



UNIL | Université de Lausanne
Ecole des sciences criminelles
bâtiment Batochime
CH-1015 Lausanne

Master of Science (MSc) in Forensic Science
Orientation Digital Investigation and Identification
&
Collaboration with UEFA

Master Thesis

Leveraging Historical Data in Digital Forensic & Open Source
Intelligence for Football Match-Fixing Investigations

Mahmut Yildirim

Under direction of Professor Thomas Souvignet

August 2024

Faculté de droit, des sciences criminelles et d'administration publique
Ecole des sciences criminelles

|||||

Tél.+41 21 692 46 00 | Fax.+41 21 692 46 05 | info.esc@unil.ch | www.unil.ch/esc

Abstract

Match-fixing incidents in football cause significant damage to the integrity of the sport. Match-fixing can be done in many ways. With the development of technology, online betting and its evolution, online betting exchange platforms have developed today. This development has also opened new areas for fixers who aim to make money from match-fixing. These betting exchange platforms, which contain significant liquidity, can also be used for crime detection by investigators who aim to detect and obtain clues.

With this study, various anomalies were tried to be detected for the purpose of match-fixing investigation in open source intelligence and digital forensic intelligence. For this purpose, betfair.com, one of the most important online betting exchange platforms, and the historical data on this site were selected. Datasets from May 2020 and July 2020 were selected for data analysis. 665 football matches from May were used for testing purposes during the creation of the method. Using automated methods throughout all stages, data from approximately 3,723 football matches across various world leagues in July 2020 were converted, processed, and visualized. Anomalies were detected based on predefined criteria. Finally, suspicious events were analyzed by examining the anomalies, graphical representations, and CSV datasets to draw conclusions.

Keywords: Digital Forensic Intelligence, Open Source Intelligence, Match-fixing, Betting Exchange, Automation, Data Analysis, Data Visualisation

Acknowledgments

I would like to sincerely thank Professor Thomas Souvignet for his guidance, support, and the many helpful insights he provided throughout this project. His advice was crucial in completing this work.

I also want to thank Benoît Meylan for his technical advice and for carefully proofreading my work. His contributions were essential in polishing the final result.

My gratitude extends to Romain Le Grand from UEFA for his collaboration and support.

Finally, I am deeply thankful to my wife, Rukiye Yildirim, and my son, Talha. Their unwavering support, patience, and understanding during the many evenings and weekends I spent studying have meant everything to me. I truly appreciate the strength and encouragement they have given me.

Contents

Contents	iii
1 Introduction	1
1.1 Motivation	1
1.2 Purpose	1
1.3 Focus	1
1.4 Goals and Questions	2
2 Related Work	2
2.1 Overview	2
2.1.1 Digital Forensic Intelligence	2
2.1.2 Open Source Intelligence	3
2.1.3 Match-fixing	3
2.1.4 Online Betting Exchange	3
2.1.5 Historical Data	4
2.2 Match-fixing in Football	4
2.2.1 Match Fixing Issue	4
2.2.2 Impact of Match-Fixing (Gain, Corruptions, Tax Revenue..)	5
2.2.3 Match-Fixing Investigations	6
2.3 Exploiting Historical Data in Match-Fixing Forensic Analysis	7
2.3.1 Role of Historical Data	7
2.3.2 Leveraging Historical Data to Detect Anomalies	7
3 Methodology and Implementation	8
3.1 Methodology	8
3.2 Implementation	10
3.2.1 Data Collection	10
3.2.2 Data Conversion	10
3.2.3 Data Processing	11
3.2.4 Creating Graphs	11
3.2.5 Creating Criteria	12
3.2.6 Creating Report	12

4 Advantages and Limitations	13
4.1 Advantages	13
4.2 Limitations	13
5 Feasibility	14
6 Results	14
7 Discussion	37
8 Conclusion	39
References	41
A Ethics	45
B Appendix	46

Abbreviations list

DFINT: Digital Forensic Intelligence.

OSINT: Open-Source Intelligence.

ID: Unique identifier for the market.

UEFA: Union of European Football Associations.

B2Z: A free and open-source file compression program

CSV: Comma-Separated Values.

PNG: Portable Network Graphics.

TXT: Plain Text File.

API: Application Programming Interface.

FIFA: Fédération Internationale de Football Association.

Terminology

eventId: The unique identifier for the event associated with the market.

marketId: The identifier for the specific market.

numberOfWinners: The number of winners expected in the market type.

marketType: The type of the market as shown in Table 23 (e.g., TO_SCORE_2_OR_MORE).

inPlay: Indicates whether the market is currently in-play.

Preplay data: Refers to information or data available before the match starts.

numberOfActiveRunners: The number of active runners in the market type.

status: The status of the market type (e.g., OPEN, SUSPENDED).

runners: An array containing details of individual runners in the market type.

selection status: The status of the runner (e.g., ACTIVE).

timezone: The timezone of the market.

publishTime: The time when the market information was published.

totalMatched: The total amount of money placed in the market type.

md.eventName: The name of the event.

md.name: The name of the market type.

selection selectionId: The identifier for the market type option.

selection lastPriceTraded: The last traded odds for each market type option.

selection md.name: The amount of money placed for each market type option.

selection ex.availableToBack selection: The odds available for backing a selection.

ex.availableToBack.size: The amount of money available to back a selection.

selection ex.availableToLay.price: The odds available for laying against a selection.

selection ex.availableToLay.size: The amount of money available to lay against a selection.

selection status: The status of the market type option.

selection totalMatched: The total amount of money placed on each market type option.

1 Introduction

1.1 Motivation

Digital forensic data has been rapidly increasing for many years [30] [50] [52]. Advances in technologies have led to an expansion in the volume, variety, speed and accuracy of data available for digital forensic analysis. Analyzing large digital forensic datasets offers the benefit of consolidating case-related information from diverse sources [51].

Moreover, many organizations [41] have recognized the potential of data accumulated in the web space as a result of digital transformation, which can be transformed into valuable information and intelligence [72]. Since these contents are freely available online, they are accessible and readable by everyone [42]. This has fostered Open Source Intelligence (OSINT), a practice over half a century old used to extract meaningful intelligence from publicly available data [72], even when dealing with big data characterized by its massive size and complexity [59] [8].

Corruption in sports, particularly match-fixing, presents a significant challenge to the integrity of sports [12] and complicating efforts to detect traces or evidence of such misconduct.

One type of investigation [64] to detect anomalies, where open source and forensic intelligence [51] are used effectively, is match-fixing investigation. Historical datasets are such valuable sources of information that they can be very useful in detecting anomalies in such investigations.

1.2 Purpose

The general purpose is to utilize Digital Forensic Intelligence and Open Source Intelligence for conducting intelligence analysis, which will help in collecting information to detect any anomalies on *betfair.com*.

First of all, the importance of Digital Forensic Intelligence and Open Source Intelligence in areas such as crime investigation will be touched upon. Online betting, betting exchanges as types of online betting and historical data terms will be discussed.

Match-fixing as an issue, its impact, and the types of forensic investigations carried out on match-fixing will be discussed, followed by touching upon the role of historical data as digital forensic data in match-fixing forensic investigation.

The exploitation of historical data to detect anomalies will be discussed, mentioning the role of historical data and how it will be leveraged.

Following this, the method created and datasets used to detect anomalies will be discussed as a step in further investigations.

At the end, the data obtained will be analyzed and discussions will be held on the results.

1.3 Focus

Digital forensics was once a science used to support criminal investigations, and digital forensics services were used only in the final stages of investigations after most of the digital evidence had already been corrupted. However, digital forensic services are now used in the initial processes of investigations [4] as intelligence processes as well.

For the intelligence and investigation processes to be effective, it must use useful sources of information [67], including relevant details. At the same time, as the volume, variety, speed and accuracy of digital

forensic data increase, it is necessary to focus on the information contained in these data, with an appropriate method to process the data in a timely manner [51].

Historical data is one of the most useful sources for assisting match-fixing intelligence and investigations. The focus of this study is the historical data obtained from the *betfair.com* website, specifically all data pertaining to May and July 2020.

Betfair was selected for study because it is “World’s largest betting exchange” and the leading player in the market, in terms of bet volume, which started in Britain in 2000 [12].

1.4 Goals and Questions

The study aims to achieve the following goals:

- Exploit historical betting data obtained from the *betfair.com* website.
- Process the data for purposes such as visualization and anomaly detection.
- Visualize the betting values on the graphs
- Develop several criteria to detect anomalies.
- Conduct intelligence analysis based on the findings from the visualization and detection stages.

Additionally, the study aims to answer the following questions:

Q1: What is the importance of historical data in match-fixing intelligence and investigations?

Q2: What anomalies are detected/detectable in match-fixing cases?

Q3: What are the findings at the end of the analysis?

The fundamental principle guiding this study is the Locard exchange principle, which stipulates that “every contact leaves a trace” [71] including digital traces.

This study also aims to convert digital traces into digital evidence, following the hypothetico-deductive model as outlined by Ribaux [53], [54] in case anomalies are found.

