

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE INFORMÁTICA



Esports – Video Game Data Analysis

Rui Pedro Fernandes Roque

Mestrado em Engenharia Informática

Dissertação orientada por:
Prof^ª. Doutora Ana Paula Pereira Afonso
Prof. Doutor Luís Miguel Parreira Correia

Acknowledgements

I would like to express my deep gratitude and appreciation to all those who have supported and guided me throughout the journey of completing this project. Without their invaluable contributions, this work would not have been possible.

First and foremost, I am profoundly grateful to my thesis advisors, Dr. Paula Pereira Afonso, Dr. Luís Miguel Parreira Correia, and Dr. Maria Beatriz Duarte Pereira do Carmo, for their unwavering support, patience, and expertise. Their guidance, constructive feedback, and dedication to my academic growth have been instrumental in shaping this research.

Throughout my academic journey, I am grateful for my family's unwavering support, encouragement, and belief in me. Your unwavering faith has been a constant source of motivation for me.

To my friends and colleagues, your understanding and encouragement during difficult times were invaluable, and I appreciate your friendship, which made this journey all the more enjoyable.

I am also grateful to the research participant who generously dedicated his time and shared his insights.

Lastly, I extend my gratitude to all those who have influenced my education and personal growth, whether directly or indirectly. Your impact on my academic and intellectual development has been significant.

This thesis represents the end result of years of devoted hard work. To everyone acknowledged and unacknowledged but equally significant, please know that your contributions are greatly valued.

Thank you all for being a part of this important chapter in my life.

Acknowledgements.

Resumo

O fenómeno dos desportos eletrónicos (Esports) tem vindo a crescer e, com este, também o interesse por vídeo jogos online, por parte de jogadores e espetadores. Com a evolução tecnológica tem-se tornado cada vez mais fácil utilizar técnicas de recolha de dados sobre os eventos que decorrem durante um jogo gerando grandes volumes de dados que podem ser utilizados para análise do desempenho de jogadores e de equipas. Este tipo de análise é de grande importância tanto em contextos pessoais como em contextos profissionais. Os jogadores casuais procuram métodos que permitam perceber que erros estão a cometer e qual a forma ótima de jogar com determinadas personagens ao passo que num contexto profissional, o foco é maioritariamente perceber que estratégias são utilizadas por outras equipas e como combatê-las.

Para que a análise deste volume de dados seja eficaz é fundamental explorar mecanismos de análise de dados combinadas com técnicas de visualização (visualização analítica) aplicadas a dados espaço-temporais e aos vários tipos de eventos durante uma partida e que são de interesse para os jogadores, treinadores e analistas. Estes eventos podem variar desde a posição de um jogador (espaço) num determinado instante (tempo), a eventos mais específicos do jogo, tais como a posição onde o jogador morreu.

Atualmente, são exploradas várias técnicas de visualização e de análise de dados para vários jogos de Esports, tais como o League of Legends (LoL) e o Defense of the Ancients 2 (DOTA2) do tipo Multiplayer Online Battle Arena (MOBA). Nestes projetos são desenvolvidos e avaliados protótipos que analisam a adequação de diversas técnicas de visualização e métodos de agregação para explorar os dados fornecidos pelas interfaces de programação de aplicações (API) para análise de jogos de equipas e jogadores para que possam ser utilizadas por treinadores de equipas profissionais na análise de estratégias de jogo e evolução temporal de jogadores. Maioritariamente, estes tipos de análises focam-se na previsão de resultados, recomendações de campeões (personagens jogáveis) para comporem as equipas e análises da toxicidade presente nos jogos. São utilizadas várias técnicas de inteligência artificial (AI) como *random forests*, *naive bayes*, *support vector machines* e redes neuronais (NN), entre outros.

O principal objetivo deste projeto é dar continuidade ao trabalho de visualização analítica de vídeo jogos, explorando novos conjuntos de dados e novas técnicas de análise de dados. Uma vez que são pouco utilizados, são explorados e aplicados algoritmos de aprendizagem automática (ML - *machine learning*) sobre dados espaço-temporais em certos eventos, como nas mortes dos jogadores nas diferentes posições de modo a descobrir e prever comportamentos regulares

dos jogadores, entre outros. Foram exploradas várias APIs de jogos de diferentes géneros como MOBAs (LoL e DOTA2), jogos de tiros em primeira pessoa (FPS - First Person Shooter), como o counter strike: global offensive (CS:GO) e o Valorant e ainda jogos Battle Royale (BR), como o PlayerUnknown's Battlegrounds (PUBG). O jogo selecionado, dependeu muito dos tipos de dados que as APIs forneciam. Durante a investigação das APIs, foram descobertas que apenas três delas forneciam dados espaço-temporais, nomeadamente, as APIs associadas ao PUBG, LoL e DOTA2. Destas opções, optou-se pela recolha de dados da API do League of Legends devido a um maior conhecimento sobre este jogo.

Durante a recolha de dados da API, e visto que não foram encontrados nenhuns conjuntos de dados que disponibilizassem dados espaço temporais, foram criados dois conjuntos de dados de jogadores de LoL através dos dados recolhidos na API da Riot Games. Um conjunto de dados contendo informações do jogador por minuto e possíveis eventos que ocorressem durante a partida em jogadores amadores europeus numa das seguintes divisões: ferro, bronze, prata, ouro e platina durante o *patch* 13.3. Um *patch* é uma atualização no jogo que serve para corrigir falhas, para adicionar novas mecânicas ou até para reajustar o poder de certos campeões utilizados pelos jogadores. O outro conjunto de dados com as mesmas informações, mas para jogadores profissionais americanos, europeus ou coreanos, que competem ao mais alto nível. Estes conjuntos de dados foram posteriormente utilizados para efetuar as análises espaço-temporais dos jogadores.

A investigação englobou a análise espaço-temporal, onde inicialmente identificou-se quais as situações que poderiam ser analisadas de forma a ajudar não só os jogadores mas também os treinadores. Depois de avaliar várias situações de jogo, o foco foi direcionado pelas necessidades mostradas pelos treinadores profissionais em trabalhos anteriores. Consequentemente, a atenção centrou-se numa análise mais aprofundada da posição de *jungler* (uma das cinco posições do League of Legends), principalmente porque a localização destes jogadores permanece oculta da equipa adversária durante mais tempo, especialmente em fases iniciais do jogo. Além disso, o *jungler* percorre todo o mapa durante toda a partida, ajudando constantemente os colegas de equipa. Estas características realçam a importância das informações espaço-temporais nesta posição, distinguindo-a como primordial entre todas as outras.

A abordagem analítica envolveu a aplicação de algoritmos de agrupamento (*clustering*) a dados espaço-temporais, com foco específico na identificação de padrões recorrentes nas mortes de *junglers* amadores e profissionais durante o *patch* 13.3. Para verificar a evolução desses padrões, foi ainda realizada uma análise comparativa utilizando dados de um *patch* mais recente, o 13.9.

Foram ainda realizadas experiências paralelas, embora com menor detalhe, envolvendo as mortes dos campeões utilizados na *jungle* em cada um dos *clusters* formados, identificação de padrões nas outras posições (*top*, *mid*, *bot*, *support*) e o mapeamento das localizações de *junglers* profissionais em diferentes fases do jogo. As visualizações desempenharam um papel fundamental para facilitar a interpretação das diferentes experiências. O conjunto de técnicas de visualização incluiu gráficos de dispersão, representações tridimensionais, histogramas, gráficos de radar e donuts, entre outros. Estas ajudas visuais enriqueceram a análise, fornecendo uma perspetiva

abrangente sobre a complexidade da dinâmica espaço-temporal no panorama competitivo dos vídeo jogos.

Com base nos agrupamentos obtidos, foi possível identificar padrões significativos tanto nas mortes dos jogadores quanto nas áreas preferidas pelos jogadores em diferentes momentos das partidas. Notavelmente, os jogadores da posição *jungle* demonstraram uma tendência de morrer mais frequentemente e de passar mais tempo nas partes inferiores do mapa durante as fases iniciais do jogo. No entanto, à medida que o jogo avançava para uma fase mais intermediária, esses jogadores passavam a morrer com maior frequência e a passar mais tempo na parte superior do mapa.

É importante salientar que padrões semelhantes foram igualmente observados nas restantes posições. Em cada posição (*top*, *mid*, *bot* e *support*), como era de esperar, a maioria das mortes durante as fases iniciais ocorreram na zona do mapa correspondente à sua posição, independentemente do nível de habilidade do jogador (amador ou profissional). No entanto, à medida que as partidas entravam nas fases intermediárias, notou-se uma tendência para as posições *mid*, *bot* e *support* deslocarem-se para áreas superiores do mapa, enquanto os jogadores da posição *top* apresentaram uma maior tendência para morrer na parte inferior do mapa durante esta fase.

À medida que as partidas avançavam para as fases finais, os jogadores de todas estas posições, tal como os *junglers*, tiveram uma tendência de morrer mais na base da equipa que estivesse a perder. Adicionalmente, foram observadas algumas variações entre diferentes *patches* do jogo na posição jungler, sobretudo entre os jogadores amadores. Estas variações incluíram mudanças nos locais mais frequentes de morte durante as fases iniciais e intermediárias das partidas.

Especificamente, nas divisões inferiores, verificou-se em algumas situações uma inversão nos locais mais comuns de morte. Durante as fases iniciais, registou-se um aumento nas mortes na parte superior do mapa, enquanto nas fases intermediárias, ocorreram mais mortes na parte inferior do mapa. Esta inversão pode ter sido causada, em parte, por ajustes de equilíbrio que diminuíram poder às posições da parte inferior do mapa (*bot* e *support*) durante este novo *patch* do jogo.

Estes resultados proporcionam uma visão valiosa das dinâmicas de jogo e das estratégias adotadas pelos jogadores, bem como das influências das atualizações do jogo nas escolhas e desempenho dos jogadores em diferentes divisões e posições.

Para concluir, os resultados foram ainda validados, através de um treinador profissional.

Palavras-chave: Esports, Visualização analítica, Dados espaço-temporais, Algoritmos de aprendizagem automática, Análise de dados

Abstract

The phenomenon of Esports has been growing and, with it, the interest in online video games by players and spectators. With technological evolution, it has become increasingly easier to use data collection techniques about the events that take place during a match, generating large volumes of data that can be used to analyze the performance of players and teams. This analysis is of great importance in both personal and professional contexts. Casual players look for methods to understand what mistakes they are making and the optimal way to play certain characters, while in a professional context, the focus is mostly on understanding what strategies are used by other teams and how to counter them.

For the analysis of this volume of data to be effective, it is fundamental to explore data analysis mechanisms combined with visualization techniques (visual analytics) applied to spatio-temporal data and the various types of events during a match that are of interest to players, coaches, and analysts. These events can range from a player's position (space) at a given instant (time) to more game-specific events, such as where the player died.

The goal of this project is to explore and apply machine learning algorithms to spatio-temporal data to discover patterns in player behaviors while continuing the work on visual analytics of video game data. The investigation extends to exploring datasets from new games, ultimately leading to the selection of League of Legends (LoL) as the focal point for in-depth analysis. One significant challenge in this pursuit is the scarcity of readily available datasets featuring spatio-temporal data. To overcome this obstacle, the research project involves the creation of spatio-temporal datasets from the selected game, LoL, through data collection facilitated by the Riot API.

In summary, this research project not only builds on previous work but also introduces new data analysis techniques, notably the clustering of spatio-temporal data, to uncover possibly hidden patterns of player behaviors in the world of League of Legends. The results obtained provide valuable insights on the players, particularly focusing on the jungler role. They provide information regarding potential death patterns and the most frequently visited locations on the map as the game progresses. Additionally, these results make it possible to observe differences in spatio-temporal data across various game patches. The culmination of these efforts promises valuable insights into the gaming ecosystem, with potential applications in game design, player engagement, and beyond.

Keywords: Esports, Visual Analytics, Spatio-temporal data, Machine learning algorithms, Data analysis

Contents

List of Figures	xvi
List of Tables	xix
Acronyms	xxii
1 Introduction	1
1.1 Motivation	1
1.2 Goals	2
1.3 Structure of the document	2
2 Background and Related Work	5
2.1 Background	5
2.1.1 Game data science	5
2.1.2 Game data	5
2.1.3 Machine learning	6
2.1.4 Spatio-temporal analysis	8
2.1.5 League of Legends	8
2.2 Related Work	11
2.2.1 Non-spatio-temporal data analysis	11
2.2.2 Spatio-temporal data analysis and clustering	11
2.2.3 League of Legends game	13
2.2.4 League of Legends data analysis applications	14
2.3 Summary	16
3 Methodology, Technologies and Tools	19
3.1 Methodology	19
3.2 Technologies and Tools	21
3.2.1 OPGG	21
3.2.2 Tracking the professionals	21
3.2.3 Jupyter Notebook	21
3.2.4 Libraries	22

3.3 Summary	23
4 League of Legend Game and Datasets	25
4.1 Game selection	25
4.2 Game datasets	26
4.2.1 Players' Information	26
4.2.2 Data collection	27
4.2.3 Data preprocessing	27
4.2.4 Dataset analysis	28
4.2.5 Additional datasets	30
4.3 Summary	31
5 Cluster Analysis	33
5.1 Type of analyses	33
5.2 Data preprocessing	33
5.3 Distance metric	34
5.4 Clustering algorithm	34
5.5 Death patterns in the Jungle Role	35
5.5.1 Jungle role	36
5.5.2 Professional junglers' results interpretation	37
5.5.3 Amateur junglers' results interpretation	40
5.5.4 Professional Junglers vs Amateur Junglers	43
5.5.5 Other experiments in the jungle role	43
5.6 Death patterns in other roles	51
5.7 Locating professional junglers	55
5.8 Summary	57
6 Validation of Results	59
6.1 Participant and procedures	59
6.2 Interview results	59
6.2.1 Validation of cluster analysis	59
6.2.2 Surprising results	60
6.2.3 Champion analyses	60
6.2.4 Utility of data analysis	61
6.2.5 Relevant events in the game	61
7 Conclusion and Future Work	63
7.1 Conclusion	63
7.2 Future Work	64
References	69

A Data collection	71
A.1 MOBA games APIs	71
A.2 First-Person Shooter (FPS) games APIs	72
A.3 Battle Royale (BR) game API	72
A.4 Overwolf	73
B League of Legends datasets features	75
C Clustering analysis plots	79
C.1 Death pattern on professional junglers	79
C.2 Death Pattern On Patch 13.9	80
C.3 Locating professional junglers	80

List of Figures

2.1	League of Legends map with important locations (Riot Games, 2022 [21])	10
2.2	Blitz.gg in-game overlays of a League of Legends match (Blitz.gg, 2022 [13]) . .	15
2.3	Interactive time-based visualization in VisualLeague III (Rafael, 2020 [9])	15
2.4	Clustering filter in VisualLeague III (Rafael, 2020 [9])	16
3.1	Steps of clustering process	20
4.1	Steps for creating the datasets	28
4.2	Patch 13.3 - Datasets win rate	30
4.3	Patch 13.9 - Datasets win rate	31
5.1	Algorithms' best <i>silhouette scores</i> (Left: K-Means; Middle: K-Medoids; Right: Spectral)	35
5.2	DBSCAN best <i>silhouette scores</i>	35
5.3	Best k value according to <i>silhouette score</i>	37
5.4	Clustering professional jungler players' deaths in red team's victory	38
5.5	Clustering professional jungler players' deaths in red team's defeat	39
5.6	Low elo divisions	40
5.7	Example of clustering amateur jungler players' deaths	41
5.8	Example of clustering amateur jungler players' deaths	42
5.9	Example of clusters obtained using only deaths' coordinates	44
5.10	Example of clusters obtained using only deaths' timing	45
5.11	Top 6 most played jungle champions by professional players and their win rate . .	46
5.12	Top 6 most played jungle champions by amateur players and their win rate	46
5.13	Radar plots with champions deaths for the Red team: Victory (left); and Defeat (Right)	47
5.14	Radar plots with champions deaths for the Blue team: Victory (left); and Defeat (Right)	48
5.15	Comparison between different patches in red team's victory on professional players	49
5.16	Comparison between different patches in red team's defeat on professional players	50
5.17	Example of the best K values according to <i>silhouette score</i> in the red team's victory	51
5.18	Deaths clusters of top laners in red team's defeat	52
5.19	Deaths clusters of top laners in blue team's victory	53

5.20 Deaths clusters of mid laners in red team's victory	54
5.21 Deaths clusters of mid laners in blue team's defeat	55
5.22 Deaths clusters of professional bot laners and supports in red team's victory	55
5.23 Best K values according to <i>silhouette score</i>	56
C.1 Clustering professional jungler players' deaths in blue team's victory	79
C.2 Clustering professional jungler players' deaths in blue team's defeat	79
C.3 Comparison between different patches in blue team's victory on professional players	80
C.4 Comparison between different patches in blue team's defeat on professional players	80
C.5 Location clusters in red team's victory on professional players	81
C.6 Location clusters in red team's defeat on professional players	81
C.7 Location clusters in blue team's victory on professional players	81
C.8 Location clusters in blue team's defeat on professional players	82

List of Tables

A.1 APIs explored for different games and which type of data can be collected from	
them	73
B.1 Description of some important features that can be collected from Riot's API . .	77
B.2 Description of some important features that can be collected from Riot's API . .	77

Acronyms

3D three-dimensional. 8, 12, 13

AI Artificial Intelligence. v, 12

API Application Programming Interface. v, vi, ix, 19, 21, 22, 25–28, 64, 71–73, 75

BR Battle Royale. vi, 25, 72

CS:GO Counter Strike : Global Offensive. vi, 1, 25, 72, 73

DBSCAN Density-based spatial clustering and application with noise. 7, 15, 34, 35

DEDICOM Decomposing Directional Components. 12, 13

DESICOM Decomposition into Simple Components. 12, 13

DOTA2 Defense of the Ancients 2. v, vi, 1, 12, 25, 65, 71, 73

ERL European Regional Leagues. 59

Esports Electronic sports. v, ix, 1, 2, 11, 14, 16, 17, 61, 63–65

FPS First-Person Shooter. vi, 25, 72

GA Genetic Algorithms. 12

HC Hierarchical clustering. 7

JSON JavaScript Object Notation. 27

K Number of clusters. xv, xvi, 7, 8, 20, 36, 51, 56

KDE Kernel Density Estimate. 38, 39

LoL League of Legends. v, vi, ix, 1, 2, 5, 8, 9, 13, 25, 26, 31, 33, 59, 60, 63

LP League points. 9, 10

ML Machine Learning. v, 5, 6, 11

MMR Match Making Rating. 9, 10

MOBA Multiplayer Online Battle Arena. v, vi, 2, 25, 71

NaN Not a Number. 27, 56

NN Neural Networks. v, 12

PUBG PlayerUnknown's Battleground. vi, 25, 72, 73

VTK Visualization Toolkit. 22

Chapter 1

Introduction

This chapter provides an overview of the project’s main motivations, goals, and the structure of the document. In Section [1.1](#), the primary motivation behind this research project is explored, emphasizing the significance of video game analysis as a central research topic. Section [1.2](#) describes specific goals that will be addressed throughout this study. Finally, in Section [1.3](#), a detailed overview of the document’s structure is provided.

1.1 Motivation

In the past few years, Esports has become one of the world’s most popular forms of entertainment, rivaling traditional sporting events. In an article published early last year by Newzoo [\[40\]](#), an analytics company, the Esports industry was estimated to generate nearly \$1.38 billion in revenues globally and a worldwide audience of 532 million people by the end of 2022.

Games like LoL [\[20\]](#), CS:GO [\[41\]](#), Valorant [\[23\]](#), and DOTA2 [\[42\]](#) are among the most played competitive 5vs5 team games worldwide. These games need a large amount of knowledge to be played competitively and have several mechanics and strategies that lead one team to victory. Many companies and research groups have developed computational tools for Esports (e.g. Blitz.gg [\[13\]](#), OP.GG [\[26\]](#), and Mobalytics [\[25\]](#), among others) to help players overcome obstacles and learn to play faster and more efficiently.

Much like the world of traditional sports analytics, data plays an indispensable role in the realm of Esports. Like in conventional sports, winning games depends on understanding which strategies work and which do not. By having access to a large volume of information, data analysts can find different gameplays, techniques, and strategies that facilitate the decision-making that can guarantee victory. It allows researchers and the Esports industry to move away from guesswork and, alternatively, make informed decisions based on pre-processed and analyzed data [\[16\]](#).

While there are already numerous ways for acquiring game data, predominantly comprising statistical information from a player’s history, there remains a significant challenge: the scarcity of readily accessible spatio-temporal data. The inclusion of spatio-temporal data promises to propel Esports even further forward by offering insights into a player’s actions in both time and space,

which can deliver more insights and complements than just statistical data. For example, gaining knowledge of where the player is dying when he died and whether he is dying alone or with nearby allies can provide additional information over just knowing the number of times the player has died.

With these considerations in mind, Esports teams and individual players are constantly looking for ways to improve their performance. By harnessing advanced data analytics and spatio-temporal insights, this project aims to contribute to the development of tailored training regimens and strategies that elevate player skills and overall team performance. The ability to dissect and analyze spatio-temporal data can offer crucial insights into player movement patterns, decision-making under pressure, and the effectiveness of in-game strategies, enabling them to refine their skills, adapt strategies, and ultimately achieve higher levels of performance in the world of Esports.

1.2 Goals

The main goals of this thesis encompass a diverse and comprehensive exploration of datasets sourced from various games while incorporating clustering techniques specifically designed for the analysis of spatio-temporal data derived from players while continuing the work on visual analytics of video game data.

In the course of the datasets exploration, it was recognized the significance of selecting a Multiplayer Online Battle Arena (MOBA) game, with a particular focus on League of Legends (LoL). This strategic decision was driven by the accessibility of data, especially spatio-temporal data, which is crucial for the analysis and research objectives. By narrowing the focus to LoL, the research can harness the wealth of readily available data to enhance insights and findings.

In summary, the overarching goals of this thesis are to delve into diverse gaming datasets, leverage state-of-the-art clustering techniques for spatio-temporal data, and advance the field of visual analytics in the context of video game data analysis. The selection of LoL as the primary game of interest aligns with the mission of facilitating comprehensive and insightful research facilitated by the accessibility of pertinent data.

1.3 Structure of the document

This document is organized as follows:

- Chapter 2 - describes some concepts in the context of game data analysis, focusing on information that can help understand this project, and presents the literature review.
- Chapter 3 - describes the methodology, technologies and tools used in this work.
- Chapter 4 - presents the reasons for selecting a specific game to analyze, and describes how the datasets were created, how it is constituted and what differs from others that can be found online.

- Chapter 5 - describes the cluster analyses performed.
- Chapter 6 - presents the validation of the results made by a professional coach.
- Chapter 7 - final considerations on the work done.

Chapter 2

Background and Related Work

This chapter provides the essential context for the research, covering key concepts of the subject in Section 2.1 and an in-depth review of previous academic research in the Section 2.2. Through this exploration, the foundations on which the research questions are built are established.

2.1 Background

This section describes some concepts that will be useful in developing the work, such as game data science, game data, Machine Learning Algorithms ML, and spatio-temporal analysis. For ease of reading, this section will also include Section 2.1.5 with a brief explanation of LoL [20] and how to play it since it is the game that will be the focus of the analysis.

2.1.1 Game data science

This subsection describes the concept of game data science.

Game data science is a field that involves using data science techniques to analyze and understand player behavior in video games [16]. This includes learning how players interact with the game, identifying patterns in data to inform decision making in different domains, and using that information to improve the game design. As such, game data science includes many analyses, such as predicting when players will stop playing the game, understanding the distribution of player actions, or user segmentation by dividing players into groups based on their characteristics or behavior.

