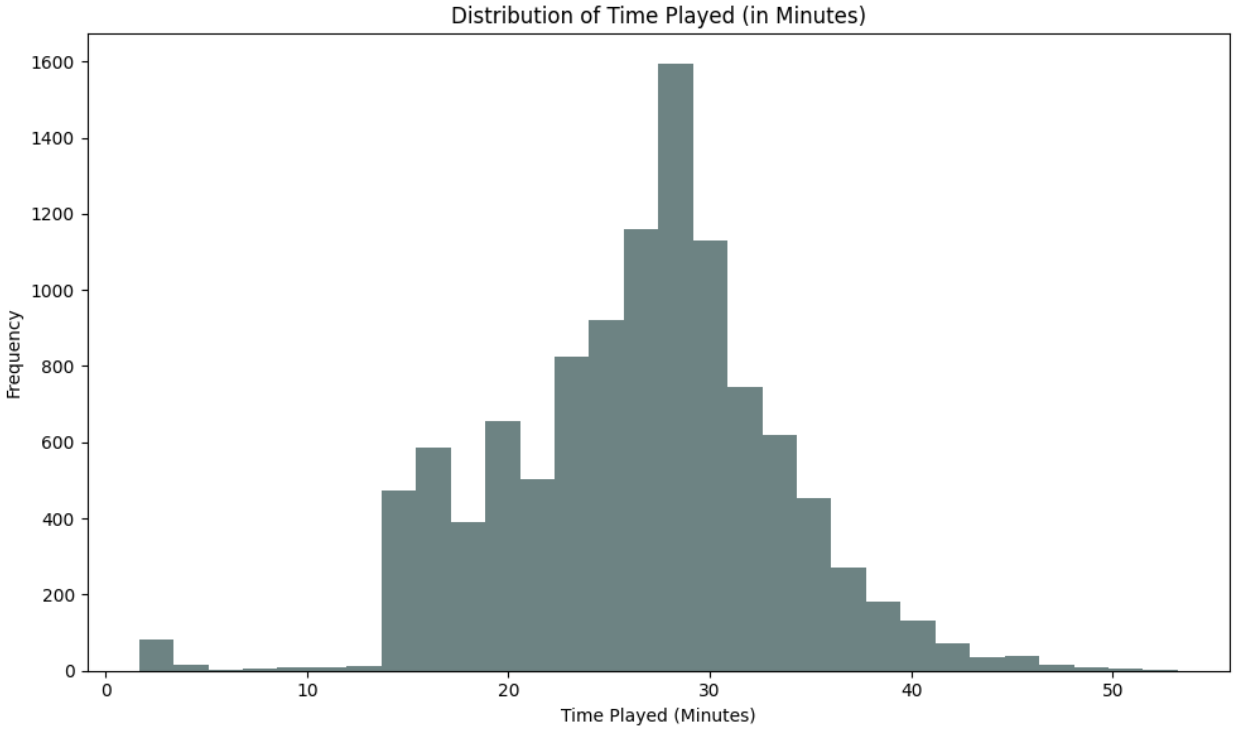
This scatter plot highlights the trend between gold spent and gold earned by the player. You can see that the two closely follow each other. I believe the reason gold earned is higher than gold spent is due to the amount of gold you receive from ending then game (kills, turret takedowns, etc.).