2 Related Work

2.1 Overview

2.1.1 Digital Forensic Intelligence

The world’s growing dependence on digital sources of information, alongside the complex computerized systems and networks used for data storage, processing, and transmission [10], underscores the significance of intelligence-led policing. This philosophy prioritizes data analysis and criminal intelligence to inform and enhance decision-making processes [51].

While the role of intelligence is to provide decision-makers with an independent and impartial understanding of issues [33], according to Ribaux, forensic intelligence is the accurate, timely and useful product of logically processing (analysis of) forensic case data (information) for investigation and/or intelligence purposes [55].

Digital forensic data is collected and analyzed to be presented in court as evidence of illegal activities [11]. Digital Forensic Intelligence is intelligence and information obtained from digital forensic data and it is suggested that digital forensic should encompass the information potential contained in digital forensic data storage [51]. So, digital forensic intelligence is a forward-looking analysis process to predict what may happen in the future by using digital data obtained from previous events.

2.1.2 Open Source Intelligence

Open source intelligence involves collecting data from publicly accessible sources [58] [56] [9] [46] organizing it, and then turning it into useful knowledge that OSINT experts and other businesses need [72]. Open source, unlike open data, which is a subset of it [48] [47] [1] [20] [68], has usage restrictions, such as requiring login credentials.

Open-source intelligence (OSINT) began as a discipline around the start of the Second World War with the establishment of the BBC Monitoring Service in Great Britain in 1939 and the Foreign Broadcast Monitoring Service (FBMS) in the United States in 1941 [6].

OSINT was initially considered unnecessary because of the low reliability of sources, which often included misinformation, secret messages, disinformation, and nonsense content. However, when properly used, it is believed to be as crucial as the most secretive and expensive aspects of intelligence collection. Ultimately, the most significant factor is the system's ability to provide sound judgments to decision makers [36]. It is known that a large proportion of the information is available to the public, and the importance of OSINT becomes even more evident.

The quality and use of intelligence depends on the quality of information. New technologies bring opportunities to obtain better quality information by supporting intelligence, thus increasing the importance of open source intelligence (OSINT) [5]. However, one challenge with OSINT is dealing with data in different languages, which may require translation services, analyst training, or translation software [67].

2.1.3 Match-fixing

Match-fixing is a global problem that threatens the fairness of sports. It happens when sports people are tempted or pressured to change the outcome of games, often for financial reasons. This dishonest practice is frequently connected to betting, with criminal networks [24] taking advantage of unregulated betting markets to make money [19]. [35]

Match-fixing, also referred to as sport manipulation, sporting fraud, sport-fixing [18] or rigging bets [34] and makes people doubt the honesty and fairness of sports teams and events [15]. People love sports for many reasons, but the main reason is the excitement and unpredictability of the outcomes [44].

The general awareness arises from scandals, frauds, rumors, and judicial investigations involving various sports across many countries. The noticeable rise in sports fraud cases, especially those linked to betting activities, is endangering the values of sport [40]. This increases the probability of connections between betting activities particularly online betting and match fixing cases aimed at gaining financially in the sports betting market.

2.1.4 Online Betting Exchange

Online betting is growing quickly in the EU, impacting the economy through innovation and boosting tax revenues for Member States. Advances in online technologies have made it possible to offer betting services through various remote channels like the internet and mobile phones [16].

In a traditional betting model, bets are matched against the bookmaker's odds. If you place a bet on a selection, the bookmaker takes the opposing side of the bet. If you win, the bookmaker pays you out of their own pocket.

The betting exchange is another development of the online era [25]. One of the most well-known examples of this is Betfair. Betfair is a betting exchange that connects people who want to bet or take bets on a specific outcome.

In a betting exchange like Betfair, customers can both back (bet on an outcome to happen) and lay (bet against an outcome to happen). When a customer backs an outcome, they effectively place a bet with another customer who is laying that outcome, and vice versa. This creates a marketplace where odds are determined by supply and demand rather than set by the bookmaker.

In other words, unlike traditional bookmakers, Betfair doesn't take bets directly but instead earns a commission by matching people who want to bet for (backers) and against (layers) a particular outcome, like "Manchester United will beat Chelsea." Clients can offer odds either to support (back) or oppose (lay) an outcome. Those offering to lay act like bookmakers but without needing a license or physical premises. Their presence in the market has contributed to lower overall betting margins because the odds they offer are often more competitive than those traditionally offered by bookmakers [26].

Betfair receives a commission from every bet placed on this platform. In this way, he completely excludes himself from any possible cheating incident. To bet on this site, it is sufficient to open an account on the website and verify your identity and address. Betfair also requires official documents for account verification in order to reduce integrity concerns. All bets placed on this website can be monitored instantly.

There is another terms must be known which is *Cross-Matching*. Cross-matching allows unmatched bets to be matched against each other, even if they were placed at different odds. For example, if one customer wants to back "match odds" at odds of 3.0 (2/1) and another customer wants to lay the same "match odds" at odds of 2.9 (9/10), the exchange may cross-match these bets if they are compatible, resulting in both bets being matched at odds somewhere between 2.9 and 3.0. Cross-matching is a key feature of exchange betting platforms like Betfair because it helps ensure that customers can find matches for their bets even if there isn't an exact opposite bet available at the desired odds.

2.1.5 Historical Data

Historical data, broadly speaking, refers to information collected about past events and circumstances related to a specific subject [65]. Considering that big data is defined by the characteristics of high volume, diversity, speed and accuracy in information [38], the importance of the fact that historical data collection can be configured only for relevant data sets, depending on the needs, [37] becomes even more important.

Historical data allows us to uncover the past of a site, revealing potential reasons for and types of pollutants [63]. However, delving into history is often complex and sometimes constrained by limited data availability.

A general strategy for pattern matching in large databases can be invaluable for locating similar patterns across extensive historical data sets [43]. Historical data plays a crucial role in digital forensic and open-source intelligence for identifying patterns and detecting anomalies in online betting investigations.

2.2 Match-fixing in Football

2.2.1 Match Fixing Issue

Betting on events dates back to ancient times [57]. Football betting has been practiced since as early as 1872 [13]. Markets in which participants take a financial position on the outcome of a sporting event, such as a football game, have long existed [57].

Even though betting has the potential to increase interest in sports [25] [2] and generate significant

income in the future, it also carries risks [27]. Corruption through match-fixing in sport is among the most important of these risks [24]. Although there are individual cases of betting corruption in sports, it is difficult to assess how widespread this is.

The rise of the betting exchange is greatly increasing the amount of money in the sports betting market and it looks set to continue to do so. This situation can be problematic for sports because high liquidity will increase the appetite of match-fixers. It can allow match-fixers to place large bets without being detected or affecting the odds. It is obvious that this situation will make corruption even easier.

The significant increase in interest in sports betting has made the markets more liquid. However, the rapid development of technology allows punters to place bets globally and operators to manage risk by transferring bets to others in different locations. This globalization is clearly seen in football, where odds in Europe quickly reflect changes in Asia, as Asian markets are larger even for European matches [25].

Football matches can be fixed [40] in many different ways [45] and may involve managers, coaches, referees or players. Match-fixers make profits by using their influence on the results of sporting events [12]. This manipulation is often linked to the emergence of person-to-person betting exchanges, a new market format in sports betting. The introduction of these exchanges has sparked debate and led to various policy responses, ranging from nuanced regulations to outright bans [60]. But banning betting does not seem practical in the context of the global market. Moreover, live betting is popular among sports lovers who like to bet while watching matches on television or on the internet, increasing their experience [25].

The current situation raises concerns about the threat to integrity posed by the existence of bets being placed. Because in these markets, there is no obstacle to prevent some bets from being manipulated by one or a small number of people.

2.2.2 Impact of Match-Fixing (Gain, Corruptions, Tax Revenue..)

The betting industry has changed significantly, initially with the growth of online platforms and more recently with the growth of mobile technology [49]. If growth in Europe is given as an example, Europe's betting market revenue increased 7.5% in 2021 and 23% in 2022 [21] [23]. Revenue in the online sports betting market is projected to reach \$49.96 billion in 2024. This market is expected to show an annual growth rate of 7.68%, resulting in a projected market volume of \$67.17 billion by 2028. Additionally, the number of users in the online sports betting market is expected to reach 181.9 million by 2028 [62]. These changes have forced governments in Europe and elsewhere to find effective ways to regulate and tax the expanding online betting industry [49].

Instances of betting-related match-fixing have been attracting increased interest [17]. Manipulation of competitions for financial gain is an important problem in the field of football. This problem poses a potential threat to the football industry, primarily by deterring football fans, sponsors and broadcasters [25].

The existence of transnational criminal organizations poses a challenge in the fight against corruption in sports. These organizations have evolved from local, close-knit groups to international organizations that use a variety of operational strategies and engage in a variety of legal and illegal activities. They take advantage of the weaknesses in legal systems, the liberalization of borders, and the expansion of free trade brought about by globalization [40]. Moreover, limiting access to the betting platform to a single supplier can create opportunities for monopoly profits, some of which could potentially benefit football [27].

In various jurisdictions around the world, governments are allowing forms of betting while addressing the concerns of those who view betting negatively. They do this by transferring tax revenues from

betting to sectors with a more positive public image, such as education, health, culture and especially sports [27].

