League of Legends Regional Differences

By

Thamiliniyan Aravinthan

Submitted to

**The University of Roehampton**

In partial fulfilment of the requirements

for the degree of

**BACHELOR OF SCIENCE IN COMPUTER SCIENCE**

Abstract

This is a data visualisation project using D3js, a JavaScript library, that focuses on the regional difference in professional League of Legends from three main regions (LEC – Europe/Middle East/Africa; LCS – North America; and LCK – South Korea). The data was sourced from Kaggle [1] which is an online community platform that hosts various datasets, competitions, code, and forums. This dataset was perfect for the project as it had data from many regions big and small that covered many aspects of the game over a 4-year period. What was especially interesting for this project was creating various interactive visualisations that use the same type of data and evaluating their effectiveness. The method of evaluation was a survey that was conducted by fellow peers and family and aimed to determine which visualisation best represented the data, and for any changes or additions that would further improve them. During this project I’ve learned a lot including self-managing a large project using various appropriate project management tools and techniques (see Section 2).

Signed (apply signature below)



**Declaration**

I hereby certify that this report constitutes my own work, that where the language of others is used, quotation marks so indicate, and that appropriate credit is given where I have used the language, ideas, expressions, or writings of others.

I declare that this report describes the original work that has not been previously presented for the award of any other degree of any other institution.

Shape, arrow

Description automatically generated

**Date:** 12/05/2023

**Thamiliniyan Aravinthan**

Acknowledgements

I would first like to express my heartfelt gratitude to Dr Charles for their guidance and support at the start of this project in helping me start a project based on something I had an interest in. I would like to extend my gratitude to Alex, my project supervisor, who provided invaluable insights, experience, and constructive criticism throughout the whole project and without his help this project would not have been possible. Furthermore, I’m grateful to Ms Kimia, my secondary marker, for their valuable feedback and insights that helped evaluate, manage, and plan the future of the project. Finally, I’m thankful to my peers who helped in proving useful feedback and insights that helped evaluate and improve the project.

**Thank you everyone for their support and guidance throughout this journey.**

Table of Contents

[**1.** **Introduction** 7](#_Toc134822380)

[**Research Question or Problem that will be Addressed** 7](#_Toc134822381)

[**Aims** 8](#_Toc134822382)

[**Objectives** 8](#_Toc134822383)

[**Legal, Social, Ethical and Professional Considerations** 9](#_Toc134822384)

[**Background** 10](#_Toc134822385)

[**Lanes and Map Structure** 10](#_Toc134822386)

[**Roles** 11](#_Toc134822387)

[**Game’s Objective** 12](#_Toc134822388)

[**Report Overview** 13](#_Toc134822389)

[**2.** **Literature & Technology Review** 14](#_Toc134822390)

[**Literature Review / Related Work** 14](#_Toc134822391)

[**Technology Review** 15](#_Toc134822392)

[**Project Management Tools** 15](#_Toc134822393)

[**Programming Languages** 16](#_Toc134822394)

[**3.** **Design** 18](#_Toc134822395)

[**Use Case** 18](#_Toc134822396)

[**Requirements** 19](#_Toc134822397)

[**Application Architecture** 20](#_Toc134822398)

[**Alternative Approaches** 20](#_Toc134822399)

[**Project Management** 21](#_Toc134822400)

[**4.** **Implementation or Results** 23](#_Toc134822401)

[**Data Sourcing and Cleansing** 23](#_Toc134822402)

[**Scraping Data** 23](#_Toc134822403)

[**Transforming Data** 24](#_Toc134822404)

[**Testing Initial Visualisations** 26](#_Toc134822405)

[**Visualisations** 27](#_Toc134822406)

[**Visualising Gold Difference Over Time Using a Line Chart** 27](#_Toc134822407)

[**Improvements to gold difference visualisation** 29](#_Toc134822408)

[**Visualising Top Gold Over Time Using a Bar Chart Race** 30](#_Toc134822409)

[**Visualising Jungle Gold Over Time Using A Sortable Bar Chart** 31](#_Toc134822410)

[**Visualising Mid Gold Over Time Using A Heatmap** 32](#_Toc134822411)

[**Visualising Bot Gold Over Time Using A Lollipop Chart** 33](#_Toc134822412)

[**Visualising Support Gold Over Time Using A Stacked Bar Chart** 33](#_Toc134822413)

[**Global Settings and Page Styling** 34](#_Toc134822414)

[Evaluation 35](#_Toc134822415)

[**5.** **Conclusion** 37](#_Toc134822416)

[**Reflection** 38](#_Toc134822417)

[Future Work 38](#_Toc134822418)

[**6.** **References** 39](#_Toc134822419)

[**7.** **Appendices** 42](#_Toc134822420)

[**Appendix A – Milestone 2 Project Proposal Form** 42](#_Toc134822421)

[**Appendix B – Milestone 3 Mark Sheet** 1](#_Toc134822422)

[**Appendix C – List of Survey Questions** i](#_Toc134822430)

[**Appendix D – Github Link** i](#_Toc134822431)

# **Introduction**

The project is based on League of Legends which is a multiplayer online battle arena (MOBA) game developed and published by Riot Games back in 2009, and is now one of the most popular games in the world with over 153 million active monthly players worldwide as of March 2023 with a daily peak of over 10 million players [2].

The focus of this project is to visualise data from three regions (LEC – Europe/Middle East/Africa; LCS – North America; and LCK – South Korea) in the professional League of Legends eSports scene from 2015-2017 and see where teams and regions outperformed one another in aspects of the game. These aspects are:

1. Total gold difference in games
2. The individual gold differences between the roles of the game (top, jungle, mid, bot, and support)
3. Total kills in games
4. Monster objectives (Baron Nashor, Dragons, Rift Herald).

The initial data that was sourced from Kaggle [1] will be used to compare these various aspects and present them on an interactive web page that will allow users to explore the visualisations in different manners that may affect the visualisations.

## **Research Question or Problem that will be Addressed**

Listed below are the research questions of the project and they serve a purpose, as the answers found will show how relevant and impactful the factors of the game as visualised are within the game. It will show whether there is a correlation between these factors and the results of teams and regions or whether there is a need for more factors of the game to be taken into consideration.

1. A focus of this project would be to determine which team performs the best within a region based on factors of the game and individual roles.
2. The next question would be which is the best team out of the 3 regions to be compared, which will be judged on the same factors.
3. With this information, the question now would be does it match with the results of the individual region’s rankings and their ranking in the global world championship Worlds.

## **Aims**

For this project, I have three aims which are as follows:

1. To implement useful visualisations that users and the game’s community can use.
2. The ability to determine the strengths and weaknesses of teams and regions.
3. The ability for pro players/teams to identify areas they are weak compared to others to focus on what to improve upon.

These aims are important as they will be used to determine the success of the project as they serve to answer the research questions above.

The first aim would answer the first two research questions of this project to determine which is the best team in a region and the best region overall. This will allow users to make judgements about teams and regions using visualised data that is factual and will encourage them to talk and engage with others within the game’s community.

The next aim of identifying strengths and weakness can relate to all the research questions as by identifying these factors you can select the best team or region. Also, when looking at the third research question the strengths or weaknesses identified can be used to determine the disparities between it and the results.

The last aim relates to the second aim where identified weaknesses then become the aims of improvement for professional players or teams. It can also relate to the third research question as it will aim on lowering the disparity from regional and global achievements, but this is assuming that such weaknesses are improved upon.

## **Objectives**

To achieve the aims of this project the following objectives are to be achieved:

1. Firstly, I need to source the data that consists of data covering the required regions and aspects of the game.
2. Secondly, I would need to scrape more data into the dataset that focuses on other aspects of the game that may not be in the initial dataset.
3. Thirdly, I need to understand and manipulate the data from the dataset to prepare it for use.
4. Fourthly, I would need to research which method of visualisation is best to present the data.
5. Finally, I need to create an interactive webpage that will present the visualisations and data in a clear format that is easy to interpret.