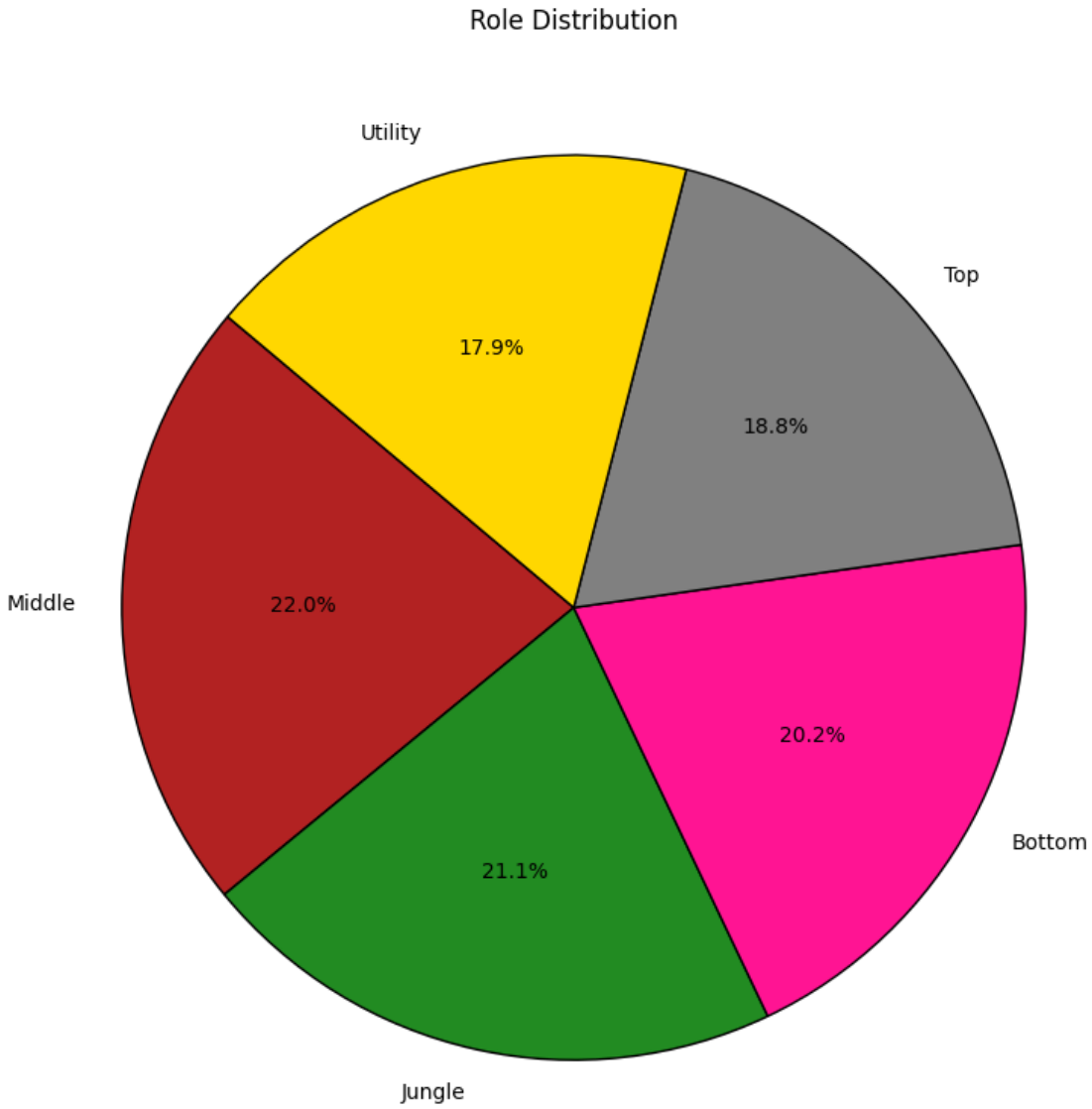
The histograms for both are interesting in that they are both normally distributed. However, the amount of gold spent has a much smaller variation. This means that most players are going to be spending the same amount of gold in their games. Potentially hinting at efficient item building to be a winning strategy.