In the past, some match-fixing scandals have occurred [34] [24] [22], and a number of large and/or high-profile betting scandals have affected football.

In 2009, the largest European football betting scandal to date was uncovered by an international investigation led by the German Public Prosecution Department in Bochum. Around 200 games across nine countries were implicated in this significant match-fixing scandal, with cash and property worth more than €1 million (£900,000) seized. At least 200 individuals, including 32 players, are suspected of being involved [32].

Perhaps the most high-profile match-fixing scandal until now was seen in Turkey in July 2011, where 93 officials and players were indicted on charges ranging from match-fixing to the payment of bribes in relation to 19 matches [7]. Such scandals threaten the integrity of the sport, football officials have implemented significant measures to prevent further incidents [14]. Unfortunately, these measures may not always be successful.

In countries with lax legal enforcement, match-fixing investigations can face obstacles. For instance, in Turkey, a prominent media personality [69] with close ties to a reputed money launderer acquired television channels and football clubs in countries like England (Hull City), Ireland (Shelbourne F.C.), and Slovenia (NK Maribor) in a short period of time. This situation exemplifies potential corruption.

2.2.3 Match-Fixing Investigations

It is necessary to fight against corruption through match-fixing in every field. It is an indispensable requirement to make regulations at the national or international level regarding the issue of fixing. In addition, departments to fight against this corruption can be established by civil authorities.

The international nature of many conspiracies requires global law enforcement cooperation. FIFA has allocated funds to Interpol for a special unit to combat corruption in sport. At the national level, laws must be clear and enforceable so that criminals can be prosecuted, but this is not always the case. Like other sports, football aims to protect themselves by lobbying legislators and regulators to limit their activities. If successful, this could reduce the number of overtures from fixers to players and officials [25].

To combat corruption in football, UEFA implemented its Betting Fraud Detection System (BFDS) at the start of the 2009 season. This system monitors all matches organized by UEFA, as well as games in the top two domestic divisions and cup competitions of member national associations. Approximately 29,000 matches are scanned annually for potentially suspicious betting patterns. Upon detecting irregularities, the BFDS generates an initial report, marking the first step in subsequent investigation phases [66].

Monitoring betting could be useful to detect the anomalies on match-fixing. Sportradar's Integrity Services has developed cutting-edge technology to detect betting-related match manipulation, supported by our AI-driven Universal Fraud Detection System (UFDS). Designed for sports federations, leagues, and governmental authorities, the UFDS is part of a comprehensive suite of products aimed at combating match-fixing [61].

Monitoring betting activity is a key component of anti-corruption efforts, although the information gathered is limited. In the dominant Asian market, surveillance can only track price movements, while European surveillance has traditionally focused on transaction volumes. Nevertheless, monitoring has gained credibility, with alerts increasingly seen as strong indicators of match-fixing. For example, evidence from Sportradar's CEO was admitted in the Bochum trial. However, the system's sensitivity is low, with many false negatives; only six out of nearly 200 monitored matches identified as fixed in

the Bochum trial had triggered alerts. This low sensitivity may be inherent, as even large bets in a highly liquid market may not cause significant odds movements [25].

Signs of match manipulation often appear as unusual patterns in data from betting markets, as those involved make financial gains by betting on matches with predetermined outcomes [27]. Every match-fixing has similar patterns [34].

Matches with an unusually large bet volume of bets or unusual betting patterns might be fixed, indicating that the bets were placed by the match fixers [12]. The majority of reports are related to football, which also generates by far the largest betting volumes [31].

Alerts do not necessarily indicate that a match-fixing incident has definitely occurred with absolute certainty. For instance, during the reporting year in 2023, Gespa reported one case to law enforcement to investigate potential criminal offenses in Switzerland. However, as far as can be determined, there were no convictions for match-fixing [31]. Therefore, while alerts serve as important signals for potential wrongdoing, conclusive evidence and legal outcomes are required to confirm the occurrence of match-fixing incidents.

2.3 Exploiting Historical Data in Match-Fixing Forensic Analysis

2.3.1 Role of Historical Data

Historical data is crucial in forensic analysis, particularly in digital forensic intelligence (DFINT) and open-source intelligence (OSINT) for football match-fixing investigations.

By examining historical digital forensic data, patterns and anomalies can be revealed. Detecting and stopping these anomalies is vital to protecting the future of sports [44]. Using open sources alongside historical data helps identify trends and patterns, providing essential context and background for current events.

Investigating online betting patterns is also important. Understanding how people place bets, the types of bets they make, and the results of those bets can shed light on betting behaviors [28]. Moreover, historical data can highlight trends in match-fixing and allowing for the identification of unusual betting behaviors and patterns, can help investigators identify links to financial gains in the sports betting market. This comprehensive information assists decision makers by providing comprehensive information on potential future scenarios as well.

Overall, historical data plays a vital role in forensic analysis by revealing historical trends, helping in anomaly detection, and supporting the investigation of complex match-fixing cases.

2.3.2 Leveraging Historical Data to Detect Anomalies

Fundamentally, the historical data in this study refers to historical digital forensic data [50]. This data is useful for detecting anomalies, which are patterns that differ from what is expected and are considered outside the ordinary.

Anomalies indicating suspicious activity may suggest match-fixing. Detecting match-fixing aims to find and stop actions that harm the fairness of sports. Anomalies are important for finding match-fixing because they can appear as strange betting patterns, unexpected changes in performance, or odd player actions. These anomalies act as warning signs of possible match-fixing, and identifying them is crucial. By looking at these anomalies, we can find cases of manipulation and take steps to keep sports competitions fair [44]. For instance, significant changes in betting amounts within a short time, such as a 20% increase or decrease in the amount of money wagered within 10 minutes.

Different from possible betting [3], sudden and drastic changes in betting odds, especially if they deviate significantly, can be seen as anomalies as well for instance.

Some information emerging from historical data allows us to detect anomalies related to match-fixing in football, in general, as follows:

- Data Anomalies: Identification of data anomalies or irregularities that may indicate manipulation or tampering with betting data. For instance, one market type having significantly larger data size compared to other types, such as the *Under 2.5 Goals* market being significantly larger than all other markets in a specific football match.
- Suspicious Betting Behavior: Anomalies in betting behavior, such as a large number of bets being placed on unexpected outcomes or sudden spikes in betting activity on a particular event. For instance, for the *half_time_score* market type, if the match score was 0-0 until the 38th minute, and for a 1-0 score, there were sudden increases in betting activity between the 40th and 45th minutes and the score became 1-0 at the 44th minute.
- Unusual Payouts: Anomalies in payouts, such as unusually large payouts for low-probability events or a high number of winning bets on suspicious outcomes.
- Inconsistent Betting Patterns: Discrepancies between betting patterns observed in the data and typical betting behavior for similar events or matches.
- Pattern of Previous Suspicious Activity: Identification of recurring patterns or similarities between current data and past instances of known fixed games or suspicious activity.
- Behavioral Analysis of Bettors: Examination of the behavior and characteristics of bettors involved in placing suspicious bets, such as multiple accounts being used by the same individual or a high volume of bets placed by a single bettor.

It was not possible to obtain data throughout the study to discuss *Pattern of Previous Suspicious Activity* and *Behavioral Analysis of Bettors* options. !!!

3 Methodology and Implementation

3.1 Methodology

Betfair is a British betting company established in 2000. It operates the world's largest online betting exchange and offers a range of products including sports betting, online casino, online poker, and online bingo. The company's business operations are managed from its headquarters in London, with additional offices in Ceuta, Dublin, Leeds, and Malta. [70]. This website offers API and historical data that include events (football competition) from the different countries and leagues. API needs to be obtained API key and session key to be executes after be logging in.

Historical data only needs to be logging in to obtained these data. 3 kinds of data: Basic, Advanced and Pro. The differences between them is the time frequency and some column information such as *totalmatched*.

Gathering Betting Data: To achieve the data collection, an automated methodology will be developed. This approach not only facilitates the repetition of data collection at different points in time but also effectively handles the large, complex datasets, streamlining the processes of data acquisition, processing, analysis, and anomaly detection.

From the three available plans (BASIC, ADVANCED, and PRO), useful plan is chosen. Among the various sports types, soccer is chosen.

Converting File to a Readable Format: The Historic Data Processor page is used to convert files from the *BZ2* format to the *CSV* format. For each event, the relevant file from the *event files* folder is uploaded to this web page. During conversion, only the necessary columns are selected. Only in-play data is included, not pre-play data. A suitable time frequency is chosen for creating graphs and for the detection part.

Processing Data: During the processing phase, generally, the CSV files were loaded into a pandas DataFrame, making the desired modifications, and then saving the modified DataFrame back to a CSV file.

Creating Timelines for Specific Events: Since there is no information related to match events such as goals, yellow cards, red cards, and their corresponding minutes during the match, this information, along with winner charts for each market type as shown in Table 23, was obtained from another web page. It is possible to observe some exceptional match events, such as goals scored in the final seconds.

Creating Graphs to Visualize Data: To visualize the data and serve as a crucial part of the analysis, graphs are created for each market type and market type option.

Utilizing Anomaly Detection Criteria: To detect potential anomalies during match events, specific patterns are established as criteria. These patterns serve as benchmarks for identifying irregularities that may occur throughout the matches.