## **Legal, Social, Ethical and Professional Considerations**

The legal issue this project may face is the sourcing of the data from Kaggle [1], and the issues with this might be that the data may be poorly documented with errors and missing values. However, this should not pose a problem as this dataset has been sourced and upvoted a lot by the community where it has also been licensed preventing any legal issues [3]. Also, based on objective 4, if there is a process of scraping further data, there is a process of making sure that it is processed legally.

The main ethical issue that links with legality is the fact that initially, the dataset can be considered to have personal data and would therefore have to abide by the GDA (Data Protection Act). Where personal data is information that can be used to identify an individual (like name, address, phone number, and other identifiers). The only sort of data that can be considered as such are the players’ names, but these are for the most part pseudonyms and publicly available data. However, pseudonymisation is not the same as anonymisation and is considered as ‘the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information’ [4]. As such, with the use of additional information it can lead to an individual’s information and therefore this data is categorised as personal data rather than anonymous data. However, as the data is being used for a legitimate purpose as it refers to the esports matches it should be considered lawful and following of the GPA’s principles.

A social and ethical issue to consider is that these visualisations may lead users to make judgements about the game and misuse it to gamble, which is not the intention of the project. However, as the data covered spans from 2015-2017, there will be no such issues unless the project is expanded to use more recent data, where this factor will be taken into consideration.

A professional consideration is to ensure that the visualisations are correct and that there are no errors made during the manipulating stage to provide factual information. Another thing to consider is that there may be teams and players who have performed poorly and wish for their data not to be used. If that is the case, it will be implemented so that it follows their wishes, and we must be sympathetic as it may misused to cause online harassment otherwise.

## **Background**

League of legends is one of the most popular online MOBA (multiplayer online battle arena) where the main game mode consists of two teams of five players fighting each other of a map, Summoner’s Rift, to defeat each other by achieving the objectives of the game.

In the game you as a player take on the role of a summoner who controls a champion with unique abilities and characteristics where there are currently 163 champions as of the recent patch 13.6 [5]. As you progress through the game your champion gains experiences and levels up to become stronger in terms of base stats and gives you the ability to level up a single ability. The objective of the game is to destroy the enemy team's nexus which is in each team’s respective bases.

Figure 1 - Map of Summoner's Rift [29]

### **Lanes and Map Structure**

The three lanes in Summoner's Rift are called top lane, mid lane, and bottom lane where each lane is guarded by a series of towers and an inhibitor which protect the team's nexus from attack. In addition to your fellow teammates, you also have the help of AI-controlled minions that can help to push into the enemy team’s base or defend your base. Your team’s minions spawn from your nexus every 30 seconds (starting from 1:05) where a wave of them is sent to each of the three lanes [6]. Killing enemy minions give summoner’s gold and experience and hence are a crucial part of the game that helps one build up a lead, where the kill count is tracked and called cs (creep score) or farm.

The towers can be destroyed by attacking them with champion abilities or basic attacks and destroying the enemy team's towers is a key part of gaining an advantage in the game, as it allows players to push into the enemy team's territory and gain access to their base. Inhibitors are also key structures in the game where they are located at the end of each lane and when destroyed spawn super minions for the enemy team. Therefore, destroying the enemy team's inhibitors is an important step towards pushing into their base and potentially ending the game.

Figure 2- Map of a base in Summoner's Rift [30]

### **Roles**

The lanes in Summoner's Rift are also divided into roles, which determine the kind of champion that players should pick to play in that lane. The top laner is usually a champion who can sustain themselves in a solo lane and has strong duelling abilities or is very tanky (very strong in terms of defence) and play to support the other members of the team. They are responsible for controlling the top half of the map and pressuring the enemy's objectives in order to gain leads for the team.

The jungler is a versatile role that moves around the map and farms the neutral monsters in the jungle and for the most part play a supportive role in the game to help gain leads in lanes. This is done by ganking (leaving the jungle and surprising an enemy in one of the lanes with a surprise attack with the intent to kill or pressure them) and providing map control to their team. They are also responsible for securing neutral objectives, such as dragon, rift herald, and Baron Nashor, which give significant bonuses in gold, experience, and buffs to the team that slays them. This responsibility is due to the fact they have a smite ability that deals a lot of damage to neutral monsters and so can be used to last hit objectives to secure them.

The mid laner is typically a champion with high damage abilities and is responsible for controlling the middle of the map and they usually have strong minion wave clear abilities, which allows them to push and roam to other lanes (with or without the jungler) to support the team and to gain leads.

The bot lane is split into two roles the Attack Damage Carry (ADC) and the support where the ADC is responsible for dealing consistent damage from a range distance, whilst the support assists the ADC and the rest of the team by providing crowd control, shielding, healing, and other buffs. Together, the ADC and support have the responsibility for controlling the bottom half of the map, taking objectives, and working to carry the team to victory.

Each role requires a different playstyle and set of skills and knowledge, and players often have a preferred role that they specialize in whilst some play as a jack of all trades. However, to be a successful team, it's important for players to be able to play multiple roles and have a good understanding of the game as a whole as you can then determine win conditions during a game.

### **Game’s Objective**

The objective of the game is to destroy the enemy team's nexus and there are several other objectives that players can work towards to gain an advantage. One of these objectives is the dragon, which appears in the alcove of the bottom river (see Figure 1) and provides a team-wide buff, gold, and experience when killed [7].

Another objective is the Baron Nashor, a powerful monster that appears in the alcove of the top river (see Figure 1). Killing the Baron provides a powerful team-wide buff, gold, and experience that can help push towards the enemy team's base and potentially end the game [8]. However, the Baron is also a difficult and risky objective to take as it is strong and is not something that can be killed quickly allowing enemies to pressure you or potentially steal it. Therefore, for the most part if a team kills the enemy jungler they will look to take the Baron as the odds are much higher in killing it successfully.

What makes this project suitable and challenging is the need to research further into modules previously studied (data science, software development 3, data visualisation, and software engineering) and into new areas within them and are as follows:

1. The research of Python and its libraries in particularly pandas and NumPy where the aim was to manipulate and prepare the data for use. The main challenge being the ability to work arithmetically with different sized arrays that is discussed further in section 3 and 4.
2. The research and learning of d3js which is a JavaScript library and focuses of making interactive visualisations. This was a very new and different experience as it was unlike any of the languages I’ve worked with and proved very challenging but its ability to make interactive visualisations proved its worth.
3. The process of learning to scrape data to improve upon the dataset in a legal and ethical manner.
4. The process of putting together the visualisations to create and style a webpage that would best reflect the visualisations.

## **Report Overview**

The following below will explain and point out the structure for the rest of the report:

* **Section 2: Literature & Technology Review**, will focus on the research gone into this project along with the technologies that were selected to use.
* **Section 3: Design**, will talk about the design of the project along with the steps taken to achieve it. It will also talk about what alternative approaches could have been taken/used in various aspects of the design. Followed by it will be how the project was managed and this will discuss the Kanban method implemented and the tools used and their usefulness and effectiveness.
* **Section 4: Implementation**, will discuss and go into further details in aspects of the methodology giving a more in-depth analysis. Followed by it will be the final results of the project that will show and present the final product and discuss and evaluate it.
* **Section 5: Conclusion**, will conclude the project as a whole and have a section to reflect upon the project and any future work that may or can be done.
* **Section 6: References**, displays all the references and sources that were used throughout the project.
* **Section 7: Appendices**, will contain any further documentation that relates to the project and its management.

# **Literature & Technology Review**

Whilst this project is more focused on the technology there was still research done into the literature side to see whether there were similar cases done for this or other similar games, as the research could help towards this project. When taking a look into papers relating to League of legends data visualisation there were a few interesting papers that had some similar aims and research questions to this project such as [9] [10] [11].

