

National College of Ireland  
BSc in Computing  
2016/2017

Alexander Hii  
13564383  
alexander.hii@student.ncirl.ie

## **League of Legends: A Detailed Analysis**

Technical Report



## Declaration Cover Sheet for Project Submission

### SECTION 1 *Student to complete*

<b>Name:</b>  
<b>Student ID:</b>  
<b>Supervisor:</b>  

### SECTION 2 Confirmation of Authorship

*The acceptance of your work is subject to your signature on the following declaration:*

I confirm that I have read the College statement on plagiarism (summarised overleaf and printed in full in the Student Handbook) and that the work I have submitted for assessment is entirely my own work.

Signature: \_\_\_\_\_ Date: \_\_\_\_\_

NB. If it is suspected that your assignment contains the work of others falsely represented as your own, it will be referred to the College's Disciplinary Committee. Should the Committee be satisfied that plagiarism has occurred this is likely to lead to your failing the module and possibly to your being suspended or expelled from college.

**Complete the sections above and attach it to the front of one of the copies of your assignment,**

## **What constitutes plagiarism or cheating?**

The following is extracted from the college's formal statement on plagiarism as quoted in the Student Handbooks. References to "assignments" should be taken to include any piece of work submitted for assessment.

Paraphrasing refers to taking the ideas, words or work of another, putting it into your own words and crediting the source. This is acceptable academic practice provided you ensure that credit is given to the author. Plagiarism refers to copying the ideas and work of another and misrepresenting it as your own. This is completely unacceptable and is prohibited in all academic institutions. It is a serious offence and may result in a fail grade and/or disciplinary action. All sources that you use in your writing must be acknowledged and included in the reference or bibliography section. If a particular piece of writing proves difficult to paraphrase, or you want to include it in its original form, it must be enclosed in quotation marks and credit given to the author.

When referring to the work of another author within the text of your project you must give the author's surname and the date the work was published. Full details for each source must then be given in the bibliography at the end of the project

## **Penalties for Plagiarism**

If it is suspected that your assignment contains the work of others falsely represented as your own, it will be referred to the college's Disciplinary Committee. Where the Disciplinary Committee makes a finding that there has been plagiarism, the Disciplinary Committee may recommend

- that a student's marks shall be reduced
- that the student be deemed not to have passed the assignment
- that other forms of assessment undertaken in that academic year by the same student be declared void
- that other examinations sat by the same student at the same sitting be declared void

Further penalties are also possible including

- suspending a student college for a specified time,
- expelling a student from college,
- prohibiting a student from sitting any examination or assessment.,
- the imposition of a fine and
- the requirement that a student to attend additional or other lectures or courses or undertake additional academic work.

## Contents

League of Legends: A Detailed Analysis.....	1
1 Introduction .....	7
1.1 Background.....	7
1.2 Aims .....	7
1.3 Technologies.....	8
2 System.....	10
2.1 Requirements.....	10
2.1.1 Functional requirements .....	10
2.1.2 Data requirements .....	11
2.1.3 User requirements .....	12
2.1.4 Environmental requirements.....	12
2.2 Design and Architecture .....	13
2.3 Graphical User Interface (GUI) Layout.....	14
2.4 Testing .....	14
3 Methodology .....	15
4 Implementation and Results.....	17
5 Evaluations and Conclusions .....	21
6 Further Development or Research.....	22
7 References.....	23
8 Appendix .....	24
8.1 Project Proposal.....	24
8.1.1 Objectives.....	24
8.1.2 Background .....	24
8.1.3 Technical Approach.....	25
8.1.4 Special Resources Needed .....	25
8.1.5 Technical Details .....	25

8.1.6	Evaluation.....	25
8.2	Project Plan.....	26
8.3	Requirements Specification .....	26
8.3.1	Introduction.....	26
8.3.2	User Requirements Definition.....	27
8.3.3	Functional Requirements.....	27
8.3.4	Non-Functional Requirements.....	32
8.3.5	Interface requirements.....	33
8.3.6	System Architecture .....	33
8.3.7	System Evolution.....	34
8.4	Monthly Journals.....	35
8.4.1	Reflective Journal - September .....	35
8.4.2	Reflective Journal - October .....	36
8.4.3	Reflective Journal - November .....	37
8.4.4	Reflective Journal – December.....	38
8.4.5	Reflective Journal – January .....	39
8.4.6	Reflective Journal – February.....	40
8.4.7	Reflective Journal – March .....	41

## Executive Summary

The Riot API was created by Riot Games, the creator of the massive MOBA (Massive Online Battle Arena) game League of Legends, to provide the public exclusive access to data and information regarding the game for developmental purposes. The objective of this research is to successfully extract useful data from the API to apply the KDD (Knowledge Discovery in Databases) process to find and analyze some patterns within the game. This project will aim to find similarities in play styles that gamers use between the different regions that the game is being played in, specifically NA (North America) and EUW (Europe West). Some of these styles include how aggressive they are (how often they go for kills) or how much gold they spend, that player's use in the game in order to optimize their win potential.

The research will be carried out using R and Python exclusively mainly because that Python can handle the JSON format that the data is provided in very efficiently. Once the data has been extracted and saved, R will be used for pre-processing and analysis.

In effect, this research is carried out to find the differences between the games regions and their play styles, and find out if a certain method of playing will lead to more successful wins in the game. This project will give us results of methods of play that players can implement to grant them a higher chance of winning in their games.

# **1 Introduction**

## **1.1 Background**

League of Legends was released by Riot Games in 2009 and the organisation continues to hit new milestones in leaps and bounds. The game is now regularly exceeding 5 million concurrent players globally, which is a far greater number than any other PC game currently. That number even exceeds concurrent users on Steam, which is a platform with almost 800 million registered games. That stat alone is mind blowing and should be enough reason to catch someone's attention.

I've been playing the game since the later stages of 2011 and have seen so many successful changes within it over the years, like the consistent visual and mechanical aspects of the game which resulted in having more active players and a generally more fun in game experience. These are just a few updates that Riot keep consistent and is part of the reason why the game is still so successful to this day.

Much of its success comes down to the huge competitive aspect of the game and the endless possibilities of dominating your opponent's.

## **1.2 Aims**

There are numerous ways for a player to improve in League of Legends, whether it comes down to the quick decision making of an individual, to their 'mechanics', the playing skill of a person. This research focuses on the less 'exciting' part of the game, but still a very important aspect to consider, and that is an individual's game knowledge.

The in game knowledge is a factor that plays a huge part to the success of a player that many people overlook sometimes. This includes map awareness, having an idea on where everyone is on the map at all times, to the items people build or even the behaviour and characteristics people show in game. Taking a look at how passive or aggressive a player is when playing the game will be interesting to see if it affects their win rate or not.