2.1.2 Game data

This subsection presents the data collected from different games and the influence of this information, not only on casual players, teams that compete at the highest level, and their viewers/fans but also on the companies that create those games.

Game data refers to the data generated by players as they interact with a video game. This large volume of game data is collected, stored, analyzed, and used to capture interesting information throughout a game. Typical data sources include behavioral data from the game, information from

advertising partners, and third parties like social media platforms. Another data source is collected from infrastructures such as servers, development processes, marketing, and user research [16].

These sources of data are typically collected by game developers and can be used for a variety of purposes, such as improving the game design and development, analyzing player behavior to help their players master their games, or making business decisions.

The player data is the most commonly used and the most available [16]. This data has different forms, such as behavioral data, also called behavioral telemetry, collected in real-time by the player. Statistical data, such as how many games they played, number of kills, deaths, and assists, and spatio-temporal data, represented in both space and time, like a player's location at a specific time of the match.

2.1.3 Machine learning

This subsection describes the concept of ML, its types, and some clustering algorithms.

Machine Learning is the science of programming computers so they can learn from data [24]. There are several types of ML from data, and for this project, it is important to distinguish three of them: supervised learning, semi-supervised learning, and unsupervised learning [24].

Supervised learning: Supervised learning is where the machine (computer) learns from known data sets, not only the independent variables of each example are known, which can be considered as the cause but also its dependent variable, the variable that can be considered as the effect, the one that we want to know/predict when new examples are given. Supervised algorithms identify patterns in known data, learn from known examples and make predictions on new, unseen data. Some common supervised learning applications include image classification, speech recognition, natural language processing, and fraud detection, among others.

Semi-supervised learning: Semi-supervised learning is similar to supervised learning but uses labeled and unlabeled data. Labeled data has the values of the dependent variables, while unlabeled data lacks that information. By using this combination, machine learning algorithms can learn to label unlabeled data. Semi-supervised learning can be useful when it is difficult or expensive to label a large dataset, but some labeled data can still be used to train a model [19]. It can be applied in disease detection, and customer segmentation, among others.

Unsupervised learning: Lastly, and the most important for this project, is unsupervised learning. This project uses clustering algorithms to solve this type of problem. Here, the model is not given any labeled data or guidance on what it is trying to learn [19]. The analysis of the available data is expected to determine correlations and other relationships. Some typical applications of unsupervised learning include dimensionality reduction, data compression, and density estimation.

Clustering in game data science

One of the challenges in game data analysis is the high dimensionality of behavioral telemetry. Many features can be extracted for each player or team [7], making it hard to find behavioral patterns. One way to overcome this problem is to use clustering methods to explore datasets and discover patterns reducing the complexity of the data [16]. Clustering is a technique used to group together data points that are similar to each other.

One common application of clustering in game data science is segmenting players into groups based on their in-game behavior or characteristics. For example, a developer might use clustering to group players who died alone without the team being around or to identify groups of players who are more likely to stop playing the game after a certain point.

Clustering can also be useful for identifying patterns in game data that can inform game design or in-game strategies. For example, clustering might be used to identify areas of the game where players are most likely to die.

Several different clustering algorithms can be used, including K-Means clustering, hierarchical clustering, and density-based clustering [16]. Each of these algorithms works in a slightly different way, but they all aim to partition the data into groups or clusters such that data points within a cluster are more similar to each other than they are to data points in other clusters.

- **k-Means** - K-Means is a simple but effective method that groups by the distance to the cluster's center. As mentioned by Magy et al. [16] this method is often used in game data science because it is conceptually easy to understand. The main goal is to find a set of centroids (centers of the clusters) and based on the provided distance metric, each data point is assigned to the centroid closest to it. Since k-Means assume clusters of spherical shape, it is important to scale or normalize the data.
- **Density-based spatial clustering and application with noise (DBSCAN)** - DBSCAN is a density-based clustering algorithm that defines clusters based on identifying areas of higher density. Can have an arbitrary shape, and the points within can be arbitrarily distributed. Density clustering can also handle noise if the noisy data is in areas of sparse data space. As Magy et al. [16] described, Density-based models have rarely been used on behavioral data from games but might be used in the future due to their ability to handle noise (a common feature in games). DBSCAN is deterministic, does not depend on randomness, and does not require the number of clusters (K) as input.
- **Hierarchical clustering (HC)** - A clustering method that builds a hierarchy of clusters instead of a set of clusters. In this technique, the objects are more related to nearby objects than those further away. Does not scale well with a large dataset. These are among the earliest techniques developed, and care must be taken concerning outliers that can cause the merging of clusters (chaining). HC goal is to construct a tree, denoting how data can be

merged or split into nesting clusters. It does not require the knowledge of the number of clusters before running. It is not very robust against noise and outliers, with outliers points ending up in clusters of their own. It can be difficult to select the right K.

2.1.4 Spatio-temporal analysis

This subsection describes what spatio-temporal analysis means.

Playing games involves action in both time and space [16]. Spatio-temporal analysis is a type of analysis that involves the study of these two components. Spatio data refers to data that has a geographic or physical location. Temporal data refers to data with a time component, such as the time an event occurred or the duration of an activity. As noted by Aung et al. [10], in game data science, this analysis can be important when analyzing player behavior, especially in 3D environments, as it helps to understand why players move the way they do.

2.1.5 League of Legends

This subsection briefly explains the LoL game [20] and how to play it to facilitate understanding some expressions used in the following chapters.

The selected game, LoL [20], is a team-based strategy game where two teams of five powerful champions face off to destroy the other's base (Nexus). There are over 140 champions that can be used by the player, each with unique abilities to make epic plays, secure kills, and take down structures as the player tries to achieve victory [20].

Map configuration: The main map shown in Figure 2.1, also called summoner's rift, is symmetrical with both sides containing three lanes, and between them, the jungle. It is covered by a fog of war, limiting the team's field of vision, and each player is associated with one of these five roles: top laner, jungler, mid laner, bot laner and support. However, it is important to note that although the map is symmetrical in terms of its core structure, there are differences in the terrain between the two sides. These variations in terrain, such as the placement of brush, the shape of walls, and other map features, contribute to strategic diversity and gameplay dynamics.

Team vision: Vision is an essential feature in the game, as it helps the team make or remake strategies based on the information gathered. Turrets, champions, wards (champion item), and minions provide limited vision to the team.

Structures and minions: Each lane has three turrets and an inhibitor, and two turrets guard each team's nexus. Nexus is where minions spawn (give birth). The good minions help reach the enemy nexus while killing enemy minions gives gold and experience to the champion and open the path for the good minions to reach enemy turrets and, consequently, the enemy nexus. Turrets damage minions and champions and provide limited vision from the fog of war for their team. Turrets prioritize minions, so they must be attacked with minions ahead of the player to avoid

damage. When an inhibitor is destroyed, super minions (more powerful minions with melee range that can more easily take down enemy minion waves and defense structures) will spawn in that lane for several minutes.

Win condition: One team wins the game by clearing at least one lane, destroying 5 turrets and an inhibitor, to get to the structure that needs to be destroyed to win the game, which is the enemy nexus.

Jungle: The jungle is where neutral minions and jungle plants reside. The three most essential monsters in the jungle are the baron nashor, the drakes, and the rift herald. Killing these units grants unique buffs to the team (champion improvements, such as increased damage, decreased abilities cooldowns, etc). Baron nashor is the most powerful monster in the jungle, it is located at the top side of the map in the river, and killing it grants the team bonus attack damage, ability power, empowered recall (the teleportation time to the base is reduced), and the power of nearby minions. Drakes are located at the bottom side of the map in the river and give unique bonuses depending on the element of the drake (elements: Infernal, Ocean, Mountain, Cloud, Hextech). On the second drake killed, the map changes according to the element of the upcoming drake. Rift herald spawns in the same pit as Baron but before it. Slaying rift herald drops the eye of herald, which grants empowered recall and gives the ability to summon the rift herald to push a lane by destroying turrets.

Team base: The base behind each team nexus (shown in gray in the Figure 2.1) is where the champions spawn and the shop's location. The shop is where the players spend the gold they have earned on items to empower their champions. It is also the location where players regenerate the champion's health.

Players' rankings: In LoL, players engage in competitive gameplay within a mode commonly known as solo/duo queue. In this mode, individuals try to climb through various divisions, showcasing their competitive skills on a personal level. LoL features a total of ten divisions that players can occupy. The lower divisions, often referred to as the "low elo" divisions, are designed for more novice players and include Iron, Bronze, Silver, Gold, and Platinum divisions. In contrast, the "high elo" divisions consist of more skilled players and include the Emerald, Diamond, Master, Grandmaster, and Challenger divisions.

A player's rank depends on their performance, particularly their win-loss ratio. Each victory in a ranked match gives the player additional League Points (LP), while a loss results in a deduction of LP. To advance to higher leagues requires accumulating one hundred LP for each tier.

League of Legends also incorporates a crucial metric known as Match Making Rating (MMR), which serves as a reflection of a player's skill level. MMR takes into account a player's average performance within their division and serves as a valuable tool for self-assessment. Players with

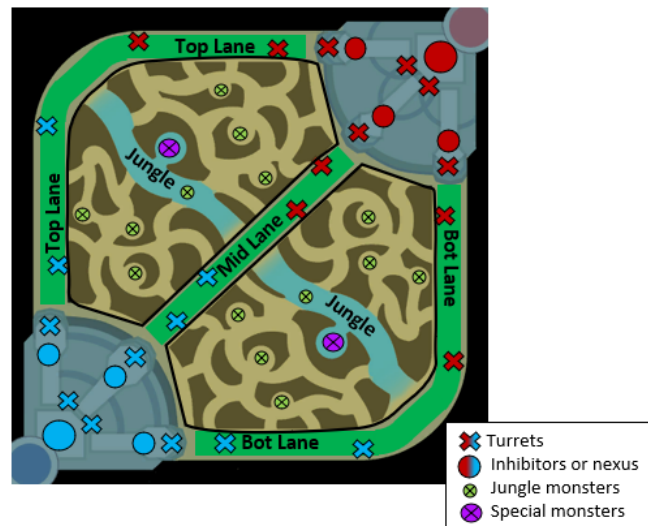


Figure 2.1: League of Legends map with important locations (Riot Games, 2022 [21])

MMR surpassing their current division will gain more LP, whereas those with MMR falling below their division will experience more substantial LP losses. This MMR system not only encourages players to evaluate their true skill level but also ensures a balanced and competitive gaming experience.

2.2 Related Work

This section presents the most important projects related to this work.

In Esports, games tend to be very complex with fast-paced action, where events can happen in multiple areas simultaneously. Having this in mind and the continuous evolution of games over the years, both in terms of players and spectators, research is becoming more and more common to help players, companies, and spectators.

In Section 2.2.1, research focusing on the analysis of non-spatio-temporal data is described. Moving on to Section 2.2.2, projects centered around spatio-temporal data analysis are explored. In Section 2.2.3, the focus changed to League of Legends, one of the most prominent games in the Esports scene. Here, are highlighted some of the most relevant information that can be explored within the context of this popular game. Lastly, in Section 2.2.4, are provided insights into existing data analysis applications in the area of Esports.

2.2.1 Non-spatio-temporal data analysis

This subsection presents some research on non-spatio-temporal data analysis.

Although research work in the field of Esports is becoming more common, many of the analyses have still focused on non-spatio-temporal data.

As an example, Fénix et al. [33] tried to understand the features that help to discriminate between successful and unsuccessful teams to improve their strategies. They identified and characterized team behavior patterns based on historical matches by applying ML and statistical analysis. The teams' performances were clustered and for each cluster, it was investigated how and what features influenced the teams' success or failure. They used only statistical data and the selected algorithm for clustering was the K-Means. To profile each cluster they analyzed the win and loss rate of the teams, the centroids of each cluster, since they summarize the features of the profiles, and to what extent these features have influence or relevance in the profiles. The profiles were shown in radar plots where each axis represented a metric, and the axis length indicated the profile score in a specific metric. Profiles were split into 4 levels regarding the win rate: very low, moderate, high, and very high. 28% of teams fall into the very low level, 36% into the moderate level, 11% into the high level, and 25% into the very high level. They identified deaths, killingSprees and neutralMinionsKilledEnemyJungle as the features that have more influence on the behavior of the profiles.

2.2.2 Spatio-temporal data analysis and clustering

This subsection describes several spatio-temporal data analyses.

Regarding spatio-temporal data, although not yet widely used, there are attempts to make predictions and cluster players based on their behavioral patterns.

Katona et al. [27], for example, aimed to improve the comprehensibility of matches for the

audience through micro-prediction, which consist of predicting events within the match. These micro-predictions offer valuable insights that can be of interest to commentators and audiences, capturing events that might otherwise go unnoticed. This not only captivates commentators and viewers but also serves as a valuable resource for esports team coaches. By using deep learning networks that take into account players' relative strengths and their real-time map locations in the DOTA2 game, Katona et al. [27] achieved accurate death predictions within a five-second timeframe. This innovative approach allows coaches to identify critical moments and understand why the model flagged their players as being potentially in danger, even if they ended up surviving.

Schubert et al. [38], have also tried to make predictions based on spatio-temporal data, more specifically, they presented a technique for segmenting matches into spatio-temporally defined components referred to as encounters. An encounter is when two or more players from opposing teams are in range to attack each other. They applied this technique to DOTA2 and presented win probability predictions based on encounters.

Another interesting work was done by Batsford [11], where he used neural networks (NN) and genetic algorithms (GA) to find an optimal path around the jungle that both players and Artificial intelligence (AI) bots can then use in the DOTA2 game. While the NN was implemented to calculate the next position for the player to move to given any current game state, GAs included a single point crossover and mutation to generate new populations of weights to be used in the NN.

Drachen et al. [15] presented three data-driven measures of spatio-temporal behavior in DOTA2 and investigated how behavior changes across these measures as a function of the skill level of teams from novice to professional players. The first experiment focussed on general movement on DOTA2 map, where the goal was to evaluate whether highly skilled players move more than the other players. The second experiment aimed to assess the team distribution on a map and its correlation with skill level and win/loss outcomes. It focused on the distances between team members. The third experiment was particularly intriguing as it sought to determine the underlying factors that contribute to different movement patterns among players. To achieve this, the researchers employed unsupervised learning, using time series clustering of the average distance between players per second. They utilized permutation distribution of time series as Euclidean and Manhattan distances were unsuitable for matches with diverse sizes and high dimensionality. Dynamic Time Warp would have been computationally inefficient. The clustering algorithms used in the study were K-Medoids and Fuzzy clustering. The researchers compared the silhouette plots from the two clustering solutions and analyzed the clusters. They determined that skill difference is more evident in shorter matches than in longer ones.

Bauchage et al. [12] also show the usage of advanced spatial clustering techniques to evaluate player behavior in 3D worlds, where they compared 4 different algorithms: K-Means, Spectral, DEDICOM, DESICOM. This paper addresses the need for a meaningful partition of maps based on player behavior in order to identify movement preferences. They concluded that the usage of

DEDICOM and DESICOM were good techniques to game analytics since the mechanics of 3D game worlds typically cause asymmetric relations between map locations. According to this paper, these algorithms, compared to the others, consistently produced meaningful and interpretable clusters from player trajectories whereas spectral clustering worked well for symmetric, spatial similarity data.

A work more directed to the game companies than to the players, using spatio-temporal data, was done by Aung et al. [10] that had as a core objective the investigation of behavioral differences between different player types defined by early abandonment and commitment in the game Just Cause 2. They applied cluster analysis and the DEDICOM decompositional model, adapted from the work of Bauchage et al. [12], to profile the behavior of players of the game Just Cause 2 by integrating both spatio-temporal trails and behavioral metrics. Firstly, DEDICOM is employed to build pattern groups for the spatio-temporal trails of players. After that, spatio-temporal profiles are constructed for the spatio-temporal trails and combined with behavioral features across the play history of the player. They extracted central waypoints from the clusters generated by the K-Means algorithm, which was used to cluster the locational information of players, and later created a waypoint graph that encodes the movement information between these extracted waypoints. Finally, for each cluster, they examined nine key non-spatio-temporal statistics such as Kill/Death ratio, playtime, game progress, etc to profile them.

Additionally, note that no research work about spatio-temporal analysis in LoL [20] was identified.

2.2.3 League of Legends game

This subsection presents the most relevant information that can be extracted from LoL [20].

League of Legends, with its rich gameplay and dedicated player base, provides a fertile ground for data analysis and research. Knowing what information needs to be analyzed in a given game can sometimes be tricky. In LoL [20], there is a lot of information that can be collected, but there are some that may be more important than others. In VisualLeague III [9], and according to the interview conducted by Afonso with two professional League of Legends coaches, it can be seen what kind of information and analysis the coaches prefer. From that, and in accordance with the objective of the project, the most relevant information extracted from the interview was the following:

- Know when a player returns to the base, whether or not they return at the same time as the other players. This information can be essential as it can impact when a team fight starts, more specifically, whether or not it delays the start of a team fight;
- The most important early timestamps are when rift herald and drakes spawn;
- Know where and when wards are placed;

- Know the path the jungler usually takes, as he is typically the main shot-caller of the team;
- Know the players' locations at the mid game because this is when the map is more open (risk of more team fights);
- Know the location and the time when a team fight happens, as it gives information on whether or not a team is good because typically good teams just fight for objectives like baron, and drakes;
- What happens after an important objective is killed like turrets, baron, drakes, rift herald;
- Know whether or not a champion has prio or not. This is related to the state of the minions' waves or the champions' abilities (champion clearing ability).

2.2.4 League of Legends data analysis applications

This subsection describes some data analysis applications.

The following applications have been developed to take advantage of the wealth of data generated by Esports events and provide valuable tools and information for players, teams, and organizations involved in the Esports industry.

Regarding currently used applications that provide game data analysis to consumers, the following applications were found: OP.GG [26], Mobalytics [25], Blitz.gg [13], porofessor.gg [43]. All these applications provide several players and teams statistics for a given match. These statistics are presented in charts, histograms, radars, and even map representations to identify, for example, spatio-temporal data like death locations. More particularly, OP.GG [26], one of the most popular applications, is often used to obtain statistical data for a single player, making it a very efficient tool for finding specific player strengths and weaknesses.

The blitz.gg application [13] is another example of one of the most used applications for gaming coaches, providing everything the others offer plus real-time data analysis represented in in-game overlays. Has a partnership with Riot Games [22], and simplifies the process of mastering difficult games. This app gives the users auto pre-game setup and information by auto-importing optimal setups into the player's game. In real-time, it also offers helpful in-game overlays that tell the player which goals to achieve and the progress towards achieving them in-game, as shown in Figure 2.2. It also gives personalized post-match feedback, helping the player to learn faster by analyzing how he played and delivering targeted feedback.

One of the best in-game overlays of blitz.gg is shown in the top right of Figure 2.2. This overlay helps players know which order they must take to kill the jungle camps in the League of Legends game [20], according to the players' average, and how long they should take compared to the average time the players take to kill them. This overlay can be very useful since it informs a player by comparing his jungle clearing timing to the average player of a selected ranking.

Another noteworthy contribution in this field is the work by Rafael [9], who developed a tool called VisualLeague III. This tool offers a range of valuable visualization techniques designed to

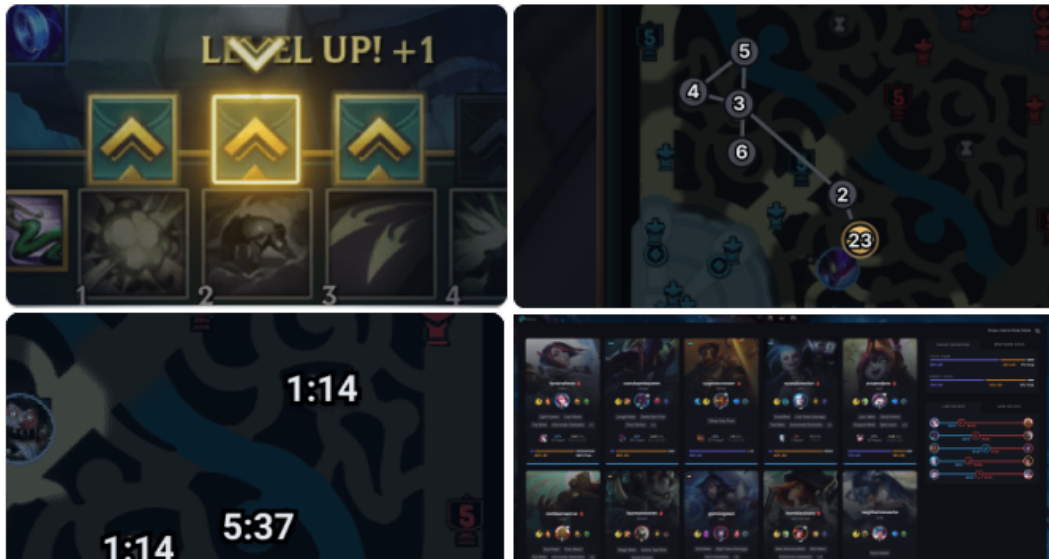


Figure 2.2: Blitz.gg in-game overlays of a League of Legends match (Blitz.gg, 2022 [13])

assist players and coaches in analyzing player matches more effectively. One particularly notable feature, as can be seen in Figure 2.3, is the interactive time-based visualization, which allows users to explore the evolution of a player’s movements throughout a match.

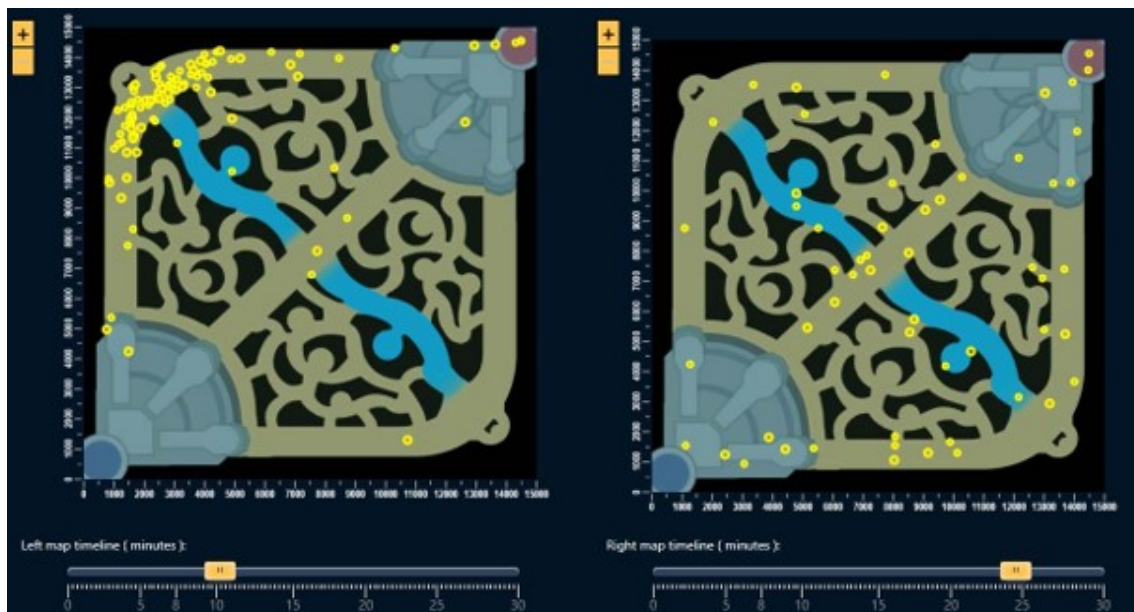


Figure 2.3: Interactive time-based visualization in VisualLeague III (Rafael, 2020 [9])

In addition, VisualLeague III includes a powerful clustering filter that leverages modifications to the DBSCAN algorithm. This feature is particularly useful for players and coaches when dealing with spatio-temporal data from numerous matches. Its primary purpose, as can be seen in Figure 2.4, is to address issues related to cluttering and overlapping data, which can become particularly problematic in analyses involving twenty or more matches. This clustering filter plays a

key role in increasing the clarity and utility of the analytical insights obtained with the tool.

