THIS PAGE CAN BE DELETED FROM YOUR FINAL REPORT

GUIDANCE (text in blue can be deleted from your final submission)

**Report Layout and Style**

T­he following guidelines **must** be adhered to:

* All **text will be black** in the report unless in very exceptional circumstance.
* Main body text must be at least **11pt font** using either **Arial or Calibri** font.
* Main body text will have **1.5 line spacing**.
* **Margins** will be a minimum of **2 cm on each side**.
* All **pages** will be **numbered consecutively**.
* **Figures** must have **captions** and be **numbered** (e.g., Figure 1).
* **Tables** must have **captions** and be **numbered** (e.g., Table 1).
* **Figures** may be **black and white**, or **colour**.

**Word Count**

There is a **10,000-word limit** for the project. Do not see this as a target, but rather a limit to the number of words we expect a project to have. All words in the main body, excluding words in figures and tables, will count to your word count. If you think you will go over 10,000 words, you should consider what can be removed from the main body and placed in an appendix.

**Copyright and Intellectual Property Rights (IPR)**

Your report should be written considering that it will be within the public domain. Normally, you retain copyright over your written work and Intellectual Property Rights (IPR) over any technical work. There are situations where this might not be as simple, for example when working with a company or on a larger university project. There are strategies you can utilise:

* You can inform your supervisory team that the project cannot be made available to other students undertaking an BSc Project.
* You can provide a shortened report for sharing that does not contain the protected information.
* You can assign IPR to the external collaborator, although you should take great care when doing so. It is best to speak to your supervisor.
* Non-disclosure agreements can be made between the external collaborator and the supervision team.

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 3 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. Where 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 (further details in Section 2).

Learning outcomes summary

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:** Enter the date here

**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** vii](#_Toc134461340)

[Research Question or Problem that will be Addressed vii](#_Toc134461341)

[Aims viii](#_Toc134461342)

[Objectives viii](#_Toc134461343)

[Legal, Social, Ethical and Professional Considerations ix](#_Toc134461344)

[Background x](#_Toc134461345)

[Report overview xiii](#_Toc134461346)

[**2.** **Literature & Technology Review** xv](#_Toc134461347)

[**3.** **Design** xix](#_Toc134461348)

[**4.** **Implementation or Results** xxi](#_Toc134461349)

[Evaluation xxiii](#_Toc134461350)

[Related Work xxiii](#_Toc134461351)

[**5.** **Conclusion** xxiv](#_Toc134461352)

[Reflection xxiv](#_Toc134461353)

[Future Work xxiv](#_Toc134461354)

[**6.** **References** xxv](#_Toc134461355)

[**7.** **Appendices** xxviii](#_Toc134461356)

# **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 regions 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 on the assumption 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. But 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 give 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 - Map of Summoner's Rift

**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 - Map of a base in Summoner's Rift

**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 the 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 ADC (Attack Damage Carry) 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. So 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**

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 doesn’t 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 doesn’t 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.

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. But 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.

**Summary Table**You might include a table that summarises the benefits and limitations of what you have reviewed, highlighting opportunities for project novelty where identified.

# **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. |

**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 |

**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.

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’ve 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.

**Your Project Management Approach**Under this heading, state your project management strategy and justify your choices compared to other alternatives. Describe the benefits and limitations of your strategy and refer to your supporting project management documentation in the appropriate area of the appendix.

**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 - Screenshot of ClickUp taskboard

A screenshot of a computer

Description automatically generated with medium confidence

Figure - Screenshot of ClickUp Gantt board

For the task management ClickUp was used (see add heading to subtitles to cross ref them) because

its list view layout was personally preferred over Trello’s card system as shown in Figure 3. 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.

* + - What planning did you do?
    - What did your initial plan look like?
    - What dependencies did your plan have?
    - When did you need to have key activities complete by to hit deadlines?
    - How did you monitor progress?
    - How would you adapt your plan?