The research aims to provide statistics on the different ways of play styles to increase a player's success rate. The best approach to achieve this is to take a look at a sample size of player data in their respective regions and how they play the game i.e. taking a look at how much gold they spend to get stronger, or how much kills they have in the game (how aggressive their play style is).

### **1.3 Technologies**

The technologies being used in this research is Python and R. A python script was written using the default IDLE editor to extract the necessary data from the Riot API provided by Riot Games Company. In this script, the data is extracted in a JSON format and saved into a data frame in order to be saved as a csv file. This will provide the data in a table format with named attributes to better understand the data and their values.

The cleaning process and analysis work is applied with the R language using the RStudio IDE as it provides a very powerful user interface. Descriptive and predictive analysis is applied in this research to give readers new found knowledge on how to increase their win rates in the game by looking at statistics from different types of play styles.

#### **Python**

Python is a high level programming language which is concise and easy to read, which pertains to the philosophy of the design of python, which is code readability. The reason I have chosen to use the python language for my research is because of the data analysis library, pandas. It provides easy to use data structures and data analysis tools for the language, which is exactly what I need for extracting my data. The Riot API that I used to extract my data is provided in a JSON format, which is a format which pandas handles very easily and efficiently.

#### **R**

R is a programming language used for statistical programming and graphics, and is a great resource for data mining analysis and statistical work. The RStudio IDE is the software tool chosen in this project because of its cross platform interface, which allows it to be used in a number of different operating systems. Another reason is that



it is free and open source. I'm using the R language because of the ease of data manipulation and the graphical displays and calculations it produces. Its ability to handle large amounts of data with little to no effort is the huge selling point in using it for my project. In conclusion, once the data has been extracted from python and the dataset has been produced, R will be used during the cleaning and analysis process.

## **2 System**

### **2.1 Requirements**

#### **2.1.1 Functional requirements**

Functional requirements is the step by step process on how the project functions, how the system works and the results and outputs are produced. The functional requirements are listed under the headings below.

#### **Data Extraction**

This requirement is the most important part of the project. The Riot API gives access to millions of data with lots of attributes and variables. What was specified was the data that had to be analysed to provide conclusions to the research I set out, which was to give readers access to information that would help their in game knowledge. So the specified data values I chose to extract were the kills, gold spent, and items that players used to get the most number of wins. The datasets were split by region, EUW and NA because the project compares the differences in play styles between the regions.

#### **Cleaning Process / Pre-processing data**

Once the data has been specified, error handling was applied during the extraction phase in order to avoid disruption when obtaining the data. Reasons being that some players may not have any items or kills at all (this can be caused by a player leaving the game prematurely), which would produce an error during extraction. To tackle this problem, we set a task to ignore the empty values and continue to extract the data until we have retrieved all the values from the player list that we have of random players. The data was also pre-processed during the extraction phase in the python script. This was carried out when I applied names to the variables that was being retrieved in order to read and understand it.

#### **Analysis**

Analysis was applied to the pre-processed data, the result of the scripts using R and Python. The data was then distributed on graphs and charts so that comparisons can be made between the results. In this project, we compare player aggression by looking

at the number of kills they rack up between their server regions EUW and NA (the continents they play at), and take a look at their win/success rate based on their behaviour in the game. Once we retrieved the results of our analysis, I was then able to evaluate and make some conclusions based on the information gathered.

## **Evaluation/Conclusion**

Once the data had been analysed, I had enough evidence to display informative conclusions to the reader. This entire research was written for the sole purpose of the reader to learn more about the game League of Legends, and how this project went about doing just that. It will inform the reader on how the results and findings were concluded from the methods that were implemented.

### **2.1.2 Data requirements**

Data requirements are the software requirements, the data processing tools where the data extracted is used to carry out analysis are explained. These tools include R and Python which are elaborated upon down below.

#### **Python**

The python script that was used to extract the necessary data needed in this project also had another function initially, which was to convert some data that returned integer values into string values readable to the user. Reason being is for the ease of analysis and displaying of results when being distributed on the graph. Giving meaning to the integer values allow readers to understand what is being tested which makes the research easier to grasp and for them to make use of to improve their success rate. Unfortunately there were problems with trying to convert the values from the Riot API so that task had to be dropped. However, the variables are still named and explained from the developer website so we can still refer back to that to comprehend the integer values that the script has extracted.

#### **R**

The R language was used solely for its ability to compute very large amounts of data with little to no effort. With the dataset being extremely large in this research, it is important that the language used for analysis is highly extensible, which R has the environment to do just that. With its huge variety of statistical and graphical

techniques, R was the prime choice for performing this kind of analysis with such a large dataset. The dataset is large mainly due to the fact that when querying one player for their builds and itemizations, we extract more than one match that that specific player has played, which means that there can be hundreds of data specific to only one player. Because we're comparing success rates from play styles, we need to gather data from a sufficient number of players, and also from more than one region. For this reason alone, I chose to use R because of its sole ability to handle such large datasets.

### **2.1.3 User requirements**

#### **Computing Device**

Readers will need to have access to a computing device, as the report journal will be in a soft copy for him/her to read.

#### **Basic Knowledge**

Having a basic knowledge of the game League of Legends or a MOBA (Massive Online Battle Arena) equivalent is also a small requirement, but is not entirely necessary as the research is explained in detail as to what is being analysed and tested.

#### **Opinion**

Once users have access to the journal, they are entitled to read it and also make their own evaluations and conclusions from the results displayed in the report. They are free to have their own opinion and do not necessarily have to agree with my conclusion written in the journal.

### **2.1.4 Environmental requirements**

These requirements refers to the environments/software tools needed to carry out this research.

## **Internet Browser**

A stable internet connection is needed in order to extract the data from the Riot API. An internet connection is also needed because the mappings of the specific data needed in this project are available online.

## **IDLE**

Idle (Integrated Development and Learning Environment) is an IDE for Python which has two main window types, the Shell and Editor Window. The shell window is where the functions of the python script are outputted. For instance, when you run a function with python, the shell window will display an output of the function applied. The editor window is where the script can be created and written. These are where functions are created and edited according to the coder. Python is not limited to the IDLE editor window and can be edited using almost any editing tool i.e. Notepad++.

The python script was also created to extract the raw data from the Riot API into more meaningful data that is understood by the viewer in IDLE.

## **RStudio**

RStudio was the IDE software tool used to pre-process the dataset extracted to carry out specific analysis tasks. The language being used was R.

## ***2.2 Design and Architecture***

As outlined previously, this project was carried out using a range of different programming languages and software development tools in order to evaluate results and finalize conclusions. The datasets extracted using the python script between the two regions (EUW and NA) is the core of this project as it holds all the information that's needed for the analysis work to be performed. For this research, the calls being made to extract the data from the Riot API was saved onto a local cache file. This saves a lot of time by not having to make the same calls and collect the same data from the API.

