

E-Sport Analytics and Predictions

League of Legends

Ritwik Katiyar

November 30th 2022

Introduction

Currently, in e-sports one of the most highly competitive games is League of Legends (LoL). The following video link goes in greater detail about what the game is all about [What is League of Legends?](#). However, as a quick introduction League of Legends is defined as a multiplayer online battle arena (MOBA). Within league two teams of five champions (each player controlling their champion in a team of five) go head-to-head in a match to destroy the enemy nexus. Each champion has a unique set of abilities, which identifies to the champion identity to them. Moreover, there are 164 champions that a player can play. To achieve victory a team has to work together to acquire gold by killing the enemy team and dominating various objectives throughout the map. Traditionally, based on the location on the map each player is assigned a specific role, with the exception of bottom lane where there are two players. In other words, one player from the team is assigned to the top lane, mid lane, jungle and two players go to bottom lane. Often the second player in the bottom lane is referred to as support.



Figure 1: Map of Summoners Rift

Every year from October to November, a World Championship event (worlds) is held where teams from different regions compete in order to win the grand prize of \$2.5 million. There are 24 teams that attend worlds. This year alone a new peak viewership record was boasted by Riot Games (the company behind League of Legends) with the final game having 5.1 million live viewers.

The worlds event is structured in multiple parts. Teams from the lower tiers first play in a qualifying tournament known as play-ins. To understand worlds we first need to understand the region tier system. If a region performed better at worlds previously then that region is given a higher tier. The tiers for this years world's event are as follows:

- I. China and S.Korea: Three teams from each region in the main event and one in play-ins.
- II. Europe: Two teams in the main event and two in play-ins.
- III. North America: Two teams in the main event and one in play-ins.
- IV. Vietnam and Pacific Championship Series (Hong Kong, Macau and Taiwan): One team from each region in the main event and in play-ins.
- V. Japan, Brazil, Australia, Latin America and Turkey: One team from each region playing in play-ins.

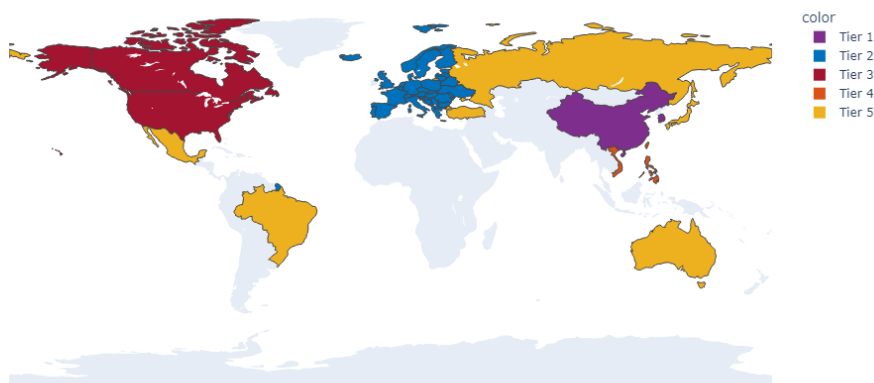


Figure 2: Tiers by Regions

Only four teams qualify out of the play-ins tournament into the worlds main event. In which four groups of four teams play a double round robin (Each team in a group plays against each other twice) and the top two teams from each group advance to a single elimination, best of five games all the way to the finals. To assist with reaching the finals at the worlds event each team consists of various coaches and staff members that have an important job of drafting the champions for their players.

Drafting champions before a game in League of Legends is an important aspect of determining if a team would win or lose. Before the game begins the players get an opportunity to pick the champions they will be playing and ban champions they wish to not play against. This is a crucial part of the game since by banning or picking away a champion the enemy player is skilled at can result in putting the enemy team at a disadvantage. Equally, picking a champion a player is skilled at can make the difference between a win and a loss. Hence, entire teams of coaching staff and data analytics spend hours compiling data to generate an optimal champion draft for their teams. Therefore, some of the most important aspects the teams pay a close attention to are the champions that were played in the past, champions that are generally banned, and champions that had a high win rate. We will be analyzing the data from worlds to get a glimpse on what champions the players would be playing at the start of the new season after worlds. By looking at champions that were highly contested or were generally considered good during the event.