## **Literature Review / Related Work**

One such paper was ‘Comparison of Visualization Tools for Matches Analysis of a MOBA Game’ [9] which investigated using different visualisation tools to analyse League of Legends matches in order to determine which was best to support and analyse player performance. This paper correlates to the third aim of this project and looks at using VisuaLeague II (a tool that analyses data using animated maps and other visualisations that the researchers developed), LoL replay system (the game’s built in match replay system), and OP.GG (the or one of the most used LoL stats search engine) to identify which provides more insightful visualisations. This study determined this by using three tasks to analyse three stages of the game:

1. Early game, usually the first 15 minutes where players are mostly in their lane.
2. Mid-game, where the teams started grouping for objectives.
3. Late-game, where team fights are more frequent and crucial and leads to one team’s victory.

Where it was concluded survey participants preferred VisuaLeague II and the Lol replay system as the visualisations were more adequate than OP.GG in displaying spatio-temporal data. This was a very insightful paper as the results clearly showed participants preferred animated maps to display the data and leads the idea to use similar methods to present data around kills and objectives. Even more so as the dataset includes map coordinates for kills which further reinforces the idea to implement an interactive map visualisation to present the data.

Another paper [10] investigated the idea of live feedback during a game to help improve player’s performances rather than relying on post-game feedback to analyse their performance. This paper shared a similar aim, where it aims to help identify player’s weaknesses to improve their performance, but for newcomers of the game rather than professionals. An interesting fact that shows in the community survey (section 4.1) is that in table 2 it shows that cs per minute is the largest factor players set as a goal for improvement which relates to objective 2 of this project. Where the dataset does not have data regarding cs and so there is the objective of scraping this data as it factors into a core part of the game.

The last paper looked at [11] focused on the ADC role and how a player’s familiarity with the role affects the team’s performance. Whilst this paper does not directly correlate to any of this project’s aims it is similar in looking at a player’s strength or weakness based on their familiarity with the ADC role. Where it concludes that having a role familiarity can have a significant impact of the team’s performance, showing that experience matters.

## **Technology Review**

This next section investigates reviewing the technology used for the projects and the reason why including alternatives that were considered when choosing them. The key things considered when choosing these technologies was how effective they were and if there was any experience with them or similar alternatives. For project management tools, it was factored more to if there was experience with it as that is more important than to learn to use a new alternative as it would take time away from working on the project. Whilst for technologies based on the coding, the main factor was at how effective they were where past experience didn’t matter as a factor as the coding aspect is meant to be the challenge and learning curve to ensure the suitability of the project.

### **Project Management Tools**

GitHub is used for version control, so it allows for the storage and tracking of the project and was chosen as it is a widely used and industrially recognised platform for code hosting. Due to this, integrating a project is very straightforward, with a lot of tools and resources that can be easily accessed by anyone. Also, having previous knowledge and experience working with GitHub in previous modules makes it the best choice compared to others. Whilst there are other alternatives that could be used, like GitLab and Bitbucket, there is no advantage to considering them, so it makes GitHub the best choice. As previously mentioned by resorting to use a new technology to store and version control the project it unnecessarily takes time away from the project.

ClickUp was used to manage the project in a Kanban style, where it kept track of all the tasks needed to complete the project. It also has an option that can display all the ongoing tasks in a Gantt chart, which is very useful, and it provides a clear overview of the project and its timeline. There were other project management tools that could have been used, such as Trello, but they were simply chosen over because of the layout preference. Trello uses a card layout, whereas ClickUp uses a simple list layout, which I simply prefer and hence why it was chosen. Whilst there is previous experience with Trello, working with ClickUp posed little to no challenge to use as with most of these tools being quite straightforward to learn and use with many similar aspects.

### **Programming Languages**

Python was chosen for testing, manipulating, and preparing the data for use due to its numerous libraries, which allow it to be very flexible. Due to having previous experience working with Python in several previous modules, there isn’t a steep learning curve. Apart from manipulating data, it can also be used to visualise data using other libraries, which makes it perfect for starting the project. There were a few alternative languages that could’ve been considered alternatives, like JavaScript or C/C++, but Python has a much lower learning curve because it’s much easier to learn. Jupyter Notebook was used as the environment to code Python in, and this was because it made executing bits of code easy to observe and debug. It is also heavily used for data exploration, as it is simple to perform iterations while observing the outcome, which is ideal for this project. Also, having used it in previous modules makes it easy to work with as there is familiarity with it and no steep learning curve to use the environment.

Structured Query Language (SQL) would’ve been the best alternative to use as its main focus is working with datasets, but while there is some experience with it in previous models, it was the bare fundamentals. However, this project could’ve used SQL as an alternative approach, where the results would’ve been the same, but the time taken would be more drastic. The reason being is that when looking at the dataset for this project it is considered a small dataset, and so to connect to the database to read the data wouldn’t improve the execute time than to read from a CSV or JSON file. Also, as there are only simple queries that will be used, there isn’t a need to use SQL as JSON has as advantage in parsing the data quicker on the server side [12].

ParseHub would be used to scrape more data for the dataset regarding other aspects of the game. The reason it was considered was due to it being one of the best free data scraping tools available that required little to no programming knowledge [13]. As data scrapping is a very prominent part of the industry, there are only a few free alternatives like Octoparse, Crawly, and ScrapingBot which can be used. Where the reason ParseHub was chosen over the alternatives was because it was described to be able to extract data from the most complex of webpages, which was useful after briefing looking into the sites to source the data.

D3.js which is a JavaScript library, was chosen as the method of visualisation because of its ability to create dynamic and interactive visualisations. Due to it being very flexible, it allows for one to create very interesting visualisations, but the learning curve is very steep and that is what makes this project challenging. However, because it is integrated with JavaScript it makes web page implementation much easier in the long run. An alternative approach could’ve been to use Tableau, but it wouldn’t make the project worthwhile as it’s easy to learn and use tool. The reason it was chosen over it was due to d3js being much more flexible and integrates with websites better making the styling much easier.

HTML and CSS will be used for the front end of the webpage, and this is simply because it is the most standard and core languages when creating webpages. HTML is used to create the structure of the webpage whilst CSS creates the styling where an alternative approach would be to use Pug which is HTML is a more readable style. Whilst it makes the code look cleaner it would be better to work with HTML and CSS as there is a lot more documentation around it which makes bug fixing easier.

# **Design**

This section will investigate the design of the product built and the foundations of how it was built up from including the use case, requirements, application architecture, alternative approaches, and project management.

## **Use Case**

|  |  |
| --- | --- |
| Actors | Player base of the game or fans of the Esports scene are the targeted audience for this project. The reason for using it would be to simply help satiate their interests towards their favourite player, team, region or the whole Esports scene. |
| System boundary | The web application where the actors can view and interact with visualisations. |
| Goals | 1. Allow the actor to reach a clear conclusion on which is the best performing team in a certain region in a certain aspect of the game. 2. Determine which is the best region and what aspects of the game they are the best in. 3. Make comparisons with other teams in same or different regions and see which is the best overall or in certain aspects or roles of the game. 4. See if whether these derived conclusions change when side selection is taken into consideration. 5. Identify a team/region’s strengths and weaknesses to identify aspects of the game that need to be worked on. |
| Preconditions | Having a basic understanding of the game. |
| Basic Flow | The actor to look and interact with the visualisations to make judgements of which team/region is the strongest at the selected time condition. |
| Alternative Flows | 1. Check whether the determined strongest team placed 1st in rankings. 2. Check a team/region’s strength or weakness changes over a selected period of time. |

Table 1- Use case

## **Requirements**

|  |  |  |  |
| --- | --- | --- | --- |
| Reference | Requirement | Description | Priority |
| R-01 | Accessible in modern web browsers | Ensure it is supported for the latest versions of Chrome, Firefox, Safari, Opera, etc. So that the runtime is consistent regardless of browser | Must have |
| R-02 | Ability to select desired portions of data | Allow users to filter data to meet their desires | Must have |
| R-03 | Accessibility for users | Ensure the site is well formatted so users can use it as easily as possible | Should have |
| R-04 | Data format | The data presented should be clear and concise to users | Should have |
| R-05 | Performance | No bugs or errors and that the site renders and loads data quickly | Should have |
| R-06 | Ease of updating | Make updating or adding additional data as easy as possible | Could have |
| R-07 | Adding further data | Add data covering other aspects of the game that isn’t within the original dataset | Would like |
| R-08 | Making comparisons | The ability to view a team or region’s standing in the league to compare the data shown to their performance. | Would like |

Table 2- Requirements for project

## **Application Architecture**

A screenshot of a computer screen

Description automatically generated with low confidence

The textbox at the top will briefly introduce the webpage and its intended purpose.