The data was extracted using the Riot API available to us from Riot's developer website into a dataset saved in a .csv format. The initial plan was to specify the data extracted directly from the Riot API, but due to the difficulty in specificity during

extraction the entire data fields joined to a player and their matches was extracted. For this reason, the dataset was then pre-processed using RStudio where the specific data needed for analysis was kept and the rest of the columns of data that wasn't needed was deleted.

The analysis were carried out by taking advantage of some of the libraries that are available in RStudio, which proved to be beneficial to gather and output results.

### ***2.3 Graphical User Interface (GUI) Layout***

The GUI is in the implementation and evaluations/conclusions where there will be graphs and charts plotted for the reader's to comprehend what is going on in the project. The KDD process is also laid out in the methodology section and how the process works on this project is explained.

### ***2.4 Testing***

The testing phase in this project involved handling some request errors during the initial stages of the project. The reason for this was because there were rate limits set on the Riot API so there were problems fetching the player information in the beginning. But after some testing by requesting from just one or two players, once the requests were made successfully, I filled the name list with a bunch of players and set a request limit of 100 players to fetch information from in the python script.

Testing during the analysis phase was tedious because every variable that was chosen by me had to be tested with each other to find some sort of pattern or correlation. Because the aim of the project was to try and find a variable that may achieve higher success rates when playing the game, each variable had to be tested against the win rates of the match, and the dataset which I extracted had 70 variables. My implementation and results are explained further down in the report.

### 3 Methodology

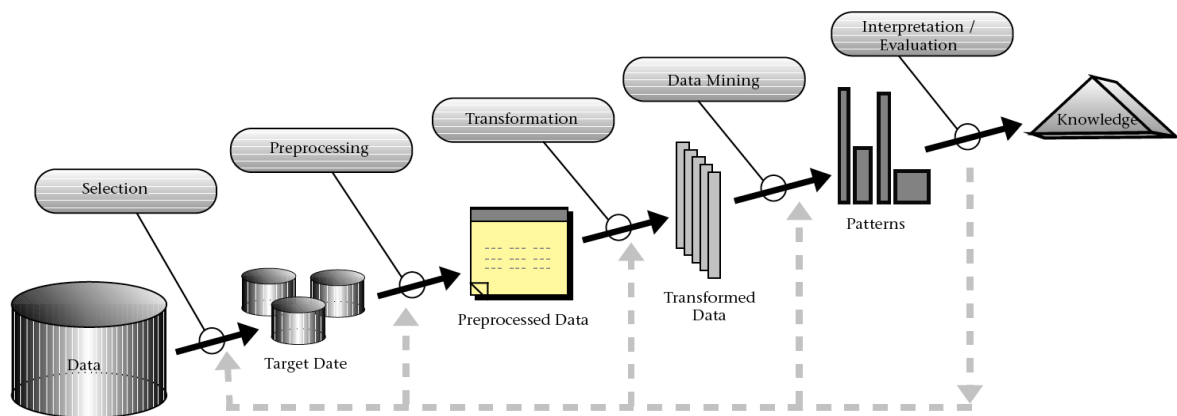


Fig. 1. KDD Process.

The analysis process of this report essentially followed the KDD process (Knowledge Discovery in Databases). Once the dataset was selected and successfully extracted from the Riot API, it was the best approach to take because it “emphasized the high-level application of data mining methods” that the KDD process performed. The data collected from the Riot API were referenced in detail from the developer website (<https://developer.riotgames.com/>). This gave an understanding to the data which allowed me to specify the kind of data I could analyse from the beginning. The diagram in Fig.1 shows the KDD process and this section will explain how the KDD process was implemented into this project.

We’ve established that the data we are using for this research are datasets extracted from the Riot developer website using the Riot API, so the next step is to select the target data. This is also known as the selection stage. The goal of the project is to find the variables which give a player the highest chance of winning. By using my own knowledge of the game I was then able to hand pick certain variables that I think might lead to a player winning a game. Once the variables have been decided, I started to remove the rest of the data that were not needed for the analysis.

The next step we enter into is the pre-processing stage. This stage is to remove the outliers or noise, some of which are the null values that the Riot API returned and the different size of datasets. So I cleaned the data by removing the null values first. To tackle the problem of different sized datasets, I randomly sampled a sample size of

1000 so that the analysis between both the EU and NA datasets will be analysed from the same number of data. This will provide a more accurate comparison between the two when evaluating. The reason I chose a smaller sample size too is so that the analysis script would run more efficiently. By performing this step, it saved me a lot of time by not having to wait for the script to compute and iterate through every single data one by one. Using a sample of 1000 will still give an accurate result as long as it's random.

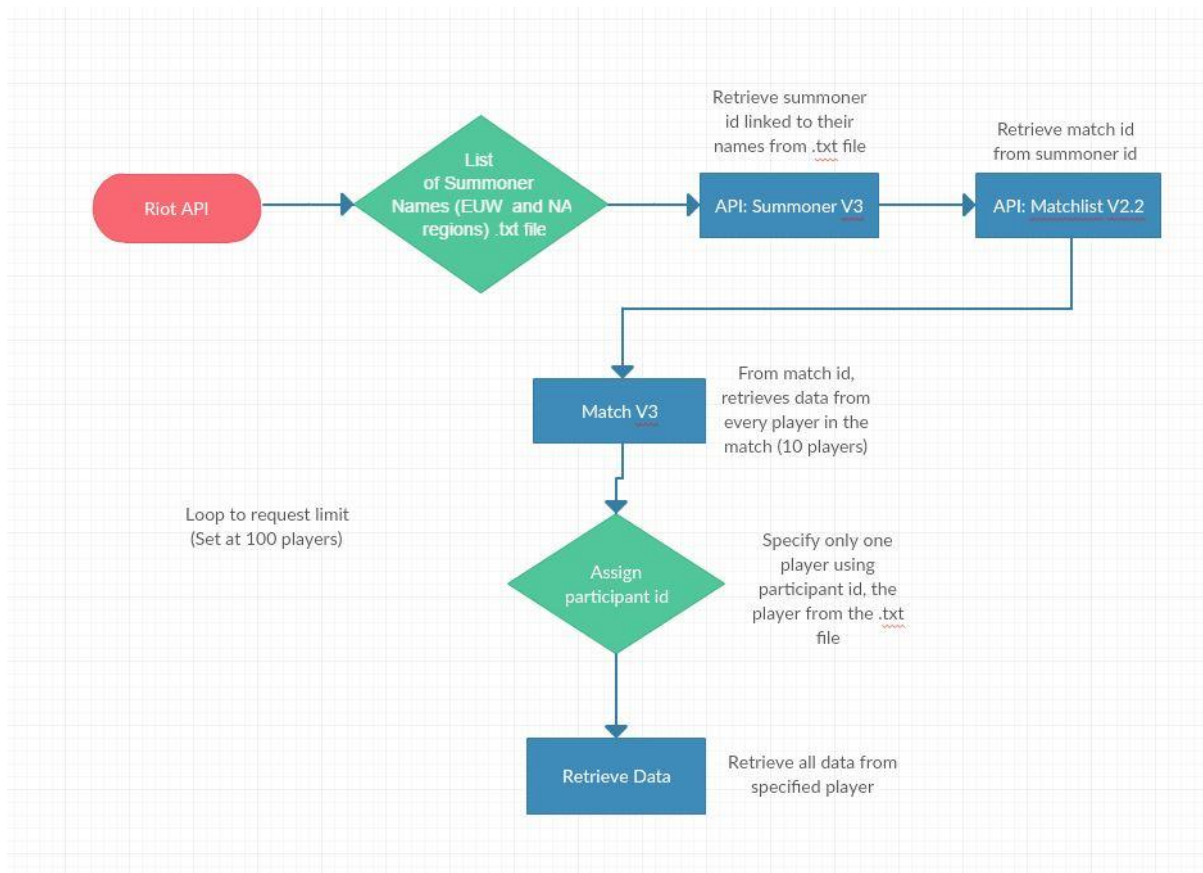
So once the variables have been specified in the data and it has been pre-processed, the next procedure was to find how I want to represent the data in the transformation stage. So useful features had to be analysed and this was carried out in RStudio using the ggplot2 library, which allowed me to plot some graphs the library provides. The project will be applying clustered analysis because the aim is to find relationships between the variables. We are trying to find if certain variables in the game give players a greater chance at winning. This analysis is carried out in the data mining stage, whereby the R code was used to search for patterns in the data by applying different methods and plotting different graphs and charts.

