**Project Report –**

**2023 LoL Worlds Winner Prediction**

# **[GitHub URL]**

[SzeWaiHo/UCDPA-SzeWaiHo (github.com)](https://github.com/SzeWaiHo/UCDPA-SzeWaiHo)

# **[Abstract]**

The aim of this project is to predict the next team to win 2023 League of Legends (‘LoL’) Worlds tournament.

This esports game consists of two teams with two sides (blue and red) with the winning objective being to kill the opponent’s towers and inhibitors – and ultimately their nexus. During the game, the players can kill their opponents, achieve objectives such as killing dragons and barons. These factors will be further explored.

There are 3 stages in the Worlds tournament are as follows:

1. PlayIns Stage – all teams to play for a chance to compete in the Swiss Stage
2. Swiss Stage - teams will be qualified to enter the Knockout (Final) Stage
3. Knockout Stage - the winner will be determined in the finals

As an example and to demonstrate visualization skills, I drafted the 2022 Stages below:

A picture containing text, diagram, plan, screenshot

Description automatically generated

A picture containing text, diagram, screenshot, plan

Description automatically generated

A picture containing text, diagram, screenshot, plan

Description automatically generated

In 2023, there are 22 regional teams pre-determined to enter the Swiss Round

A picture containing text, screenshot, font

Description automatically generated

# **[Introduction]**

I chose this project as there is a lot of data statistics online in which I can perform the functions I’ve learned in DataCamp. It would also be interesting to analyse for modelling and considering for machine learning.

Please note that many of these leagues are on-going at the moment so the API may pull new matches and so the figures may not match the report.

# **[Dataset]**

Most of my dataset is imported data from Leaguepedia (<https://lol.fandom.com>)

This is done by importing through API:

**Scoreboard Games Data**

A screenshot of a computer

Description automatically generated with medium confidence

This dataset gives detailed information for each match played in the listed leagues (including but not limited to the Worlds Tournament). This helps me analyse each team and different factors that may help them win.

**Teams/Regions Data**

**A picture containing text, screenshot, font

Description automatically generated**

This dataset imports the team name, the team name abbreviation and their region for each team in the listed leagues. This helps me analyse each team and each region.

\*I made some manual adjustments that aren’t updated in Leaguepedia which I won’t include in this report but are noted in the Jupyter Notebook.

# **[Implementation Process]**

DATA MANIPULATION

First, I import the packages needed and initialize some factors.

A screenshot of a computer program

Description automatically generated with medium confidence

Then, I import the data through API and make minor adjustments.

I convert the datasets to pandas dataframes for data manipulation.

A screenshot of a computer code

Description automatically generated with low confidence

Next, I merge the scoreboard and team datasets to find out which teams and regions won each match:

A screenshot of a computer code

Description automatically generated with low confidence

For my custom model later, I need to find out if the winning team of each match was on the blue or red side:

A screen shot of a computer

Description automatically generated with low confidence

# **[Results]**

ANALYSIS– 2011 – 2022 WORLDS TOURNAMENT

The Worlds Tournament data started in 2011 and ends in 2022.

I want to find the winners of the matches in the Worlds Tournament only.

A screen shot of a computer code

Description automatically generated with low confidence

**The top 10 teams who won the most matches in Worlds Tournament:**

Only 4 regions are shown for the top 10 teams. We can already find the strong regions within the tournament.

A picture containing text, screenshot, plot, diagram

Description automatically generated

**The top 10 regions who won the most matches in Worlds Tournament:**

Korea, China, EMEA and North America are predominantly strong teams compared to the other regions when comparing winning teams in all matches in the Worlds Tournaments.

A picture containing text, screenshot, display, diagram

Description automatically generated

**The teams who won the Worlds Title:**

T1 is the only team that has won the Worlds Title more than once.

T1, FNC, C9, DK is also in the top 10 teams list.

A picture containing text, screenshot, plot, line

Description automatically generated

**The regions who won the Worlds Title:**

Only 4 regions has won the Worlds Title in 2011 – 2022 (12 years).

Korea won 6 out of 12 World Tournaments – making them a very strong candidate for the next Worlds Winner.

A picture containing text, screenshot, diagram, rectangle

Description automatically generated

ANALYSIS – REGIONAL – WIN RATE%

Other than “who won the most matches” and “who won the Worlds Title”, I am also interested in how many teams were able to participate in Worlds.