The 3 buttons will each direct to a new page that will have the same visualisations but for a specific region only.

The year and season will be drop-down list that can filter the data for all visualisations.

Followed by this will be a visualisation that is accompanied with a textbox that gives instructions for the visualisations as they are interactive. Where this will repeat for all visualisations.

Figure 3- Webpage Design

The visualisations will be created using the D3.js library that will utilise HTML, SVG, and CSS to pull data from JSON files to create interactive visualisations. Whilst the rest of the page (textboxes, buttons, drop-down list, any text, and background) will be coded using HTML and CSS. The background will be a related League of Legends wallpaper art that best suits the design without overshadowing the visualisations.

## **Alternative Approaches**

An alternative approach that could have been taken was to pull the data from an SQL database using queries instead of JSON files where there are advantages and disadvantages to this.

The advantages would be that the scalability ensures that SQL databases can handle large amounts of data and scale as the data increases whilst it becomes harder for a JSON file. It is also more secure as it provides built in security features, role access control and encryption that will help protect the data which JSON files may not have. Also, using a SQL database allows for higher flexibility when querying as there are many ways to select and aggregate the data whilst with JSON file there may be a need to go through additional processing.

The disadvantages are that is comes with a learning curve as you need knowledge of the SQL language and database structures, which will require time and resources. Another is that when querying a large database, the performance can be slow and hence there may be a need to apply optimisation techniques like indexing.

The design for the project is quite simple with the visualisations centered so that it stands out as much as possible as the central point of view. Similarly, the title, textboxes, buttons, and drop-down lists are centered and evenly spaced for a clear and concise design. There could’ve been changes regarding the design but in my opinion, it wouldn’t affect the weight of the visualisations which is the key point of this project.

## **Project Management**

This project adopted an Agile methodology using a scrum framework where there was a focus of breaking down the project into smaller goals and with there being regular meetings with the project supervisor.

The first reason this methodology was chosen, was due to there being regular weekly meetings with the project supervisor it follows the scrum framework of regular communication so that everyone is on the same page with any issues being addressed as soon as possible. This allows for flexibility for any issues or circumstances that may occur to be dealt with in a timely manner, so the project stays on track. This is important as discussing with the project supervisor can be useful in gaining valuable feedback and insights to ensure the priorities are adjusted to keep the project on track (see section 4 scraping data section). Whilst there was no team there was still a collaboration between me and the project supervisor who as mentioned provides insightful feedback and advice. This allows for the project to stay on track and produce tangible results regularly to ultimately meet the requirements set.

A screenshot of a computer

Description automatically generated

Figure 4- Screenshot of ClickUp taskboard

A screenshot of a computer

Description automatically generated with medium confidence

Figure 5- Screenshot of ClickUp Gantt board

For the task management ClickUp was used (see Project Management Tools) because its list view layout

was personally preferred over Trello’s card system as shown in Figure 4. It also came with an integrated

Gantt board which was first implemented after feedback from Milestone 3 from my secondary marker. It

was useful when planning out the schedule for the remainder of the project and was easier to visualise the

process in doing so.

# **Implementation or Results**

## **Data Sourcing and Cleansing**

The first step in implementing the design of this project was to source the data from Kaggle dataset [1] which was downloaded as a CSV file and then proceeded to be cleansed. This was done in excel by identifying and getting rid of unnecessary columns of data (towers, champions, bans, and address); that data did not correlate to the visualisations intended.

This dataset had over 7000 rows and covered data from over ten regions and so for the purpose of this project was narrowed down to three main regions (LEC, LCS, and LCK). After identifying these values with the ‘League’ column the remaining data was filtered out, excluding data with the values ‘MSI’ and ‘WC’ as they referred to the two global international tournaments. This was done as the scope of the project hoped to include this data as well and visualise teams and regions performance in global tournaments as well.

The final part of the data cleansed was the year period the dataset covered as it ranged from 2014-2018, where the data for 2014 and 2018 were only partial and therefore filtered out. Now the CSV file was ready for any transformation where the original CSV file was kept in the case of scaling the project in the future if there was time to do so.

## **Scraping Data**

The next step was to scale the dataset’s coverage of game aspects and in particular scrape data regarding creep score (cs) as it is a crucial part of the game. To do this a scraping tool ParseHub was used on sites like [14] [15] [16] where the method would be to navigate through the pages and scrape the data regarding cs into a CSV file that would later be merged.

However, this managed to be very challenging as there was a need to navigate through many pages to select the cs data which was too much for the scraping tool. As this was a free version the need to navigate through many pages to selected data for a single game caused the software to repeatedly crash. Therefore, after discussing with my supervisor, it was easily concluded that scraping data should be put side and returned to later depending on the progress of the project.

Another way to scrape the data as suggested by my supervisor would be to find the source of the data using inspect element tool to identify and connect to the API source to scrape the data.

## **Transforming Data**

From here, the data is ready to be transformed in Python and is therefore first read and arranged into a pandas dataframe using the ‘read\_csv’ function. Then a subset of the dataset of chosen which was the 2017 Summer LCK team SKT where the focus was the ‘golddiff’ column to get an element wise average of the games. The ‘golddiff’ column had rows of data in a list format that showed the gold difference in games where the index of the lists represented each minute of a game. Where the challenge was that as not every game lasts the same amount of time and so the lists were all various lengths making arithmetic operations hard.

A screenshot of a computer

Description automatically generated with medium confidence

Figure 6- Transforming the arrays to element wise arrays.

Next to get an element wise average the ‘golddiff’ data was pushed into a NumPy array where it was then used to transform the data, from lists for each game’s gold difference to lists for each game’s minute of gold difference (see Figure 6). This solution to work with different size elements to get an element wise average was found on Stack Overflow [17] but ran into an error in handling the data.

It was found that the NumPy arrays that were initially created were being formatted as a list of lists rather than a NumPy array (see Figure 7). After discussing with my supervisor, he suggested it may be due to the data type of the read data and so after checking it was found the data was being read with tuple type. To fix this it was clear there needed to be changes when the data was read.

A picture containing text

Description automatically generated

Figure 7- Issue with NumPy array

The solution to fix this issue was found after checking the documentation for the ‘read\_csv’ function [18] was to use the convertors property. The converters argument allows you to parse a column and apply a function or value when reading the CSV file, and so the ‘golddiff’ column was given the value ‘pd.eval’. Where the ‘pd.eval’ value converts the string/tuple representation of expression to an Python expression that can handle arithmetic operations [19], and so fixes the issue shown from Figure 7 to Figure 8.

Text

Description automatically generated

Figure 8- NumPy array after fixing data type error.

Now, when looking at the solution to get the element wise average it uses the list function with the ‘zip\_longest’ function from the itertools library [20], and its purpose is described in Figure 6. The function also has an argument called ‘fillvalue’, which by default is set as None but in this circumstance where the arrays are of different lengths and arithmetic operations will be performed, it is given the value ‘np.nan’ [21]. This fills the empty indexes with a value of NaN instead of 0, as a 0 will affect the results when calculating the average.