Generating the Report: The findings were documented and saved in a comprehensive file.

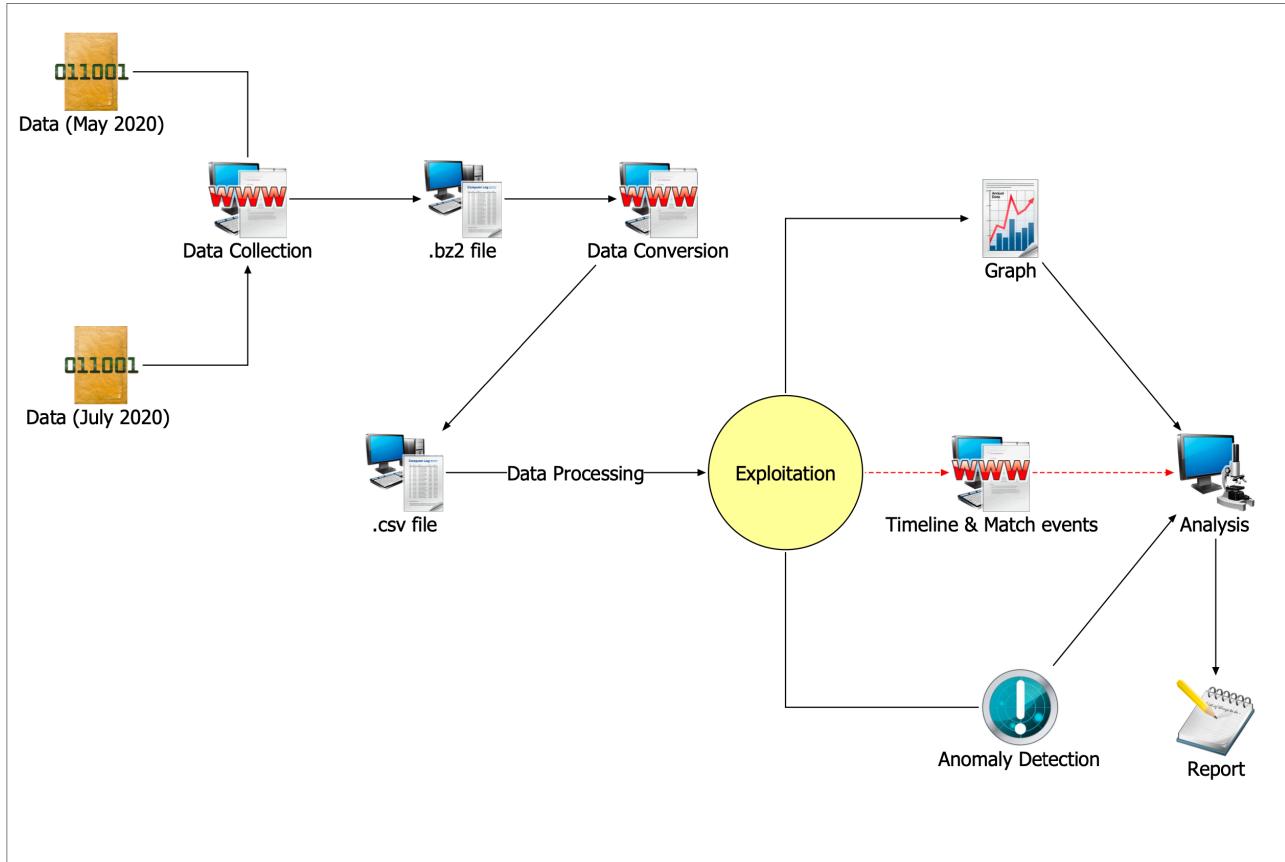


Figure 1: Scheme

3.2 Implementation

The implementation of the methodology outlined above and on *Figure 1* involved several steps, including data collection, data conversion, data processing, data visualization and anomaly detection. All codes created can be found in this link¹.

To be ready to collect data, an account must be created as logging in is required to access the historical data page. An API key and session token were also obtained in preparation to collect the data, accommodating any additional requests that might be necessary for further investigation.

Due to the large amount of information that needed to be obtained and analyzed, and the significant time required to do this using an API, it was preferable to use historical data for this project.

3.2.1 Data Collection

An account was created on the required platform, and an API key along with a session token was obtained to facilitate data collection. The *selenium* tool was used to automate the data collection process.

Among the three available plans (BASIC, ADVANCED, and PRO), the ADVANCED plan was chosen for its comprehensive and adequate information.

Examiners often fail to obtain data in a forensically accurate and intact manner, or once they obtain it, they fail to process it to completion [30]. Therefore, the data was obtained from official sources.

Data from May 2020 (Advanced Plan, soccer) was collected for free from Betfair.com. This dataset was used only in the testing phase to develop and improve analysis methods. Additionally, Betfair, in collaboration with UEFA, provided data from July 2020 for the Soccer Advanced Plan. This data set was used to determine the results and conclusions of the study.

Even though it may not be directly useful for this project, a timeline including match events as a *PNG* file, match events such as goals and cards in a *TXT* file, and the winner chart for each market type as a *PNG* file were created to obtain information related to football matches from the last 2 years 25

3.2.2 Data Conversion

Historical data for each football match is organized into folders 24, with each folder generally containing 25 files in *BZ2* file format. The structure of the files as follows:

- EventID File: One of these files is named with the match ID (marketId). This file contains all-inclusive data about all specific match bet types.
- MarketId Files: The remaining files are named with bet type IDs (marketId). These files detail the various types of bets placed on the match.

The exact number of files in each folder can vary depending on the number of market types 23 placed. Larger matches with more bet types will have more files, while smaller matches with fewer bet types will have fewer files.

Historic data processor page was accessed using *selenium*. For each event, only one file with *EventId* name from 25 files that contains all market information was uploaded to be converted. These event files have the same name as the folder they are in, and they are the largest files in their respective folders.

¹All codes created can be found at <https://github.com/mahomut41/betfair>

When converting the historical data to *CSV* file format, specific columns were selected in the data processor. The selected columns include:

marketId, publishTime, md.eventName, md.name, numberActiveRunners, numberWinners, status, totalMatched, selection_selectionId, selection_lastPriceTraded, selection_md.name, selection_ex.availableToBack.price, selection_ex.availableToBack.size, selection_ex.availableToLay.price, selection_ex.availableToLay.size, selection_status, selection_totalMatched

From Output preplay data and Output inplay data, only *Output inplay data* was selected. As the time frequency, *60s* was selected to be converted to *zip* format file.

Each event's relevant data, encompassing market and selection details, was compiled into separate *zip* files. Generally, 24 *zip* file with *MarketId* names were created for each football competition. Then these zip files were transformed into *CSV* format for ease of processing and saved in a folder named *team1_team2..*

3.2.3 Data Processing

During the processing phase, the CSV files were loaded into pandas DataFrames. Various modifications were applied to the data, including:

- The columns *totalMatched, selection_ex.availableToBack.price, selection_ex.availableToBack.size, selection_ex.availableToLay.price, selection_ex.availableToLay.size, and selection_totalMatched* were converted to the float data type, with non-numeric values handled appropriately.
- The column *publishTime* was converted to date time format
- The *percent_money_on_market* column was created by dividing the *selection_totalMatched* values by the *totalMatched* values. This column will display the ratio of the market type option within the overall market type.
- The *Minute* column was created by defining the start time and then subtracting the start time from *publishTime*. This column will display the match minutes up to the end of the match.
- The column *StatusMatch* was created and in this column end of match was defined.

Processing times are increasing with the increase in the amount of data required to be analysed [50].

3.2.4 Creating Graphs

To analyze these large data sets [29] [39], graphical representations were generated for each market type and option. These visual insights into the data serve as vital components of the analysis process.

From each market (bet), 3 kinds of visualisation created to assist analysis phase:

- Percent_money_graphs
- Total_money_graphs
- Price_and_Volume_graphs

The *Percent-money-graph* shows the distribution of the total money placed across different outcomes on each bet type option during football matches, until the matches end and bets are closed. One graph was created for each market type, with additional graphs for each market type option indicating WINNER or LOSER labels 41. The vertical axis shows the percentage of total money placed on each bet type option. The horizontal axis represents time.

The *Price_and_Volume_graph* displays the market type option price and size over time during the match. One graph was created for each market type option 42. *Price and Volume Graphs* show the relationship between the betting odds (price) and the volume of money bet over time. The prices (odds) are depicted on the vertical axis. Lower odds indicate a higher probability of the event happening (e.g., home team winning). The volume of money bet at each price point is typically shown with bars or shading behind the price line. Higher bars indicate more money placed at that price.

The *Total_money_graph* shows the changing of the total amount of money placed on each bet type option during football matches. Similar to the *Percent_Money_graph*, one graph was created for each market type, with additional graphs for each market type option indicating WINNER or LOSER labels 43. The vertical axis represents the cumulative amount of money placed for bet type options. The horizontal axis shows the progression of time from the start of the match.

0 and *NaN* values are excluded for each kind of graph.