**Previous text to edit/remove to section 4**

The first step of the design of this project is to source the data from Kaggle dataset [1] which is then cleaned in excel by getting rid of unnecessary columns of data (towers, champions, bans, and address) where then the data is narrowed down to focus on 3 of the main regions (LEC, LCS, and LCK). Now that we have a dataset cleansed and downsized to focus of the project’s aim the next step would be to scrape more data from sites that host data for professional games that focus on other aspects of the game such as cs (otherwise known as creep score/farm, is the count of monsters killed that give gold and experience). To do this a scraping tool, like ParseHub, will be used on sites like [15] [16] [17] where the method would be to simply scrape the data displayed on the site. Another way would be to find the source of the data using inspect element tool to directly connect to the API to source the data.

Once the data has been cleansed, we can then move onto manipulating and preparing the data to use when making the visualisations. To do this Python and its libraries pandas, NumPy, and itertools will be used to perform the manipulations using their respective functions. The pandas library will be used for creating a data frame of the data whilst, the NumPy library will focus on the arithmetical and transformation operations with itertools being used to merge different sized arrays.

The sort of manipulation would be getting average scores of all games played by a team and this data will then be used to explore basic visualisations using the matplotlib library. This will be done to see the shape of the data and see if can clearly visualise the data and to potentially see if there is anything that can be spotted that relates to the aims of the project. Followed by this will be a more advanced visualisation using the bar\_chart\_race library to create an advance visualisation to test its effectiveness in portraying the data. This project could’ve alternatively been done solely within python and its libraries for visualisation however, d3js allows for visualisations that are more interactive, flexible, and better overall.

Once this has be done the next step will be to export the manipulated data into a new csv file that will then be transformed into a json file, so the data is stored in array and object format. Once the json file is ready then next step will be to make the first d3js visualisation and start developing the

# **Implementation or Results**

GUIDANCE (text in blue can be deleted from your final submission)

Once the examination team know what you planned to do, you must tell them what happened -- **What was the outcome of the work you undertook in the project?**

A build or investigative project will discuss the implementation. **Do not just paste in lines of code to your report and call that an implementation! Your report should feature minimum code to only discuss points.** The idea for implementation is to describe how the design has actually turned out.

A research or investigative project will present the results from performing the methodology. These results must be correctly presented, using appropriate tables, charts, and statistical tests that suit the nature of the project. Results should be summarised, and any findings clearly presented.

When cleansing the original dataset within Excel, this was done by selecting the unnecessary columns: towers, champions, bans, and address (broken link for match’s history). Next, by filtering the region column by its value, three of the main regions (LEC, LCS, and LCK) were selected, with all the data regarding the other regions being deleted. The final thing to do was to filter the data once again for its year and delete any data from 2018, and this was done due to the data for 2018 only being partial data from the whole year.

From here, the data is ready to be manipulated in Python and is therefore first read and arranged into a dataframe using the pandas library. Then a variable is created that selects a subset of the dataset and focuses on a single team’s data, where the first focus is the ‘golddiff’ column, which shows the gold difference in a game. The main step is now to implement an element-wise average of the ‘golddiff’ column to show the average gold difference for all the games a team plays in one season. But while it might be easy to get an element wise average of arrays normally, due to the fact that the arrays within the ‘golddiff’ column are of different lengths, as some games are longer than others. The solution to this was found to be using the NumPy and itertools libraries, where the first step is to select the column and make a NumPy array. But this led to an issue where the arrays were being read as strings (see Figure 3).

A picture containing text

Description automatically generated

Figure - Issue with NumPy array

This meant after some type testing that the data being read through the csv file was of the string type, and so to fix this after checking the documentation for the ‘read\_csv’ function [18], using the converters property was the solution. 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 is given the value ‘pd.eval’. And what the ‘pd.eval’ value does is that it evaluates the string expression, returning it as an ndarray, scalar, pandas object, or None [19], and together with the converters argument, the column is read as an ndarray now rather than a string.

Text

Description automatically generated

Figure - NumPy array after fixing data type error.

Now, after fixing the data type error when creating the NumPy arrays, they are created in the desired format (see Figure 4). After this, the list function is used with the ‘zip\_longest’ function from the itertools library [20], and what this did was that it created lists based on the indexes of the array. 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, where each array will represent each minute in a game starting from 0 (see Figure 5 for the result).

Text, letter

Description automatically generated

Figure - Array of element wise average

* Basic matplotlib line chart
* Bar chart race
* First d3 visualisation
* Add values (coordinates) to the hover function.
* Adding the game counts to the hover function.
* Mentioned feedback from supervisor and what it changes? Why is it better? Does it help?