Now it is converted back into a NumPy array, where all the arrays are of the same length and are ready to be averaged. To work out the average, the ’nanmean’ function from the NumPy library is used, which calculates the mean of the arrays while taking into consideration of any NaN values [22]. The argument axis is set to 1 so that the mean calculation works out the average of each array and outputs an array that represents the average for each game minute (see Figure 9).

Text, letter

Description automatically generated

Figure 9 - Array of element wise average

## **Testing Initial Visualisations**

With the transformed data, there was now the stage of testing the data and making simple visualisations to see if it correlated with the project’s aims and research questions. To do this, data for three teams were used to create a simple line chart using the matplotlib library in Python (see Figure 10).

A picture containing text, line, plot, diagram

Description automatically generated

Figure 10- Basic Line Chart Visualisation

With the intent to create interactive visualisations there were also some testing done using Python specifically the ‘bar\_chart\_race’ library to create an animated visualisation. The results did not meet expectations, and this was due to limitations such as being unable to clearly present negative values, choppy rendering, and inability to clearly identify the team and its value. However, this testing was not in vain as it showed that to create the visualisations envisioned there was a need to research further.

One such alternative considered was Tableau, but I felt that using it would not be challenging and hence unsuitable for this project. When looking into other alternatives, my project supervisor mentioned D3.js where after researching into it and seeing its complex but amazing and flexible visualisations it was perfect for meeting aim 1 of this project.

## **Visualisations**

### **Visualising Gold Difference Over Time Using a Line Chart**

This first visualisation will be created using the transformed data from the ‘golddiff’ column for all 10 teams in the LCK Summer 2017 league representing the average gold difference. There will also be a drop-down list that will be used to filter and update the visualisation based on the game’s side selection which can be either both, blue and red. The first thing done before creating the first D3 visualisation was in changing the data’s format from CSV to JSON and the reason for this was that the data handles arrays which JSON files support. This saves time in cutting down on any additional parsing required when reading the data as a CSV file and hence improves the performance meeting require R-05.

Once the data has been loaded and parsed into a variable the next step is to setup the Scalable Vector Graphics (SVG) container which will display the chart. This is done by assign the containers dimensions to variables and then using the ‘d3.select’ function to assign them as shown in Figure 11. Where the SVG is also appended a value ‘g’ which is discussed later and represents the combination of SVG group elements which in this case are the lines of the chart and its features.

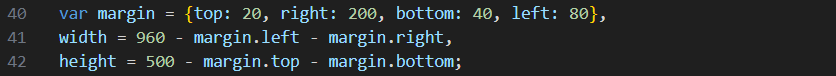
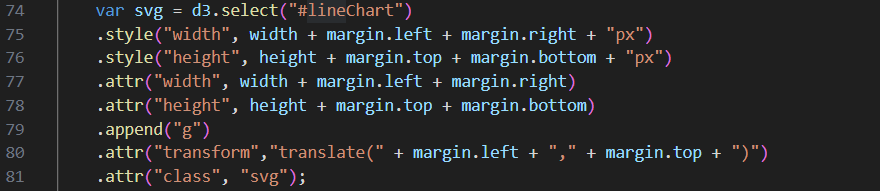


Figure 11- Setup SVG container

The next step is to set up the scales of the line chart where the x-axis will represent the time of the game in minutes and the y-axis will represent the gold difference value. As one might tell from looking at the line chart created in Python the ranges for the values are significant, and so rather than using predefined values an accurate range is calculated. This was done by calculating the maximum value of the x-axis hence the length of the longest array and the maximum and minimum value of the y-axis as is ranges from high positive and negative values. Once these values are calculated, the range of the chart is set according to these values and the axis are finally setup along with their labels.

The next stage for this visualisation is to define and setup the lines for the chart which was done with the help of a similar found chart [23] where it uses the ‘d3.line’ function to do so. From here some further attributes are appended such as the lines colour, which is unique for each team, the lines width, and its mouseover function.

A screen shot of a computer code

Description automatically generated with low confidence

Figure 12- Mouseover function

Above shows the mouseover function which when a mouse hovers over a line will decrease the opacity of the remaining line, and the tooltip is selected where it will show the team’s name. It calculates where the mouse is using the ‘event.pageX/Y’ function where as you can see the y coordinates are decreased by 60 pixels and this is due to the space above the chart taken for the title and drop-down list. The tooltip contains CSS style properties that will be applied when a line is hovered over such as colour, font, padding, and etc. Below in Figure 13 you can see the final output of the first visualisation covering gold difference with the mouseover function being displayed as well.

A picture containing text, diagram, plot, line

Description automatically generated

Figure 13- Initial first visualisation

### **Improvements to gold difference visualisation**

After successfully creating the first visualisation there were both points of self-evaluation and feedback from my project supervisor that were addressed in our meeting. The points of self-evaluation were that it was hard to differentiate when the line dipped into the negative especially in the first 20 minutes as shown in Figure 13. The solution to this was to bring the axis up to start where the y value is zero to clearly differentiate between positive and negative values. Also, where many lines overlap it causes the mouseover function to keep jumping from one line to another so by adding a function to select a single line it would make the visualisation better. The point my supervisor made was similar in that it was hard to tell the values at certain points and that it would be ideal to add the coordinates to the mouseover function. This feedback was very insightful as by doing so I would make the visualisations clearer and more useful reaching towards aim 1.

A picture containing text, diagram, line, plot

Description automatically generated

Figure 14- Golddiff visualisation after improvements

As shown above in Figure 14 the changes were implemented and there is a stark contrast in the visualisation in both design and functionality.

### **Visualising Top Gold Over Time Using a Bar Chart Race**

The second visualisation method was a bar chart race which was first tested upon in Python and was chosen because examples online showed a much smoother and more expressive visualisation. A bar chart race is when the bars are horizontal and move to represent the change in data where an online source reserved as a reference when building it [24]. Following the same basic principles, the data was parsed and then the SVG dimensions and margins were created and set. For this visualisation the y-axis is set to range to ten to represent each team whereas the maximum and minimum values are calculated for the range of the x-axis representing the gold value. A variable called ‘tickDuration’ is created and given a value of 1000 where this will affect the speed of the animation and can be adjusted based on user feedback. Where one possible future feature would be the ability for users to change the value themselves to a speed that suits them best. Similar to the line chart it has a mouseover function that performs the same functions in showing the team’s name, gold value and minute in game.

The points for this chart to work were for the animation transitions of the bars as the data was updated from one value to the next. For the transition from one value to another the function ‘d3.transition’ was key and it allows for smooth change rather than an instantaneous change which would be very choppy and confusing. These transitions function are wrapped under a function called ‘d3.interval’ which is used to repeatedly execute a function with a set interval. Together, the two functions are key in allowing the bar chart race to smoothly transition and display the data. The chart also comes with a reset button that will restart the animation and this is for users to save time from reloading a page to restart it instead.

A screenshot of a computer

Description automatically generated with medium confidence

Figure 15- Bar Chart Race visualisation

### **Visualising Jungle Gold Over Time Using A Sortable Bar Chart**

The third visualisation was for a sortable bar chart that allowed for users to sort from the highest or lowest values where the idea was sourced and referenced from [25]. It follows the same principle of all visualisations in that it sets up the SVG’s dimensions, axis, and mouseover function using the same methods as the line chart. The unique aspect of this chart is that it has a sort function that can update and sort the data from either the highest or lowest value.

A picture containing text, screenshot, plot, font

Description automatically generated

Figure 16- Sortable Bar Chart

### **Visualising Mid Gold Over Time Using A Heatmap**

The fourth visualisation was a heatmap that allowed for users to see a colour gradient change of the mid gold value where the idea was sourced and referenced from [26]. It also follows the same principle of all visualisations in that it sets up the SVG’s dimensions, axis, and mouseover function using the same methods as previous visualisations. The challenge was similar to the problem faced when transforming the data in that as the arrays were of different lengths there were many empty squares in the heatmap that needed to be filled. This was done by normalising the arrays to the length of the largest array where the gold was set to zero (further details in Global Settings and Page Styling). The method to create the squares of the heatmap was the use of the ‘bandwith’ function that will evenly divide the charts area by the domain value which in this scenario is the largest index value of all the array’s data.