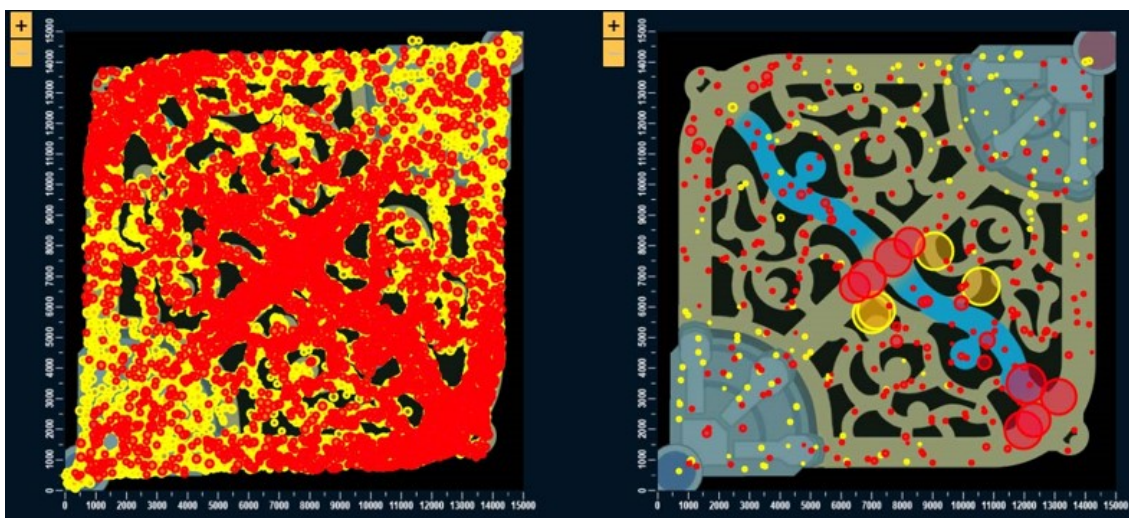


Figure 2.4: Clustering filter in VisualLeague III (Rafael, 2020 [9])

2.3 Summary

The Background and Related Work chapter provides essential context for understanding the research's foundation and its place within the existing body of knowledge.

In the background section, the chapter delves into fundamental concepts of game data science. It begins by defining the field of game data science, emphasizing its role in unraveling the intricacies of competitive gaming. The section progresses to the significance of game data and its far-reaching influence, highlighting how it underpins decision-making, player performance analysis, and the evolution of Esports as an industry.

Furthermore, the chapter explores the notion of spatio-temporal analysis in the context of game data science. It elucidates the meaning and relevance of spatio-temporal analysis, emphasizing its ability to reveal hidden knowledge within the spatial and temporal dimensions of gameplay. By studying player movements, map control, and dynamic interactions, this analytical approach promises to unlock a deeper understanding of League of Legends and Esports in general.

The chapter also explores the domain of machine learning, highlighting its applicability in analyzing Esports data. By laying this foundations, a clear understanding is gained of the interaction between Esports and analytical methods.

Additionally, the chapter introduces one of the most prestigious Esports titles, League of Legends, providing an in-depth explanation of its mechanics, gameplay dynamics, and map configuration, among others. This section serves to describe the specific game under analysis, ensuring a solid comprehension of its details and the spatio-temporal data generation processes unique to it.

In the related work section, prior research in the field of game data science and Esports data analysis is reviewed. A notable observation is that existing studies predominantly focus on non-spatio-temporal data, such as player statistics, team performance metrics, and match outcomes.

These analyses have contributed significantly to the Esports research scene but tend to ignore the spatio-temporal data generated during gameplay.

A significant gap in the literature is identified – the lack of comprehensive analyses on spatio-temporal data in League of Legends from the perspective of game data science. Although there is information available in this field, previous research has not focused on exploring the spatial and temporal dimensions of the game, including player movements, map control, and dynamic interactions. This omission highlights the novelty and potential contributions to filling this analytical gap.

The next chapter will delve into the methodologies, technologies, and tools required for conducting this type of analysis.

Chapter 3

Methodology, Technologies and Tools

This chapter provides an overview of the methodology adopted for the development of this project, as well as the set of technologies, and tools used during its execution. In Section 3.1, an exploration of the methodology employed is presented, explaining the sequential steps that underpin the entire project. Next, the Section 3.2 discusses the set of resources used to improve the framework and execution of the project.

3.1 Methodology

The methodology adopted relies on the knowledge discover process presented in the Game Data Science book [16], with some additional elements such as dataset creation, and results' validation steps, as illustrated in Figure 3.1. This process comprises eight steps:

1. **Data collection** - This is where the data is collected from relevant sources, typically through the use of APIs or other data retrieval methods. This data is crucial as it forms the foundation for the entire analysis.
2. **Datasets creation** - Following data collection, the next step is to create datasets using the collected data. These datasets should be organized and structured in a way that facilitates subsequent analysis. Clearly define the variables included in the datasets and document any data cleaning or transformation steps that are applied to ensure data quality and consistency.
3. **Variable selection and/or transformation (Data pre-processing)** - This step is where it is decided which data features will be used as input for the algorithm. In some cases, especially when having too many features and some being irrelevant, data transformation is needed to get the correct set of variables for the analysis.
4. **Selection of metric** - Selecting an appropriate similarity or distance metric is pivotal to the success of clustering analysis. This metric defines the “similarity” in the data. There are

some popular choices, such as the Minkowski distance function for data represented as numeric vectors or the Euclidean distance.

5. **Selection of clustering** - There are many clustering algorithms, so when selecting an algorithm, it is necessary to consider its computational demand and ease of use and be aware of its limitations. Silhouette score can be a good metric to help select the clustering algorithm to use.
6. **Model tuning** - In this step is determined the right values for hyperparameters. Some algorithms require the number of clusters (K), as an input. It is necessary to determine a suitable K to use in the algorithm. The elbow method and the silhouette score can be good ways to help select the best K .
7. **Result evaluation and interpretation** - Upon successful clustering, evaluation of the obtained clusters is conducted using established metrics such as the silhouette score. Visualization techniques are applied to gain insights and facilitate the interpretation of the clusters. This step involves a comprehensive exploration of the implications and relevance of the clustering results in relation to the research objectives.

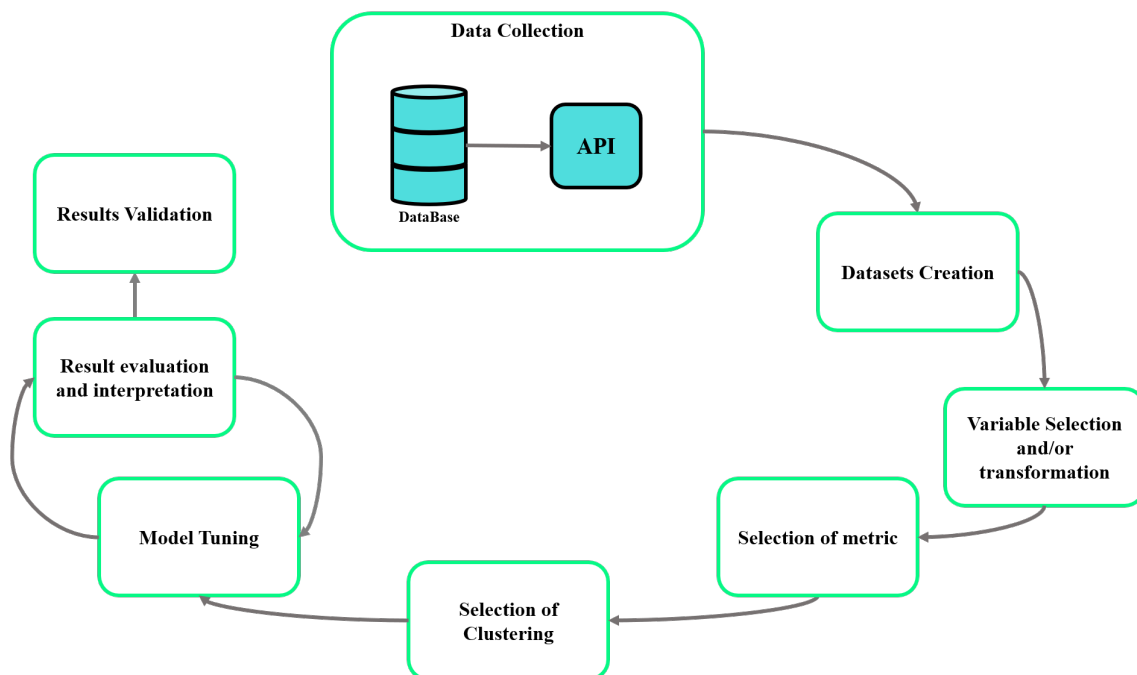


Figure 3.1: Steps of clustering process

8. **Results validation** - The final step encompasses the critical process of results validation. Collaboration with domain experts or professionals is instrumental in confirming the prac-

tical significance of the clusters. The feedback received from an expert is documented, and any disparities or ambiguities are addressed to ensure the validity and reliability of the findings. This phase serves to enhance the robustness and real-world applicability of the clustering results.

In Chapter 4 and next, it is discussed each of the steps instantiated to the goals of this work.

3.2 Technologies and Tools

All the work was carried out using the Python programming language, either to collect data through request modules or to analyze the video game data.

3.2.1 OPGG

OP.GG [26] proved to be an indispensable resource for the project, serving as a reliable and accessible repository of player statistics across multiple online gaming titles. Its primary function was to aid in selecting players who met predetermined criteria set by coaches and analysts in previous works [9] to be included in the datasets. The selection process involved various criteria, with the most prominent being the minimum number of games a player needed to have played during the patches to be considered for analysis.

OP.GG offered an easy way to access various pieces of information about the players, such as their current divisions and rankings within the game. This feature provided a valuable context for the process of creating datasets, making it possible to consider not only the number of games played but also the competitive level at which players were participating. This depth of information allowed to make more informed decisions when selecting players for analysis, ensuring that the datasets were not only comprehensive but also reflective of the players' competitive standings.

3.2.2 Tracking the professionals

Trackingthepros [31] also assumed a crucial role in data acquisition strategy, focusing on professional League of Legends players. Its primary function was to assist in identifying these players' accounts, retrieving their unique IDs, and subsequently accessing their match histories. This process was instrumental in determining whether these professionals met the predetermined criteria set for inclusion in the dataset. Its specialized focus on this specific gaming community, combined with its efficiency in data retrieval, made it an ideal complement to OP.GG, and Riot API.

3.2.3 Jupyter Notebook

The code for analysis was written in the Jupyter Notebook [1] as it offers a way to mix Python code blocks with formatted text, allowing data scientists to create and share documents that integrate live code, equations, computational output, visualizations, and other multimedia resources, along with explanatory text in a single document. It is used for all sorts of data science tasks, including data cleaning and transformation, numerical simulation, exploratory data analysis, data

visualization, statistical modeling, machine learning, and deep learning, among others. It is very interactive, as users run the code, see what happens, modify it, and repeat it in a kind of iterative conversation between the data scientist and the data.

3.2.4 Libraries

RiotWatcher: The RiotWatcher [6] is a Python wrapper for the Riot Games API, which allows developers to access data from Riot Games' various gaming platforms, most notably the League of Legends game. RiotWatcher [6] was chosen for its ease of use, providing a clean and intuitive interface for interacting with the Riot Games API, for its simplicity, abstracting away many of the low-level details, for its documentation, as it is very comprehensive and well organized, and for being actively maintained, which means it is more likely to adapt to any updates or changes made by Riot Games.

Scikit-Learn: The Scikit-Learn [7], also known as sklearn, is a Python library to implement machine learning models and statistical modeling. Through Scikit-Learn [7], various machine learning models, particularly clustering algorithms, were used to analyze the game's spatio-temporal data.

Pandas and Numpy: Pandas [4] and Numpy [3] libraries are two of the most popular libraries in data science and were used to analyze and manipulate the data. Pandas [4] was primarily used for data manipulation and analysis, particularly for structured or tabular data. Pandas [4] provides a powerful tool for data cleaning, transformation, aggregation, filtering, and merging. It's widely used for data preprocessing tasks, exploratory data analysis, and creating summary statistics. Pandas [4] allows to load data from various formats (CSV, Excel, databases), perform calculations, handle missing values, and perform complex data manipulations efficiently.

Numpy [3] was primarily used for tasks such as:

- Array manipulation - to find, for example, the unique elements of an array while preserving their order, or find the maximum value along a specified axis in an array.
- Numerical operations and mathematical functions - to calculate, for example, the arithmetic mean (average) of the elements in an array, or to compute the standard deviation of the elements in an array.

Matplotlib and Seaborn: Matplotlib [2] and Seaborn [8] libraries were used to create different types of visualizations of the results obtained in the analyses performed, as they are two of the most widely used libraries for creating visualizations in Python.

PyVista and VTK: Visualization Toolkit (VTK) is a complex and versatile library that allows the creation of interactive 3D graphics applications, visualization of scientific data, rendering of 3D scenes, and performing various data processing tasks. PyVista [5] library was used to simplify the usage of VTK by offering a more user-friendly Pythonic API while maintaining access to the underlying capabilities of VTK. This was used in particular to create the 3D visualization of the spatio-temporal analysis.

3.3 Summary

In summary, Chapter 3 provides a detailed description of the methodology employed, emphasizing the steps taken in the project's execution. Additionally, it highlights the essential technologies and tools used to collect, analyze, and visualize data for the project.

In the following chapters, the document will delve into the various stages of the knowledge discovery process. It begins by discussing the choice of League of Legends as the game for analysis and the development of datasets for this game.

Chapter 4

League of Legend Game and Datasets

This chapter briefly explains how the LoL game was chosen for analysis over other games, with a detailed explanation provided in Section [4.1](#). Additionally, it discusses the process of creating datasets for this game and the underlying motivations, which are elaborated on in Section [4.2](#).

4.1 Game selection

Several game APIs of different genres, including MOBA (Riot Games API [\[21\]](#), Overwolf [\[35\]](#), and OpenDota [\[39\]](#)), FPS (Riot Games API [\[21\]](#), Overwolf [\[35\]](#), Tracker.gg [\[34\]](#), and Faceit [\[18\]](#)), and even BR games (PUBG API [\[29\]](#), and Overwolf [\[35\]](#)), were explored. This comprehensive exploration included understanding how games are played, distinctions among gaming genres, and the nature of data accessible through the APIs for the developer community.

The game API's selection followed three criteria:

1. The game's API must give access to spatio-temporal data.
2. A solid comprehension of the game's core elements, such as its structure and gameplay mechanics.
3. A game that has not yet been used much in research work, particularly within the domain of spatio-temporal data analysis.

The related work section shows that DOTA2 is the most commonly used dataset for game data science research, and accordingly to Table [A.1](#), OpenDota and PUBG APIs [\[39, 29\]](#) are the APIs that give access to most of their games' data. However, it was decided to select the Riot API [\[21\]](#) to collect League of Legends [\[20\]](#) data. For the simple reason of being the most experienced game, also allowing the collection of multiple types of data, especially spatio-temporal data like in OpenDota and PUBG APIs [\[39, 29\]](#), and because no research work about spatio-temporal analysis in League of Legends [\[20\]](#) was found.

The games Valorant [\[23\]](#) and CS:GO [\[41\]](#) will therefore not be selected due to the impossibility of collecting spatio-temporal data, as can also be seen in Table [A.1](#).

As mentioned before, LoL was selected as the game to be analyzed spatiotemporally, as it was the one that fulfilled all the criteria. More details about the data that can be collected from this and other APIs, and more concrete information on the games explored can be found in the Appendix [A](#).

4.2 Game datasets

In response to the scarcity of datasets featuring spatio-temporal data in the context of League of Legends (LoL), two distinct datasets from patch 13.3 were created. These datasets have been created to meet the needs of developers looking for pre-processed LoL datasets readily available with spatio-temporal information. One dataset included a large volume of data from professional players' matches, while the other included information from amateur players' matches. They were also created to assist the spatio-temporal analyses in this study.

The subsequent sections, namely [4.2.1](#), [4.2.2](#), and [4.2.3](#), delve into the detailed methodologies employed during the creation of these datasets. Additionally, section [4.2.4](#) offers a succinct yet insightful analysis of the datasets.

Section [4.2.5](#) elucidates the rationale behind the creation of two more separate datasets, providing valuable context for researchers and developers alike.

4.2.1 Players' Information

In the process of creating the datasets for data collection from the Riot API, the research began with the initiation of targeted searches on OP.GG [\[26\]](#), and [trackingthepros \[31\]](#). Two distinct categories of League of Legends players were identified, as represented in Figure [4.1](#) (steps 1 and 2):

- Professional players - The initial search encompassed players for each of the five primary roles (top, jungle, mid, bot, and support) across the European, North American, and Korean regions. These professional players represent the highest levels of League of Legends competition.
- Amateur players - Focused searches were also conducted to identify European players for each role, specifically within the Iron, Bronze, Silver, Gold, and Platinum divisions, commonly referred to as the "low elo divisions".

From this pool of identified players, specific criteria were employed based on insights collected from the interview conducted by Afonso with two professional League of Legends coaches [\[9\]](#). Specifically, players who had accumulated a minimum of 20 games played in the context of patch 13.3 were selected, ensuring that the dataset would reflect recent gameplay experiences.

To facilitate data collection and simplify subsequent processes, essential player information was recorded. This included their usernames, respective gaming regions, designated roles, unique player IDs, and any historical usernames they may have had.

All of this valuable player information was organized and stored in a JavaScript Object Notation (JSON) file. This JSON file serves as a crucial reference point, enabling the research to keep player data up-to-date and readily accessible for future data collection endeavors from the Riot API, specifically within the context of each player's games.

4.2.2 Data collection

After careful player selection, a script (represented in Figure 4.1, step 3) was created to use the Riot API (represented in Figure 4.1, step 4). This script was designed to retrieve essential player match history data and to facilitate continuously updating player information in the JSON file whenever changes occurred, such as changes to a player's username. This script was continuously executed until the introduction of a new patch, ensuring the relevance of the datasets.

The fundamental operation of this script centered on the use of the players' unique IDs, which had been systematically recorded in the JSON file. These IDs served as the key to accessing the corresponding match IDs associated with each player. Consequently, this access allowed for an in-depth examination of the match timeline, a comprehensive source of minute-by-minute statistical insights related to champion performance. These statistics covered a wide range of attributes, including health, armor, magic damage, attack damage, magic resistance, and more. Furthermore, the match timeline provided valuable information about the in-game events. These events included details such as the player's time and location of death, the time and lane of the structures' demolition, and the timing of epic monster eliminations, among other key occurrences.

Significantly, in the datasets created when an event occurs, all event-related columns will have values, and the others will not (columns not related to events will have NaN values). Similarly, when collecting minute-by-minute data on champions, the opposite is true: all columns related to events will not have values (NaN values), and the champion's information features will have.

4.2.3 Data preprocessing

After collecting all matches' data, a clean and checking process of the datasets began (represented in Figure 4.1, step 5). This required an examination of each feature to verify that values were consistent with their expected types and that any instances of Not a Number (NaN) were appropriately situated within the dataset. Additionally, special attention was given to features with discrete values, such as events, which were relatively easy to validate due to their limited range of possibilities.

To enhance data quality, rows with any duplicate games were deleted, ensuring that each match was unique and representative of distinct gameplay instances. Furthermore, matches with a duration of less than 10 minutes were excluded from the dataset, as were the corresponding players who had participated in those games. This curation process was performed to maintain a higher level of data reliability and statistical significance.

In addition, five features were removed from consideration due to their consistent null values across all records. This step was taken to organize the dataset and remove superfluous information

that would not contribute to meaningful analysis.

Lastly, to safeguard the privacy of individuals involved, all usernames were anonymized. This precaution was implemented to protect the identities of players while still allowing for comprehensive analysis of the data.

The final dataset with professional players includes 55 features, while the dataset with amateur players includes an additional feature that represents the player's division, categorized as iron, bronze, silver, gold, or platinum. For a more detailed description of the features collected from the Riot Games API, please refer to Appendix [B](#).

This rigorous process of data collection and refinement ensures that the datasets are reliable and suitable for the subsequent analyses carried out in this study.

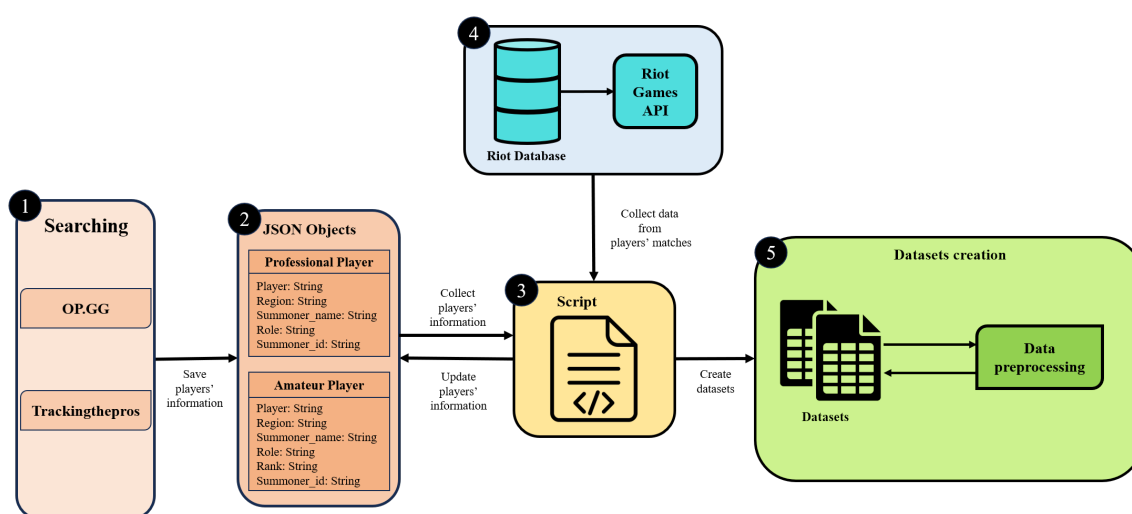


Figure 4.1: Steps for creating the datasets

4.2.4 Dataset analysis

In the following subsections, a concise analysis of the datasets created will be presented, offering a general overview of their key characteristics. The exploratory data analysis (EDA) report, which was created with a comprehensive summary of the dataset's structure, statistics, distributions, and more, is available on the project's github repository for reference [\[37\]](#).

This analysis provides insights into the number of players involved, the number of games, and other pertinent details regarding the datasets, setting the stage for a broader understanding of the data's content and significance.

Professional dataset

In this subsection dedicated to the analysis of the professional dataset, the gathered data from a comprehensive sample of 100 players, spanning a total of 3,946 matches, is presented. These

matches were distributed across different roles within the game, with 836 matches featuring players in the top role, 847 in the jungle role, 1,120 in the mid role, 1,080 in the bot role, and 853 in the utility/support role. This diverse dataset provides a robust foundation for the analysis.

Furthermore, the players have been categorized into their respective roles, with 17 players in the top role, 20 in the Jungle role, 19 each in the mid and support roles, and 25 in the bot role. This categorization allows for a more detailed examination of performance and trends within each role.