In each market (football match bet), a specific graph is created for each option in the bet type. The number of these graphs depends on the options within each bet type. For instance, the *DoubleChance* bet type requires three graphs to be created, but the *CorrectScore* bet type needs several graphs, depending on the goals scored. For a 0-2 final score, there would be 19 options, so 19 bets could be placed. The reason for creating a graph for each option is to provide more detailed information if necessary.

For each bet type, to see all bet type options together on one graph, an all-inclusive graph was created. Therefore, the number of all-inclusive graphs is the same as the number of bet types. Legends were created on the graphs to indicate each bet type option. For the *Price_and_Volume_graph*, it was not possible to create an all-inclusive graph because there are already plots and bars on the graph, leaving no place to add another option.

Legends created on graphs to mention each bet type option.

The time interval was not set to a specific interval, so the time axis on the graphs will adjust according to the amount of information in the *CSV* files. This was automated because some files contain varying amounts of information, making it impractical to set a fixed interval of 3 or 5 minutes for all graphs.

3.2.5 Creating Criteria

Criteria were created to detect anomalies. Totally, 8 criteria were created as shown in Table 1.

This implementation facilitated a systematic approach to data collection, processing, and analysis, providing valuable insights into the market dynamics and match events. Moreover, these criteria, defined as specific patterns, aided in the detection of irregularities throughout the matches.

3.2.6 Creating Report

The findings were documented and saved in a comprehensive database file, serving as a report.

The database file named *Anomaly_detection_results* was generated to identify potential anomalies during match events. This report comprises the name of the criteria found, the name of the market, the name of the market option, the winner or loser status, and details about the finding.

No	Criteria
1	Percent of money on market is above 95%, total matched exceeds 10,000, and selection price is above 4.
2	Throughout the match, excluding the last 5 minutes, the total matched exceeds 11,000, and the selection price is above 1.3.
3	During both halftimes, excluding the last 5 minutes, the percentage of money in the market is higher than 80%.
4	There is at least a 25% sharp increase during the last 10 minutes of both halftimes, excluding the last 5 minutes.
5	There are big and sharp bets at the end of the game, with a change factor of more than 3x in the last 10 minutes, excluding the last 3 minutes.
6	The price sharply increases, for instance, from 1.75 to 3 in 10 minutes, and more than 25% of the money leaves the market.
7	In the first 10 minutes, there is a disproportion with the percentage of money between 60% and 70%, the price is above 3, and the total matched money is above 15,000.
8	The selection price is above 1.5, the percentage of money is above 98%, and the total matched is above 10,000.

Table 1: Criteria

4 Advantages and Limitations

4.1 Advantages

For the Betfair website’s developer mode, creating an account is sufficient; account verification is not required to obtain the API key and session token.

Thanks to the website’s reliable infrastructure (Betfair Historical Data Processor), especially during the data conversion, the enormous amount of existing data [59] was converted reliably without issues such as data loss or corruption.

The chosen methodology employs automation for data conversion, processing, visualization, and detection, enabling the process to be repeated efficiently. The effectiveness of automation in these processes has ensured that the datasets required to detect anomalies and analyze the detection results are obtained in a standard format. This methodology helps save time and human resources by avoiding manual data exploitation.

Structured data was manipulated with Pandas, a Python Data Analysis Library, thus facilitating data processing.

All datasets obtained using Matplotlib, a library for visualization in Python, were visualized by creating lineplot and barplot to facilitate data analysis and to be used in verifying anomaly detection results.

The data was made available for free thanks to UEFA. UEFA’s contribution helped in understanding the problems related to match-fixing cases and identifying the key areas to focus on.

4.2 Limitations

To visit Betfair.com and its subpages, a Virtual Private Network (VPN) needs to be used. This is because access from certain countries where bets are not accepted, including Switzerland, is restricted by Betfair. When an attempt is made to access Betfair.com from Switzerland, the connection is blocked by the website due to the identification of the location based on the IP address.

Even if an account is successfully created using a VPN, difficulties will be encountered during the verification process. All users are required to have their accounts verified by Betfair. The verification process requires that proof of address, which must be from the country where Betfair operates, be provided. Verification usually involves the submission of a *utility bill (less than 6 months old)* or a *copy of a bank statement (less than 6 months old)*.

As for the automation process, a *sleep time* was necessary in order to execute their corresponding codes correctly. As web page loading times, saving files or manipulate data can vary considerably, the predefined sleep time can occasionally result in long waste of time.

It is not possible to convert files with a file size of 5MB or larger in the data processor, because the website does not convert files larger than this size. Additionally, some files do not contain enough information for analysis and some files cannot be converted at the conversion stage because they contain bad lines. These have caused the available historical data not to be analyzed 100%.

Data analysis can be difficult due to reasons such as format incompatibilities or simply lack of training. Due to data management issues, even analyzing the data can take a long time before review [30]. All operations in this study were tested once on the datasets of May 2020 until the results were obtained. Since the study was conducted on the datasets of July, the data were in different quantities and the contents of some of them were different, requiring modifications on the created codes, which caused loss of time.

Since the datasets belong to the year 2020 and only the timeline and market events information related to football match competition for the last two years can be obtained with the created codes - codes have been created for further purposes anyway - but timeline and event information could not be added to the results of this study.

5 Feasibility

Historical Data Processor website were chosen specifically for their accessibility as they don't require any Captcha, which facilitates the automation of data conversion.

Datasets and created plots were selected from only *inplay* market types information to make visualisation better. However, it restricts us to see the bets before the football match competitions.

Since choosing the criteria and number of criteria is chosen completely subjectively, naturally the results obtained will also be subjective. Although this detects most anomalies, it is not clear whether it will detect anomalies in specific match-fixing events.

In addition, since data is only from one website and only July 2020 data was examined, the amount of data collected is kept at a manageable level. However, it restricts us from conducting a definitive results since the sample datasets might not be representative of the betting market.

6 Results

Historical data was collected on football matches played during two different months:

- May 2020: Data was obtained for a total of 665 football matches.
- July 2020: Data was collected for a total of 3723 football matches.

Additionally, 63 files contained no enough information, and 35 files were too large to convert, so they could not be converted.

In a converted CSV file, there are 20 columns as explained in terminology, data conversion and data

processing parts and is shown in Figure *Sample CSV file processed 44*. The number of the rows depend on the market type and market type options.

According to anomaly detection results, there is no market option to match with 2nd, 3rd, and 5th criteria!

Only the 4th, 6th, or 8th criteria were not accepted as anomalies. These criteria alone cannot define if there is an anomaly; they rather promote other criteria such as the 1st, 2nd, and 3rd ones.

According to the criteria defined, CSV files processed and the graphs 19 market/market options of the events were found suspicious:

No	Criteria	Event	League Name
1	4, 6, 7	Atlanta Utd vs New York Red Bulls	USA Major League Soccer
2	4, 6, 7	Betis vs Villarreal	Spain La Liga
3	4, 6, 7	Cagliari vs Udinese	Italy Serie A
4	4, 6, 7	Entella vs Perugia	Italy Serie B
5	4, 6, 7	Haugesund vs Rosenborg	Norway Eliteserien
6	1, 4, 6, 7	Henan vs Jiangsu Suning	Chinese Super League
7	4, 7	Huddersfield vs Preston	English Football League
8	4, 6, 7	Hull vs Middlesbrough	English Football League
9	4, 6, 7	Juve Stabia vs Entella	Italy Serie B
10	4, 6, 7	Oita Kobe	Japan J1 League
11	4, 6, 7	Pescara vs Frosinone	Italy Serie B
12	4, 6, 7	Real Salt Lake vs Kansas City	USA Major League Soccer
13	4, 6, 7	Rio Ave vs Portimonense	Portugal Primeira Liga
14	4, 6, 7	Rubin Kazan vs Rostov	Russia Premier League
15	1, 4, 8	Sampdoria vs AC Milan	Italy Serie A
16	1, 4, 6, 8	Vikingur Reykjavik vs Stjarnan	Iceland Cup
17	4, 6, 7	Vikingur Reykjavik vs Valur	Iceland Premier League
18	4, 6, 7	Watford vs Norwich	England Premier League
19	4, 6, 7	Yokohama FM vs FC Tokyo	Japan J1 League

Table 2: Detected Events

Henan vs Jiangsu Suning (Chinese Super League)

Date: 26 July 2020

Score: 3-4

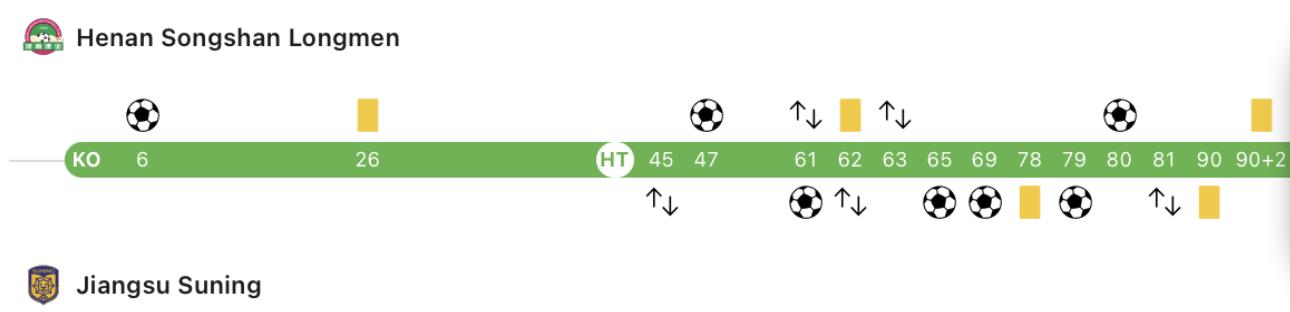


Figure 2: Match Timeline (Henan vs Jiangsu Suning)

Even if in *Double Chance*² market, *Draw or Away* market option has more possibility to gain (2/3 possibility), in this match until the 65th minute, there is less possibility for a *Draw or Away* option. Between the 65th and 79th minutes, the away team scored 3 goals and this odd came to be certain.