A screen shot of a computer

Description automatically generated with low confidence

Figure 17- Heatmap Visualisation

### **Visualising Bot Gold Over Time Using A Lollipop Chart**

The fifth visualisation utilised near similar attributes and functions of the sortable bar chart in the fact that this lollipop chart also had a sort function. This visualisation was chosen to test and compare the effectiveness of the visualisation later when conducting the survey by changing the bars to a lollipop. The method of creating the lollipop was by creating a line and circle SVG shape and joining them to replicate a lollipop shape and this idea came from [27]. Where a Cleveland dot plot [28] was initially considered for the ability to compare two teams similar to other visualisations, however was decided against as there may be a case of many close and overlapping circles affecting the visualisation.

A picture containing text, screenshot, font, line

Description automatically generated

Figure 18- Lollipop Chart Visualisation

### **Visualising Support Gold Over Time Using A Stacked Bar Chart**

The final visualisation was near similar to the sortable bar chart in the fact that the difference was that as the name suggest it stacked two bars on top of one another to compare two teams. This visualisation did not have any extra functionality like the sort function and is a straightforward visualisation that gives you the ability to compare are two unique teams. The method of creating the stacked bar chart was similar to the bar chart where the difference lies in the ‘d3.stack’, ‘d3.nest’, and ‘d3.rollup’ functions.

The stack function takes arrays of data and joins them together in layers where from one layer it calculates the cumulative value when adding another layer. Where the nest function allows you to group an array of data by using keys and with the rollup function it calculates aggregate values for each index of data. These three functions together ensure the data is correctly joined according to index with the total height being accurate against the axis.

A screenshot of a computer

Description automatically generated with low confidence

Figure 19- Stacked Bar Chart Visualisation

### **Global Settings and Page Styling**

This section will cover the data transformation, data filters, additional features, and webpage styling and present the final webpage pre survey evaluation (where there may be changes due to the feedback).

As seen with some of the visualisations, most clearly in the heatmap, there were areas of filled in data with values of zero. The reason the data was normalised to the size of the largest array was because it would not be possible to work with the data in creating the visualisations and therefore was crucial. One piece of feedback my project supervisor previously gave when initially creating the first visualisation was for the mouseover function to show the coordinate values. Along with this feedback it was also mentioned that by adding the counts of the game for each index it could present the data in a more informative matter. This is because as the longer the game is the lower the game count hence why when you look at the golddiff visualisation the lines get quite erratic past 40 minutes. The game count data was gathered in Python by summing the count of the array in a separate variable before the ‘nanmean’ function was applied. This data was also normalised before being added to the mouseover function to provide users more information about the data used.

All the visualisations have a filter for ‘side’ which filters the data to either present data for all games regardless of the game’s side or for either red or blue side. At the top of the webpage there had been two global filters added which can filter the data by the year and the season (spring or summer). Where the plan is to move the side filter to a global filter as it would sense for the data filters to be all located in one space rather than being separated.

The design of the webpage managed to reach the expectations as shown in Figure 3 where I am satisfied with the outcome. The design was quite straightforward with everything being centered in the middle and evenly spaced where there may be changes based on the survey feedback.

A screenshot of a computer

Description automatically generated with medium confidence

Figure 20- Webpage pre-survey

## Evaluation

The method to evaluate the outcome of this project and to see if it met the aims and objectives set was to conduct a survey between peers and family. The reason for this is because it contains a mix of those who play or are familiar with the game and people who aren’t. Although it is a small sample to evaluate from it is the best method to due to the size of project and timeframe as by including members of the public there would be a need to conduct a lot of work to follow the GPA.

The aim will be to have a survey roughly by the end of the March and this will consist of all visualisations for a specific region to test the visualisation’s effectiveness. Where then hopefully a week or two later there can be another similar survey and this will be done to see how effective the changes made were after the initial feedback from the first survey. For the list of survey questions look at Appendix C – List of Survey Questions where on Github (Appendix D – Github Link) you can find a CSV file of the results.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Visualisation | Highest answer | Percentage | Correct | Average Time Taken |
| 1 – Line Chart | ESC | 62.5 | Yes | 0-10s |
| 2 – Bar Chart Race | SSG | 75 | Yes | 21-30s |
| 3 – Sortable Bar Chart | AFS | 75 | Yes | 11-20s |
| 4 – Heatmap | ESC | 87.5 | Yes | 11-20s |
| 5 – Lollipop Chart | SSG | 87.5 | Yes | 0-10s / 11-20s |
| 6 – Stackable Bar Chart | SKT | 100 | Yes | 0-10s / 11-20s |

Table 3- Results of survey

The table above shows the results of the survey where the sample size was quite small at 8 it still shows some evidence to the visualisations effectiveness. All the visualisations were answered correctly for the majority where the first visualisation had the weakest result, and this can be easily identified as the hardest question as the lines are overlapping one another making it hard to judge. When looking at the average time taken to reach an answer based on frequency the average was 11-20s which to me feels like it meets aim 1 in providing an effective and useful visualisation. The oddity being the bar chart race taking from 21-30s can be explained as time is spent waiting for the animation to progress to the point of the question. This also meets aim 2 of this project as the users were able to compare and identify the strengths of a team clearly and correctly.

A picture containing screenshot, line, plot, diagram

Description automatically generated

Figure 21- Ranking of visualisations.

Figure 21 shows how the survey users ranked the visualisations from best to worst and there are clear takeaways that can be derived from this. The line chart proved to be the best visualisation by far with only one user of the seven disliking it, proving it to be the most effective visualisation of the six. When comparing the sortable bar chart and lollipop chart which can both be said to be similar but different in design the results show users did not like either very much compared to the others. The outcome I was expecting was for one of these two to be rated higher than the other to determine which was a more effective visualisation, but unfortunately it seems both were not the right choice to represent this data.

Overall, I’d say that I was partially successful in meeting the first two aims of this project as not all the visualisations created were as effective as one another. With the additional written feedback at the end of the survey there is now rooms of improvement to make such as the graph titles, background image and moving global filters (see Appendix D – Github Link for survey results ) that is now the area of focus.

A screenshot of a computer

Description automatically generated with medium confidence

Figure 22- Changes based on user feedback

Looking at Figure 22 above, there were changes made to the styling especially the background to a dark colour as suggested which when comparing to Figure 20 looks much better and makes the text clearer to read. There were also titles added to the visualisations to identify them more easily where there was also a missing axis label added to the heatmap. The final thing changed was that the year and season drop-down lists now stick to the top of the page as you scroll, which makes it much more convenient to use.

# **Conclusion**

In concluding this project, I have managed to create a webpage with six visualisations that provide insightful information for users about the LoL Esports Scene. When looking back at my original aims, I can confidently say that I have met the first two aims but was unfortunately unable to meet the last one. This statement can be backed from evidence provided in the Evaluation, where all the visualisations were answered correctly for the majority and answered in a short timeframe showing their effectiveness.

## **Reflection**

I feel very satisfied with the output created where it has been an invaluable and insightful project and was my most challenging piece of work. What went well was that I managed to create six insightful visualisations and learned a lot about visualising data and learnt an expressive library for it. As mentioned before the first two aims of this project were met were unfortunately the last was not where in hindsight, I feel this aim would be unachievable unless I had started coding the visualisations in December. The reason for this was because by separating and querying data based on the individual player it would require a more complex data transformation and structuring than currently used.