In terms of regional distribution, the dataset encompasses matches from three major regions: Europe, North America, and Korea. Specifically, there were 1,847 matches from Europe, 855 matches from North America, and 1,244 matches from Korea. These regional distinctions enable an exploration of potential variations in player performance across different gaming environments.

Additionally, within the dataset analysis, valuable insights were uncovered regarding match duration, with an average match time of 25 minutes.

Notably, a surprising phenomenon is revealed: a significant difference in win rates between the red and blue sides of the game. Figure 4.2(A) illustrates that players on the red side have a notably higher win rate of 60.1% compared to the 39.9% win rate on the blue side. This observation caused some surprises, but with the course of the analyses that will be detailed in the following chapter, some possible reasons for this difference may have been found.

Amateur dataset

In the context of the amateur player's dataset analysis, a repository of 9,617 matches was gathered, revealing a distribution across various in-game roles. Specifically, there were 2,054 matches in the top role, 1,870 in the jungle role, 1,788 in the mid role, 1,881 in the bot role, and 2,060 in the utility/support role.

Furthermore, the 167 players within the dataset have been classified into their respective roles, comprising 35 in the top role, 34 in both the jungle and mid roles, and 32 in both the bot and support roles.

The dataset also presents matches categorized by division, showcasing 1,121 matches in the Iron division, 2,399 in the Bronze division, 2,088 in the Silver division, 1,670 in the Gold division, and 2,340 in the Platinum division. This stratification by division is instrumental in exploring potential disparities in gameplay and outcomes across different skill levels.

Moreover, the analysis has unveiled a noteworthy observation: the average match duration is 30 minutes.

Additionally, as Figure 4.2(B) shows, there is a slight difference in the win rates between the red and blue sides of the game, with a win rate of 50.9% on the red side compared to 49.1% on the blue side. This discovery is of interest and requires a justification that will be given in the following chapters.

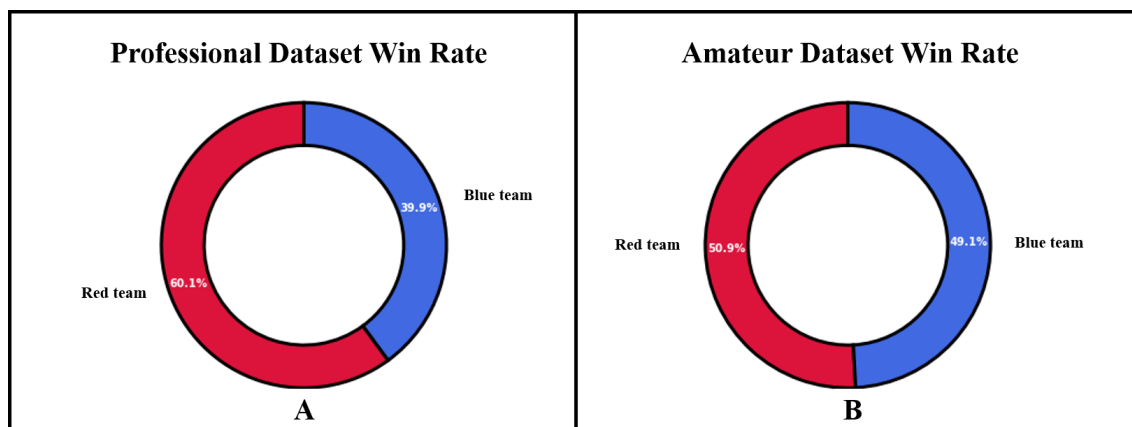


Figure 4.2: Patch 13.3 - Datasets win rate

4.2.5 Additional datasets

In addition to the creation of datasets that facilitated spatio-temporal data analysis for each of the designated roles (Top, Jungle, Mid, Bot, and Support) in patch 13.3, additional datasets were created for patch 13.9, following the same data collection methodology. However, for patch 13.9, the focus was narrowed to encompass only data from professional and amateur junglers, enabling comparisons between the analyses conducted under these distinct scenarios for the Jungle role.

Within the dataset related to professional players, data was gathered from 680 jungle matches, with 19 professional players. The regional distribution included 340 matches in the Europe region, 50 in North America, and 290 in Korea. Notably, these matches had an average duration of 25 minutes. Figure 4.3 (A) illustrates that players are more likely to secure victories while on the red side, with a win rate of 57.1%, as opposed to 42.9% on the blue side. It's worth noting that this reflects an increased win rate on the blue side compared to patch 13.3.

On the other hand, the dataset featuring amateur players includes 757 jungle matches, with 24 amateur players. These matches had an average duration of 29 minutes. Remarkably, Figure 4.3 (B) indicates a distinct change, with a higher probability of winning on the blue side, with a win rate of 55.5%, compared to 44.5% on the red side. This marks another increase in the blue side's win rate compared to the previous patch.

Overall, this exploration of the datasets reveals several noteworthy observations. Firstly, it is evident that the datasets for patch 13.9 are characterized by smaller player and game counts compared to patch 13.3. Additionally, a new trend emerges, wherein the blue team's win rate has experienced an increase in patch 13.9 when contrasted with the win rates observed in patch 13.3. Further insights into the potential factors contributing to these changes will be described further on, as the analysis unfolds.

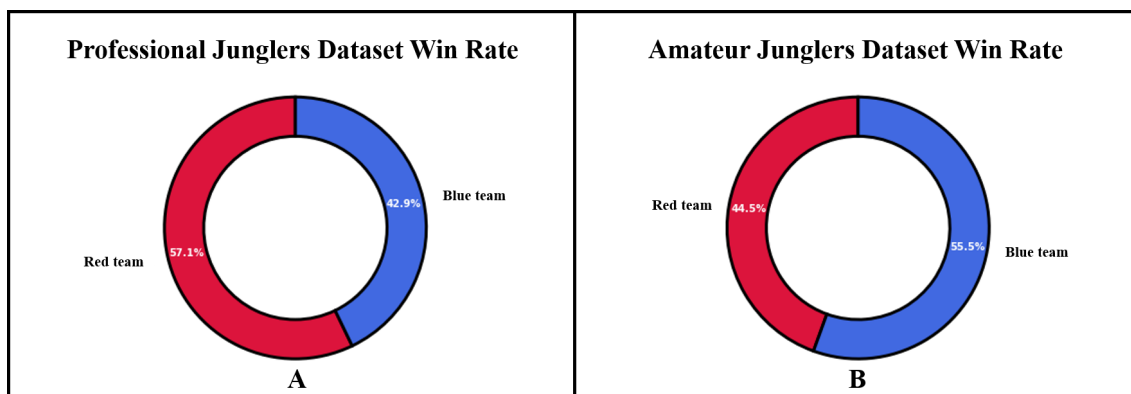


Figure 4.3: Patch 13.9 - Datasets win rate

4.3 Summary

This chapter describes the process of selecting LoL as the game for spatio-temporal data analysis, as well as the creation of datasets for this purpose.

The chapter begins by discussing the criteria for selecting the game for analysis. These criteria include the availability of spatio-temporal data, a deep understanding of the game's core elements, and the relative lack of previous research on spatio-temporal analysis in the chosen game.

The chapter then delves into the creation of two distinct datasets from patch 13.3: one featuring professional players and the other featuring amateur players. These datasets are designed to meet the needs of developers and researchers seeking readily available LoL datasets with spatio-temporal information. The subsequent sections described the methodologies employed in creating these datasets.

In the following chapter, the focus changes towards the cluster analysis of spatio-temporal data obtained from the League of Legends datasets. This chapter describes the specific analyses conducted, details the data preprocessing steps for each analysis, discusses the chosen distance metric and clustering algorithm, and provides an interpretation of the results obtained.

Chapter 5

Cluster Analysis

In this chapter, spatio-temporal data from the LoL datasets are analyzed using clustering algorithms in different situations. Section 5.1 presents an overview of the types of analyses conducted. Section 5.2 provides a detailed explanation of the necessary data preprocessing steps required to carry out these analyses. Sections 5.3 and 5.4 describe the process of selecting the appropriate distance metric and clustering algorithm, respectively. In Sections 5.5, 5.6, and 5.7, interpretations of the clustering results obtained for each analysis are presented.

5.1 Type of analyses

Two distinct sets of analyses were performed, focusing on spatio-temporal data clustering:

- Clustering the locations and stage (early, mid, or late game) of players' deaths. This study encompasses both amateur and professional players, considering their wins and losses within matches in both teams (red and blue). The primary objective here is to identify patterns in the spatial distribution of player deaths throughout various stages of the game, with special attention to different ranking divisions.
- Clustering the locations of junglers at each minute of gameplay. This investigation seeks to identify the trends and favorite areas on the game map during the different stages of the match. By closely examining the spatio-temporal behavior of junglers, the aim is to discover valuable information about map control, strategic decisions, and the use of key map resources in gameplay scenarios.

5.2 Data preprocessing

The datasets used were those detailed in the "League of Legends Datasets" section. These datasets were divided into several subsets, each designed to address distinct research objectives. Within the analysis, primary importance was accorded to three main features: the `coordinate_x`, `coordinate_y`, and `timestamp`. These attributes were chosen due to their ability to effectively represent the dynamic interplay between player movements (spatial data) and the progression of time (temporal

data). The x and y coordinates encapsulated the spatial aspects, enabling the tracking of player positions and map interactions, while the timestamp provided the temporal dimension, making it possible to identify when specific events occurred during the match.

Recognizing that these features had different scales, the *StandardScaler()* method was applied to normalize the data. This transformation aimed to centralize the data distribution around a mean of zero and a standard deviation close to one, ensuring that all features contributed equally to the subsequent analyses.

To improve the quality of the initial results, with a specific focus on elevating the *silhouette score*, a discretization process was introduced for the timestamp feature. This temporal discretization was implemented in three distinct phases, each corresponding to a specific stage of League of Legends gameplay:

- The early game, also known as the laning phase, corresponds to the first 15 minutes of the match, it is the stage of the game where each player goes to the lane related to his role and stays there to fight the enemy "laner";
- The mid game, approximately between minutes 15 and 25 of the match, is the stage of the game where the first structures are destroyed and players usually swap lanes, making the map more open;
- The late game, from minute 25 until the end of the match, is the phase in which there is no more laning phase but is more focused on the team fighting for objectives like baron, drakes, or base structures.

The incorporation of these temporal stages not only contributed to an overall enhancement of the results but also yielded a notable increase in *silhouette scores* across various scenarios. In some instances, a remarkable improvement was observed, with the *silhouette score* rising from an initial 0.30 to a significantly enhanced 0.43.

5.3 Distance metric

To define the similarity that could exist between the player's locations, the Euclidean distance was used as the distance metric since it is one of the most used distance metrics, is one of the simplest to understand, and because it is, by default, the distance function of the selected algorithm.

5.4 Clustering algorithm

This section presents the reasons for the selection of the clustering algorithm used.

Based on the literature, the k-Means clustering algorithm is more straightforward than the others, has good results, and is one of the most used algorithms in game data science [16]. Despite this, several algorithms, such as K-Means, DBSCAN, K-Medoids, and Spectral were evaluated.

For each algorithm, the *silhouette score* was calculated for different numbers of clusters to see how well the data points were clustered.

After the evaluation, it became evident that the K-Means algorithm consistently presented the best *silhouette score* results, as illustrated in Figure 5.1. In the picture, K-Means is the leftmost, followed by Spectral clustering, and finally, K-Medoids. Consequently, the following clusters' procedures and interpretations were obtained using the K-Means algorithm.

It is also worth noting that the DBSCAN algorithm produced the worst *silhouette score* among the evaluated algorithms, with most of them being negative. This poor performance can be attributed to DBSCAN's tendency to create a cluster that includes potential noise points, thereby reducing the overall *silhouette score*. For a visual representation of the best *silhouette scores* with the DBSCAN algorithm, see Figure 5.2.

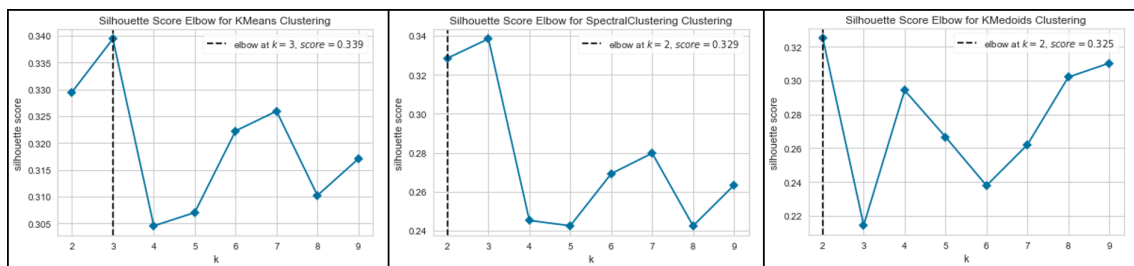


Figure 5.1: Algorithms' best *silhouette scores* (Left: K-Means; Middle: K-Medoids; Right: Spectral)

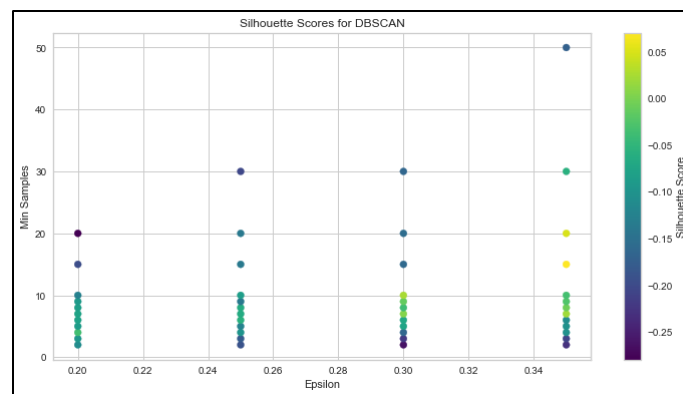


Figure 5.2: DBSCAN best *silhouette scores*

5.5 Death patterns in the Jungle Role

In this section, the localization and timing of junglers' deaths are analyzed to find possible patterns in players' deaths.

5.5.1 Jungle role

Jungler's analysis is very important for managers and players. They are not only all over the map trying to help their teammates get leads (advantages over the enemy laner), but they also spend most of the time hidden from the enemy team's vision in the early stages of the game and are typically, like the supports, the teams' main shot-callers. These are the reasons that make the jungle role the main research focus.

This analysis is about finding patterns in the players' deaths (professional and amateur) who have the jungle as their main role. Knowing the places and moments of the match where players usually die more often can help players and coaches identify in which zones and times of the match players are more at risk, making them change strategies or alerting them to be more cautious in these moments of the match. Thus, it was selected from the datasets of professional and amateur players all the events of the type "Death" that occurred during the games in the jungle role. Then, the dataframes derived from this selection were divided into four, one with all the deaths of a jungler in the red team's victory, one with all the deaths in the red team's defeat, another with all the deaths in the blue team's victory, and finally one with all the deaths in the blue team's defeat.

The determination of the optimal number of clusters (K) for analysis involved a multi-faceted approach. Firstly, the *silhouette score* results were considered, evaluating a range of cluster numbers from 2 to 20. Given the complexity of the game under study, and the personal experience as a player, the incorporation of personal semantic insights enriched the decision-making process. This includes a deep understanding of player-specific terminology and an awareness of the information they require during gameplay.

Among the tested cluster numbers, it was observed that three clusters consistently yielded the highest *silhouette scores*. When this wasn't the case, additional factors were taken into account. Specifically, an examination of the spatial distribution of clusters on the game map was undertaken to assess if these divisions correspond to meaningful zones within the game environment. In instances where the *silhouette score* did not favor three clusters but the geographical distribution offered interpretive potential, the choice of three clusters was maintained.

To facilitate the interpretation of the resulting clusters, various visualization techniques were used:

- Scatter plots with deaths' coordinates: These plots incorporated the summoner's rift map as a background, providing a spatial context for the cluster locations;
- Histograms: Histograms were used to count the deaths' number in different game stages (early, mid, and late game) for each cluster;
- Silhouette values' plots: These plots offered insights into the quality of the clustering by displaying silhouette values for each cluster;
- Space-time cube (3D visualization of time and space): This visual representation allowed exploration of the temporal and spatial dimensions of the data in a single display;

- **Dataframe statistics:** A structured summary of each cluster’s characteristics, including standard deviation, maximum, minimum, and average values, was provided for a comprehensive understanding of cluster attributes.

To enhance clarity and comprehensibility, each cluster’s geographical location was color-coded in the visualizations: blue denoting clusters situated at the blue team’s base, red for clusters at the red team’s base, yellow/gold for clusters positioned on the top side of the map, and green for clusters located on the bottom side of the map.

These comprehensive visualization techniques were used to facilitate an in-depth interpretation of the clustered data and enable the derivation of meaningful insights within the context of this thesis.

5.5.2 Professional junglers’ results interpretation

Clustering professional players’ deaths involved a comprehensive analysis of several game scenarios, including red team victories, red team defeats, blue team victories, and blue team defeats. This study resulted in the identification of three distinct clusters in all of these situations. The selection of these clusters was based on superior *silhouette scores* in the majority of scenarios, with an overall *silhouette score* of 0.34, as shown in Figure 5.3, and the insights obtained from the spatial and temporal patterns of the clusters.

The importance of these three clusters lies in the rich insights they offer into the dynamics of professional player deaths during gameplay. By categorizing these events into distinct clusters, valuable information was discovered that may previously have gone unnoticed.

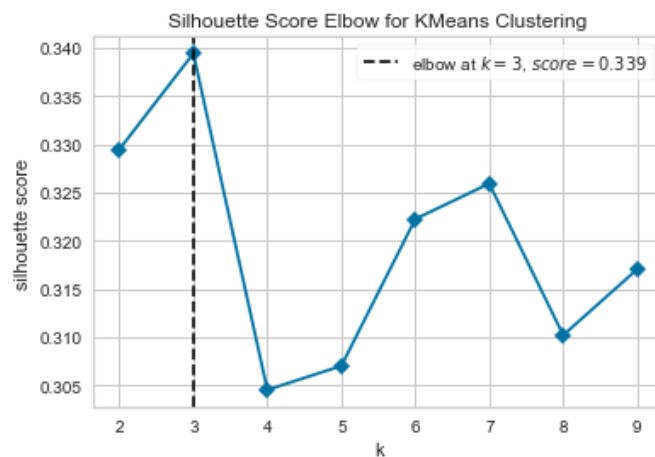


Figure 5.3: Best k value according to *silhouette score*

Red team’s victory

In the context of analyzing red team victories, as illustrated in Figure 5.4, a distinct pattern emerges among professional junglers’ deaths. Specifically, it is observed that during the early stages of the game, professional junglers tend to experience a higher frequency of deaths on the bottom side

of the map (green color). However, as the game progresses into the mid stages, the focus shifts and deaths tend to occur more frequently on the top side (yellow color). During the final stages of the game, characterized by the red team's victorious phase, a substantial part of deaths takes place within the enemy team's base (belonging to the blue team).

When examining the distribution of deaths among professional junglers in red team victories, the top side of the map records the highest count of deaths, totaling 452. This is followed by the bottom side, with 426 deaths, and the blue team's base, where 240 deaths are recorded.

To provide further insights into the temporal aspects of these deaths, the Kernel Density Estimate (KDE), as presented in Figure 5.4, reveals specific time intervals associated with each cluster. Deaths on the top side of the map are predominantly concentrated between minute 0 and minute 30 of the matches, with a notable peak occurring around minute 15. On the other hand, deaths on the bottom side are most frequent between 0 and 27 minutes into the game, with a significant peak at around minute 5. Finally, deaths within the blue team's base are primarily clustered between minute 12 and minute 40, with a notable peak observed at minute 25.

A three-dimensional representation of time and space was also tested to provide an alternative perspective to the visualization techniques already in use. An example of this representation can be found in the video created [36].

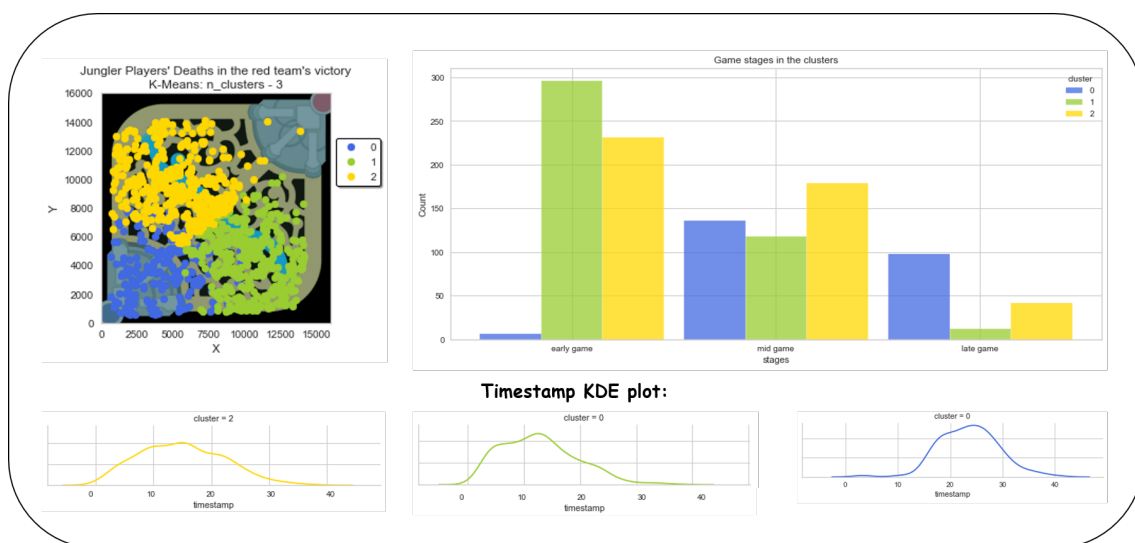


Figure 5.4: Clustering professional jungler players' deaths in red team's victory

Red team's defeat

In the red team's defeat, as shown in Figure 5.5, a trend emerges in the distribution of deaths among professional junglers. During the early stages of the game, these players tend to experience a higher frequency of deaths on the bottom side of the map. As the game progresses into the mid stages, there is a noticeable change in the distribution, with deaths becoming more prevalent on the top side. It's worth noting that, unlike red team victories, there is an increase in deaths occurring within the red team's own base during the mid-game phase. In the final stages of the

game, characterized by a defeat for the red team, a substantial majority of deaths take place within their team's base.

When examining the distribution of deaths among professional junglers in red team defeats, the bottom side of the map records the highest count of deaths, totaling 462. This is followed by the top side, with 449 deaths, and the red team's base, where 330 deaths are recorded.

The KDE plot, as presented in Figure 5.5 reveals that deaths on the top side of the map (yellow color) are predominantly concentrated between minute 0 and minute 35 of the matches, with a notable peak occurring around minute 22. Conversely, deaths on the bottom side (green color) are most frequent between 0 and 30 minutes into the game, with a significant peak at around minute 14. Finally, deaths within the red team's base are primarily clustered between minute 13 and minute 40, with a notable peak observed at minute 25.