²All market types are shown in Table 23: Football Betting Market Types on Betfair

Criteria	Market	Market option	Details
1. criteria	DoubleChance	DraworAway	Percent of money on market is above 95%, total matched exceeds 10,000, and selection price is above 4.
1. criteria	DoubleChance	DraworAway	Percent of money on market is above 95%, total matched exceeds 10,000, and selection price is above 4.
1. criteria	DoubleChance	DraworAway	Percent of money on market is above 95%, total matched exceeds 10,000, and selection price is above 4.
4. criteria	HalfTimeScore	1 - 0	Percentage change between minute 11 and 20: 55.04%
4. criteria	DrawnoBet	Jiangsu Suning	Percentage change between minute 4 and 5: 40.92%
4. criteria	HalfTime	Henan	Percentage change between minute 4 and 6: 38.34%
4. criteria	HalfTime	Henan	Percentage change between minute 6 and 9: 91.53%
4. criteria	FirstHalfGoals1.5	Under1.5Goals	Percentage change between minute 5 and 6: 60.87%
6. criteria	FirstHalfGoals1.5	Under1.5Goals	The odds raised in 10 minutes and 67.38% of the money left the market between minute 9 and 15
7. criteria	MatchOdds	Henan	In first 10 minutes, there is a disproportion of bets (percentage: 66.45% and price: 3.75)
7. criteria	MatchOdds	Henan	In first 10 minutes, there is a disproportion of bets (percentage: 66.33% and price: 3.50)
7. criteria	MatchOdds	Henan	In first 10 minutes, there is a disproportion of bets (percentage: 65.78% and price: 3.45)
7. criteria	MatchOdds	Henan	In first 10 minutes, there is a disproportion of bets (percentage: 63.62% and price: 3.35)

Table 3: Anomalies (Henan vs Jiangsu Suning)

In this market, 3 times more than \$10,000 was matched, with more than 95% of the money on this option and the price was above 4 45.

In the *Double Chance* market, the *Draw or Away* option has a higher probability of winning (2 out of 3). However, in this specific match, until the 65th minute, the likelihood of a *Draw or Away* outcome was low. Between the 65th and 79th minutes, the away team scored 3 goals, making the *Draw or Away* option almost certain to win. In this market, more than \$10,000 was matched three times, with over 95% of the money placed on this option. The odds for this option were above 4.

For the *1-0* market option in the *Half Time Score* market, there is no suspicious activity; an increasing percentage for the actual half-time score can be considered normal. Nonetheless, a more than 50% increase at the 6th minute is notably high 46.

In the *1-0* market option for the *Half Time Score* market, there is nothing suspicious about the increasing percentage reflecting the actual half-time score. However, between the 11th and 20th minutes, an increase of more than 50% is unusually high.

In the *DrawnoBet* market for the *Jiangsu Suning* option, a 40% increase before a goal could be considered normal if the amount of money placed was not too high.

In the *DrawNoBet* market for the *Jiangsu Suning* option, a 40% increase in the percentage of money

placed on *Jiangsu Suning* before a goal is scored can be considered normal, provided that the total amount of money placed is not excessively high. This implies that the change in betting volume is within expected ranges for this type of market 47.

In the *Half-Time* market for the *Henan* option, a 38% increase in the percentage of money placed on Henan before their goal is observed. However, after Henan scores a goal, between the 6th and 9th minutes, the percentage of money placed on Henan rises by more than 90%. This significant increase in betting volume shortly after the goal is noteworthy and could indicate a strong reaction from bettors, suggesting heightened confidence or a shift in expectations for the match outcome 48.

In the *First Half Goals 1.5* market, where *Under 1.5 Goals* is the option, a 60% increase in the percentage of money placed on this option between the 5th and 6th minutes can be considered normal, provided that the amount of money placed before this period was relatively low. This means that, depending on the initial amount of money wagered, a 60% increase in betting volume is expected and falls within normal fluctuations 49.

In the *Match Odds* market for the *Henan* option, there were four instances detected within the first 10 minutes where a disproportionate amount of bets were observed. The percentages of money placed ranged between 66.33% and 63.62%, while the odds varied between 3.75 and 3.35. The frequency and nature of these betting patterns within the initial 10 minutes of the match indicate an unusually high level of interest.

Match stats show that Jiangsu Suning had almost twice the ball possession compared to the home team throughout the match. These stats cover both halves, so they do not provide specific information about the first half 26. Interestingly, Henan made three shots on goal, scoring with all three, while Jiangsu Suning scored four goals from seven shots on goal.

Sampdoria v AC Milan (Italie Serie A)

Date: 29 July 2020 Score: 1-4

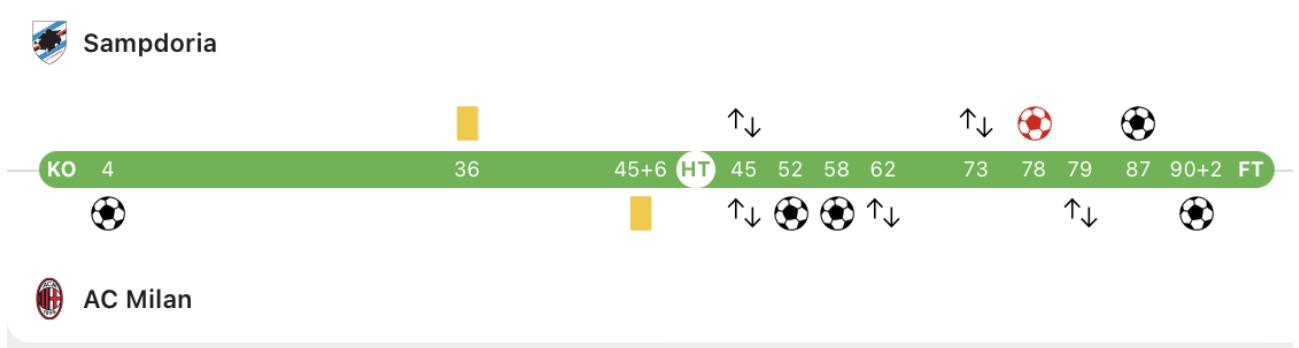


Figure 3: Match Timeline (Sampdoria v AC Milan)

Almost all anomaly detections are related to the *Match Odds and Both Teams to Score* market, except for two instances.

In *Match Odds and Both teams to Score* there market for *ACMilanYes* option, 4 times more than 10000 USD money placed when Percent of money on market is above 95% and price (odds) is above 4. Since AC Milan football match is very popular, it is normal to 10000 usd money placed, when price is above than 4, percentage money is above 95% and 4 times , it takes attention. There is too much gain .

In the *Match Odds and Both Teams to Score* market for the *AC Milan Yes* option, more than 10,000 USD was placed four times when the percentage of money in the market was above 95% and the price (odds) was above 4. Since AC Milan matches are very popular, it is normal for large amounts of money to be placed when the odds are above 4 and the percentage of money in the market is above

95%. However, the fact that this happened four times is noteworthy due to the significant potential gains.

In the *Correct Score* market for the *Any Other Away Win* option and the *Half Time Score* market for the *0 - 1* option, bets were placed after the goal occurred. These instances happened only once and thus do not appear very interesting as anomalies.

In the *Match Odds and Both Teams to Score* market for the *Sampdoria Yes* option, bets were placed a total of 78 times. The frequency of these bets is important when evaluating this market option; however, the number of wins and the amount of money gained are particularly noteworthy. For a win, both teams must score at least one goal. In this case, Sampdoria scored their goal in the 87th minute, after missing a penalty in the 78th minute.

In the match Stats, there are no notable anomalies. AC Milan had slightly more ball possession than the home team and successfully scored 4 goals from 9 shots on target²⁷.