Another factor that did not go well and something I have very much learned upon was planning the future of a work. This was first brought up during the Milestone 3 meeting (see Abstract asasd) where the future stages of the project were simply listed where the time aspect was not considered. This in my opinion lead for me to start coding the D3 visualisation much later than I initially expected and eventually meant that I was even unable to get a second set of feedback which was planned. Therefore, one useful skill that I’ve learnt is to plan and organise my work in a more professional manner whether than be to use Gantt charts or another method.

## Future Work

There is a lot I would like to work on in the future to scale this project further in factors that were not met during this project. One factor is the aforementioned aim three that was not met where whilst there would be a lot planning involved to create an efficient database it would be very fun. Another would be to scale the data and scrape further data, which was first envisioned but dropped due to difficulties where this is a field I would like to delve more in depth to as I only managed to scrape the surface of it. The final factor which was briefly mentioned (Legal, Social, Ethical and Professional Considerations)was to scale the data from 2017 to more recent years as there would be more interest from users to this data than older data.

# **References**

1. <https://www.kaggle.com/datasets/chuckephron/leagueoflegends>
2. <https://activeplayer.io/league-of-legends/>

1. <https://creativecommons.org/publicdomain/zero/1.0/>
2. <https://edps.europa.eu/system/files/2021-04/21-04-27_aepd-edps_anonymisation_en_5.pdf>

1. <https://www.esports.net/news/lol/how-many-champions-are-in-league-of-legends/#:~:text=How%20many%20Champions%20are%20there,live%20servers%20in%20Patch%2013.6>.
2. <https://leagueoflegends.fandom.com/wiki/Minion_(League_of_Legends)>
3. <https://leagueoflegends.fandom.com/wiki/Dragon_pit_(League_of_Legends)>
4. <https://leagueoflegends.fandom.com/wiki/Baron_Nashor_(League_of_Legends)?so=search>
5. A. P. Afonso, M. B. Carmo and T. Moucho, "Comparison of Visualization Tools for Matches Analysis of a MOBA Game," 2019 23rd International Conference Information Visualisation (IV), Paris, France, 2019, pp. 118-126, doi: 10.1109/IV.2019.00029. URL: <https://ieeexplore-ieee-org.roe.idm.oclc.org/stamp/stamp.jsp?tp=&arnumber=8812067&isnumber=8811888>
6. Frans Rijnders, Günter Wallner, and Regina Bernhaupt. 2022. Live Feedback for Training Through Real-Time Data Visualizations: A Study with League of Legends. Proc. ACM Hum.-Comput. Interact. 6, CHI PLAY, Article 243 (October 2022), 23 pages. <https://doi.org/10.1145/3549506>
7. Joshua A. Eaton, David J. Mendonça, Matthew-Donald D. Sangster, “Attack, Damage and Carry: Role Familiarity and Team Performance in League of Legends”. SN 2169-5067, 2018/09/01. doi: 10.1177/1541931218621030. URL: <https://doi-org.roe.idm.oclc.org/10.1177/1541931218621030>
8. <https://www.geekboots.com/story/advantages-and-disadvantages-of-json-over-sql>
9. <https://builtin.com/data-science/free-web-scraping-tools>
10. <https://gol.gg/esports/home/>
11. <https://www.factor.gg>
12. <https://oracleselixir.com>
13. <https://stackoverflow.com/questions/60366450/numpy-mean-std-dev-of-arrays-of-various-lengths>

1. <https://pandas.pydata.org/docs/reference/api/pandas.read_csv.html>
2. <https://pandas.pydata.org/docs/reference/api/pandas.eval.html>
3. <https://docs.python.org/3/library/itertools.html#itertools.zip_longest>
4. <https://numpy.org/doc/stable/reference/constants.html#numpy.nan>
5. <https://numpy.org/doc/stable/reference/generated/numpy.nanmean.html>
6. <https://d3-graph-gallery.com/graph/line_several_group.html>
7. <https://observablehq.com/@d3/bar-chart-race-explained>
8. <https://observablehq.com/@avis-n/d3-sortable-bar-chart>
9. <https://d3-graph-gallery.com/graph/heatmap_tooltip.html>
10. <https://d3-graph-gallery.com/graph/lollipop_basic.html>
11. <https://d3-graph-gallery.com/graph/lollipop_cleveland.html>
12. figure 1 <https://www.deviantart.com/narishm/art/League-of-Legends-Summoner-s-Rift-Regions-Map-413345149>
13. figure 2 <https://www.riftherald.com/2016/9/29/13027318/lol-guide-how-to-watch-play-intro>

# **Appendices**

## **Appendix A – Milestone 2 Project Proposal Form**

**Note:** Project Supervisor was changed in November

MILESTONE 02: Project Proposal

CMP040X301Y: Final Year Project

Your first and last name

Thamiliniyan Aravinthan

Your Student ID

ARA20480536

Your Programme

BSc Computer Science

What is the working title of your project (this can be changed at a later date)?

Regional Differences in Pro League of Legends

First Supervisor

Momina Shaheen

Second Supervisor (second marker)

Kimia Aksir

Classify your project as a technology theme (i.e. How you want your project to be scrutinsed)

Data Analysis/Visualisation

If you selected "Other" above, please specify your theme below.

Describe your project

This project will based on League of Legends which is one of the largest and popular online game in

the world currently and the focus will be on the esports scene cited as the world's largest. A common theme as with many sports is the competitiveness by everyone as to who is the best team and who is the best country/region. The same can be said about the League of Legends esports scene where the community is just as passionate.

The goal of this project is to gather data from across several regions of the league worldwide and then to visualise and analyse the data to show which is the best region and their rankings compared to others. This will be presented on a web application that will give users the option to compare regions with various aspects of the game.

This project is envisioned for the game's community that are very passionate with regards to their region's strength especially so in international tournaments. Where the hope in presenting this is for the community to use the web application to make and get clear understandings and judgements from the data presented rather than using the current tabular format that is available but obscure.

List up to 3 aims of your project.

NOTE: An aim is an expected outcome of your project (e.g., issues it will address, how it might improve or enhance a situation for stakeholders, etc.)

1)

To visualise and analyse the data of regions and show differences between them in various

aspects

2)

Generate useful general data for the game's community

3)

Can allow pro players to see areas they are weak in compared to the players on top so that

they can focus on what to improve upon.

List up to 5 objectives of your project.

NOTE: Objectives are tangible tasks that you will complete. They are typically steps/activit- ies that you must complete in order to deliver your project aims successfully.

1)

Obtain the data on the pro league matches.

2)

Sanitise the data and format it.

3)

Research into the tools that will be used to visualise the data like Grafana.

4)

Investigate development environment to find what's best suited to present the data like

using elk stack.

5)

Have users test the web app to get feedback and make changes where required.

List background/literature/technology review sources, that have been used to inform your project.

Background sources and data sources

https://en.wikipedia.org/wiki/League\_of\_Legends https://[www.kaggle.com/datasets/chuckephron/leagueoflegends](http://www.kaggle.com/datasets/chuckephron/leagueoflegends) https://gol.gg/esports/home/

Technology sources https://[www.elastic.co/elastic-stack/](http://www.elastic.co/elastic-stack/) https://nodejs.org/en/ https://expressjs.com https://grafana.com/grafana/ https://[www.mysql.com](http://www.mysql.com/)

Describe any risks, ethical issues or other factors that have been considered as part of developing this project proposal.

An ethical issue to consider is that the project follows the gdpr (general data protection regualtion)

even thoough its public data to ensure data is keep securely and that no pii (Personally Identifiable Information) is made available or used in project.

A risk to consider is that there may be steep learning curves when learning to use the visualisation tools. If such a situation does occur it will have to result in using much more basic tools to visualise the data despite lower quality.

