League of Legends Proplay Analysis Report

DS 2500 - Intermediate Programming with Data
Jaden Chin - chin.jade@northeastern.edu,
Fanta Kebe - Kebe.f@northeastern.edu
Rukia Nur - nur.r@northeastern.edu

Introduction:

League of Legends, created on October 27th 2009, has been one of the largest video games worldwide with over 100 million users. League of Legends is a 5v5 strategy based MOBA(Multiplayer Online Battle Arena) where the goal of the game is to destroy the opposing team's base. Players must choose from a variety of playable characters, each with their own unique set of skills and abilities. Players must work together to get kills and objectives in order to destroy the opposing base and win the game.

Esports has been emerging all over the world. With the Olympics passing a new Esports section in the games, we wanted to see what it takes to go pro in some popular competitive video game. We decided to study League of Legends, which is considered one of the hardest games to master as there are so many factors that are intertwined into gameplay. Our goal is to understand what factors are the most impactful to one's performance. We hope that with our findings, we can help those who want to improve and take their playing to the next level.

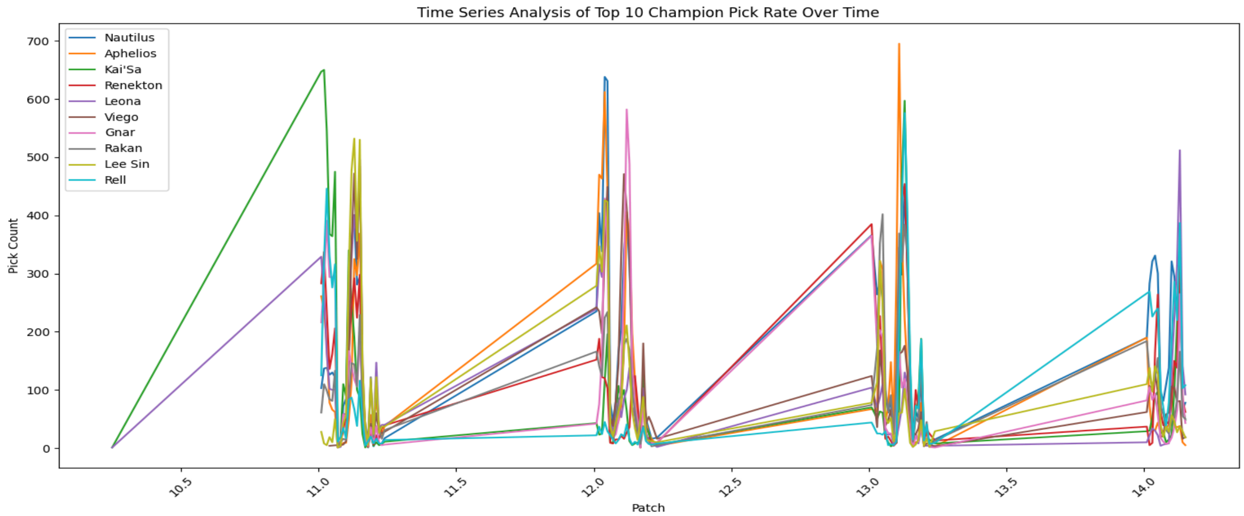
Introduction to your Data

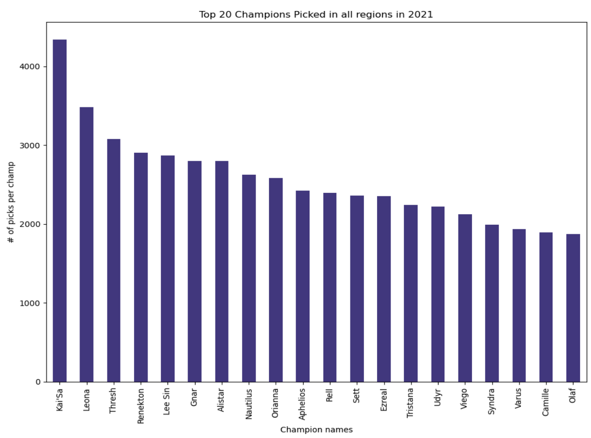
All data has been sourced from 'Tim Sevenhuysen of *OraclesElixir.com* who aggregated and released them for public use': https://oracleselixir.com/tools/downloads. We investigated competitive matches from all regions from 2014-2024. Our data includes detailed match statistics, player performance metrics, and champion picks and bans. Though the source of the data is from a secondary website, the source claims that the data collected is from several different sources, including Match History pages, lolesports.com, lpl.QQ.com, Leaguepedia, as well as the Riot Games solo queue APIs. Despite the fact that this data is derived from public esports matches it is important to consider the privacy of individual players. The data does contain sensitive personal information though. However, typically players tend to use pseudonyms as their display names offering further anonymity.