Criteria	Market	Market option	Details
1. criteria	Match Odds and Both teams to Score	ACMilanYes	Percent of money on market is above 95
1. criteria	Match Odds and Both teams to Score	ACMilanYes	Percent of money on market is above 95
1. criteria	Match Odds and Both teams to Score	ACMilanYes	Percent of money on market is above 95
1. criteria	Match Odds and Both teams to Score	ACMilanYes	Percent of money on market is above 95
1. criteria	Match Odds and Both teams to Score	ACMilanYes	Percent of money on market is above 95
1. criteria	Match Odds and Both teams to Score	ACMilanYes	Percent of money on market is above 95
1. criteria	Match Odds and Both teams to Score	ACMilanYes	Percent of money on market is above 95
1. criteria	Match Odds and Both teams to Score	ACMilanYes	Percent of money on market is above 95
1. criteria	Match Odds and Both teams to Score	ACMilanYes	Percent of money on market is above 95
4. criteria	CorrectScore	Any Other Away Win	Percentage change between minute 19 and 37: 43.38%
4. criteria	HalfTimeScore	0 - 1	Percentage change between minute 9 and 18: 108.56%
8. criteria	Match Odds and Both teams to Score	SampdoriaYes	Win rate is greater than 1.5, percentage of money is greater than 98%, and total matched is more than 10,000 (78 times)

Table 4: Anomalies (Sampdoria v AC Milan)

Vikingur Reykjavik vs Stjarnan (Iceland Cup)

Date: 25 July 2020 Score: 1-2

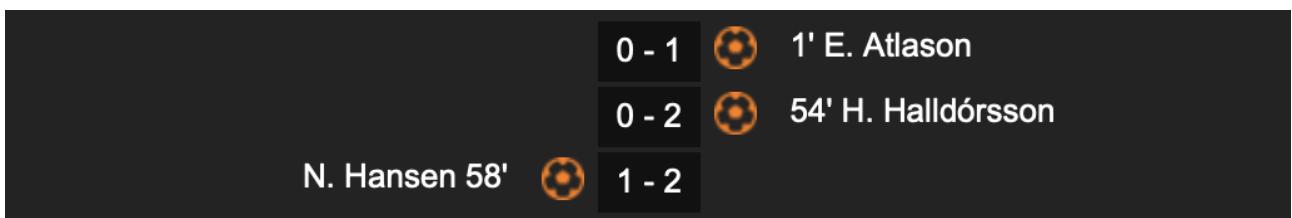


Figure 4: Match Timeline (Vikingur Reykjavik vs Stjarnan)

Asian Handicap works as follows: In this market, if Vikingur Reykjavik had a +1.5 handicap, they would be given an additional 1.5 goals. The adjusted score would then be Vikingur Reykjavik 2.5 - 2

Stjarnan. As a result, Vikingur Reykjavik would win the bet because 2.5 is greater than 2.

The data does not specify whether the Asian Handicap is +1, +1.5, or another value. However, since Vikingur Reykjavik is indicated as the winner, the handicap could be +1.5. If the handicap were +1, the money would be refunded, but in the data, there are only two options: loser and winner.

Given that the money is not refunded and Vikingur Reykjavik is indicated as the winner, it implies a gain of more than 10,000 USD on 47 occasions when the price was above 4 and the percentage of money was 95%. This is highly suspicious due to the substantial gain. If the goal difference were 2, it is quite plausible that the bettor would lose, making this situation extremely risky for bettors.

In the *Double Chance* market for the *Draw or Away* option, the change in the percentage of money between the first and second minutes, given that the away team scored in the first minute, is quite normal.

Criteria	Market	Market option	Details
1. criteria	AsianHandicap	VikingurReykjavik	Percent of money on market is above 95%, total matched exceeds 10,000, and selection price is above 4. (47 times)
4. criteria	DoubleChance	Draw or Away	Percentage change between minute 1 and 2: 78.38%
4. criteria	FirstHalfGoals1.5	Under 1.5 Goals	Percentage change between minute 1 and 18: 1910.24%
4. criteria	FirstHalfGoals1.5	Under 1.5 Goals	Percentage change between minute 18 and 20: 95.93%
4. criteria	OverUnder5.5Goals	Under 5.5 Goals	Percentage change between minute 17 and 19: 29.51%
4. criteria	HalfTimeScore	0 - 1	Percentage change between minute 27 and 36: 3036.30%
4. criteria	MatchOdds	Stjarnan	Percentage change between minute 1 and 19: 98.36%
4. criteria	MatchOdds	Stjarnan	Percentage change between minute 19 and 22: 121.74%
4. criteria	FirstHalfGoals2.5	Under 2.5 Goals	Percentage change between minute 18 and 20: 165.61%
6. criteria	OverUnder3.5Goals	Under 3.5 Goals	The odds raise in 10 minutes and 36.33% of the money left the market between minute 91 and 97
6. criteria	FirstHalfGoals2.5	Under 2.5 Goals	The odds raise in 10 minutes and 66.70% of the money left the market between minute 20 and 27
8. criteria	AsianHandicap	VikingurReykjavik	Win rate is greater than 1.5, percentage of money on market is greater than 98%, and total matched is more than 10,000. (5 times)

Table 5: Anomalies (Vikingur Reykjavik vs Stjarnan)

After a goal is scored in the first minute, the increase in the percentage of money placed on the *Under 1.5 Goals* option in the *First Half Goals 1.5* market is noteworthy. This is interesting because the probability of winning this bet decreases with each goal scored. If another goal is scored by either team, the bettor would lose the bet.

In the *Over/Under 5.5 Goals* market, the percentage chance for the *Under 5.5 Goals* option between the 17th and 19th minutes when score is 0-1, showing a change of 29.51%, is not unusual if the amount

of money placed is not too high.

In the *Half Time Score* market for the *0 - 1* option, betting on the actual score appears quite normal. However, a 3036.30% change in the percentage of money between the 27th and 36th minutes is highly suspicious.

In the *Match Odds* market for the *Stjarnan* option, a significant change in the percentage of money during the first half is notable, with a 121.74% change observed between the 19th and 22nd minutes.

In the *First Half Goals 2.5* market for the *Under 2.5 Goals* option, a 165.61% change in the percentage of money appears interesting, depending on the amount of money placed.

In *OverUnder3.5Goals* market for *Under 3.5 Goals* market option, 36.33% of the money left the market between minute 91th and 97th minutes. This can show sometimes that odds dropping and the spread of these games in closed / open telegram groups, bets on a different market (total under) go very sharply (excapper.com). it is the case in *FirstHalfGoals2.5* market for *Under 2.5 Goals* market option. The odds raise between minute 20 and 27 and -66.70% of the money left the market. It is value that be taken attention. Thanks to that bettor before close the market they leave, they gain close the maximum gain anyway.

In the *Over/Under 3.5 Goals* market for the *Under 3.5 Goals* option, 36.33% of the money left the market between the 91st and 97th minutes. This can sometimes indicate that odds are dropping and information is being shared for instance in closed or open Telegram groups, leading to sharp bets on a different market, as observed on excapper.com. Similarly, in the *First Half Goals 2.5* market for the *Under 2.5 Goals* option, the odds rose between the 20th and 27th minutes, and 66.70% of the money left the market. This significant change is noteworthy. Thanks to bettors leaving before the market closes, they secure their gains, even if not the maximum (An explanatory YouTube video here?). This way, malicious bettors will draw less attention to themselves.

In the *Asian Handicap* market for the *Vikingur Reykjavik* option, even though the price is not very high (above 1.5), the percentage of money in the market is greater than 98%. The amount of money placed is also significant, with more than 10,000 USD being placed 5 times. This situation is noteworthy and warrants attention.

Atlanta Utd vs New York Red Bulls (USA Major League Soccer)

Date : 12 July 2020

Score: 0-1

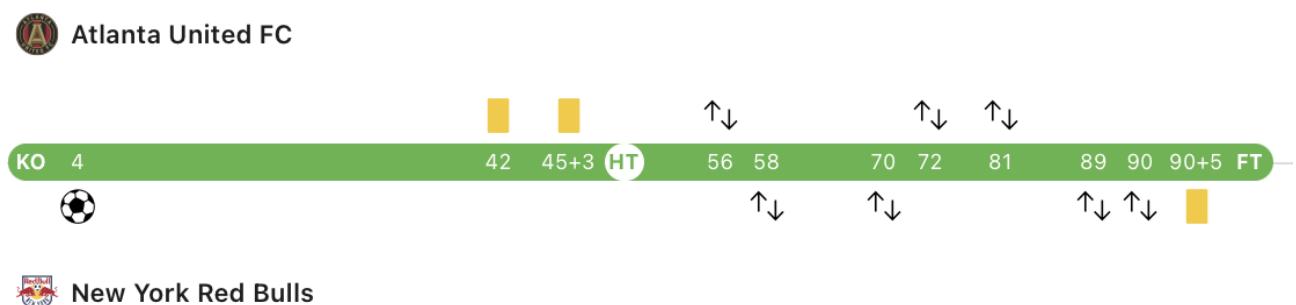


Figure 5: Match Timeline (Atlanta Utd vs New York Red Bulls)

Even with a goal scored in the 4th minute, the percentage of money increased 28.61% between the 4th and 6th minute in the *FirstHalfGoals1.5* market for the *Under 1.5 Goals* option. This goal actually decreases the likelihood of winning this market option.

In the *MatchOdds* market for the *New York Red Bulls* market option, the percentage of money

increased by 70.53% between the 2nd and 4th minute. This is a significant increase. However, since this increase occurred in the 2nd minute, it is not necessarily very suspicious.

In first 10 minutes, there is a disproportion of bets in *OverUnder2.5Goals* market for *Under2.5Goals* market option, the bet placed when money percentage is around 65% and odds is between 3.50 and 3.80. and it placed 6 times. it is surprise to bet for this option even if the score is 0-1 in 4th minute.