Once the different patterns are found from our results, we interpret the discovered knowledge and make an evaluation. These results will be displayed using a number of different graphs and charts which will be used as evidence towards our conclusion. How this was carried out will all be explained in more detail in the following Implementation and Results section.



## 4 Implementation and Results

The implementation of this entire project was a tedious task solely because of the difficulty in extracting the dataset from the Riot API. The developer website had no issues with querying and retrieving the data online but the mapping of API's was very long winded and took quite a lot of time to comprehend. This was definitely the elephant in the room as it proved a problem from the very beginning of the project. Once I had a grasp of the mapping of the API, the next step was to implement a Python script to connect the components and map each API to another to retrieve the data I needed for this project. The biggest hurdle by far was trying to perform this task because it required a lot of knowledge from me to use Python and also, I only had one reference from a YouTube video to lead me in the right direction (<https://www.youtube.com/watch?v=ml0IKDU5JvY&t=27s>). The guy's python script helped me to understand how to link the API's to python and his script also made a single request to call very specific variables. The workflow of the API can be seen in the table down below, where I successfully extracted data from the API called Match V.3.



The python script was set to scrape 10 matches from 100 players from each region but there was a rate limit set on the developer site which led to the extraction returning request errors. So in the end, instead of scraping for a huge amount of data, I decided to just grab as much data as I could until I reached the limit and from there just make a sample size of 1000 to run my analysis. This proved to be crucial to the completion of my project otherwise I would have wasted too much of my time trying to perform error handling tasks for my script. This was how I implemented my python script to successfully extract data from the Riot API.

Once the data has been extracted successfully from the Riot API and saved into a .csv format, I was able to use the read.csv () function in R to download the dataset and look at the variables I want to use for analysis. The type of analysis that was carried out on these datasets were descriptive and exploratory analysis. The descriptive analysis was used to find out information on the collected dataset. This will then help us decide the approach to take to retrieve the results we need. Exploratory analysis was used to find the unknown relationships between the variables in the datasets because so far they are only values with named variables. Only from this type of analysis can we try to find patterns from the data to make our statements and evaluations based off the results.

During the data mining process, I tried to find if the win conditions were affected by the number of kills players had in the game. The results EUW region are displayed in the graph in Fig.2 down below.

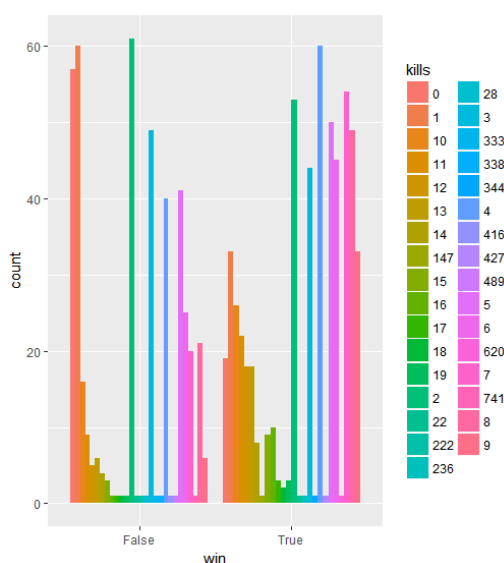


Fig. 2. EUW Wins based on Kills.

From Fig.2, it shows that the number of kills does in fact affect a player's win rate. The orange bar (bar to the very left of the plot) represents the low kill counts from 0-10 and we can see it's so much bigger in the false (indicates a loss in the game) heading under the win variable, while the orange bar in the true heading has a much smaller orange bar. This in fact tells us that low kill counts generally means a higher chance of losing. The bars that represent higher kill counts fluctuate among the two headings true and false, which tells us that more kills does not necessarily mean more wins, it just means that there's a higher chance in succeeding with more kills.

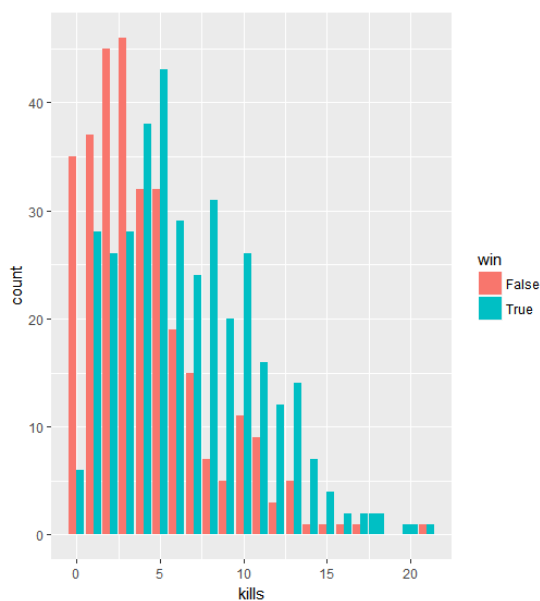


Fig. 3. NA Wins based on Kills.