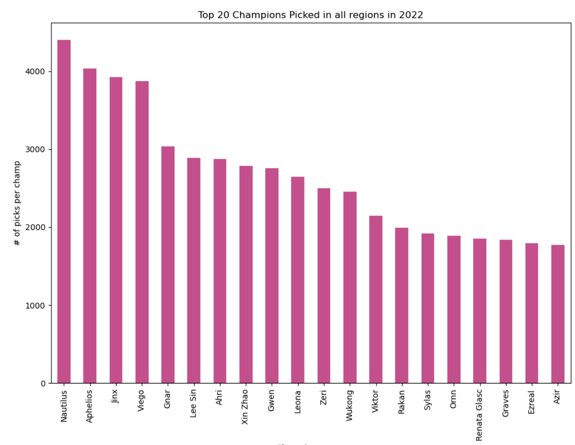
Data Science Approaches:

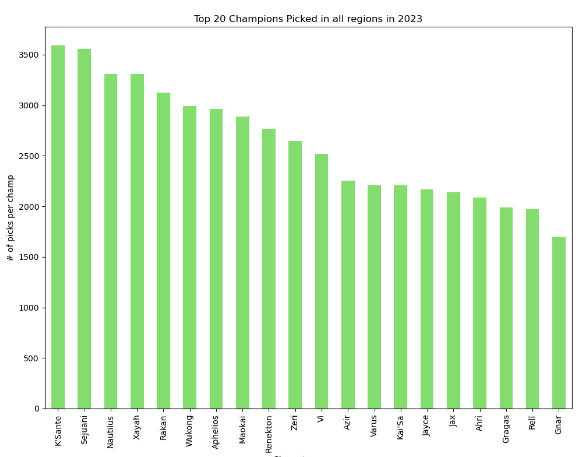
To analyze the data, we applied the following data science techniques we learned:

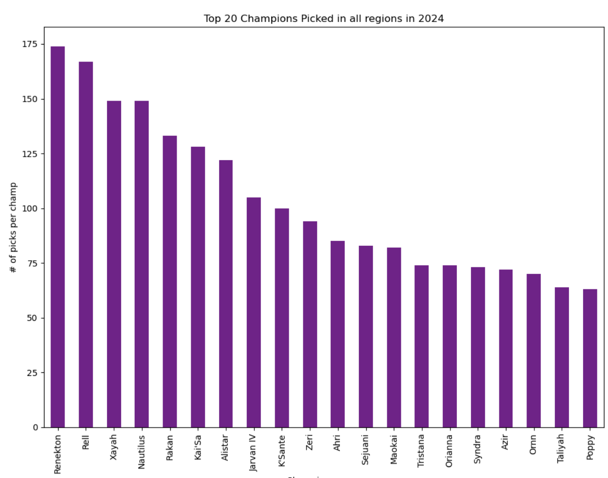
1. <u>Linear regression:</u>
    - Player Performance Metrics: We used linear regression to understand the impact of various player performance on team success specifically. By determining the link between these metrics and match results, we hoped to determine which aspects of player performance are most important for winning. We focused on the following player performance metrics: Kills, Deaths, Assists, Total Gold, Damage to Champions, Visions Score.
2. <u>Visualizations Techniques:</u>
    - Champion Popularity: We utilized line charts to show champion popularity rates over several patches between 2021 and 2024. This method enabled us to properly track changes in champion pick rates, detecting major spikes, continuous trends, and reductions over time. The analysis showed how game updates affect champion selection in professional matches. Additionally we used multiple bar graphs for multiple years to see the different distribution of champion picks for varying years(2020-2024).