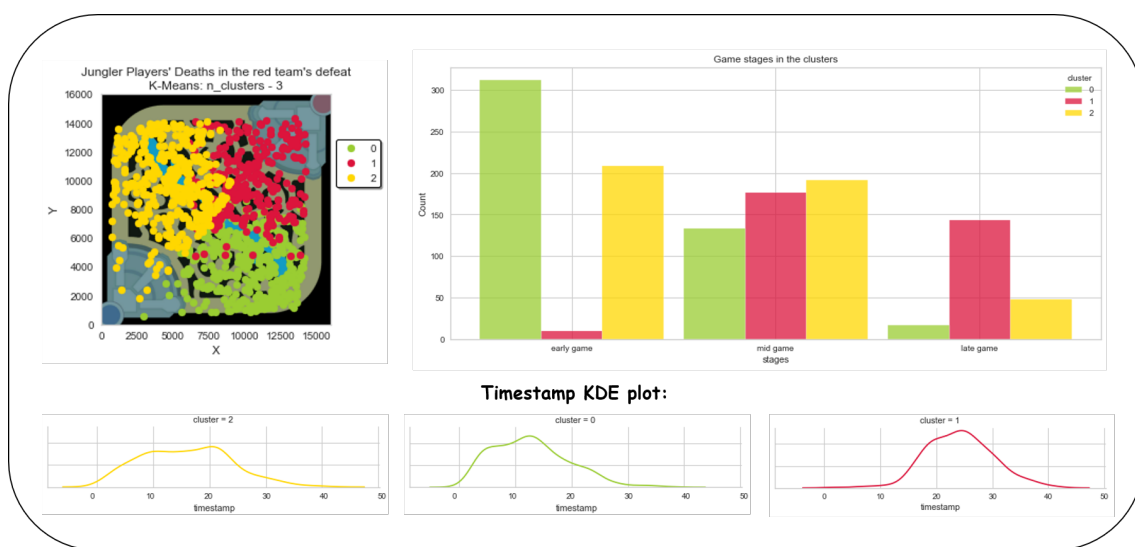


Figure 5.5: Clustering professional jungler players' deaths in red team's defeat

Blue team

Similar patterns are consistently observed within the context of the blue team's gameplay. Professional junglers on the blue team tend to show identical behaviors, with a tendency for more deaths on the bottom side during the early game, transitioning to an increased frequency on the top side in the mid-game phase, and a notable concentration of deaths within each team's respective base during the late stages of the game (winning - red team's base, losing - blue team's base).

It is noteworthy that a divergence in this pattern occurs when professional junglers are on the blue team and face defeat. In this scenario, as opposed to an overall tendency for more deaths on the bottom side, there is a shift, resulting in a higher total of deaths on the top side.

In summary, across various game outcomes (victory, and defeat), professional players consistently experience the majority of deaths on the bottom side of the map.

For a more detailed visual representation of the clustering results for the blue team, consult the materials provided in Appendix C

5.5.3 Amateur junglers' results interpretation

The analysis conducted in this project extends beyond the clustering of spatio-temporal data related only to professional players, also including an exploration of spatio-temporal data related to amateur players as well.

This particular subsection delves into the examination of the other League of Legends dataset described in the "League of Legends Datasets" section, which exclusively comprises data from amateur players representing various amateur divisions, commonly referred to as low elo divisions. Figure 5.6 provides an overview of the amateur divisions.

Remarkably, for this subset of data, in line with the findings from the professional player dataset, the application of the k-means algorithm continued to yield the most compelling results when employing 3 clusters. This selection was based on both the *silhouette score* and the valuable insights gained from the spatial representations generated by the clusters. The overall *silhouette score* for this analysis was 0.33.

The decision to maintain 3 clusters across all scenarios was driven by the desire to ensure uniformity and comparability with the professional player dataset. This consistency facilitates a more coherent interpretation of results, ultimately reinforcing the research's robustness and clarity.



Figure 5.6: Low elo divisions

Early game

At the beginning of the match, players tended to die more on the bottom side of the map, across all divisions, and in most game scenarios. However, it is noteworthy that a few exceptions to this pattern exist:

- In the Iron division, when the red team secures victory, amateur junglers tend to exhibit a distinct behavior, with a tendency for more deaths on the top side of the map;
- Similarly, in the Bronze division, when the blue team emerges victorious, amateur junglers also deviate from the typical pattern, having more deaths on the top side of the map;
- In cases of a blue team's defeat across the Iron, and Silver divisions, amateur junglers exhibit a tendency to experience more deaths on the top side.

Some of these exceptions can be seen in Figure 5.7.

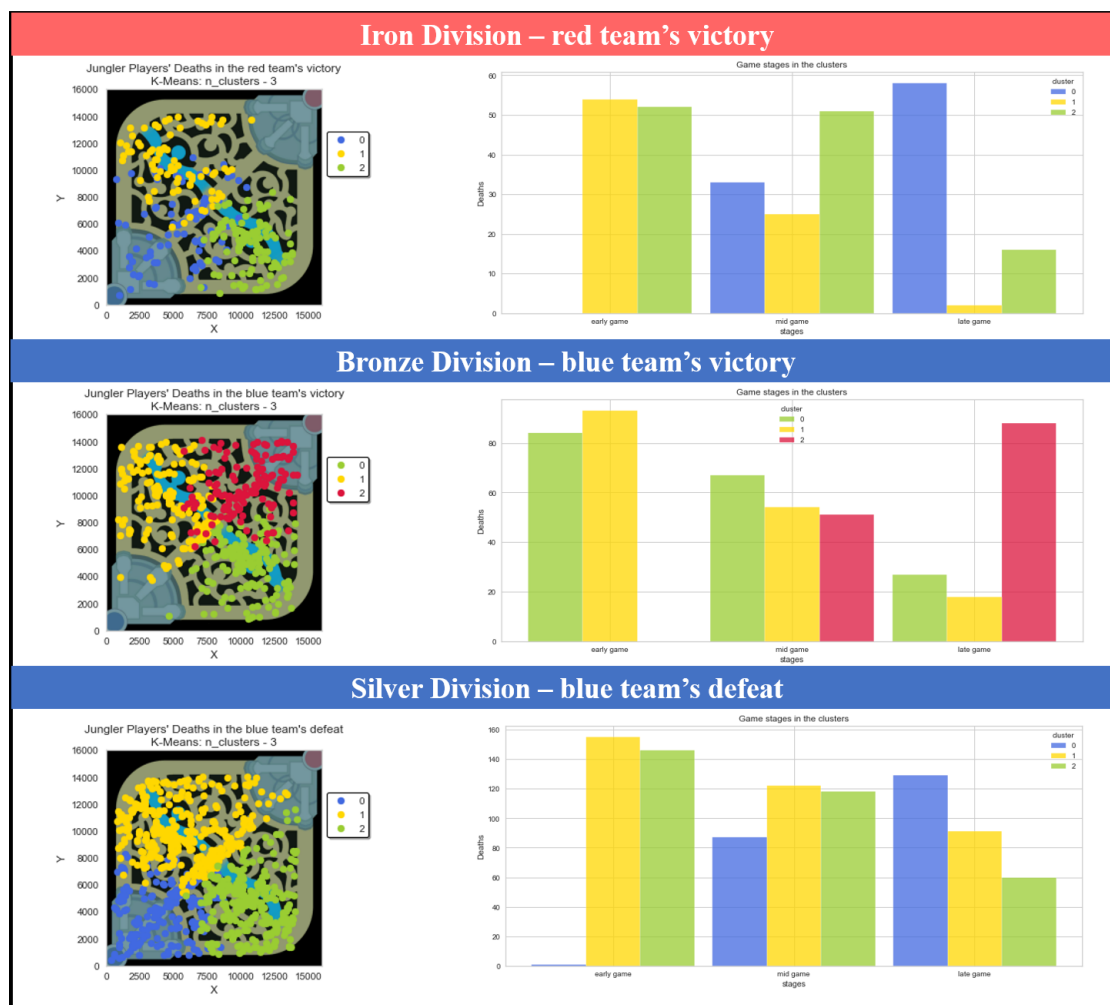


Figure 5.7: Example of clustering amateur jungler players' deaths

Mid game

During the mid stage, significant changes in the distribution of player deaths across all divisions become evident. At this stage, deaths occur predominantly at the top and bottom sides of the map, with a slight preference for the top side.

The primary locations for the most frequent deaths during this mid-game stage, across different divisions and game outcomes (victory and defeat for both teams), can be summarized as follows:

- In the blue team's victory, only silver players die more on the top side. The players from the other divisions usually died more on the bottom side;
- In the blue team's defeat, only iron players died more on the bottom side. The players from the other divisions usually die more on the top side;
- In the red team's victory, iron and gold players die more on the bottom side, while the players from the other divisions (bronze, silver, and platinum) usually die more on the top side.

- Finally, in the red team's defeat, the distribution of deaths diverges among several divisions. Bronze, silver, and gold players are more likely to die on the top side, whereas players from the iron division tend to die more on the bottom side. Notably, platinum players exhibit a unique pattern, with a greater number of deaths occurring on the red team's base.

Some of these clusters can be seen in Figure 5.8.

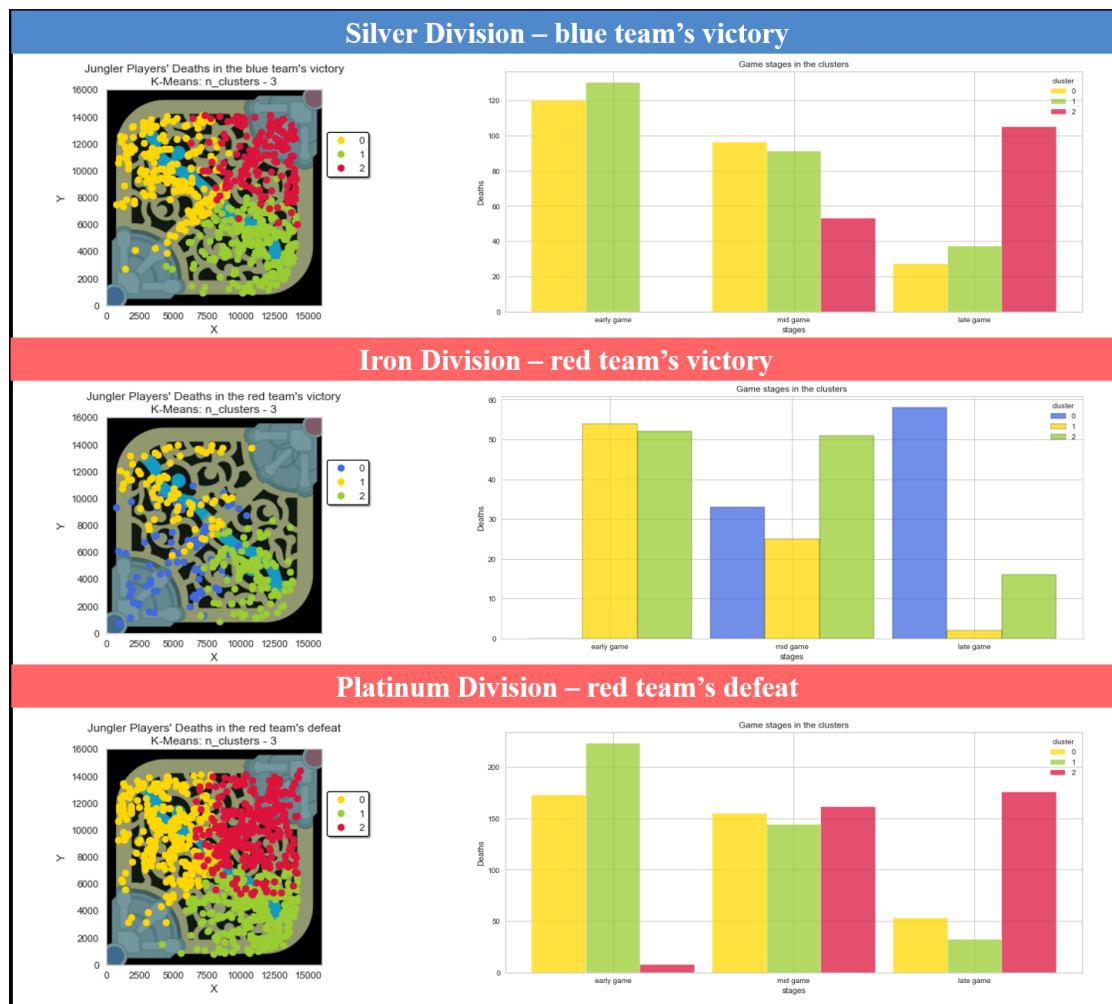


Figure 5.8: Example of clustering amateur jungler players' deaths

Late game

In the late game, amateur junglers' deaths tend to concentrate primarily at each respective team's base. Specifically, this occurs at the red team's base when the blue team is winning, or if the amateur junglers are facing defeat while on the red team. Conversely, deaths shift to the blue team's base when the red team is winning, or if the amateur junglers are on the blue team and facing defeat.

Interestingly, the bases on the map typically record the fewest instances of player deaths, with three exceptions. Firstly, in the Iron division, both during red team victories and defeats, there are

fewer deaths on the top side. Secondly, in the Gold division, during blue team victories, there are again fewer deaths on the top side.

Overall, it becomes evident that the majority of deaths in a game occur on the bottom side when amateur players are winning the game, while during moments of disadvantage, deaths tend to shift to the top side. This observation underlines the strategic significance of these locations during the late-game phase, highlighting how gameplay dynamics evolve as the match progresses.

Additional insights regarding the cluster and its representation are available in the notebooks accessible through the GitHub repository created for this research [37].

5.5.4 Professional Junglers vs Amateur Junglers

Comparing the results obtained from the analysis of professional and amateur players allowed to make several pertinent observations. One notable trend is the correlation between a player's division and the average number of deaths per game. As players progress through the divisions, a decline is observed in the number of deaths per game. In the Iron, Bronze, and Silver divisions, players exhibit an average of around seven deaths per game, which decreases to six deaths per game in the Gold and Platinum divisions. Professional players, on the other hand, have an even lower average of just five deaths per match. Additionally, all players tend to die more when playing for the red team, and when they are losing.

Location-wise, the analysis reveals intriguing disparities in death occurrences between professional and amateur players. Professional junglers exhibit a consistent tendency for the bottom side of the map, particularly during the early stages of the game. Conversely, amateur players in the Iron, Bronze, Silver, and Gold divisions tend to concentrate their deaths more on the top side during this phase.

During the mid-game stage, the divergence in death locations becomes more pronounced across divisions. While professional junglers consistently gravitate towards the top side of the map, the death locations of amateur junglers are influenced by both their division and the outcome of the match, oscillating more frequently between the top and bottom sides.

In the late game, a uniform pattern emerges among both amateur and professional junglers: deaths predominantly occur within the losing team's base.

Another noteworthy finding is the difference in game duration between professional and amateur players. Professional players' games tend to be shorter, suggesting that the higher level of play makes it increasingly challenging to reverse the course of a match, consequently reducing its duration. This trend underscores the importance of gaining advantages and maintaining control throughout the game.

5.5.5 Other experiments in the jungle role

In this subsection, the focus will shift to designing additional experiments conducted to enrich the analysis of the jungler's behavior, based on the knowledge previously outlined. These experiments

not only serve to improve the understanding of the jungler's role but also help in other analyses derived from the clusters obtained in the initial analysis.

Clustering only deaths' coordinates and clustering only deaths' timing

To gain a deeper understanding of temporal effects on specific data points, a series of separate clustering analyses were conducted based on just two key variables: the coordinates of player deaths and the timing of those deaths.

Figure 5.9 shows the outcomes of clustering using only the spatial coordinates of players' deaths. This approach divided the game map into distinct regions:

- the top side in yellow;
- The bottom side, represented in green;
- The red team's base and its surroundings are colored red;
- The blue team's base and its surroundings are marked in blue.

This division suggests that player deaths tend to concentrate within specific zones based on their spatial coordinates during the game.

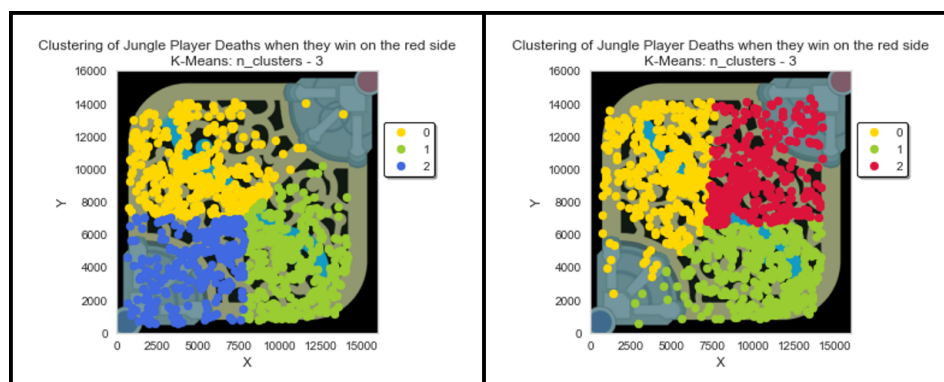


Figure 5.9: Example of clusters obtained using only deaths' coordinates

In contrast, Figure 5.10 explores the impact of death timing on clustering. Here, it is observed that the moment players die may not have such a significant influence on the division of clusters. Despite some variations, deaths occur in all three phases of the game (early, mid, and late game) across a wide variety of locations. However, a recognizable trend emerges over time, indicating a funneling effect.

Specific observations include:

- During the early game, jungler deaths tend to cluster around each lane (top, mid, and bottom lanes), with some occurring in the jungle itself.
- In the mid game, deaths are distributed across a more extensive range of locations but are more dispersed.

- In the late game, deaths become even more tapered, particularly within the losing team's base.

Histograms further underscore that the majority of deaths occur during the laning phase, corresponding to the early game.

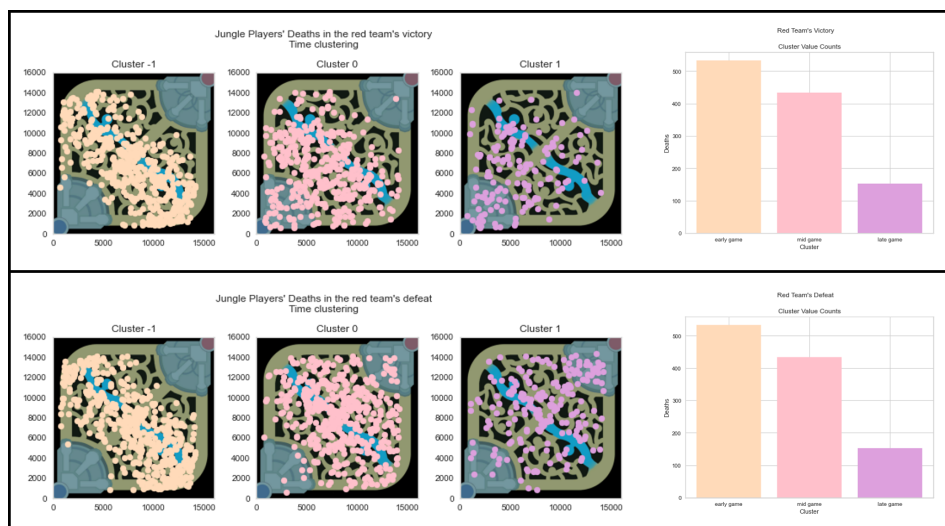


Figure 5.10: Example of clusters obtained using only deaths' timing

Champion's deaths in each cluster

Champions are an essential part of the game, serving as the avatars with which the players play. Each champion possesses a unique array of abilities and gameplay styles, which makes them essential for making strategic decisions. The game is in a constant state of evolution, with some champions increasing in power while others decrease. These ongoing adjustments are orchestrated by Riot Games to rectify win rate imbalances from the previous patch. Consequently, maintaining an in-depth knowledge of the prevailing champions in the current patch, as well as their usual positions and vulnerabilities, is indispensable for players striving to perfect their strategies.

In the pursuit of a deeper understanding of this ever-evolving gaming landscape, an analysis of spatio-temporal data related to champions' deaths was conducted. This analysis unearthed crucial insights, including the popularity of specific champions among players, their corresponding win rates, and the frequency of deaths associated with each jungle champion within various clusters. To facilitate the interpretation of these findings, a range of visualization techniques was employed.

The first step involved identifying the most frequently selected champions in the jungle and their corresponding win rates during patch 13.3. Figure 5.11 presents an overview of the six most favored champions among professional players and their respective win rates. The champion Lee Sin emerges as the most played champion in the jungle, featuring in 106 games and an impressive 56% win rate. It is noteworthy that all six of these frequently chosen champions had positive win rates, which reinforces their effectiveness in high elo.

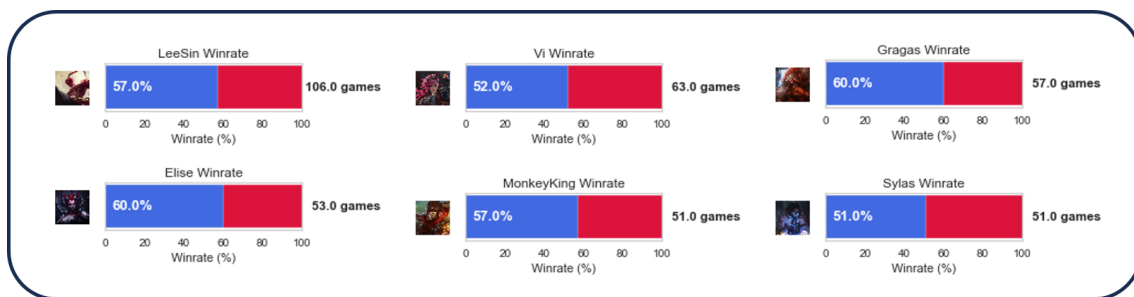


Figure 5.11: Top 6 most played jungle champions by professional players and their win rate

Figure 5.12 changes the focus to the preferences of amateur players, revealing a distinctive set of champion choices compared to those observed in Figure 5.11. Interestingly, while Sylas remains a common selection, Lee Sin and other champions preferred by professional players do not maintain the same level of popularity among amateurs. Furthermore, it becomes clear that the win rates of these most-played champions among amateur players are less favorable, with two of the six champions (Nunu, and Sylas) experiencing negative win rates. This contrast highlights the divergence in dynamics between professional and amateur gameplay.

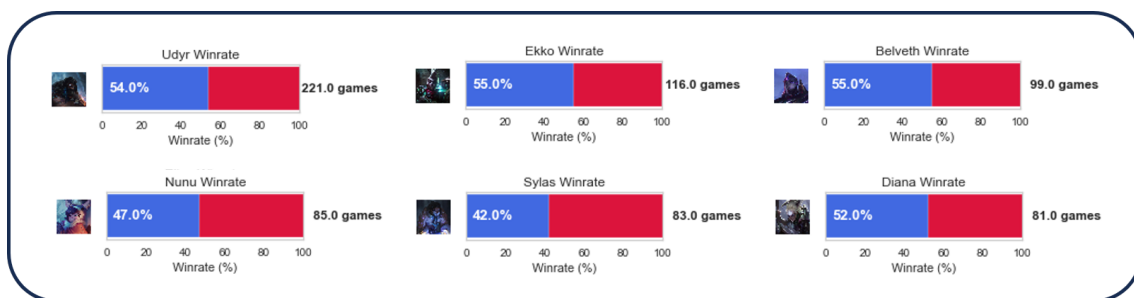


Figure 5.12: Top 6 most played jungle champions by amateur players and their win rate

Once the champions' win rates are calculated, it is important to investigate the occurrence of champion deaths within each cluster. Figure 5.13 shows the distribution of champion deaths within each cluster for both the red team's victories (A) and defeats (B). Using radar plots, the variations in champion deaths among professional junglers are revealed, offering a unique perspective on the game's complexities.

In the case of red team victories (A), Lee Sin stands out as the champion with the highest number of deaths, predominantly concentrated on the bottom side with over 50 deaths. This pattern suggests that these deaths likely occurred in the early stages of the game, aligning with the majority of deaths on the bottom side. When other frequently played champions are in use, deaths tend to occur more on the top side, indicating a greater frequency during the mid game, when most deaths in this location occur.