From Fig.3 we see a graph from the NA region representing the average of the number of kills and whether they win or not. This result clearly shows that there are that players have a higher chance of losing the less kills they have. It is represented on the graph (the highest bars to the left represent the games that players lose in – Boolean is false). The stats show that the higher the kills, the more likely a player is going to win. When you compare it with EUW, we see that they share the same pattern whereby, the lower the kill count, the higher the chance of losing, the higher the kill count, the higher the chance is of winning.

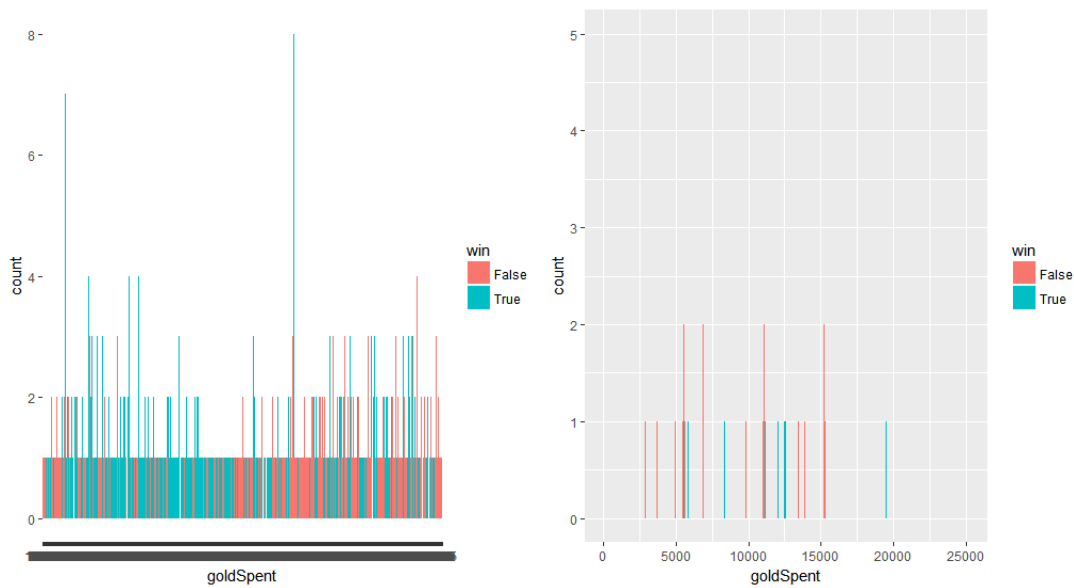


Fig. 4. EUW Gold Spent Win & NA Gold Spent Win respectively

In Fig.4 and Fig.5 the graphs represent the amount of gold spent by a player and whether it increases their chances of winning. But the results show that they have no correlation, indicating that spending more gold in game does not necessarily grant you a win.

When comparing both regions together, it seems that they both share the same results of not having a pattern between gold spent and win rate. This means that this finding is inconclusive and more specific analysis needs to be performed on these two variables if we want to find a pattern.

## 5 Evaluations and Conclusions

The objective of this project was to try and find comparisons in play styles between the two regions I chose, Europe West (EUW) and North America (NA). I was hoping to find some differences like European player's play styles are more aggressive compared to the more passive players from America, but sadly the results show similar traits where the players are split quite evenly, almost 50/50. What the two regions have in common is that the more aggressive players who score more kills will have a higher chance of winning. So this is a huge revelation for readers who want to take something from this research. If you decide to adopt a more aggressive style of play and try to score more kills in the game, it will raise your chances of winning the game.

For the second analysis, it was quite disappointing as the results showed no visible pattern between gold spent and win rate. It's a shame because from my game knowledge experience, usually the player with more items will come out on top. But that doesn't seem to be the case statistically.

To conclude, we have taken a closer look at the success rates of random players and have somewhat achieved the goal set out for this project which was to recommend reader's a way to increase their win potential in League of Legends. Unfortunately there is only the one conclusion, which is to adopt a more aggressive play style, but its advice that users can confidently take on board with the evidence backed up behind the analysis.

## **6 Further Development or Research**

Further development of this project would definitely include testing more variables that showed visible patterns with the goal in mind. With that said, the big problem in the beginning was the selection phase. If there was more development in error handling on my part during extraction phase, specifying target data and finding visible patterns between different variables would have been a lot easier. Because there were 70 variables to analyze with, it was very difficult to come up with a way to find good and interesting data to analyze apart from my own knowledge and experiences of the game. More analysis work would be carried among other regions to make it a more interesting research.

## 7 References

- *Summoner's ladder ranking : LoL Stats, Record Replay, Database, Guide, MMR - OP.GG.* [ONLINE] Available at: <http://www.infovis-wiki.net/images/4/4d/Fayyad96kdd-process.png>
- *[R] random sampling inside a dataset .* [ONLINE] Available at: <https://stat.ethz.ch/pipermail/r-help/2012-October/336866.html>
- *University DBD: KDD Process/Overview* [ONLINE] Available at: [http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1\\_kdd.html](http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html)
- *Learn to use Riot's API in 15 Minutes using Python! :* YouTube. [ONLINE] Available at: <https://www.youtube.com/watch?v=ml0IKDU5JvY&t=27s>
- *Python Data Analysis Library — pandas: Python Data Analysis Library.* [ONLINE] Available at: <http://pandas.pydata.org/>

## **8 Appendix**

### ***8.1 Project Proposal***

#### **8.1.1 Objectives**

The objective of my project is to extract a sufficient dataset where I will be able to retrieve the necessary information from the collected data, to determine the success of League of Legends. This research will specifically aim to figure out what makes this PC game so much more successful than the other similar games currently out there. I will look over a few ideas as to why the game may be so popular, from the actual in game experience, the companies marketing and manifesto, the time period of its release, and making comparisons with other games alike.

The theme of the project may be in relation to a game, but the prime goal of this project however, is to create a process, a reusable set of steps for organisations to take advantage of and find information on aspects of their business that are succeeding. This system would be something similar to the Knowledge Discovery in Databases (KDD) except in a more business sense. I want to create a process that will benefit users immensely with minimal effort.

#### **8.1.2 Background**

League of Legends was released by Riot Games in 2009 and the organisation continues to hit new milestones in leaps and bounds. The game is now regularly exceeding 5 million concurrent players globally, which is a far greater number than any other PC game currently. That number even exceeds concurrent users on Steam, which is a platform with almost 800 million registered games. That stat alone is mind blowing and should be enough reason to catch someone's attention.