As mentioned, a key factor the coaches focus on while drafting is the player's ability to play a certain champion. If the probability of winning a game increases when a player plays a specific champion then that champion becomes a priority for the team. To mimic the team staff we will utilize a simple logistic model in order to predict what champion a player would prioritize during the champion drafting phase.

Finally, to predict a winner before the game has occurred is a sport's analytics ultimate goal. However, making an accurate model to predict a winner can take months or years of work. Therefore, we will just be scratching the surface by generating a model that can provide predictions on a team's odds of winning based on the amount of gold a team earns. To answer the above questions we will be utiliz-

ing the data set obtained from [Oracleselixir](#)

The [Oracleselixir](#) dataset contains end game information for every game that was played this year from January to November across all regions, across all leagues (including amateur leagues), and across all events. The information was collected by taking the end game results from [LoL Esports](#) from every region and compiling the data into one dataset. For every game the datasets first presents performance of each player per game. Then the dataset shows a cumulative team data for the two teams that played that game. This way we see the player's performance for a game, as well as, the overall team's performance. Since, the data comprises of complete end game results, it means that the data has all the end game variables and statistics that are displayed at the end of each game.

There a total of 116 variables and 147,865 rows of observations in the data set. Therefore, to get a general idea on what the data set looks like, here are the first five rows and some of the variables from the data set:

	league	year	split	playoffs	game	patch	side	position	playername	\
0	LCK	CL	2022	Spring	0	1	12.01	Blue	top	Soboro
1	LCK	CL	2022	Spring	0	1	12.01	Blue	jng	Raptor
2	LCK	CL	2022	Spring	0	1	12.01	Blue	mid	Feisty
3	LCK	CL	2022	Spring	0	1	12.01	Blue	bot	Gamin
4	LCK	CL	2022	Spring	0	1	12.01	Blue	sup	Loopy

	teamname	...	opp_csat15	golddiffat15	xpdiffat15	\
0	Fredit BRION Challengers	...	121.0	391.0	345.0	
1	Fredit BRION Challengers	...	100.0	541.0	-275.0	
2	Fredit BRION Challengers	...	119.0	-475.0	153.0	
3	Fredit BRION Challengers	...	149.0	-793.0	-1343.0	
4	Fredit BRION Challengers	...	21.0	443.0	-497.0	

	csdiffat15	killsat15	assistsat15	deathsat15	opp_killsat15	\
0	14.0	0.0	1.0	0.0	0.0	
1	-11.0	2.0	3.0	2.0	0.0	
2	1.0	0.0	3.0	0.0	3.0	
3	-34.0	2.0	1.0	2.0	3.0	
4	7.0	1.0	2.0	2.0	0.0	

	opp_assistsat15	opp_deathsat15
0	1.0	0.0
1	5.0	1.0
2	3.0	2.0
3	3.0	0.0
4	6.0	2.0

[5 rows x 116 columns]

Methodology

Considering every question is unique, we will try and answer each independently from one another with minor overlaps. The project goals we will try and accomplish are as follows:

Gaol-1

- Which champions had the highest pick rate?
 - Filter the data and obtain games from the worlds event. The champion columns is part of the rows that contain individual player data.
 - The champion column would need to then be aggregated on the number of times a champion was picked.
 - Then, the pick rate would be calculated by dividing the amount of times a champion was picked over the total number of games at worlds.
- Which champions had the highest ban rate?
 - The data set would need to be filtered by the cumulative team data rows since the bans are organized by columns from ban1 to ban5. There a total of 10 bans per game and each team gets to ban 5 champions.
 - Once the data is filtered each ban column would need to be aggregated on champions and the number of times a champion was banned.
 - Since there are 5 columns all 5 columns would need to be summed for every champion.
 - To acquire the ban rate the number of times a champion was banned would need to be divided by the total number of games.
- Which champions had the highest win rate?
 - For this part of the question we can utilize the champion data that was filtered previously for pick rate. However, we also add results from the games, where the result was a win.
 - Then the data would need to be aggregated on champions and the number of times the champion won the game.
 - To calculate the win rate, the champion's total wins were divided by the number of games played at worlds.
- Heat Map
 - The data from pick rate, win rate and ban rate can then be combined into one.
 - To better visualize the date and get the champions that had the highest pick rate and win rate but a low ban rate a heat map can be constructed by sorting the data as mentioned.