Conversely, in the red team's defeats (B), there is a noticeable increase in deaths near the red team's base, indicating a prevalence of late-game deaths in close proximity to their base. Unlike the victorious scenarios, Lee Sin registers more deaths on the top side during these defeats,

possibly occurring in the mid game. Meanwhile, Sylas, Elise, and Vi register higher death rates on the bottom side, suggesting a propensity for early game deaths. These findings can potentially convey various implications; for instance, champions like Sylas, Elise, and Vi may struggle to recover if they fall behind in the early game, or excessive deaths on the bottom side may lead to eventual defeat on the red side. On the other hand, if Lee Sin accumulates more deaths in the mid and late game, particularly on the top side and near the red team's base, it appears to significantly impact the champion's overall performance.

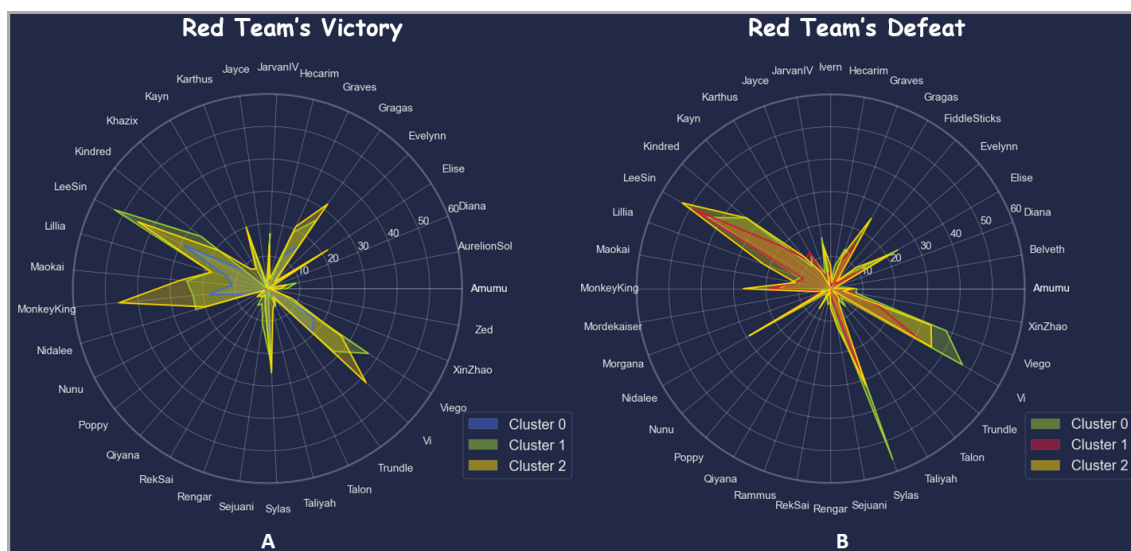


Figure 5.13: Radar plots with champions deaths for the Red team: Victory (left); and Defeat (Right)

In the analysis of the blue team's results, some similar patterns emerge to those observed in the red team's scenarios. Once again, Lee Sin stands out, displaying a tendency to accumulate more deaths in both victorious and defeated matches. When the blue team secures victory, Lee Sin's deaths tend to be concentrated on the top side during the mid game phase, suggesting that even in the context of victory, this champion may encounter vulnerabilities in the mid game period. Conversely, in blue team defeats, Lee Sin's death pattern mirrors this mid game trend, with an increased number of deaths on the top side. This recurring trend could imply that Lee Sin's mid game performance significantly influences the overall match outcome.

Changing attention to Sylas and Elise, their death patterns maintain consistency across blue team victories and defeats. Elise tends to face early game deaths on the bottom side, indicating potential vulnerabilities during the game's initial stages. Sylas, on the other hand, experiences a surge in deaths on the top side during the mid game in both scenarios. This suggests that Sylas players might need to exercise caution and adapt their strategies during this crucial phase of the game.

Interestingly, Vi demonstrates different patterns of death distribution based on the outcome of the match, both when playing on the blue and red teams. In victories for both teams, Vi's death distribution remains relatively consistent, indicating a stable and adaptable playstyle, which

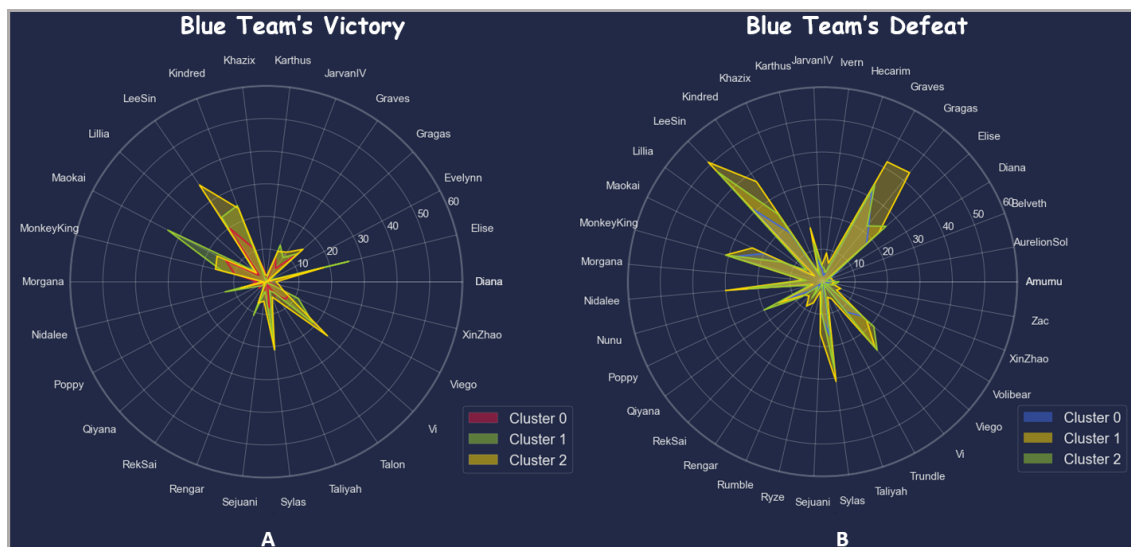


Figure 5.14: Radar plots with champions deaths for the Blue team: Victory (left); and Defeat (Right)

contributes to her effectiveness in securing wins. However, in losses for both teams, Vi's death pattern takes on a different form, suggesting that her performance may face particular challenges or vulnerabilities when she dies in the early stages of the game or on the bottom side.

One noteworthy observation in the analysis of champions played by amateur players is the distinct pattern associated with Nunu's performance. When playing with Nunu and securing a victory, amateur players tend to experience more deaths on the top side during the mid game phase. This phenomenon suggests that despite emerging victorious, Nunu players may face vulnerability in the mid game, possibly due to specific strategies or playstyle adaptations.

On the other hand, when amateur players are on the losing side of a match with Nunu, a different trend emerges. In these cases, Nunu players tend to accumulate more deaths on the bottom side during the early stages of the game. This pattern implies that Nunu players may struggle to establish control and stability in the early game, potentially leading to unfavorable outcomes for their team.

In summary, this knowledge contributes to a deeper understanding of how specific champions perform under diverse circumstances, enriching the continuous search for strategies in the competitive game.

Comparing junglers in a newer patch (13.9)

One key aspect of this study involves comparing the performance of junglers in the League of Legends between two different patches: patch 13.3 and a more recent patch 13.9. This comparative analysis allows for insights into whether the patterns of jungler deaths, in terms of location and time, remain consistent or if they adapt to the dynamic changes introduced in each patch. These changes include champion balance adjustments, where some champions are improved while others may be less effective. Such alterations can significantly impact jungler playstyles since the choice

of a champion is fundamental to their role.

Initially, a focus was placed on clustering spatio-temporal data related to jungler deaths in patch 13.9. using datasets specific to this patch, detailed in the section "League of Legends Datasets", with the same features as those from patch 13.3, the analysis was limited to professional and amateur junglers. Subsequently, the results obtained in patch 13.9 were examined.

Figure 5.15 and Figure 5.16 reveal that the location and timing of professional jungler deaths on the red team are very similar to those observed in patch 13.3. The main difference emerges in Figure 5.15, where in the early game of the red team's victories, professional junglers tend to die slightly more on the top side (yellow cluster) than the bottom side (green cluster), whereas in patch 13.3, junglers predominantly met their deaths on the bottom side. For the blue team, deaths' locations and timings remain consistent regardless of victory or defeat, as shown in Figures C.3 and C.4 in the Appendix C

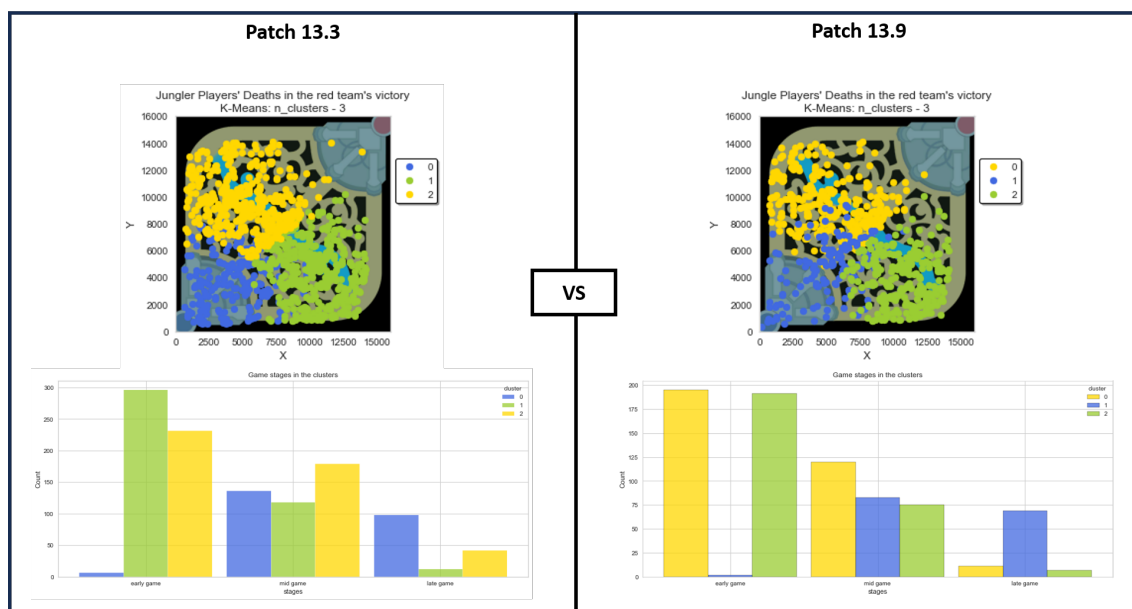


Figure 5.15: Comparison between different patches in red team's victory on professional players

In amateur players' deaths, there were more pronounced differences across patches, particularly during the early and mid game phases.

In the Iron division, during the initial phase, junglers were more prone to die on the top side in all scenarios except when winning on the red team, where deaths were more likely to occur on the bottom side. This trend persisted into the mid game, with deaths occurring more frequently on the top side in all situations, except when winning on the blue side, where deaths predominated on the bottom side, and in the red team's base.

In the Bronze division, the only change was observed in the mid game when amateur players won. In this case, deaths occurred more frequently on the bottom side when playing for the red team and more frequently on the top side when playing for the blue team.

In the Silver division, there were two changes in the early game, one when winning for the

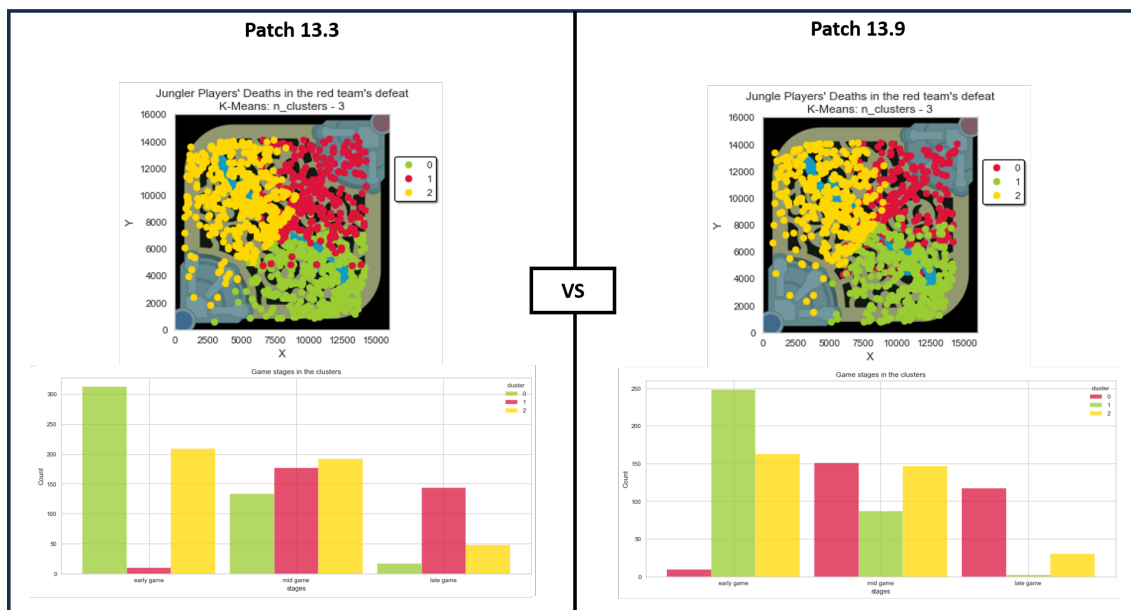


Figure 5.16: Comparison between different patches in red team's defeat on professional players

red team, with more deaths occurring on the top side, and another when losing for the blue team, with more deaths occurring on the bottom side. Similarly, there was a change in the mid game when playing for the blue team, with more deaths on the bottom side in both victory and defeat scenarios.

The Gold division showed a change in the early game consistent with the Silver division. In the mid game, there was a change when playing for the red team, with more deaths on the top side in team victories and more on the bottom side in team defeats.

In the Platinum division, most of the changes occurred during the mid game. Amateur junglers experienced more deaths on the bottom side in the red team's defeats and on the top side in the blue team's victories. Another change was noted in the early game of the blue team's defeats, with more deaths occurring on the top side. Detailed visual representations of these comparisons on amateur players can be found in the notebooks accessible through the GitHub repository created for this research [37].

These variations in amateur jungler behavior can be attributed to several factors, including a potentially less accurate understanding of the current meta (knowing which champions are good and bad in this patch) and champion strategies compared to professional players. Additionally, the smaller dataset for patch 13.9 may have played a role, potentially influenced by the conclusion of the competitive split, players reaching their desired division, or taking breaks from the game during this particular patch.

Regarding win rates, there was an increase observed on the blue side for both amateur and professional players. This increase can be largely attributed to power reductions in the bot and support roles during patch 13.9. However, it's worth noting that these roles remained influential, explaining why the win rate increase on the blue side for professional players was relatively mod-

est, as they were able to take better advantage of these small details to capitalize on these game dynamics.

5.6 Death patterns in other roles

In this section, the analysis of death patterns in roles other than the jungler within the context of patch 13.3 is explored. The preceding section highlighted the distinctive challenges posed by the jungler role, while attention is now directed toward the other roles of the game (top, mid, bot, and support laners). This section delves into how different roles exhibit diverse patterns of deaths, revealing hidden insights into gameplay in League of Legends. Through a rigorous examination of cluster analysis and *silhouette scores*, valuable information about how players in various roles navigate the game's dynamics is uncovered, ultimately contributing to a more comprehensive understanding of their gameplay strategies and positioning on the map.

In contrast to the findings within the jungler role, the analysis of death patterns in other roles in patch 13.3 revealed significantly improved *silhouette scores*, with the majority of clusters achieving a score of 0.45. This marked improvement can be attributed to the distinct characteristics associated with these roles, as previously discussed. Unlike junglers, who roam throughout the map, these roles tend to exhibit more defined and predictable positioning, which contributes to a more cohesive cluster structure.



Figure 5.17: Example of the best K values according to *silhouette score* in the red team's victory

For these other roles, a deliberate decision was made to adopt a four-cluster configuration. This selection was motivated not only by the presence of high silhouette values, as exemplified in Figure 5.17, but also by a comprehensive comprehension of the game's dynamics. The choice of four clusters has been made to facilitate a more detailed interpretation of the outcomes acquired,

aligning effectively with the intricacies characterizing each role and the distinct death patterns that are inherently linked to them.

Top laners

The role of the top laner is integral to a team's composition, as these players typically serve as the frontline tanks and primary engage initiators. As the role's name implies, the players initially go to the top lane, where they remain for a significant portion of the early game until the mid-game phase is reached. Analysis of spatio-temporal data related to top laners consistently produced specific cluster formations, as can be seen in Figures 5.18 and 5.19. These clusters delineated four significant locations: the top side (represented in yellow), the bot side (in green), the team's base (in red or blue, depending on the team), and a central location symbolizing the center of the map (in purple).

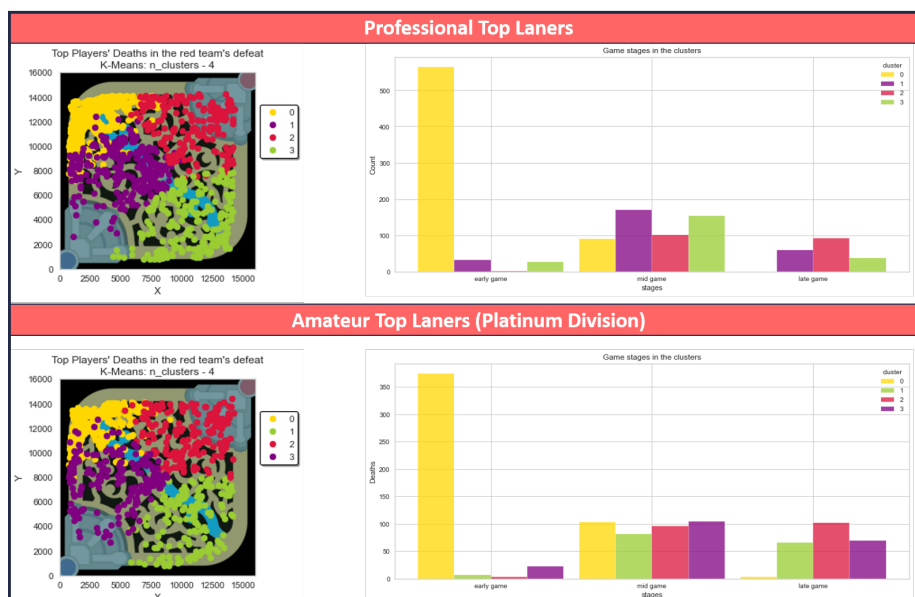


Figure 5.18: Deaths clusters of top laners in red team's defeat

During the early stages of a match, it is notable that players exhibit a higher tendency for deaths on the top side of the map.

Transitioning into the mid-game phase, distinct patterns emerged among players of varying skill levels. Amateur players tended to experience a greater frequency of deaths on the top side and, in many instances, in the mid lane. On the other hand, professional players displayed a contrasting trend, with more deaths occurring in the mid lane during defeats and an increased likelihood of deaths on the bottom side when achieving victories.

In the late game, a recurring trend was the occurrence of player deaths within the confines of the team's base. This pattern was significantly influenced by the match's outcome, depending on whether the team was in a winning or losing position. Figures 5.18 and 5.19 provide graphical representations of this phenomenon, offering insights into deaths during scenarios where the red

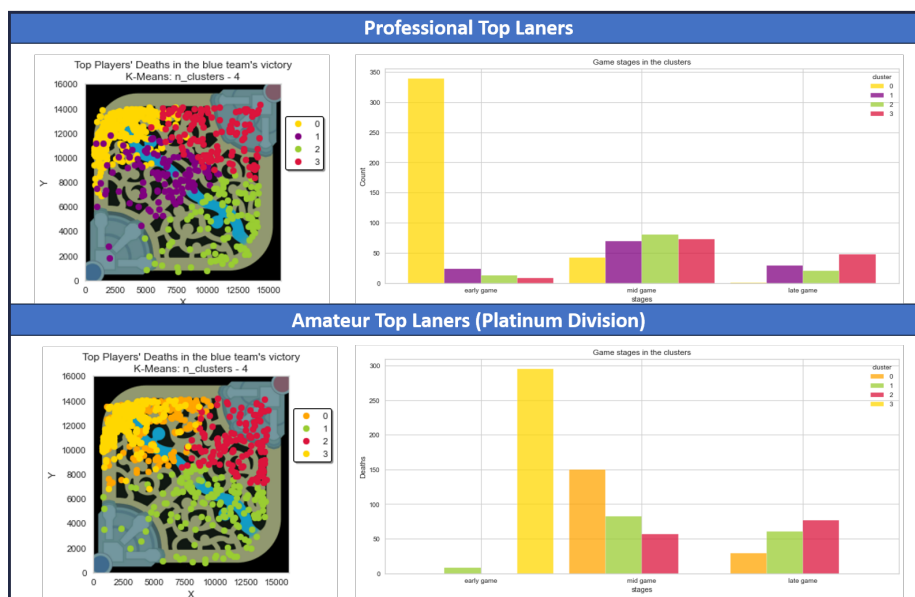


Figure 5.19: Deaths clusters of top laners in blue team's victory

team faced defeat and the blue team secured victory, across both amateur and professional player segments.

This analysis also reveals that overall, as anticipated, the majority of deaths among top laners typically occur on the top side of the map.

Mid laners

The mid laner role also holds immense significance, mainly due to its central positioning on the map. This centrality provides midlaners with a strategic advantage that allows them to exert greater control over the game, particularly when they gain an advantage over their counterpart from the opposing team. As expected, a substantial portion of deaths in matches occurred in the center of the map.

In the early stages of matches, deaths were consistently observed in the mid lane for both professional and amateur players.

In the mid game, a perceptible discrepancy became evident between amateur and professional players. Professional players showed a change in death patterns, with a higher frequency of deaths occurring on the top side of the map. On the other hand, amateur players continued to experience a higher incidence of deaths in the mid lane, despite a general increase in deaths in all other zones.

In the later stages of matches, a recurring pattern persisted, with deaths predominantly taking place within the base of the losing team, a trend applicable to both professional and amateur players. Figures 5.20 and 5.21 provide insightful examples of cluster patterns for mid laners in scenarios where the red team emerged victorious and the blue team faced defeat, respectively.

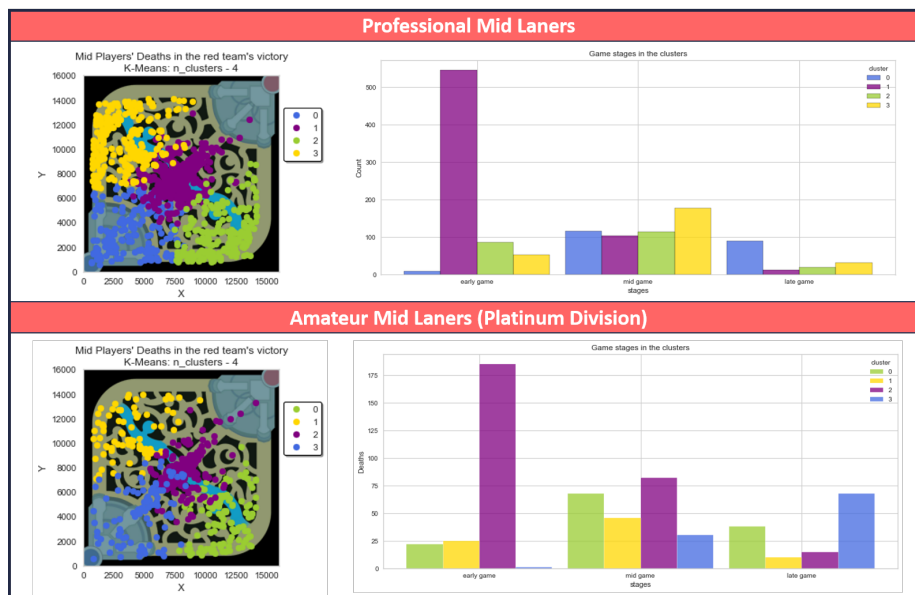


Figure 5.20: Deaths clusters of mid laners in red team's victory

Bot laners and supports

The bot laner and support roles both operate within the same lane, known as the bottom lane. While the bot laner primarily focuses on carrying the team through damage and offensive plays, the support role is oriented toward safeguarding and offering utility to teammates, with a particular emphasis on assisting the bot laner. Consequently, the clusters obtained for both roles showed remarkable similarities.