I've been playing the game since the late part of 2011 and have seen so many successful changes within it over the years, like the consistent visual and mechanical aspects of the game which resulted in having more active players and a generally more fun in game experience. These are just a few updates that Riot keep consistent and is part of the reason why the game is still so successful to this day.



I feel that this research will be an interesting endeavour and is also just a small piece of a puzzle that I hope to solve in the coming months.

### **8.1.3 Technical Approach**

The language I am planning to use will be the R programming language. This will be the most optimal language to use for the statistical analysis work that I will carry out in this project. I will be using RStudio to create a program in R, to extract the data that I need from multiple sources online. I will also be using RStudio to query my data as it produces a very powerful user interface. I will also be implementing charts and graphs into my project which I will create using both Microsoft Excel and IBM Statistical Package for the Social Sciences (SPSS a.k.a Statistical Analysis Software Package).

Another language that I will consider using is Python because it is recommended in my stream of study and will benefit the progress of my project. But being new to both R and Python, I feel that focusing on just one at this point in time would be a better approach to reach the objectives of this research.

### **8.1.4 Special Resources Needed**

- Book – Mining of Massive Datasets
- Programs/Software Tools – RStudio, Python, Word
- Online Resources

### **8.1.5 Technical Details**

R will be the primary language I will use to carry out the project for analysis. I will be using Python to develop a script to extract from the Riot API.

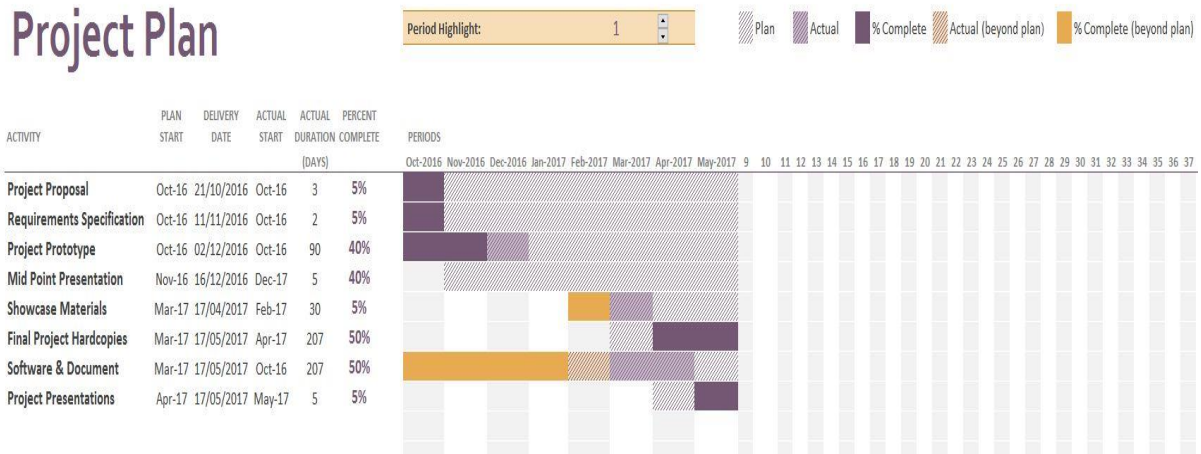
### **8.1.6 Evaluation**

I will constantly be going over the program I am creating and be thorough with the extracted data to make sure that it is credible and there is minimal amounts of redundant data i.e. when extracting similar data from multiple, credible online sources

etc. The information gathered will be displayed using graphs and charts so there will be good representation of the content that I will be discussing.

## 8.2 Project Plan

### Project Plan



## 8.3 Requirements Specification

### 8.3.1 Introduction

#### 8.3.1.1 Purpose

The purpose of this document is to set out the requirements for the development of the study on how League of Legends is such a successful game using Data Analytics.

The intended customers are other gaming companies that can use this study to compare against their own data to see where they can improve on their own successes. This study could also extend to corporations that want a guideline for attracting a massive user base that is continually growing at a rapid pace, which is what League of Legends is successfully carrying out.

### **8.3.1.2 Project Scope**

The scope of the project is to develop a research on the success of League of Legends, created by Riot Games. The research will have results and conclusions acquired from datasets that have been provided from credible online resources, and created by myself from various ways of extracting data using different software tools, which will be explained further below. The research will also show the process of how I acquired my results and my conclusions from these results.

### **8.3.1.3 Definitions, Acronyms, and Abbreviations**

KDD	Knowledge Discovery and Data Mining
LOL	League of Legends
EDA	Exploratory Data Analysis
REST	Representational State Transfer
API	Application Programming Interface

## **8.3.2 User Requirements Definition**

The research paper will draw comparisons with other games alike, and will distinctly show the differences. Readers will be able to agree with the conclusions because of the appropriate tests and results shown. The paper will show a clear process of the steps I took to come up with the results, from the extraction of the data and forming a sufficient dataset, to showing the large impact LOL has made in the gaming industry.