Goal-2

- Generating a model that can predict the probability of win for champions a player plays.
 - To accomplish this first the data set would need to be filtered by the individual player data as well as by regions or leagues. Considering that the top four regions are generally the one that are most substantial at worlds it makes sense to filter the data by those regions.
 - Once the data is filtered the three variables that will be required to answer this question would be player name, champion and result.
 - Then the data would need to be filtered again by a specific player.
 - Considering we are trying to predict the probability of a win we will need to use a logistic regression model.
 - To use a logistic regression model we will need to create dummy variables for the champions as well as the results
 - In this case our predictor would be the champions and our target variables would be the result.
 - Once the model has been constructed we would need to check if the model results are valid by find the accuracy of the model.
 - Then we can predict the probability of win for each champion and display the probability using a bar plot.

Goal-3

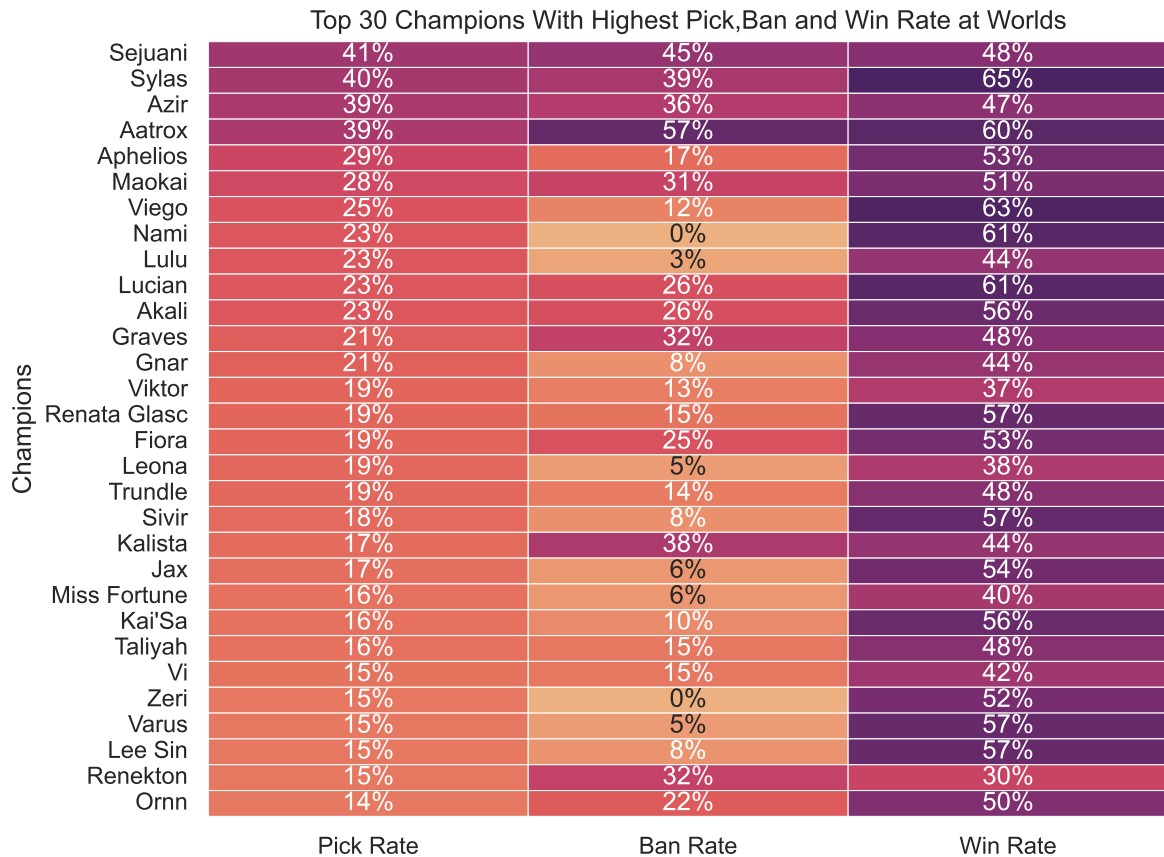
- A Model that predicts the probability of a win based on the total gold a team earns in a game.
 - Filter the data based on the total gold earned by the team from the summer and spring splits, and the regional games should be from the top three tiers at worlds. Hence, those leagues being the LCK (Korean league), LPL (Chinese league), LCS (North American league) and LEC (The European league). This is done to insure the best quality of the data. Since lower league games are highly unpredictable and contain a lot of outliers and inconsistencies.
 - Construct a logistic regression model for the total gold and the result columns from the filtered dataset.
 - Evaluate the model performance in order to determine if the model serves as a good predictor of wins.