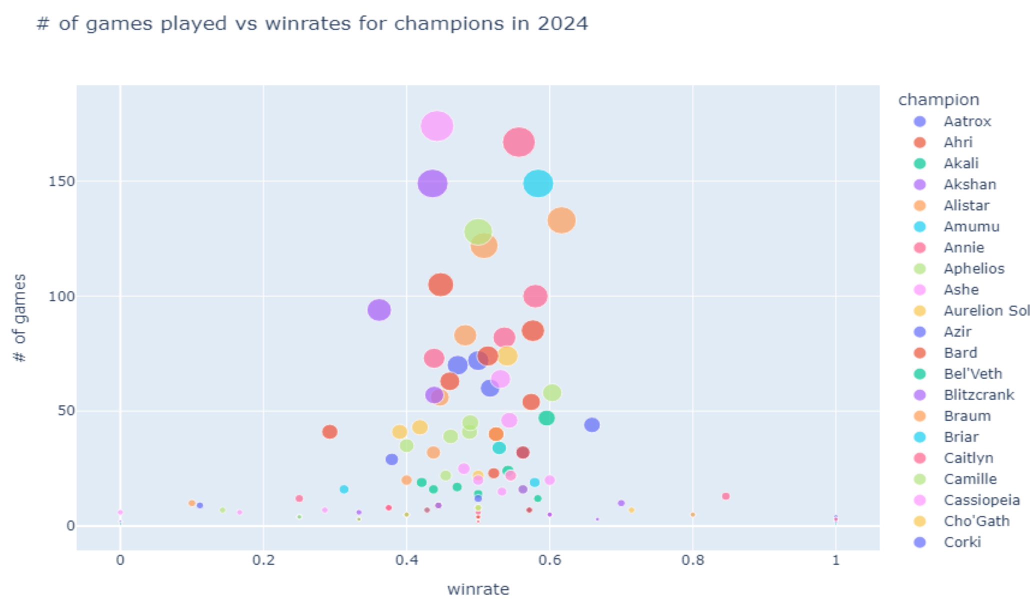








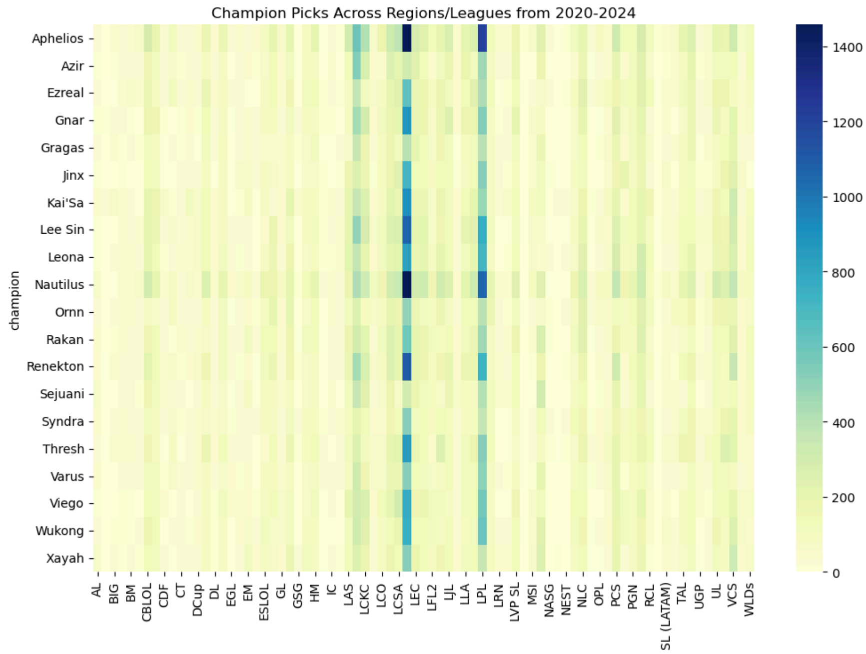
- Bubble Chart: To get a better visualization of how the top champion picks perform, we used a bubble chart. Not only did we want to consider win rates but also how many games they were picked(Since there are some characters that were picked once and either one or loss resulting in a 100% or 0%) winrate. Bubble charts were an excellent choice of visualization as the size of the bubbles represent the number of games a champion was picked, meaning the larger the bubble the more games they were picked. Using this type for each year, we can see how well the most picked characters performed in a year. Additionally we display the number of games a champion was picked on the y-axis while the x-axis represents win rates.



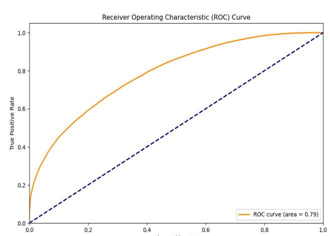
-

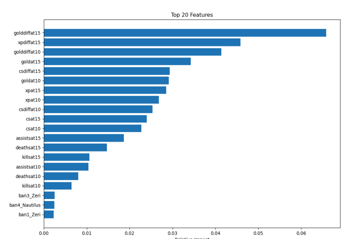
- **Heatmap:** To understand the distribution of top champs worldwide, we decided to use a heatmap for the past 5 years of different leagues' top 20 champion picks. The map not only helped see what champions were consistently picked across regions but also showed us what regions had more matches occurring in a year.