In both roles, a significant proportion of deaths occurred on the bottom side of the map. Interestingly, these roles mirrored the early game death patterns observed in the top lane, although in reverse, with the majority of deaths occurring on the bottom side.

During the mid game phase, both the bot laner and support roles showed a tendency to migrate toward the upper regions of the map. This shift was predominantly attributed to the presence of the Baron Nashor, which typically appears around the 20-minute mark. Given their crucial importance during this patch, the bot laner's proximity to the Baron often resulted in an increased number of deaths in the middle and top sides of the map.

In the late game, a recurring trend was the prevalence of deaths occurring within the team's base, a pattern consistent with other roles.

The primary differentiator between these two roles lies in their mobility and map presence. While the bot laner tends to concentrate more on a specific area of the map at each stage of the game, the support role shows greater mobility and roams across the map, helping other team members. Consequently, the support role experiences a higher incidence of deaths across various clusters, reflecting their dedication to supporting their teammates throughout the game.

Figure 5.22 provides an example of cluster patterns for professional bot laners and supports in scenarios where the red team emerged victorious. For additional details about the cluster and

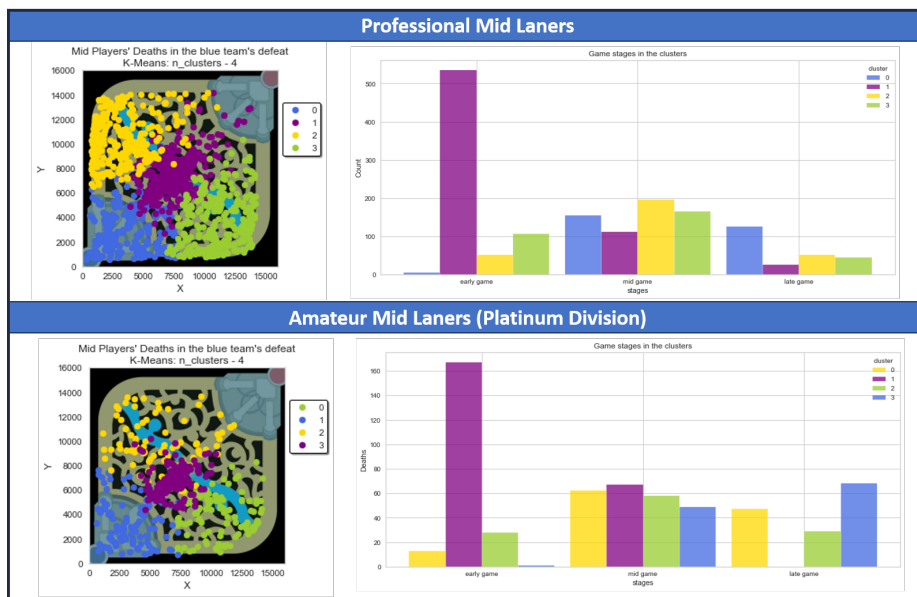


Figure 5.21: Deaths clusters of mid laners in blue team's defeat

its representation, the available notebooks in the GitHub repository provide comprehensive information and examples related to the cluster's results, functionality, and usage. Access to these notebooks is facilitated by visiting the project's GitHub repository [37].

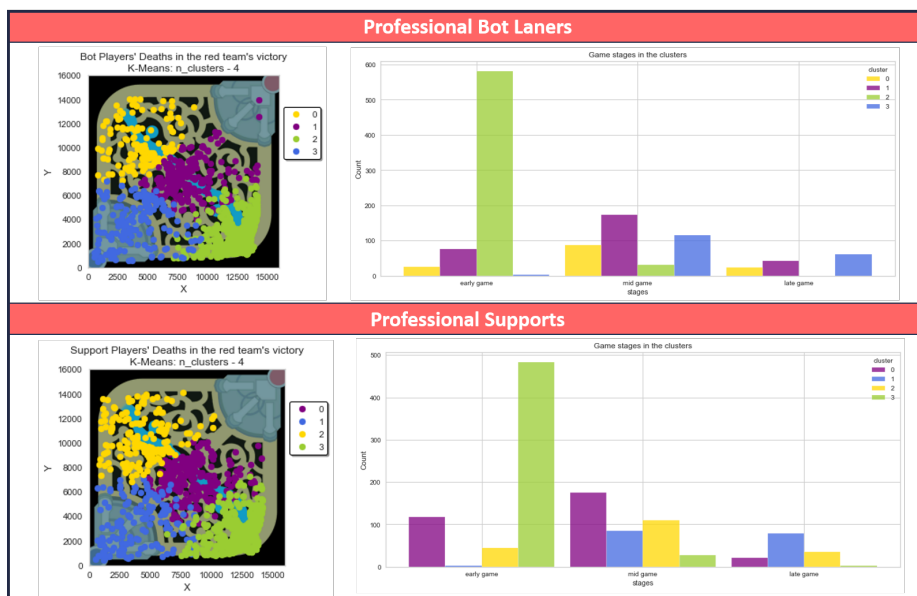


Figure 5.22: Deaths clusters of professional bot laners and supports in red team's victory

5.7 Locating professional junglers

The final analysis conducted addresses a critical concern for professional players and coaches: understanding the most likely locations of junglers during a League of Legends match. This analysis,

when combined with information about player deaths, offers valuable insights into the behavior of individual junglers. It enables professional players and coaches to see where junglers tend to position themselves throughout the game and at which stages of the match they are typically found in specific areas. Furthermore, it allows for the investigation of whether these jungler locations coincide with areas of high player deaths. If not, it serves as a warning sign for junglers to exercise caution in those specific regions during that particular stage of the game.

To initiate this investigation, the dataset covering professional matches from patch 13.3 was used. The initial step involved filtering out all non-event data points from the datasets, which essentially involved deleting rows where the "eventType" column had the value NaN.

The results from the clustering analysis, as can be seen in Figure 5.23, produced an even higher *silhouette score* compared to previous analyses, averaging 0.39. When examining matches that were won, four distinct clusters were identified, whereas in matches that ended in defeat, three clusters were observed. The variation in the number of clusters results from the fact that players inevitably cross the areas of their bases, whether they are winning (for item purchases or respawning) or losing, resulting in the presence of both the team's base cluster and the enemy team's base cluster in winning games.



Figure 5.23: Best K values according to *silhouette score*

It was noted that during the early and mid game phases, players tend to go to the bottom and top sides of the map, respectively, reflecting the clustering patterns associated with player deaths. However, when moving into the late game, the analysis reveals that players spend more time on the top side. This finding suggests that these areas could potentially represent a higher risk during the late game, as a significant amount of late-game deaths tends to occur at the teams' bases. Figures C.5, C.6, C.7, and C.8 of Appendix C confirm the interpretations made of these clusters.

5.8 Summary

In summary, this chapter delves into a comprehensive analysis of cluster patterns related to players deaths and positions in League of Legends, focusing mainly on the dynamics of patch 13.3 while also comparing it to the more recent patch 13.9. The main highlights of this chapter cover:

- **Jungler Performance:** The chapter elucidates how the performance of junglers can be intrinsically linked to the match's outcome and the ranking of players (divisions). Divergent death patterns between professional and amateur players, across various divisions, reveal the potential challenges and vulnerabilities faced by this key role.
- **Comparative Patch Analysis:** A comparison between patches 13.3 and 13.9 highlights the subtle changes in jungler behaviors as a result of champion balance adjustments and the evolution of playstyles.
- **Roles Beyond Jungling:** The chapter also investigates the death patterns of top laners, mid laners, bot laners, and supports. Each role presents distinct cluster patterns, revealing some differences between amateur and professional players within these roles.
- **Locating Professional Junglers:** An exploration of the likely positions of professional junglers throughout matches offers valuable information for players and coaches. By understanding the junglers' behaviors at different game stages, it becomes possible to identify potential risk areas on the map.

The next chapter presents the validation of the results through a professional coach.

Chapter 6

Validation of Results

This chapter presents the validation process for the cluster analysis results. Starting with a description of the expert participant involved, as well as a portrait of the procedures used (Section 9.1), the subsequent section (Section 9.2) sets the respective validation that was carried out to establish the credibility and resilience of the obtained results.

6.1 Participant and procedures

To validate the cluster analysis results, a session on Discord with a professional coach was conducted. The coach has five years of experience in European Regional Leagues (ERL), and at the time of the interview was one of the coaches of the Portuguese LoL team, *White Dragons*.

The interview was made in a session of approximately 90 minutes, during which a brief presentation outlining the key aspects of the analysis was delivered to the coach. Subsequently, a series of targeted questions were asked to obtain information and feedback regarding the obtained results. This process served the dual purpose of confirming the alignment of interpretations with real-world observations and addressing any remaining uncertainties, namely in areas such as win rates.

Throughout the session, notes were recorded, capturing the coach's responses and perceptions. Upon conclusion, the coach was further consulted regarding the utility of the analysis approach for both players and coaching staff. Additionally, information was sought on potential avenues for further analysis using spatio-temporal data, thereby enriching the overall depth of understanding within the context of esports performance analysis.

6.2 Interview results

This section will describe some of the conclusions obtained from the interview.

6.2.1 Validation of cluster analysis

According to the coach, most of the interpretations made corresponded to the reality experienced in LoL.

Regarding the division of game stages in LoL, it became evident that they are well-defined, although they are not exclusively determined by time. In particular, the mid and late game phases are more easily recognizable based on the accumulation of players' items.

In the context of identifying clusters within the game, it was observed that despite the clusters having some different forms over time and space, they were consistently categorized as bottom side, top side, team bases, and, for the analysis of other roles, the mid lane cluster was also identified.

These clusters offer valuable information about playing patterns, which can prove beneficial for players. They serve as effective indicators for player positioning, potentially influencing the enemy team's strategic decisions and prompting them to adapt or revise their strategies accordingly.

6.2.2 Surprising results

The findings resulting from the analysis produced some curious revelations, particularly the higher win rate of the red team over the blue team. However, through the use of the generated clusters, potential reasons behind this phenomenon were uncovered. Several factors emerged as contributing to this disparity.

One noteworthy factor is the pivotal role of the bottom lane throughout the year. The proximity of the jungler and mid laner to these bot laners on the red team's side suggests that this setup may facilitate more effective assistance, potentially improving the red team's performance. Moreover, the red team's geographical advantage in accessing the dragon, a crucial in-game objective, probably reinforces their winning prospects. Additionally, the red team's ability to pick more champions after the blue team during champion selection may indicate that counter-picking strategies play a more important role in the game dynamics during the patch 13.3.

On the other hand, the observed increase in the blue team's win rate during patch 13.9 can be attributed to nerfs (power reductions) affecting the bot and support roles. Despite these adjustments, these roles remain quite powerful, thereby continuing to give the red team a higher win rate, particularly in high elo, where players can effectively use these power imbalances to their advantage.

6.2.3 Champion analyses

Understanding which champion holds a prominent position in the current patch significantly influences the outcomes of matches. In high elo gameplay, champions at their state are the most played, while in low elo, the players' preferences often guide the selection of champions. This is particularly pronounced for low elo players, who may lack comprehensive knowledge of all champions. To facilitate their rapid mastery of a champion, it becomes crucial to conduct in-depth analyses that can identify critical risk zones and crucial moments that affect a team's performance.

This phenomenon was exemplified by the cases of the champions Elise, Sylas, and Vi. In-depth analysis revealed a convincing correlation between their early deaths, particularly on the

bottom side of the map, and their eventual loss of the game. This insight serves as a valuable warning sign for novice players who opt for these champions, guiding them to exercise greater vigilance during specific phases of the game in these specific areas.

In essence, the analysis not only assists players in optimizing their champion selection but also empowers them with strategic knowledge that can improve their overall gameplay. It highlights the dynamic interplay between champion picks, strategic positioning, and match outcomes, providing a valuable framework for new players.

6.2.4 Utility of data analysis

One noteworthy observation concerns the predominant absence of data analysis resources in some Esports teams, with many still relying on manual replay analysis for their game evaluations. In assessing the value of data analysis in the Esports domain, it becomes clear that the inclusion of data analysts can substantially benefit teams by uncovering patterns of player behavior patterns that might otherwise go unnoticed. It also speeds up the identification of established patterns, significantly reducing the time invested in analyzing game replays to uncover these potential patterns.

According to the knowledge provided by the coach, spatio-temporal analysis plays a key role in evaluating the performance of his team and its opponents. This analytical approach allows for the identification of critical patterns in specific areas and game scenarios, facilitating the rectification of team errors and the formulation of strategic adaptations to capitalize on these moments. Consequently, spatio-temporal analysis emerges as a key tool with the potential to significantly influence match results.

6.2.5 Relevant events in the game

In addition to the events examined within this work, the coach provided valuable information about other noteworthy events that deserve to be explored and analyzed. These events extend the scope of potential research within the field of Esports analytics. The coach's suggestions cover:

- **Patterns in Jungler Movements:** A deeper examination of the jungler's location and path during the initial 7 minutes of gameplay, clarifying early-game strategies and positioning;
- **Objective Control Analysis:** Investigating the instances when a team secures jungle special monsters, such as the Herald, Baron, and Dragons, which have significant game-altering implications;
- **Warding Patterns:** An examination of player ward placement, particularly during the early game, to identify key vision control strategies;

- **Team Death Patterns:** Analyzing patterns of team deaths, particularly when the same set of five players consistently participates as a team, offering insights into team dynamics and vulnerabilities;
- **Player-Specific Analyses:** Tailoring the analyses to focus on individual players, allowing for a more detailed understanding of their contributions and behaviors within the team;
- **Location Patterns for Junglers and Supports:** Recognizing that certain positions hold greater significance for junglers and supports, emphasizing the importance of identifying location patterns specific to these roles.

Chapter 7

Conclusion and Future Work

This chapter describes possible future work and some final thoughts on the work done.

7.1 Conclusion

The world of Esports has undergone remarkable growth and transformation in recent years, expanding its reach to encompass a global community of players, spectators, analysts, coaches, commentators, and gaming companies. This evolution has been complemented by a surge in research efforts and data analysis applications aimed at enhancing various facets of the Esports experience.

The use of data analysis in Esports has become increasingly common, with a diverse array of applications ranging from predicting match outcomes and identifying high-risk situations for players to recommending optimal in-game strategies and mitigating toxicity issues [14]. Accessible tools and platforms have also emerged, enabling players and coaches to take advantage of data-based information, visualize performance metrics, and elevate their competitive edge.

Furthermore, the longevity of Esports titles has been significantly extended, in part thanks to the contributions of gaming data science. Games like League of Legends (LoL), now 14 years old, continue to thrive with a large and dedicated player base, showcasing the enduring appeal of these virtual games.

In the course of analyzing the spatio-temporal data, a series of challenges emerged. Determining the appropriate analyses and algorithms for this data type posed significant complexities. Furthermore, while certain game companies readily provide data, access to spatio-temporal data remains limited for specific titles, often confined to historical gameplay data as opposed to real-time information. This limitation primarily arises from concerns related to cheating and the possibility of this type of data being used to take advantage of imbalances or exploits, since this data can be used to optimize strategies and tactics that the creators did not anticipate. Game companies may also have commercial agreements and partnerships that imply exclusive access to certain data, such as spatio-temporal data.

The clustering analysis of spatio-temporal data, despite its inherent value, revealed areas for potential improvement, particularly in terms of silhouette scores. Addressing these challenges

and enhancing the dataset could involve the incorporation of additional spatio-temporal data from diverse players and games, or securing access to more detailed real-time data through game APIs. These steps have the potential to mitigate the aforementioned limitations and produce more refined results.

From the perspective of the professional Esports team coach, the analyses presented valuable insights and patterns within each cluster. However, several constraints, such as infrequent updates of positional information, limitations in ward positioning data, and the absence of spatio-temporal data from esports tournaments, emphasized the unexplored potential for even more comprehensive analyses. The prospect of obtaining precise player position data over time holds the promise of revolutionizing the comprehension of player behavior and strategies, particularly within the pivotal early-game phases.

Regarding clustering analysis of spatio-temporal data, the resulting clusters offered valuable insights for players. These clusters serve as a powerful tool for players, aiding them in the identification of critical patterns within their specific roles and divisions without the arduous task of reviewing numerous game replays.

Moreover, this analysis helps players identify areas and times where specific roles in various divisions are most active. By leveraging clustering techniques, players can navigate the complexities of the game with greater precision and insight, ultimately elevating their competitive edge in the ever-changing League of Legends patches.

In summary, while significant progress has been made in the domain of Esports data analysis, numerous promising avenues for further exploration and refinement exist. The dynamic nature of Esports, combined with the ever-expanding availability of data and continuous advancements in analytical techniques, foretell a future where Esports enthusiasts can uncover deeper insights and push the boundaries of competitive gameplay. This research journey signifies just the initial step, as the Esports ecosystem continues its evolution and growth, driven by innovation based on data-driven approaches and a strong commitment to excellence in this rapidly expanding field.

7.2 Future Work

In the context of future research and development, there exists several opportunities for further exploration and innovation within the dynamic landscape of gaming and its evolving technology. As games and their underlying technologies continue to advance year by year, there remain a lot of unexplored territories, ready to be analyzed and acted upon for the benefit of the gaming community. This chapter analyses various avenues of potential exploration and contribution, each of which is poised to have a significant impact on the gaming ecosystem.

In addition to the significant events unveiled by the coach, other spatio-temporal scenarios deserve future analysis:

- Clustering spatio-temporal data from other games, like PUBG, also provides a lot of data to the developer's community.

- Spatio-temporal analysis of Esports teams - collaborating with teams to explore and analyze collected data.
- Predicting where a player might be at a given moment in the game based on patterns found with spatio-temporal data.
- Find possible patterns in players' paths (the API should provide the coordinates per second to improve the analysis results).

Other analyses of interest for the future may include:

- Detect toxicity in players through their movements and interactions in the match, with the help of chat logs to understand the moment of analysis.
- Detect toxicity in player voice chat.
- Creation of player behavior profiles, and physical metrics of the player (heart rate, and eye tracking, among others) to simplify the selection of the desired players for the teams.
- Use of ML for map generation.
- Creation of a tool to extract replay data from games other than DOTA2 (extract data from videos).

Bibliography

- [1] Jupyter. <https://jupyter.org/>. (Visited on 2022-10-30).
- [2] Matplotlib. <https://matplotlib.org/>. (Visited on 2022-11-10).
- [3] Numpy. <https://numpy.org/>. (Visited on 2022-10-30).
- [4] Pandas. <https://pandas.pydata.org/>. (Visited on 2022-10-30).
- [5] Pyvista. <https://docs.pyvista.org/version/stable/>. (Visited on 2022-11-30).
- [6] Riotwatcher documentation. <https://riot-watcher.readthedocs.io/en/latest/>. (Visited on 2022-10-25).
- [7] Scikit-learn. <https://scikit-learn.org/>. (Visited on 2022-10-30).
- [8] Seaborn. <https://seaborn.pydata.org/>. (Visited on 2022-11-10).
- [9] Rafael Afonso. Visualeague iii: Visual analytics of multiple games. Master's thesis, Universidade de Lisboa (Portugal), 2020.
- [10] Myat Aung, Simon Demediuk, Yuan Sun, Ye Tu, Yu Ang, Siva Nekkanti, Shantanu Raghav, Diego Klabjan, Rafet Sifa, and Anders Drachen. The trails of just cause 2: Spatio-temporal player profiling in open-world games. In *Proceedings of the 14th International Conference on the Foundations of Digital Games, FDG '19*, New York, NY, USA, 2019. Association for Computing Machinery.
- [11] Tom Batsford. Calculating optimal jungling routes in dota2 using neural networks and genetic algorithms. *Game Behaviour*, 1(1), 2014.
- [12] Christian Bauckhage, Rafet Sifa, Anders Drachen, Christian Thureau, and Fabian Hadiji. Beyond heatmaps: Spatio-temporal clustering using behavior-based partitioning of game levels. In *2014 IEEE conference on computational intelligence and games*, pages 1–8. IEEE, 2014.
- [13] Blitz. Blitz.gg application. <https://blitz.gg/>. (Visited on 2022-10-12).
- [14] Lincoln Magalhães Costa, Anders Drachen, Francisco Carlos Monteiro Souza, and Geraldo Xexéo. Artificial intelligence in moba games: a multivocal literature mapping. *IEEE Transactions on Games*, 2023.

- [15] Anders Drachen, Matthew Yancey, John Maguire, Derrek Chu, Iris Yuhui Wang, Tobias Mahlmann, Matthias Schubert, and Diego Klabajan. Skill-based differences in spatio-temporal team behaviour in defence of the ancients 2 (dota 2). In *2014 IEEE Games Media Entertainment*, pages 1–8, 2014.
- [16] Magy Seif El-Nasr, Truong-Huy D Nguyen, Alessandro Canossa, and Anders Drachen. *Game Data Science*. Oxford University Press, 2021.
- [17] Inc Electronic Arts. Apex legends. <https://www.ea.com/games/apex-legends>. (Visited on 2022-10-12).
- [18] Faceit. Faceit developer api. <https://developers.faceit.com/>. (Visited on 2022-10-20).
- [19] Peter Flach. *Machine learning: the art and science of algorithms that make sense of data*. Cambridge university press, 2012.
- [20] Riot Games. League of legends. <https://euw.leagueoflegends.com/en/>. (Visited on 2022-10-12).
- [21] Riot Games. Riot developer portal. <https://developer.riotgames.com/>. (Visited on 2022-10-20).
- [22] Riot games. Riot games. <https://www.riotgames.com/en>. (Visited on 2022-10-12).
- [23] Riot Games. Valorant. <https://playvalorant.com/>. (Visited on 2022-10-12).
- [24] Aurélien Géron. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. ” O’Reilly Media, Inc.”, 2022.
- [25] Gamers Net Inc. Mobalytics application. <https://mobalytics.gg/>. (Visited on 2022-10-12).
- [26] OP.GG Inc. Op.gg application. <https://www.op.gg/>. (Visited on 2022-10-12).
- [27] Adam Katona, Ryan Spick, Victoria J. Hodge, Simon Demediuk, Florian Block, Anders Drachen, and James Alfred Walker. Time to die: Death prediction in dota 2 using deep learning. In *2019 IEEE Conference on Games (CoG)*, pages 1–8, 2019.
- [28] Inc Krafton. Playerunknown’s battlegrounds (pubg). <https://emea.battlegrounds.pubg.com/en/>. (Visited on 2022-10-12).
- [29] Inc Krafton. Pubg developer api. <https://developer.pubg.com/>. (Visited on 2022-10-20).