Results

Goal-1

The following plot aims to accomplish the goal of answering which champions had the highest pick rate, lowest ban rate and highest win rate at worlds?

Text(0.5, 1.0, 'Top 30 Champions With Highest Pick,Ban and Win Rate at Worlds')



The above heat maps presents the top 30 champions with a high win and pick rate, but a low ban rate. Champions with high ban rates or low pick and win rates are most likely not going to see much game play at the games that will be played after worlds. However, champions like Sejuani, Sylas, Azir and Aatrox would be continued to be played. We may even see an increase in bans for Nami and Lulu since these champions seems to be demonstrating a relatively high pick and win rate with an extremely low ban rate.

Goal-2

Generating a model that can predict the probability of wins for champions a player plays. In this case we will look at a player known as Zeka from the South-Korean regional league (LCK) his team; DRX won worlds this year. It would be interesting to see if the model can predict what champions Zeka prioritized at worlds.

Intercept : [-0.19981425]

Coefficient: $\begin{bmatrix} 0.03486086 & -0.11260707 & 0.13318622 & 0.78672547 & -0.36291826 & -0.40916786 \\ 0.44019753 & -0.61410461 & 0.06653633 & 0.60258965 & -0.20236812 & -0.36291826 \end{bmatrix}$

The equation for our model can be given as:

$$P(\text{Win}) = \frac{e^{(-0.19+0.034*Champion_1)+(-0.19-0.112*Champion_2)...}}{1 + e^{(-0.19+0.034*Champion_1)+(-0.19-0.112*Champion_2)...}}$$

Accuracy Test Results for Zeka: 0.65

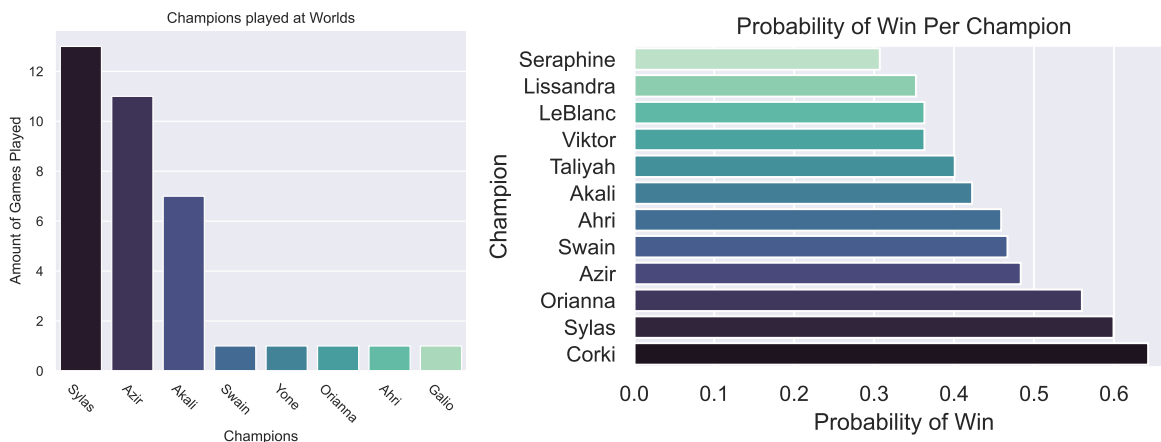
From the results above we can see the intercept and coefficients for all the champions the player plays constructed by the model. We can also see the accuracy score of the model. The accuracy score is generated by computing the confusion matrix. A confusion matrix is a matrix constructed by taking the actual values and comparing them to the values predicted by the model. The accuracy represents the amount of times the model was able to predict the result correctly. Hence, the score suggests that our model is able to predict the results correctly about 65.3% of the time. The score for this player is particularly low, although after testing a number of players the accuracy score seems to range from 59% to 100%. The reason the accuracy score is low is due to the amount of different champions the player plays. Therefore, players that play fewer champions yield a higher accuracy score. If a player plays too many different champions, then there simply is not enough data for the model to be more accurate.

For example a player named Larssen has a small pool of champions that he plays than Zeka. if we run our model on him the accuracy score should be higher.