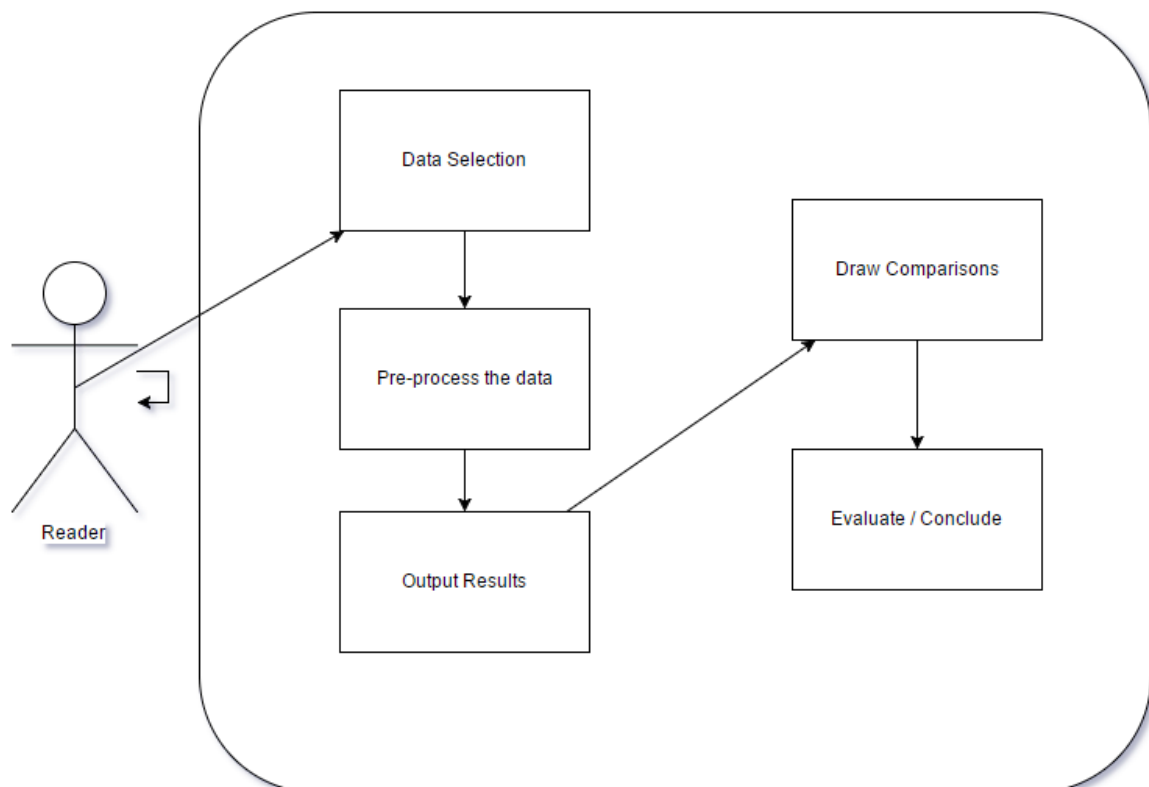
## **8.3.3 Functional Requirements**

Before getting into the system phase, one of the most important phases we have to look at is the planning phase for a usable dataset. The dataset needs to be sufficient for the project at a very professional level for it to be anyway credible for companies to use. This is a very important requirement before EDA even starts. The data collection process **cannot** be overlooked.

Once we have established a sufficient data collection process, we can now list the functional requirements of software systems that will be used.

- The languages that will be used are Python and R. These languages will also be used to develop the programs that will run on the data collected, to represent results that will be displayed on the research paper.
  - The programs created will also make comparisons with other datasets to see the difference in popularity between other similar games.
  - There will also be comparisons with other related works to see if the results correlate with each other for a better understanding of the data.
1. Languages used will be Python and R. Functions of them will be:
    - To crawl data for a sufficient datasets.
    - Clear the data of errors that will skew the result
    - Develop programs for the datasets to output results.
    - Compare against other datasets.
    - Compare with other related works
  2. Data Analysis using IBM SPSS and Microsoft Excel
    - Charts and Diagrams
    - Social Analysis of the results

### 8.3.3.1 Use Case Diagram



### **8.3.3.2 Requirement 1 – Data Selection / Collection**

#### **Description & Priority**

Using the programming languages mentioned, I will crawl the data required during the data collection process, because certain REST API's tailor fit to one or the other of the two languages.

#### **Use Case**

##### **Scope**

The scope of this use case is to extract raw data from different sources to be used for EDA.

##### **Description**

This use case describes the collection phase of the data.

##### **Flow Description**

This use case correlates with the KDD and is in the first phase of it.

##### **Main flow**

1. The system extracts the data selected.
2. The system performs actions set by the researcher.
3. Datasets are created after extraction.

##### **Alternate flow**

n/a

##### **Exceptional flow**

n/a

##### **Termination**

Datasets are formed as a result.

##### **Post condition**

The system goes into a wait state.

### **8.3.3.3 Requirement 2 – Pre-process data**

#### **Description & Priority**

The datasets are cleared of outliers and dummy data that will skew the results during testing.

## **Use Case**

### **Scope**

The scope of this use case is to clear out unnecessary data post extraction.

### **Description**

This is the cleaning phase of the data in the use case.

### **Flow Description**

This is the second phase in the KDD.

### **Main flow**

1. The system identifies the unnecessary data from the extraction.
2. This unnecessary data is defined by the researcher.
3. The system deletes the redundant data from the datasets.

### **Alternate flow**

n/a

### **Exceptional flow**

n/a

### **Termination**

The datasets are updated and now displays data that can be used for testing.

### **Post condition**

The system goes into a wait state

## **8.3.3.4 Requirement 3 – Testing**

### **Description & Priority**

*The new and updated datasets can now be tested for results.*

## **Use Case**

### **Scope**

The scope of this use case is to display outputs that can be used in the final journal.

### **Description**

This is the testing phase of the data.

### **Flow Description**

This is the third phase of the KDD and is also known as the transformation stage.

#### **Main flow**

1. The researcher creates the program for the system to query against the datasets.
2. The system runs the program against the dataset and displays the results.

#### **Alternate flow**

n/a

#### **Exceptional flow**

n/a

#### **Termination**

The results are displayed from the tests that were carried out by the system.

#### **Post condition**

The system goes into a wait state

### **8.3.3.5 Requirement 4 - Compare**

#### ***Description and Priority***

The results can now be data mined and compared with other data.

#### ***Use Case***

##### **Scope**

The scope of this use case is to mine the data from the outputted results and try to find comparisons with other data.

##### **Description**

This is the phase where comparisons are made between results outputted.

##### **Flow Description**

This is the fourth stage of the KDD whereby patterns are drawn from the data mined.

### **Main flow**

1. The researcher compares results displayed by the system.
2. The researcher tries to find patterns in the results to come up with an evaluation.

### **Alternate flow**

n/a

### **Exceptional flow**

n/a

### **Termination**

The results are evaluated by the researcher.

### **Post condition**

The system goes into a wait state

#### **8.3.3.6 Requirement 5 – Conclusion**

The researcher evaluates and concludes the results from the tests carried out by the program. This should give us a better understanding on why LOL is succeeding today.

### **8.3.4 Non-Functional Requirements**

#### **8.3.4.1 Availability requirement**

Datasets need to be available in order for this research to take place. Without a sufficient dataset, there will be no research.

#### **8.3.4.2 Robustness requirement**

The data extracted to form the datasets need to be well thought out. The researcher needs to be thorough in looking for errors in the data that will negatively affect the outcome of the results when tested.

#### **8.3.4.3 Security requirement**

Any data collected from different sources needs to be carried out legally and there cannot be any breaches of terms of use of the information collected.



#### 8.3.4.4 Reliability requirement

The extracted data needs to be from a reliable source, otherwise the research will not be credible and companies will not be able to make use of the research.

#### 8.3.4.5 Reusability requirement

This research will act as a guideline for readers who are interested in the decision-making process from a corporation that caused them to acquire a massive user base in such a short amount of time.

#### 8.3.4.6 Resource utilization requirement

The resources utilized is very minimal to carry out this research and only requires a few important software tools.