Additionally, I am also interested in the ‘win rate%’ = #wins/#matches

We can actually ignore 2011 because not all regions were invited in the Worlds tournament.

I created a custom function to graph this win rate per region.

For regions I am interested in:

**Korea**

We can see that Korea has a very high number of participation counts. They also won many matches.

A picture containing text, screenshot, plot, diagram

Description automatically generated

**China**

China is also a strong region, however the win rate is slightly lower than Korea

A picture containing text, screenshot, plot, diagram

Description automatically generated

**EMEA**

EMEA played many matches but their win rate seems to be declining.A picture containing text, screenshot, plot, diagram

Description automatically generated

To get an overview, I created a table to get the win rate% mean of all years for all regions:

Korea, China, EMEA and North America are again the 4 strongest regions according to this measure.

A picture containing text, screenshot, font, number

Description automatically generated

I also graphed the matches played by each region in the Worlds Tournament historically:

A picture containing line, diagram, plot, screenshot

Description automatically generated

Korea, China, EMEA and North America’s performance has improved and outperformed other regions in the last 3-4 years.

From the analysis above, I decide to focus on KR, CN, EMEA and NA in the rest of the analysis. I will call them the ‘top leagues’.

This also ties in with the pre-determined team quotas per region for 2023 Worlds.

ANALYSIS – TEAMS – WIN RATE%

**Top Regional Leagues – KR, CN, EMEA, NA**

In 2023 Worlds, the quota are 4, 4, 3, and 3 teams for KR, CN, EMEA and NA respectively.

To predict these teams, I created a custom function to get the top X teams per region (only counting teams that have played more than Y matches in their leagues) – by the measure of win rate%.

**Korea – LCK**

For Korea, their regional league is LCK.

I got the top 4 teams that played more than 300 matches in LCK historically.

There were 3664 matches/ 12 years so I set the Y parameter to 300 (filter teams that played more than 300 matches)

A screenshot of a computer

Description automatically generated with low confidence

**China – LPL**

For China, their regional league is LPL.

I got the top 4 teams that played more than 500 matches in LCK historically.

There were 6279 matches/ 12 years so I set the Y parameter to 500.

A screenshot of a computer

Description automatically generated with low confidence

**EMEA – LEC**

For EMEA, it is a little more complicated as some regions have merged into EMEA (mainly Europe) recently but the strongest teams are still originally from LEC so I only filter the regional league LEC.

I got the top 3 teams that played more than 100 matches in LCK historically.

There were 1130 matches/ 12 years so I set the Y parameter to 100.

A screenshot of a computer

Description automatically generated with low confidence

**NA – NA CLS**

For EMEA, it is a little more complicated as some regions have merged into EMEA (mainly Europe) recently but the strongest teams are still originally from LEC so I only filter the regional league LEC.

I got the top 3 teams that played more than 100 matches in LCK historically.

There were 1130 matches/ 12 years so I set the Y parameter to 100.

A screenshot of a computer

Description automatically generated with low confidence

The strongest team by win rate% in all the top leagues have a win rate above 65% and all the strong teams have a win rate above 50%. This is reassuring as it means that they mostly win the matches they play.

**Worlds**

I also want to examine the win rate% per team within the Worlds Tournament.

I adjusted the function to look at the Worlds Tournament only.

I put in top 10 teams as a parameter for teams that played more than 50 matches – only 8 teams came up but I kept the table as I think 50 is a good number of matches to look at.

It is interesting to see that observe that:

* T1 and DK (KR teams) perform a lot better in Worlds than in their regional league LCK. This may be because they have higher skill sets and other regions underperform in the Worlds.
* RNG and EDG (CN team) performs worse in the Worlds but very well in their regional league LPL. They may be strong teams within their region but not against other regions.
* FNC (EMEA team) performs stably similar to in their regional league LEC.

A screenshot of a computer

Description automatically generated with medium confidence

From this table, I decide to exclude NA from the top regional leagues.

MODEL – TEAMPOINTS

I created a model that would consider factors within each match to measure the performance of teams.

The factors taken into consideration were:

* Blue/Red Side
* Kills
* Gold Lead
* Barons
* Dragons
* Rift Heralds
* Towers
* Inhibitors

I first analyse the significance of each factors by seeing if there is a significant difference between the winning and losing team. I then assign ‘points’ to the factor with conditions.