Accuracy test result for Larssen: 0.7

We can see that with a smaller champion pool our model is able to perform better. Regardless, we can plot our predicted results along with the results from worlds in order to see if our model was able to provide some insight about what champions he would play at worlds.

Text(0.5, 1.0, 'Probability of Win Per Champion')



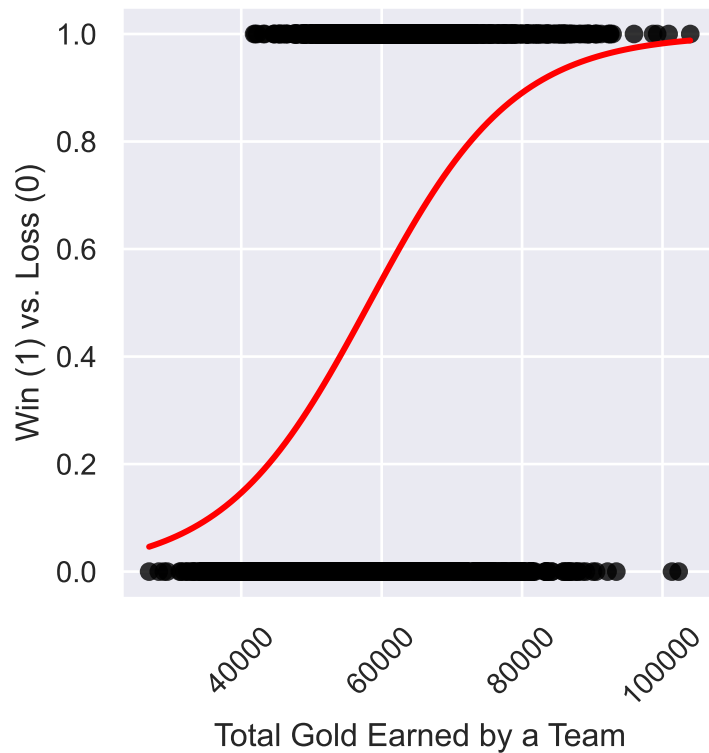
The plot on the left shows the champions played by Zeka at worlds vs. the amount of times he played each champion. The plot on the right shows the probability of win per champion as generated by the model. We can see that Zeka has a high win rate probability for Corki, Sylas, Orianna, Azir, and Swain. Looking at the plot on the left we can see he did in-fact play Azir and Sylas the most where he has a relatively high probability of win. However, he did not play champions like Orianna and Corki.

Goal-3

The following results aims to utilize a generated model that can predict a team's win probability by the amount of gold they earned on average in a game:

Text(0.5, 0, 'Total Gold Earned by a Team')

Logistic Regression Curve For Gold Earned vs. Probability of Winning



The plot above shows the regression curve for the probability of win based on the amount of gold earned in a game. We can see the probability of win increase based on the gold; capping at 10,000 gold totals. The few dots at the bottom where a team lost with 100,000+ gold could be from a game that lasted for over 45 minutes where both teams were able to acquire a copious amounts of gold. For the model the total gold earned has a p-value of approximately 0.0 which means that it is a significant

predictor for wins. The model's equation can be given as:

$$P(Win) = \frac{e^{(-5.619+9.64e^{-05}*Total_{Gold})}}{1 + e^{(-5.619+9.64e^{-05}*Total_{Gold})}}$$

Confusion Matrix:

	Positive	Negative
Positive	1249	523
Negative	551	1221

The above data shows the confusion matrix constructed by the matrix library. We can see that our model is predicting only 1249 wins and 1221 losses instead of 1772 wins and losses (Our filtered data set contains 1772 wins and losses). Using this we can hence calculate the accuracy of our matrix using the accuracy score library.

Accuracy score for the model: 0.7

We can see that our accuracy score for the model is 70%. Furthermore, to increase the accuracy we would need to determine further variables that can significantly affect the game and give us the results we desire. As a test, we can still predict what the probability of win for a game between two teams would be by calculating the mean total gold they earn from their past games and using our model to predict the probability.