**Student and First Supervisor Project Sign Off**

**STUDENT:** I agree to completing this project:

✔

Student Name: Student Signature:

✔

Thamiliniyan Aravinthan

**Date:** / /

**SUPERVISOR:** I approve this project proposal:

11

10

**Date:**

1 / 12 /

2022

2022

Supervisor Name: Supervisor Signature:

Alex Collins

Alex Collins

It is the supervisor’s responsibility to approve this project as meeting the requirements for the module. This includes professional body requirements, programme requirements, and module requirements. By signing the form, you are agreeing that you have validated the suitability of the project.

## **Appendix B – Milestone 3 Mark Sheet**

BSc Final-Year Project- CMP040X301Y

Milestone 03: Mid-Project Demonstration, Project Management and Project Report Progress

|  |  |
| --- | --- |
| Student Name | Thamiliniyan Aravinthan |
| Student ID | ARA20480536 |
| Project Title | Regional Differences in Pro League of Legends |
| Supervisor | Alex Collins |
| Second Marker | Kimia Aksir |

**For this Milestone, students will need to complete the following:**

## Contact their supervisor and second marker to arrange a date and time for their Milestone 03: Mid-Project Review.

1. During the meeting, they will need to demonstrate:
   * Latest progress with their technical artefact/implementation.

Please indicated progress using the rubric on Page 2.

## Evidence that they have been managing their project.

Please indicated progress using the rubric on Page 3.

## Latest progress with their project report

Please indicated progress using the rubric on Page 4.

## NOTE:

* This form should be completed by the **second marker**.

## It is the student’s responsibility to organise and coordinate the meeting for Milestone 03: Mid-Project Review.

SECOND MARKER FEEDBACK: **STUDENT’S ARTEFACT/IMPLEMENTATION DEMONSTRATION**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| RUBRIC FOR STUDENT’S ARTEFACT/IMPLEMENTATION DEMONSTRATION | | | | | |
| Based on the artefact and/or implementation demonstration, it is evident that excellent progress has been made for this stage of the project.  For example, in terms of project complexity, work is of an intermediate/advanced level and demonstrates an extensive level of completion. | Based on the artefact and/or implementation demonstration, it is evident that very good progress has been made for this stage of the project.  For example, in terms of project complexity, work is of an intermediate/advanced level and demonstrates a very good level of completion. | Based on the artefact and/or implementation demonstration, it is evident that good progress has been made for this stage of the project.  For example, in terms of project complexity, work is of an intermediate level and demonstrates a good level of completion. | Based on the artefact and/or implementation demonstration, it is evident that adequate progress has been made for this stage of the project.  For example, in terms of project complexity, work is of a somewhat basic level and demonstrates adequate levels of completion. | Based on the artefact and/or implementation demonstration, it is evident that inadequate progress has been made for this stage of the project.  For example, work is very basic and underdeveloped. | A demonstration was not given, OR  the artefact and/or implementation demonstrated is representative of failing work for this stage of the project. |

SECOND MARKER FEEDBACK: **STUDENT’S PROJECT MANAGEMENT EVIDENCE**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| RUBRIC FOR STUDENT’S PROJECT MANAGEMENT EVIDENCE | | | | | |
| Project management evidence is excellent.  For example:  There is excellent evidence of a project plan.  A dedicated project management tool is actively used all of the time and is fully up to date.  There is excellent evidence of how project activities are captured and tracked (e.g., milestones, timelines, tasks etc).  There is excellent evidence of how project activities are evidenced to the supervisor.  There is excellent evidence regarding considerations of project problems and how they might be mitigated. | Project management evidence is very good.  For example:  There is very good evidence of a project plan.  A dedicated project management tool is actively used most of the time.  There is very good evidence of how project activities are captured and tracked (e.g., milestones, timelines, tasks etc).  There is very good evidence of how project activities are evidenced to the supervisor.  There is very good evidence regarding considerations of project problems and how they might be mitigated. | Project management evidence is good.  For example:  There is good evidence of a project plan.  A dedicated project management tool is generally being used appropriately.  There is good evidence of how project activities are captured and tracked (e.g., milestones, timelines, tasks etc).  There is good evidence of how project activities are evidenced to the supervisor.  There is good evidence regarding considerations of project problems and how they might be mitigated. | Project management evidence is adequate.  For example:  There is limited evidence of a project plan.  A dedicated project management tool is available but is not actually being used.  There is limited evidence of how project activities are captured and tracked (e.g., milestones, timelines, tasks etc).  There is limited evidence of how project activities are evidenced to the supervisor.  There is limited evidence regarding considerations of project problems and how they might be mitigated. | Project management evidence is inadequate.  For example:  There is no clear evidence of a project plan.  A dedicated project management tool is not being used.  It is not clear to see how project activities are captured and tracked (e.g., milestones, timelines, tasks etc).  It is not clear to see how project activities are evidenced to the supervisor.  There is no clear evidence regarding considerations of project problems and how they might be mitigated. | Project management evidence was not shown or is not captured in a manner that the student can evidence. |

SECOND MARKER FEEDBACK: **STUDENT’S PROJECT REPORT PROGRESS**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| RUBRIC FOR STUDENT’S PROJECT REPORT PROGRESS | | | | | |
| The project report is at an advanced stage and a draft is already available for review. | The project report is at an intermediate stage with a substantial level of content in all sections. | The project report is at an early stage with some of the sections containing significant levels of content. | The project report is at a very early stage and is essentially the core template with some place holder content in place. | The project report is at a “scoping out” stage and is essentially the core template only. | No evidence was provided to show that the report had been given any consideration |

# Other Checkpoints

## Has the student been attending supervision meetings? (YES / NO)

1. Does the student know how to evaluate their work? (YES / NO)
2. Overall, has the student made suitable progress so far? (YES / NO)
3. Do you have any concerns regarding this student’s progress? (YES / NO) If “YES”, please indicate concerns below.

NO

Additional Comments (optional)

The choice of method should be appropriate and based on the structure/type of the data you are using. You must also think about how the validation/evaluation of your model will be done at this stage. It is important to validate your model as you are expecting to recommend the “best team” and the “best region”. It is important to establish that you model has required level of accuracy as well. I can suggest you go for confusion matrix or something similar. Do not forget to maintain Gantt chart/ Kanban board to show your consistent progress. Try to estimate how much time you actually need for the front-end development and if necessary you can think about narrowing down the “front-end” part with fewer features only. You are also advised to shape your report with all detailed information from the methodology/methodologies used, artefacts and validation/evaluation process(s).

Good Luck!

Second Marker Signoff

It is the second marker’s responsibility to approve this project as meeting the requirements for the module.

By signing the form, you are agreeing you have validated the suitability of the project.

|  |  |
| --- | --- |
| Second Marker Name | Kimia Aksir |
| Signature | Kimia Aksir |
| Date and Time | 21/02/2022 |

## **Appendix C – List of Survey Questions**

1. Have you played League of Legends before or are familiar with the game?
2. Have you previously looked at any sort of data covering the LoL Esports scene?
3. Looking at the 1st visualisation, which team has the biggest gold difference lead at 20 minutes?
4. How long did it take you to figure the answer out?
5. Looking at the 2nd visualisation, which team has the gold lead at 20 minutes?
6. How long did it take you to figure the answer out?
7. Looking at the 3rd visualisation, when comparing teams ESC and AFS who has the gold lead at 20 minutes?
8. How long did it take you to figure the answer out?
9. Looking at the 4th visualisation, which team has the gold lead at 20 minutes?
10. How long did it take you to figure the answer out?
11. Looking at the 5th visualisation, when comparing teams MVP and SSG who has the gold lead at 20 minutes?
12. How long did it take you to figure the answer out?
13. Looking at the 6th visualisation, when comparing teams SKT and ROX who has the gold lead at 20 minutes?
14. How long did it take you to figure the answer out?
15. Rank the visualisations from best to worst (1-6)
16. When looking at the visualisations is there anything that should added/changed?
17. When looking at the design of the page is there anything that should be added/changed?

## **Appendix D – Github Link**

<https://github.com/Tham-A/Final-Year-Project>