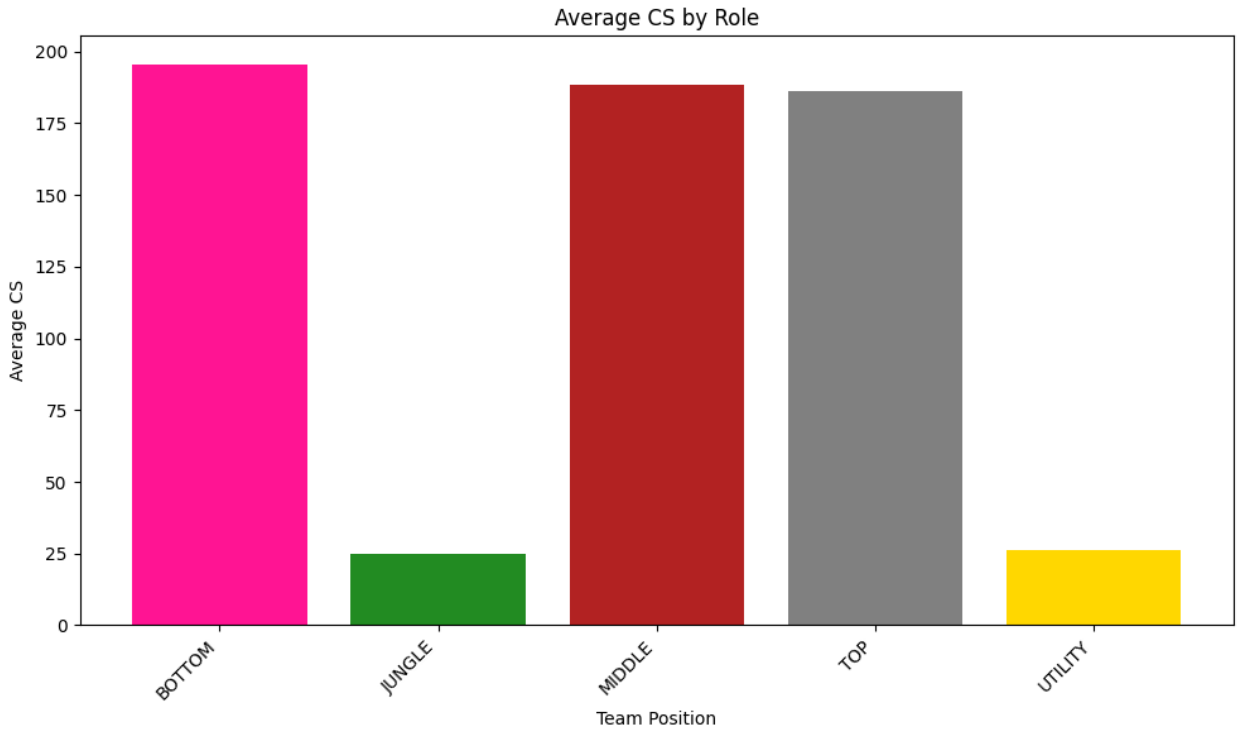
Above is a histogram of kills, deaths and assists in each of the games. All three of the variables appear to be right skewed with high numbers in each of the categories to be an outlier. Another important finding from this graph is that there are gaps in the number of deaths that players have. This could mean that players tend to either die a lot or not die much. This could mean that deaths could have a huge impact on performance.