Probabilty of win for T1: [0.57822449]

Probabilty of win for JD Gaming: [0.52758718]

Our model shows the probability of win for the team T1 to be 57% and the team JD Gaming to be about 53% so we can assume that T1 would win. At the worlds semi-finals game T1 vs. JD Gaming; T1 managed to get a narrow 3-2 victory in a best of five series. However, we are only looking at the individual team probability and we are not taking into consideration which team T1 was facing. For example, if we look at the similar probabilities between T1 and the team Cloud 9 we get the following:

Probabilty of win for T1: [0.57822449]

Probabilty of win for Cloud 9: [0.57027813]

We can see that the probabilities are extremely close. However, Cloud 9 has never won a game against T1 so our model is nearly failing since there is no accounting for which team a team is against. The probability of win for Cloud 9 should be incredibly lower compared to T1.

Discussion

For the first question we concluded that the champions that will be most impactful after the worlds event would be Sejuani, Sylas, Azir, Aatrox, and Aphelious. Some of the champions that were undoubtedly strong but were worthy of a note were champions like Nami, Lulu, Fiora and Jax that had a low pick and ban rate.

The second question aimed to answer if a model could be created to determine the probability of wins for a player based on the champion selected. After, using Multiple Logistic regression model our results varied from 59% to 100% accuracy based on the amount of champions the player played. Our model also lacked critical data from practice games and off stage games, with more data our model could predict with higher accuracy. Alternatively, there might be a model out there that can predict the probability with the amount of data that we have. Despite the low accuracy, our model was able to provide some insight on what champions a player would play. By looking at a player known as Zeka, we noticed that his most played champions at worlds were Sylas, Azir, and Akali. Based on the model the player's highest probability of wins were on Corki, Sylas, Orianna, Azir, Swain, Ahri, and Akali respectively. Our model was able to show that the drafting team for Zeka was focusing on getting him the champions he had high probability of winning on. We can also use our results from the first question in order to determine why he did not play Corki, Orianna, Swain, and Ahri. Considering those champions were not a priority by the teams at worlds it could be possible that the coaching staff decided to have Zeka focus on Sylas, Azir and Akali champions in order to simply deny them from the enemy. Moreover, since he has a decent win rate with Sylas and Azir it is a relatively safe pick.

For our final question we tried to determine the probability of win for a team. In order to determine this, we tried to simply base the probability of wins on the total amount of gold a team earns. We used logistic regression model again and were able to obtain 70% accuracy for our data set. The accuracy for our model would most likely be improved by finding other win determining factors. The high value factors can be determined by using random forest, Principle Component Analysis, and other machine learning algorithms. However, even for a low accuracy score our model is still useful and can predict wins. As an example we used it to try and predict the probability of win based on the mean total gold earned by team T1 and compared the probability to another team called JD Gaming. Since T1 had a higher probability of winning we determined that T1 would win and in fact at the worlds event the T1 managed to beat JD Gaming at the semi-finals. Considering the model is largely incomplete the model has a number of short comings. Such as the fact that the model does not take into account which team a team is playing against, champions played by the team, player's performance, other game factors such as a team's ability to obtain objectives, etc. Despite the short comings the model still has a decent accuracy for a start.

The dataset we used for this project is massive and can be used to answer a myriad of questions about the game and the players. Due to constraints some of the questions and their results were not included in this report. These results can be viewed raw within the 'Project_playground.ipynb' file along with the report. Overall, we were successfully able to display and look at the champion's pick, win and ban rates at the worlds event. We also were able to predict what champion a player would play based on their probability of winning when playing said champion, and finally we were able to successfully set up the foundations for a model that can determine the winner before the game has been played.

Bibliography

- [1] Riot Games. *League Of Legends*. <https://www.leagueoflegends.com/en-us/>. Accessed: December 13th 2022. 2022.
- [2] Jose Garcia. “League Of Legends Worlds 2022 Finals Breaks Peak Viewership Record”. In: *DualShockers* (2022). Accessed: December 13th 2022.
- [3] Riot Games. *What is League of Legends*. <https://www.nasa.gov/nh/pluto-the-other-red-planet>. Accessed: December 13th 2022. 2015.
- [4] Tim Sevenhuysen. *LoL Esports Match data from Oracles Elixir*. <https://oracleselixir.com/tools/downloads>. Game data from all the games, across all regions, and events for League of Legends Esports, Viewed: December 13th 2022. 2022.
- [5] LoL-Esports Staff. “Learn more about the 2022 League of Legends World Championship in North America”. In: *League of Legends Esports* (2022). Accessed: December 13th 2022.
- [6] Kevin Webb. “More than 100 million people watched the League of Legends World Championship, cementing its place as the most popular esports”. In: *Insider* (2019). Accessed: December 13th 2022.