## Evaluation

GUIDANCE (text in blue can be deleted from your final submission)

The examination team now need to know how well the project went -- **How good was the outcome from the project?**

Evaluation is an important element of any project. You must tell your reader how good the final deliverable is. **Your project does not have to be perfect -- indeed the outcomes might have been bad.** The point is you must evaluate the outcome and discuss its strengths and weaknesses.

A key element of this section is a reflection on the aims and objectives set out at the start of the project, and how well these have been met. **Again, it is possible not to achieve an aim or objective.** The point is you evaluate how well you did meet your goals.

## Related Work

GUIDANCE (text in blue can be deleted from your final submission)

Answer the question -- **Who else has done something similar and how does my work compare?**

Another key element of this section is evaluating your work against that of others. How good is your work when compared to other people who have undertaken similar work? It is important to be able to understand how well you have achieved your goals in relation to others, while also considering the time limitations of the project.

# **Conclusion**

GUIDANCE (text in blue can be deleted from your final submission)

The conclusion summarises the project. You need to highlight your key outputs and/or discoveries. There are some particular subsections that must appear in your conclusion.

## Reflection

GUIDANCE (text in blue can be deleted from your final submission)

You must critically reflect on the entire project process and how well you have worked on the project. What particular things have you learned during the project? Why were you able and unable to meet project goals? What would have you done differently in hindsight?

A common approach many students take in this section is to claim poor time management. **Poor time management is rarely a problem unless you had too much to do in the project.** Normally, what is called poor time management is poor organisation, planning, and motivation. Being honest in your reflection will help you understand how you can improve these issues rather than focusing on time management issues.

## Future Work

GUIDANCE (text in blue can be deleted from your final submission)

Answer the question -- **What next?**

You've completed a significant piece of work -- perhaps the largest piece of work you have ever done. But no project is ever 100% complete, and you will have found new ideas along the way. If someone were to pick up your project, what avenues should be explored next?

# **References**

GUIDANCE (text in blue can be deleted from your final submission)

In this section, you **must** reference any sources used in your work. Typically, these sources will have come up during the investigation and related work sections. Your referencing must use the IEEE referencing style [IEEE Citation Guidelines2.doc (ieee-dataport.org)](https://ieee-dataport.org/sites/default/files/analysis/27/IEEE%20Citation%20Guidelines.pdf) .

It is **highly** recommended that you use reference management software such as Mendeley or Zotero.

Many students ask how many references are required. That is like asking how long a piece of string is. Your project should have as many references as is required for it. However, having few references indicates that no thorough investigation has occurred.

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. Paper on Logic Mining in League of Legends <http://pertanika2.upm.edu.my/resources/files/Pertanika%20PAPERS/JST%20Vol.%2028%20(1)%20Jan.%202020/12%20JST-1649-2019.pdf>
11. <https://gol.gg/esports/home/>
12. <https://www.factor.gg>
13. <https://oracleselixir.com>

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>

figure 1 <https://www.deviantart.com/narishm/art/League-of-Legends-Summoner-s-Rift-Regions-Map-413345149>

figure 2 <https://www.riftherald.com/2016/9/29/13027318/lol-guide-how-to-watch-play-intro>

# **Appendices**

GUIDANCE (text in blue can be deleted from your final submission)

Appendices appear after references. Your appendices depend on the nature of your project. **Do not assume people will read your appendices.** Even if you direct them to do so in your main text, appendices are considered additional information and should not be relied upon to understand your main body of work. Refer readers to an appendix using a phrase such as *see Appendix A for further details*.

The following documents **must** be included as references:

* Your Project Proposal.
* Your Progress Review Form.
* Your original plan and revised plans as your project evolved.
* A description of how to access any technical output. **It is strongly recommended you use GitHub or something similar to do this.**

Any important communications between you and external stakeholders -- **please ensure private data is removed and communications anonymised.**

* Things that when talking with someone its not eh core part of story. Additional information
* Appendix A – TASK MANAGEMENTS AND IT LISTS THE TASK OR THE TASKS THAT WERE COMPLETED IN EACH MILESTONE
* Show whole section of code whilst in main body you can talk about specific part
* Input csv format or csv file
* Gantt chart / roadmap template
* Change of supervisor maybe???