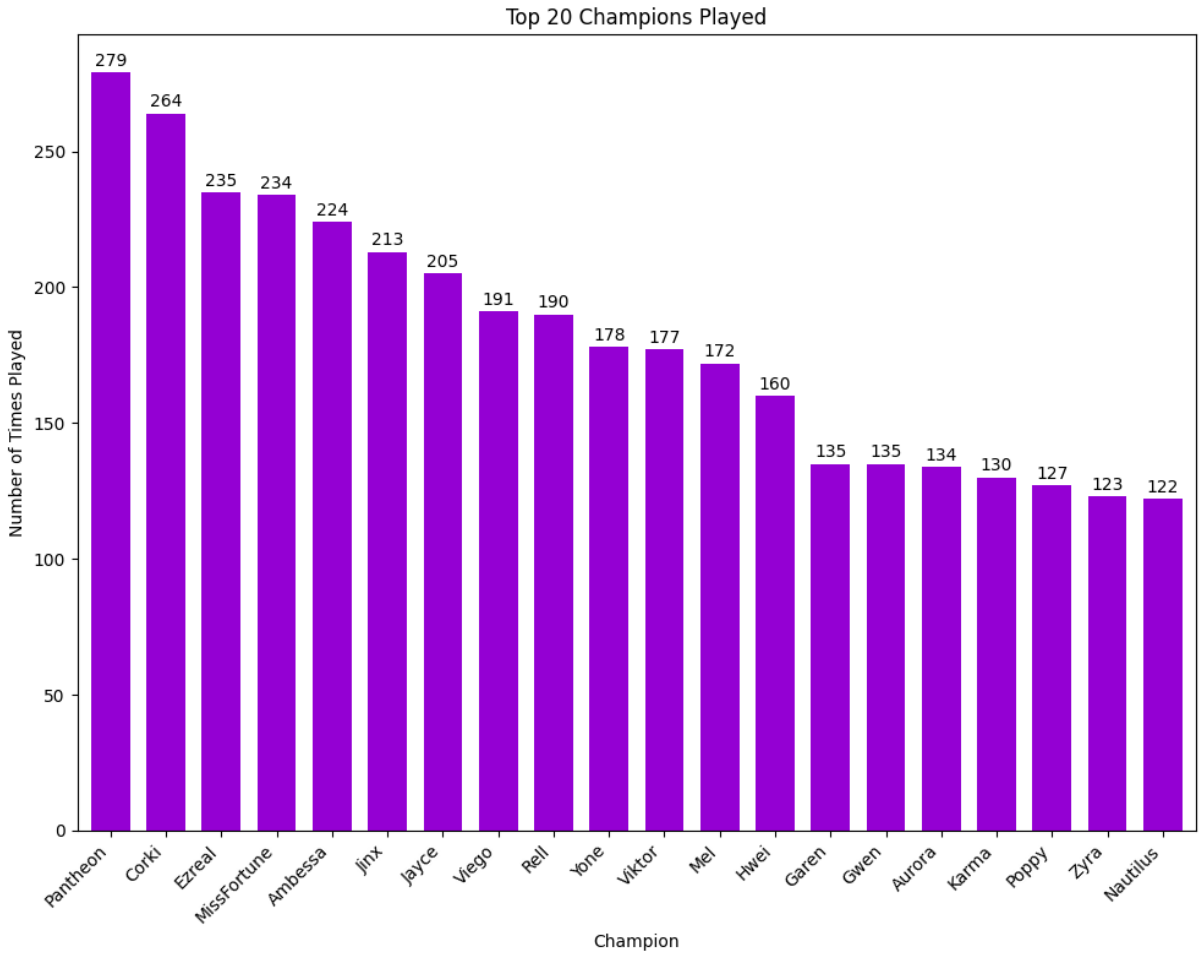
The amount of time that is spent in a game appears to be normally distributed. This surprised me as I thought more games would end in 15 minutes due to teams forfeiting matches. However, this graph remains normally distributed.



The distribution of role types across the ladder is fairly balanced, with Support being the least played, followed by Top, Bottom, Jungle, and Middle as the most played role.



Bottom players have the highest average CS, followed closely by Middle and Top players, while Jungle and Support roles have significantly lower average CS. This disparity may pose challenges when trying to predict jungler patterns in machine learning models.



The final graph is the top 20 most played champions found in games. Pantheon being the most played champion is surprising considering that he is not in the games current meta. The top champions are mainly top laners and attack damage carries or ADCs. I believe this could be because many people in those roles tend to play the same champions many times in a row.

Summary of Findings

The dataset contains 10,957 rows and 138 columns, offering detailed statistics on player performance across multiple games. Key findings include that players average 5.6 kills, 4.87 deaths, and 8.2 assists per game, with a game duration averaging 26.5 minutes. Gold earned averages at 10,741, and turret takedowns are typically around 2 per game. Team objectives like baron and dragon kills are less frequent, with players earning fewer kills on these objectives compared to the team's performance. Significant correlations were found, such as a strong link between gold earned and gold spent, and moderate correlations between kills and turret takedowns, and assists with vision score. Notably, there were no outliers in major performance categories like CS and ward scores, but there were outliers in neutral minions, dragon kills, and baron kills, especially indicating extreme performances by jungle players. Visualizations, such as heat maps and scatter plots, confirmed these relationships, showing how metrics like gold earned and gold spent are closely tied. Gold spent having a much closer range than gold earned could mean that efficient item building could be a potentially winning strategy. Additionally, role distributions show a relatively even spread, with Support being the least played role and Middle being the most popular. Bottom players had the highest average CS, while Jungle and Support roles had far less. Finally, the top 20 most played champions are dominated by top laners and ADCs, with Pantheon as the most played champion, despite not being in the current meta. Turrets being linked to deaths and kills could mean that performance is heavily tied to personal deaths and kills in game. The CS column could pose problems to my prediction system if the amount of jungle camps are not being taken into account. I believe that creating a new calculated column

called total CS score might help with fixing my model. I will add the role to the few rows that were missing the role and if I am unbale to determine the role I will just remove the rows.