Then, I use this model to apply to all the teams in the historical matches and calling the total ‘points’ they’ve achieved their ‘TeamPoints’.

Lastly, I will rank these teams by their ‘TeamPoints’.

**Blue/Red Side**

I first looked at blue/red side. There is a 50% probability that a team would be assigned the blue or red side. There is no tangible advantage to this factor so with the number of matches in my data – I expected a close to 50/50 distribution. However, I observed that there is a slightly higher number of matches won when the team is on blue side.

A screenshot of a computer program

Description automatically generated with low confidence

I looked at all the other factors by looping through the matches and analysing their summary statistics.

A screenshot of a computer

Description automatically generated with low confidence

I can see significant differentiation between the winning and losing teams for these factors:

**Kills**

For the winning team, they get a lot more kills than the losing teams.

I decide to assign the points (according to the 25%, 50% and 75% percentiles) as follows:

1 point if the team has 11 – 16 kills

2 points if the team has between 17 – 20 kills

3 points if the team has more than 20 kills

**Gold Lead**

For the winning team, the gold lead is significant.

I decide to assign the points (according to the 25%, 50% and 75% percentiles) as follows:

1 point if the team has a gold lead of 7701– 10686

2 points if the team has a gold lead of 10687– 13414

3 points if the team has a gold lead more than 13414

**Dragons**

For the winning team, the number of dragons killed is different.

I decide to assign the points (according to the 25%, 50% and 75% percentiles) as follows:

1 point if the team has killed 2 – 3 dragons

2 points if the team has killed 4 dragons

3 points if the team has killed more than 4 dragons

**Towers**

Although conquering the towers is part of the winning objective, the teams do not need to conquer all towers to win. Hence, I still examined this factor as a winning feature.

There are 11 towers the teams can conquer.

For the winning team, the number of towers conquered is significant.

I decide to assign the points (according to the 25%, 50% and 75% percentiles) as follows:

1 point if the team has conquered 8 – 9 towers

2 points if the team has conquered 10 or more towers

The other factors didn’t show significant differentiation so were ignored in my model:

A screenshot of a computer

Description automatically generated with low confidence

I applied this ‘TeamPoints’ model to all the teams in the top league (top 4 KR, top 4 CN, top 3 EMEA)

A picture containing text, screenshot, diagram, plot

Description automatically generated

I also applied this ‘TeamPoints’ model to all the teams in all leagues.

A picture containing text, screenshot, colorfulness, diagram

Description automatically generated

# **[Insights]**

* I predict that team “T1” (SK Telecom T1) from Korea (KR) will win the 2023 LoL Worlds Title.
* Korea is the strongest region by winning most of the Worlds Titles, most of the worlds matches, having the highest win rate% throughout all regions
* T1 won the Worlds Title 3 times already and being the only team who has won it more than once
* T1 having the highest win rate% within LCK (KR regional league) and Worlds Tournaments
* T1 having the highest ‘TeamPoints’ within my model
* My second estimation would be “G2” (G2 Esports) from EMEA.
* G2 never won the Worlds Title and EMEA is ranked third within region performance but they still have a high win rate% and participated in Worlds many times
* According to my ‘TeamPoints’ model, they are the second best team
* Some factors have a significant impact on the winning teams:
* If the team is on the blue side
* If the team has many kills (especially more than 17)
* If the team has more dragon kills
* If the team conquers most the towers (>10)
* Regional performance is fairly stable.
* The top leagues play well consistently
* Other leagues mostly don’t make it to the Worlds Tournaments

MACHINE LEARNING

My model and analysis is very limited as the winner of the 2023 LoL Worlds is also affected by many other factors such as: players (their age, champion pool, experience), the coach and managing team, regional league statistics and other in-game statistics.

If I use machine learning, I would pull in all these data to make a better prediction. I can use classification methods to analyse in-game factors such as the blue/red side. I can also use tree models with set parameters to predict the line up of the 3 stages up until the winner of the finals. Machine learning can help me better improve my model by dimensionality reduction and hyperparameter tuning.

# **[References]**

I only inputted datasets from Leaguepedia which is referenced in the Jupyter Notebook.