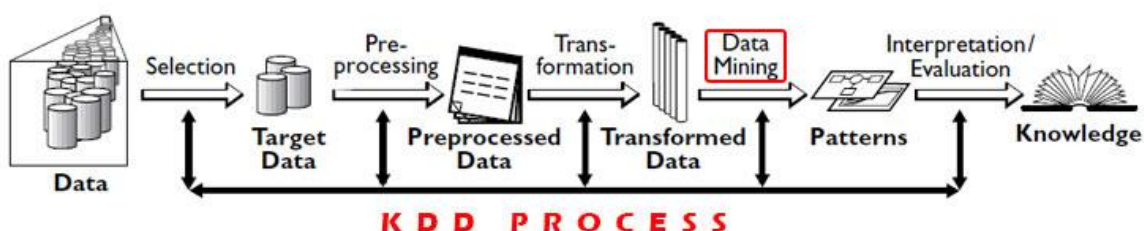
### 8.3.5 Interface requirements

API's will be used during the data collection phase to extract data from online resources. These API's will be provided mainly by the news and information website [www.programmableweb.com](http://www.programmableweb.com) and other websites alike.

#### 8.3.5.5 Application Programming Interfaces (API)

This project will use a LOL REST API provided by Riot Games to access the in game statistics. This includes most of the information in game and is provided in a secure and reliable way for the LOL developer community.

### 8.3.6 System Architecture



The KDD process is the system architecture that will be used to carry out the research. It is the foundation of any analysis work on data and is the best practice to follow in order to achieve the best results from the data.

### **8.3.7 System Evolution**

The plan for this system is to be re used by other companies and corporations to test their own products and services to make an evaluation on their own successes. This will give them a better understanding on what is, and is not working in their company. The system will evolve according to different tests that will be carried out to tailor to each user's needs.

## **8.4 Monthly Journals**

### **8.4.1 Reflective Journal - September**

Student name: Alexander Hii

Programme (e.g., BSc in Computing): BSc in Computing

Month: September 2016

#### **Progress**

This month, I was able to get my project approved by the dragon's den committee. My project will be a research on the most prominent computer game right now, League of Legends. I will be doing data analytical work to find out what makes this game so successful when compared to other games. I am currently looking for a sufficient dataset to use for my project, but I may have to extract my own data from the multiple social media sites the game as a company, takes advantage of. Meanwhile, I am also looking for the right software tools and languages to help me with my project. So far, I have been researching on R and Python, to see which of the two can produce the best results. I will also be using Excel and SPSS to produce my charts and diagrams as recommended by a dragon's den judge.

#### **Supervisor Meetings**

To be announced

## **8.4.2 Reflective Journal - October**

Student name: Alexander Hii

Programme (e.g., BSc in Computing): BSc in Computing

Month: October 2016

### **Progress**

I've managed to find out some required information that I will need for my dataset in order for it to be sufficient for my project. I am currently working on extracting this data from online resources suggested by my lecturer, from my Data Analytics class. There are websites that provide ready to use datasets that I am currently looking into to find any information that I need, but I am also looking to extract some data from social media sites in order to gather information that's needed from players of the game. I am following step by step guides provided by Simon Canton in my Data Analytics tutorials, and also from the websites that provide the ready-made datasets.

I intend to gather the required data for my project and will have to run tests and queries against it to make sure that it is sufficient. I will need to make sure that the data gathered will output plausible results with accurate information. Planning and gathering the required data is one, if not the most important process of data analytics so I need to make sure that the data is credible and I can achieve accurate results.

### **Supervisor Meetings**

I have emailed Lisa Murphy and intend to have a meeting with her on Friday, the 11th of November.

### **8.4.3 Reflective Journal - November**

Student name: Alexander Hii

Programme (e.g., BSc in Computing): BSc in Computing

Month: November 2016

#### **Progress**

My project has reached the querying phase of a dataset that I have successfully found on [www.programmableweb.com](http://www.programmableweb.com). I used some MapReduce models to process and generate the large dataset to refine what I am trying to find from it. This dataset contains information regarding the in game statistics. What I am trying to find from my queries is whether or not an individual's win rates affect how long he/she continues to play in one sitting. Hopefully from gathering this information, I can hypothesise and declare that because the majority of the people that play have a higher rate of wins than losses, is the reason why they continue to play and aid to League of Legends success.

#### **Supervisor Meetings**

I met up with Lisa recently and we made sure that my project is going in the right direction with the necessary deliverables for the presentation taking place in 2 weeks' time. She suggested to create a survey to use as a think aloud test during the presentations and also made aware of the programming languages and other software tools I can use for my project.

#### **8.4.4 Reflective Journal – December**

Student name: Alexander Hii

Programme (e.g., BSc in Computing): BSc in Computing

Month: December 2016

##### **Progress**

This month, I successfully presented my progress to the examiners during the mid-point presentation. After having discussions with them and receiving constructive criticism, it has been brought to my attention that the title and aim of my research is decreasing the productivity of my project. This is mainly due to the difficulties finding the necessary dataset to reach the goals of my report.

In this scenario, I was recommended to change my aims and goals of my project to tailor to the datasets that are already out there for public viewing, and is exactly what I've done. I've changed the aims of my project to suit the datasets that I have already extracted and a going to continue with it to apply the KDD process. This was the best thing to do after all the trouble of trying to find or extract my own datasets to tailor to my goals.

##### **Supervisor Meetings**

After my mid-point presentation, it was clear to see that the aid of my supervisor will help in the coming months and I plan to meet with Lisa as soon as the semester commences.

### **8.4.5 Reflective Journal – January**

Student name: Alexander Hii

Programme (e.g., BSc in Computing): BSc in Computing

Month: January 2017

#### **Progress**

The aims and goals of my project have now been tailored to the datasets that are readily available online. I've now successfully applied the initial stages of the KDD process and making progress in my research paper in regard to the problems addressed in project. I now also have results that I can discuss about which I have failed to achieve previously with my initial aims and goals.

#### **Supervisor Meetings**

I plan to meet with my supervisor next week to discuss the next step forward in my project.

### **8.4.6 Reflective Journal – February**

Student name: Alexander Hii

Programme (e.g., BSc in Computing): BSc in Computing

Month: February 2017

#### **Progress**

The dataset was extracted using Python into a csv file. The cleaning and analysis phase were carried out using R in the RStudio IDE. Evaluations and conclusions were made in the reporting section of the project.

The technical report will also need to be edited from future to past tense as the project is now nearing completion.

#### **Supervisor Meetings**



### **8.4.7 Reflective Journal – March**

Student name: Alexander Hii

Programme (e.g., BSc in Computing): BSc in Computing

Month: March 2017

#### **Progress**

The final edit of the technical report is near completion with slight additions of analysis work to include with proof reading to do. An IEEE format will be included in the final hard copy which will inform the reader a detailed analysis of the work that was carried out in this research. The work includes the implementation of the analysis applied with evaluations and conclusions that were made by me. The research will also have results that readers can use to their advantage when playing their game.

The poster for the showcase was also created and printed for the deliverable for the 10th of May.

#### **Supervisor Meetings**

A meeting with Lisa will be scheduled before the presentation date to make sure that I will have everything prepared come the day.