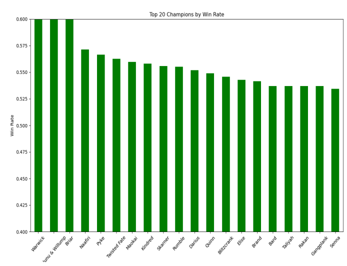
Major regions like LPL LCS and LCK had the most occurring champion picks



3. Prediction Model:







- **ROC Chart**: used to graphically depict how well our model did at predicting wins and losses in a professional match. The true positive rate is the recall and measures the proportion of actual wins against the false positive rate measuring actual losses that were incorrectly classified as wins by our model. Our chart showed an area under the curve of .79 which would indicate that our factors significantly impact match performance.
- **Top 20 Features:** used to identify and rank the most influential features in predicting the outcome of a game. In our project the plot showed that gold diff at 15 and xp diff at 15 were by far the top predictors of match outcomes in formal league games.

- **Champion Win Rate:** by examining the rate of wins by champion we can gain insights into the likelihood of winning based on champions on a team. In our analysis the champion win rate revealed Warwick Nunu Willump and Briar as the top champs. However many things go into champion win rate such as skill floors, the frequency with which a champ is picked and more as wins don't happen in vacuums so further considerations would have to be made in future works.

Results and Conclusions:

When looking at the distribution of characters picked. There are a lot of underlying factors that were not considered. For example some characters were obviously released at different dates and times and were therefore not accessible during certain years or matches. Additionally, things like updates that were purposed to balance out certain characters was another factor that fluctuated data. These factors were things that were hard to avoid since champions in-game statistics could not stay consistent over the course of several years or even in a single year. That's why through our findings, we saw so much variation among the top champions picked. We do see however regarding regional data, that Aphelios and Nautilus were the most consistent picks from 2020-2024 for all regions with the most picks being in LPL and LCS which are China and North America respectively. Additionally we found how marginally certain in game statistics can enhance performance and affect outcome of game. Here is what we found from using our models:

- Kills: +0.018 - Each additional kill slightly increases the likelihood of team success.
- Deaths: -0.052 - Each additional death decreases the likelihood of team success.
- Assists: +0.017 - Each additional assist slightly increases the likelihood of team success.

- Total Gold: +0.000011 - Higher total gold contributes positively, but its impact is relatively small.
- Damage to Champions: -0.000003 - Damage to champions has a very small negative impact.

Linear Regression: The model achieved a R^2 score of 0.399, indicating that approximately 39.9% of the variability in match outcomes can be explained by the selected features. certain champions like Blitzcrank (+0.398) and Briar (+0.498) had higher positive coefficients, suggesting a stronger association with winning outcomes when these champions are picked.

Champion Popularity Trends: Our time series analysis revealed distinct trends in champion popularity:

- **Kai'Sa, Leona, and Viego** have seen an increase in popularity over time.
- **Nautilus, Aphelios, Renekton, Gnar, Rakan, Lee Sin, and Rell** have decreased in popularity.

These trends demonstrate the influence of game updates and meta shifts on champion choices in professional play. The R^2 value and coefficients show how certain in-game actions impact team success, offering useful information for players looking to enhance their performance in competitive encounters.

Future Work:

In the future we could aim to enhance our analysis by incorporating more sophisticated techniques, potentially expanding our dataset, and applying our findings to broader contexts. Here are some key things we potentially want to explore:

1. Further Character Analysis:

In order to understand the roles, advantages, and disadvantages of each champion in varied team configurations and scenarios, we would conduct in-depth evaluations of each individual champion. Furthermore, We could determine the circumstances in which each champion performs best and offer guidance on how to reach their full potential in competitive play by analyzing each champion's performance in various settings.

2. General Player Data:

   We would examine data from an even wider range of players, including casual and semi-professional gamers, to better understand the way different skill levels influence gaming and strategy. By studying players across different levels of expertise, we could identify the strategies and behaviors that are effective for beginners, intermediate players, and experts, providing targeted recommendations for each group. This helps create an inclusive analysis for the League of Legends considering its diverse player-base.

References:

Oracle's Elixir - LoL Esports Stats https://oracleselixir.com/tools/downloads

https://scikit-learn.org/stable/auto_examples/model_selection/plot_roc.html