- [30] Yann LeCun, Yoshua Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995.
- [31] 2023 Beckstar LLC. trackingthepros. <https://www.trackingthepros.com/players/>. (Visited on 2023-4-2).
- [32] Tom M Mitchell. *Machine learning*, volume 1. McGraw-hill New York, 1997.
- [33] Fernando Felix do Nascimento Junior, Allan Sales da Costa Melo, Igor Barbosa da Costa, and Leandro Balby Marinho. Profiling successful team behaviors in league of legends. In *Proceedings of the 23rd Brazillian Symposium on Multimedia and the Web, WebMedia '17*, page 261–268, New York, NY, USA, 2017. Association for Computing Machinery.
- [34] Tracker Network. Tracker developer api. <https://tracker.gg/developers/>. (Visited on 2022-10-20).
- [35] Overwolf. Overwolf. <https://www.overwolf.com/>. (Visited on 2022-10-20).
- [36] Rui Roque. 3d visualization of time and space. <https://clipchamp.com/watch/sD3RN13pa8R>, 2023. Clipchamp video.
- [37] Rui Roque. Esports games data analysis. <https://github.com/Bambito9/Esports-Games-Data-Analysis>, 2023. Github repository.
- [38] Matthias Schubert, Anders Drachen, and Tobias Mahlmann. Esports analytics through encounter detection. In *Proceedings of the MIT Sloan Sports Analytics Conference 2016*. MIT Sloan, 2016. MIT Sloan Sports Analytics Conference ; Conference date: 11-03-2016.
- [39] Steam. Dota2 developer api. <https://docs.opendota.com/>. (Visited on 2022-10-20).
- [40] Hugo Tristão. The esports audience will pass half a billion in 2022 as revenues, engagement, & new segments flourish. *2022 Global Esports and Live Streaming Market Report*, 04 2022.
- [41] Valve. Counter-strike: Global offensive. <https://blog.counter-strike.net/>. (Visited on 2022-10-12).
- [42] Valve. Dota 2. <https://www.dota2.com/home>. (Visited on 2022-10-12).
- [43] WARGRAPHS. Porofessor application. <https://porofessor.gg/>. (Visited on 2022-10-12).

Appendix A

Data collection

Nowadays, as can be seen in Table [A.1](#), it is easier to collect game data since most game companies or commercial companies offer access to their APIs where the developers can collect data from their games. However, most of them only give access to statistical data. The following subsections (Subsection A.1 to Subsection A.4) present some of the game genres explored and what data can be extracted from games' APIs of these genres. Subsection A.5 presents the decision of the API to extract data and consequently perform an analysis of the collected data.

A.1 MOBA games APIs

MOBA games are a type of video game in which two teams of players compete against each other on a battlefield. Each player controls a single character, and the goal is to destroy the other team's base or stronghold. Players work with their teammates to defeat the enemy team, using a variety of abilities and tactics.

In the MOBA genre, the following two APIs from two different games were explored:

- Riot API [\[21\]](#) - League of Legends [\[20\]](#);
- OpenDota API [\[39\]](#) - DOTA2 [\[42\]](#).

Both offer a large volume of data as they are two of the most played games in the world. The data that can be collected in OpenDota API [\[39\]](#) is very equivalent to the data that is collected from Riot API [\[21\]](#), not only because they give access to the same types of data but also to the fact that both games are very similar.

In more detail, both APIs give access to statistical data for a given player or match, like player scores, player rank, heroes/champions stats, number of wins, number of losses, number of camps killed in a match, and total gold for each team. They also allow us to collect spatio-temporal data like a player's location at a given time and live game data.

The main difference between them is that DOTA2 API is limited to 60 requests every 1 minute, while the Riot API is limited to 50 requests per minute. The DOTA2 datasets are the most com-

monly used for game data science research as it gives more detailed data and replays from the matches.

A.2 First-Person Shooter (FPS) games APIs

First-Person Shooter games are a type of video game played from the player's perspective, with the gameplay centered around guns. The games usually have core objectives that make the team win, like planting a bomb or defusing it. FPS games often involve strategy and resource management elements in addition to fast-paced combat.

In the FPS genre, the following APIs were explored:

- Riot API [21] - Valorant [23];
- Tracker.gg API [34] - CS:GO [41];
- Faceit API [18] - CS:GO [41].

Valorant [23] has a policy where only an established brand can access the API, and only statistical data like player stats can be collected. In-game overlays that include real-time data that would improve a player's performance are not allowed. Another problem is that players need to permit data sharing, and even if all third-party sites are approved by Riot, they can not access data from a player that did not give permission, which reduces the amount of data that can be collected. These restrictions may be because the game has only recently been released but influenced the API selection used in this project.

In both CS:GO APIs, only statistical data can be collected, like the number of kills, headshots, deaths, and assists, among others. Unlike the tracker API [34], where only general statistical data can be collected from all the games played, the Faceit API [18] gives access to more specific statistical data by giving player data from each map played.

A.3 Battle Royale (BR) game API

Battle royale games are a type of video game in which many players compete against each other in a last-man-standing deathmatch. Players are dropped onto a large map and must scavenge for weapons and other equipment to survive. As the game progresses, the play area becomes shorter, forcing the remaining players to confront each other.

In the BR genre, the following API was explored:

- PUBGAPI [29], an API of the PUBG game [28].

It gives access to player stats, match stats, and telemetry data, such as a player's location on the map or item pick-up at a given time. The API is limited to 10 requests every 1 minute, but matches and telemetry endpoints are not rate-limited.

A.4 Overwolf

Overwolf [35] is an all-in-one platform for creating, sharing, and monetizing in-game apps and modifications (mods). The platform gives access to an API where developers can collect statistical data in real-time from multiple supported games such as Apex Legends [17], PUBG [28], DOTA2 [42], CS:GO [41], League of Legends [20], and Valorant [23], among others. This platform presents some advantages, like giving easier access to statistical data in real-time from the valorant game or in real-time data from CS:GO. The main disadvantage of the API is that it does not offer more detailed data than some of the others mentioned before and does not give any spatio-temporal data except for the PUBG game. This prohibition of spatio-temporal data collection is probably to prevent people from using this kind of data for cheating because it can be critical data that influence the outcome of a match. The collection of data in real-time is done through some events that occur during the game, such as a player killing another player, the purchase of an item, an item picked up, or even when a player jumps.

Game	API	Data					Request Permission
		Spatio-temporal Data	Statistical Data (number of kills, deaths, etc)		Behavior Telemetry (data in real-time)		
			Matches data	Players general Data	Statistical	Spatio-temporal	
LoL	Riot Games	✓	✓	✓	✓	X	✓
	Overwolf	X	X	X	✓	X	✓
Valorant	Riot Games	X	✓	✓	✓	X	X
	Overwolf	X	X	X	✓	X	✓
CS:GO	Tracker.gg	X	X	✓	X	X	✓
	Faceit	X	✓	✓	X	X	✓
	Overwolf	X	X	X	✓	X	✓
PUBG	PUBG	✓	✓	✓	✓	✓	✓
	Overwolf	✓	X	X	✓	✓	✓
Dota2	OpenDota	✓	✓	✓	✓	✓	✓
	Overwolf	X	X	X	✓	X	✓

Table A.1: APIs explored for different games and which type of data can be collected from them

Note: All data collected from these APIs come in JSON format

Appendix B

League of Legends datasets features

Attribute name	Description	Type
gameId	Unique Riot ID of the game. Can be used with the Riot Games API.	String
team	Team the player played for (map side). Red if played on the red side, and blue if played on the blue side	String
summoner	Anonymous summoner name	String
region	Server region where the player plays. Can be used with the Riot Games API	String
champion	Champion's name that was used by the player	String
timestamp	Specific moment of the match in milliseconds	Float
coordinate.x	x-coordinate of the player's position. Ranges from 0 to 16 000	Float
coordinate.y	y-coordinate of the player's position. Ranges from 0 to 16 000	Float
level	Champion's level. Ranges from 1 to 18	Int
minionsKilled	Number of minions killed	Int
jungleMinionsKilled	Number of neutral minions killed (monsters in the jungle)	Int
currentGold	Amount of gold the player has	Int
goldPerSecond	Amount of gold the player has per second	Int
totalGold	total gold the player had (gold spent plus current gold)	Int
xp	Total champion's experience acquired until a specific moment of the match	Int
abilityPower	Amount of magic damage the champion has	Int
armor	Amount of armor the champion has	Int
armorPen	Amount of armor penetration the champion has	Int
armorPenPercent	Percentage of armor penetration	Int
attackDamage	Amount of attack damage the champion has	Int
attackSpeed	Amount of attack speed the champion has	Int

ccReduction	Amount of crowd control reduction the champion has	Int
health	Amount of health the champion has	Int
healthMax	Champion's maximum amount of health	Int
healthRegen	Amount of health a champion regenerates over a five-second period	Int
lifesteal	Life steal is a stat that grants healing equal to a percentage of the damage dealt by basic attacks	Int
magicPen	Amount of magic resistance ignored when dealing magic damage	Int
magicPenPercent	Percentage of magic penetration	Int
magicResist	Amount of magic damage reduced/blocked	Int
movementSpeed	Rate at which a champion travels across a map. One point of movement speed translates to one game distance unit traveled per second	Int
omnivamp	grants healing equal to a percentage of the physical damage, magic damage, and true damage dealt	Int
power	Amount of mana the champion has	Int
powerMax	Champion's total mana	Int
powerRegen	Amount of mana regeneration	Int
spellVamp	percentage of champion's damage, dealt via spells/abilities, back as health	Int
magicDamageDone	Amount of magic damage done to everything (minions, turrets, champions, etc)	Int
magicDamageDoneToChampions	Amount of magic damage done to champions	Int
magicDamageTaken	Amount of magic damage taken	Int
physicalDamageDone	Amount of physical damage done to everything (minions, turrets, champions, etc)	Int
physicalDamageDoneToChampions	Amount of physical damage done to champions	Int
physicalDamageTaken	Amount of physical damage taken	Int
totalDamageDone	Total damage done to everything (minions, turrets, champions, etc). This includes all the physical, magic, and true damage done.	Int
totalDamageDoneToChampions	Total damage done to champions	Int
totalDamageTaken	Total damage taken	Int
trueDamageDone	Amount of true damage done. True damage is damage that cannot be mitigated or reduced	Int
trueDamageDoneToChampions	Amount of true damage done to champions	Int
trueDamageTaken	Amount of true damage taken	Int
victory	Indicates whether or not the player won the match. True if won, False if he didn't	String
eventType	Indicates what type of event occurred. This dataset contains 5 types of events, which are SkillLevelUp, Elite Monster Kill by his team, Assist, Kill, Death and Purchased Items	String
skillSlot	Indicates which champion's skill the player has leveled up. 1 for Q's skill, 2 for W's skill, 3 for E's skill, and 4 for R's skill	String
monster	Indicate which monster was killed by his team. There are 10 different types of monster	String
buildingType	Indicate which type of structure was destroyed. It can be a turret or an inhibitor	String

lane	Indicates in which lane the structure was destroyed	String
itemName	name of the item purchased	String
role	indicates the player's role	String

Table B.1: Description of some important features that can be collected from Riot's API

Attribute name	Description	Type
rank	indicates the player's division	String

Table B.2: Description of some important features that can be collected from Riot's API

Appendix C

Clustering analysis plots

C.1 Death pattern on professional junglers

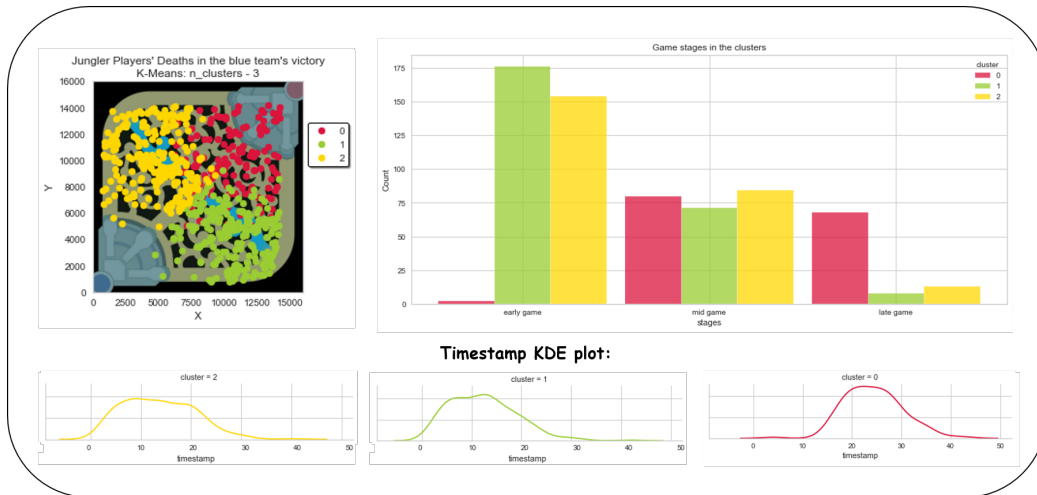


Figure C.1: Clustering professional jungler players' deaths in blue team's victory

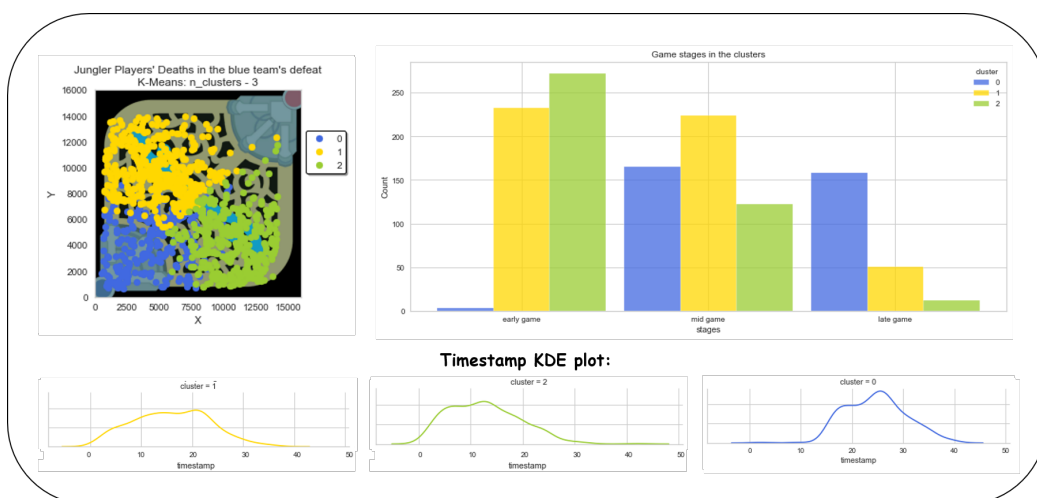


Figure C.2: Clustering professional jungler players' deaths in blue team's defeat

C.2 Death Pattern On Patch 13.9

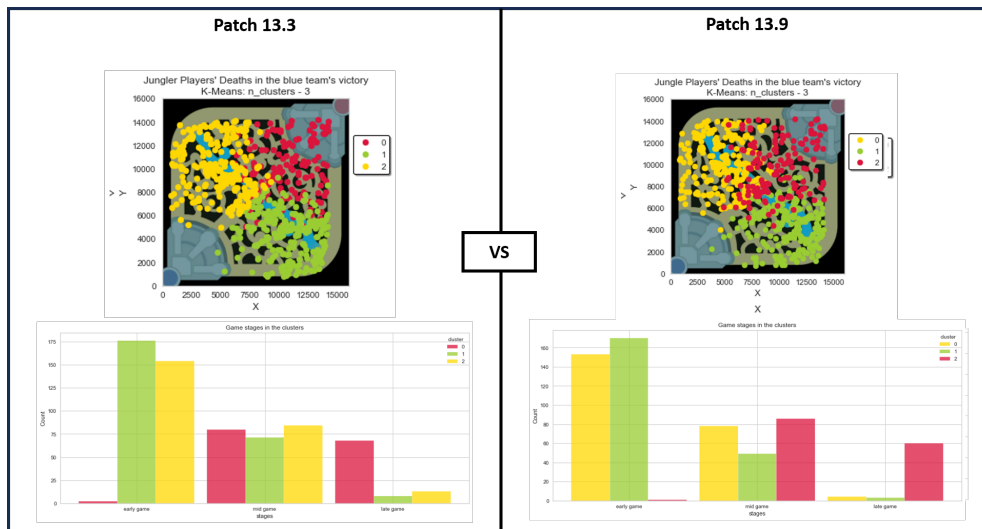


Figure C.3: Comparison between different patches in blue team's victory on professional players

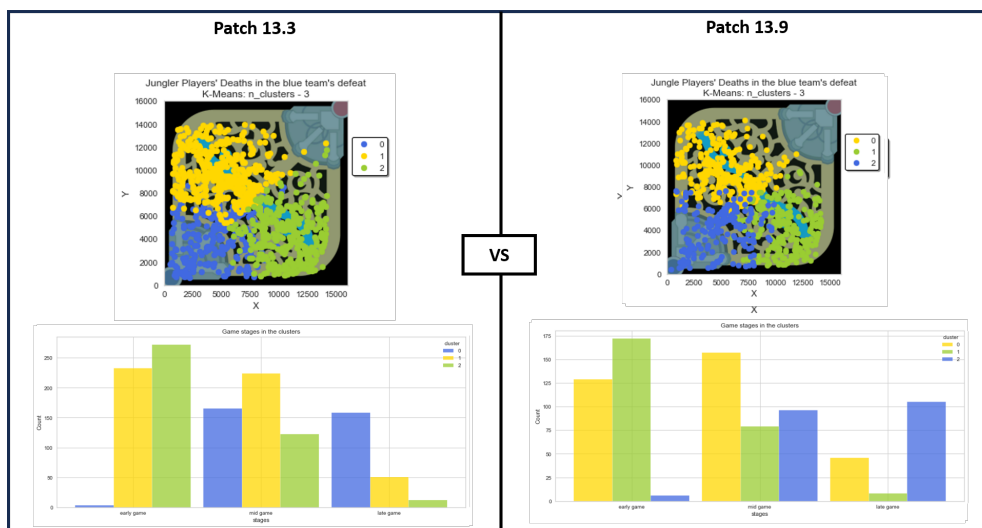


Figure C.4: Comparison between different patches in blue team's defeat on professional players

C.3 Locating professional junglers

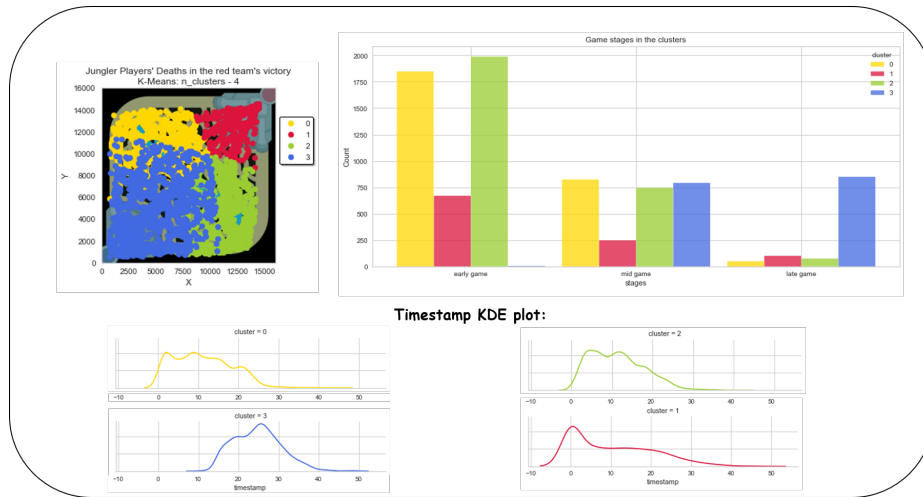


Figure C.5: Location clusters in red team's victory on professional players

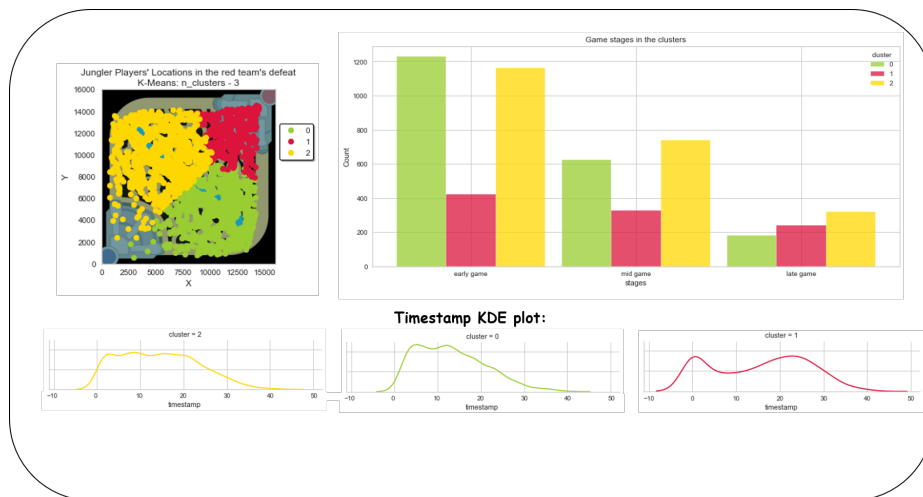


Figure C.6: Location clusters in red team's defeat on professional players

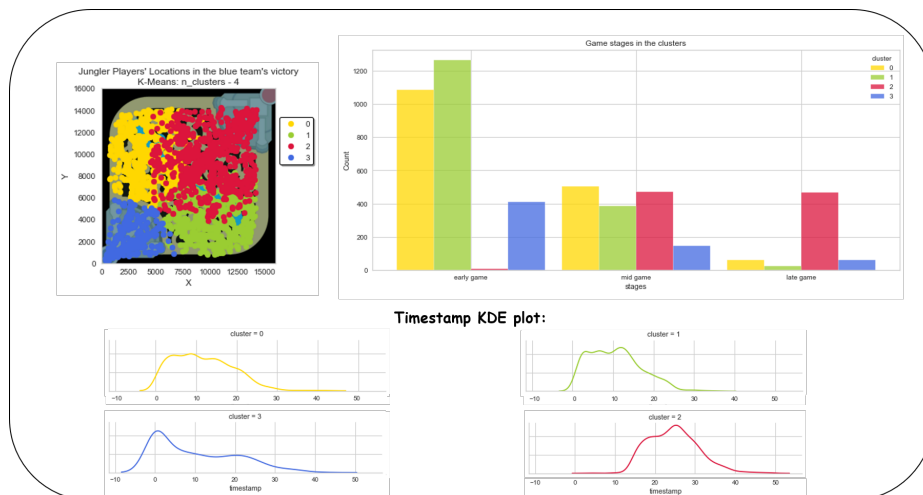


Figure C.7: Location clusters in blue team's victory on professional players

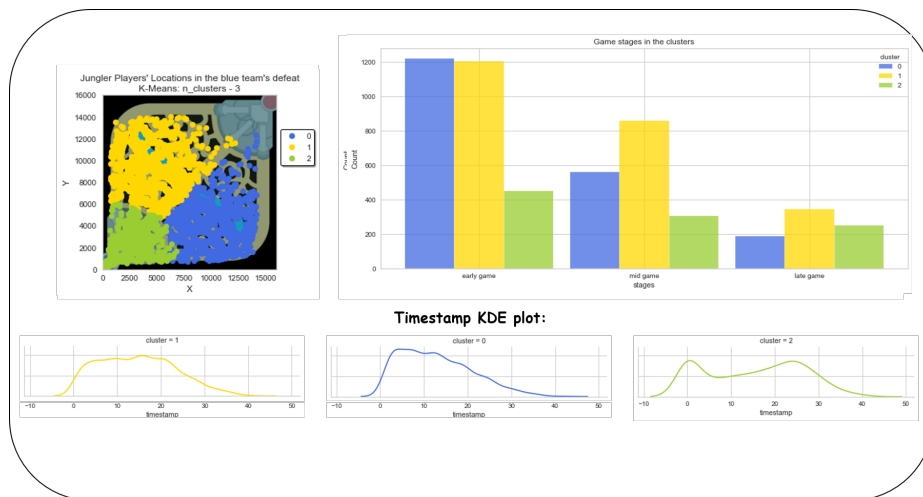


Figure C.8: Location clusters in blue team's defeat on professional players