In the first 10 minutes, there is a disproportionate number of bets in the *OverUnder2.5Goals* market for the *Under 2.5 Goals* option. The bets were placed when the money percentage was around 65% and the odds were between 3.50 and 3.80, and this happened six times. It is surprising to see bets for this option even with a score of 0-1 in the 4th minute.

Regarding the match Stats, there is nothing noteworthy. 28

Criteria	Market	Market option	Details
4. criteria	FirstHalfGoals1.5	Under 1.5 Goals	Percentage change between minute 4 and 6: 28.61%
4. criteria	MatchOdds	New York Red Bulls	Percentage change between minute 2) and 4: 70.53%
7. criteria	OverUnder2.5Goals	Under2.5Goals	In first 10 minutes, there is a disproportion of bets (percentage: 65.58 and price: 3.50)
7. criteria	OverUnder2.5Goals	Under2.5Goals	In first 10 minutes, there is a disproportion of bets (percentage: 65.59 and price: 3.75)
7. criteria	OverUnder2.5Goals	Under2.5Goals	In first 10 minutes, there is a disproportion of bets (percentage: 63.77 and price: 3.80)
7. criteria	OverUnder2.5Goals	Under2.5Goals	In first 10 minutes, there is a disproportion of bets (percentage: 63.63 and price: 3.75)
7. criteria	OverUnder2.5Goals	Under2.5Goals	In first 10 minutes, there is a disproportion of bets (percentage: 63.69 and price: 3.70)
7. criteria	OverUnder2.5Goals	Under2.5Goals	In first 10 minutes, there is a disproportion of bets (percentage: 63.78 and price: 3.60)

Table 6: Anomalies (Atlanta Utd vs New York Red Bulls)

Betis vs Villarreal (Spain Laliga)

Date: 1 July 2020 Score: 0-2

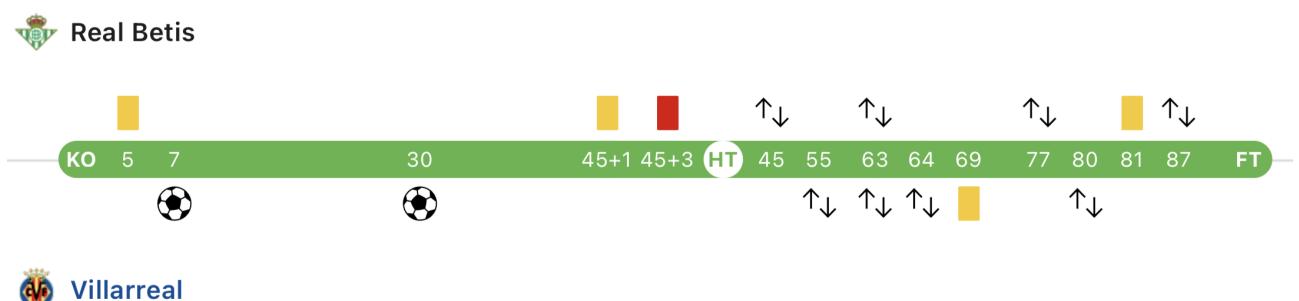


Figure 6: Match Timeline (Betis vs Villarreal)

It could be seen as normal in the *HalfTimeScore* market for the 0 - 2 market option, because there were goals in the 7th and 30th minutes, resulting in a very large increase in the percentage of money, specifically 1267.69%, between the 31st and 40th minutes. However, in the *CorrectScore* market for 0 - 2, a 49.61% change between the 21st and 39th minutes is very interesting because this bet relates to the match score, and there is a significant shift towards a 0 - 2 half-time score. Since it is a big match with high betting frequency, it is very difficult to estimate the half-time score as the final match score

as well.

In the *HalfTime* market for the *Villarreal* market option, a 52.19% increase in the percentage of money in the 5th minute, before the first goal, seems very interesting. It is even more intriguing that there was a 43.97% increase in the same market option between the 5th and 9th minutes. It is not difficult to estimate that after the goal, bets will favor the team that scored the goal.

In first 10 minutes, there are two disproportions of bet in *OverUnder2.5Goals* market for *Under2.5Goals* market option. When the money percentage in market option is 60.0 with the price (odd) is 3.45 and percentage is 60.53 with the price is 4.20, the bets placed. In a match with 2 goals scored and there is a red card in first half time, these bets seems too suspicious if the amount of the money placed is high also.

In the first 10 minutes, there are two disproportionate bets in the *OverUnder2.5Goals* market for the *Under 2.5 Goals* market option. Bets were placed when the money percentage was 60.0 with odds of 3.45 and when the percentage was 60.53 with odds of 4.20. In a match with 2 goals scored and a red card in the first half, these bets related to match score seem too suspicious if the amount of money placed is also high.

According to the match Stats, the away team appears to be more dominant in all aspects compared to the home team. Additionally, the home team, Betis, received a red card at the end of the first half. 29

Criteria	Market	Market option	Details
4. criteria	HalfTimeScore	0 - 2	Percentage change between minute 31 and 40: 1267.69%
4. criteria	CorrectScore	0 - 2	Percentage change between minute 21 and 39: 49.61%
4. criteria	HalfTime	Villarreal	Percentage change between minute 5 and 5: 52.19%
4. criteria	HalfTime	Villarreal	Percentage change between minute 5 and 9: 43.97%
7. criteria	OverUnder2.5Goals	Under2.5Goals	In first 10 minutes, there is a disproportion of bets (percentage: 60.0 and price: 3.45)
7. criteria	OverUnder2.5Goals	Under2.5Goals	In first 10 minutes, there is a disproportion of bets (percentage: 60.53 and price: 4.20)

Table 7: Anomalies (Betis vs Villarreal)

Cagliari vs Udinese (Italie Serie A)

Date: 26 July 2020 Score: 0-1

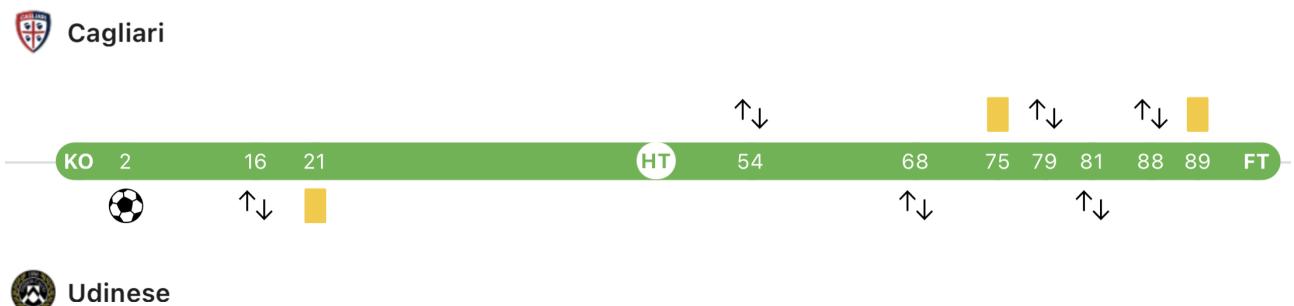


Figure 7: Match Timeline (Cagliari vs Udinese)

In the *HalfTime* market for the *Udinese* market option, an increase of more than 25% in the money percentage after the 2nd minute goal seems quite typical between the 6th and 9th minutes.

In the *HalfTimeScore* market for the *0 - 1* market option, a significant increase of 68.02% between the 20th and 29th minutes is notable, as is an increase of 34.27% between the 30th and 39th minutes.

With a +1 handicap on Cagliari, the effective score for betting purposes in a match where the actual score is Cagliari 0 - 1 Udinese would be Cagliari 1 - 1 Udinese. Similarly, the same logic applies to a *Cagliari +2* handicap.

In the *Cagliari+1* market for the *Draw* market option, a 34.35% increase in the money percentage between the 5th and 7th minutes, following the first goal, is quite risky for bettors. After the goal, the increase in the money percentage for the Draw market option is particularly notable and draws attention.

Criteria	Market	Market option	Details
4. criteria	HalfTime	Udinese	Percentage change between minute 6 and 9: 37.80%
4. criteria	HalfTimeScore	0 - 1	Percentage change between minute 20 and 29: 68.02%
4. criteria	HalfTimeScore	0 - 1	Percentage change between minute 30 and 39: 34.27%
4. criteria	Cagliari+1	Draw	Percentage change between minute 5 and 7: 34.35%
6. criteria	Cagliari+2	Cagliari +2	The odds raise in 10 minutes and 83.25% of the money left the market between minute 42 and 48
7. criteria	OverUnder2.5Goals	Under2.5Goals	In first 10 minutes, there is a disproportion of bets (percentage: between 63.46-61.65 and price: between 3.35-3.25) (9 Times)

Table 8: Anomalies (Cagliari vs Udinese)

In the first 10 minutes, there is a disproportion of bets in the *OverUnder2.5Goals* market for the *Under 2.5 Goals* market option. The bets were placed 9 times, with the percentage of money being more than 60% and the odds ranging between 3.25 and 3.40.

In the match Stats 30, there is nothing particularly noteworthy.

Entella vs Perugia (Italie Serie B)

Date: 24 July 2020 Score: 0-2

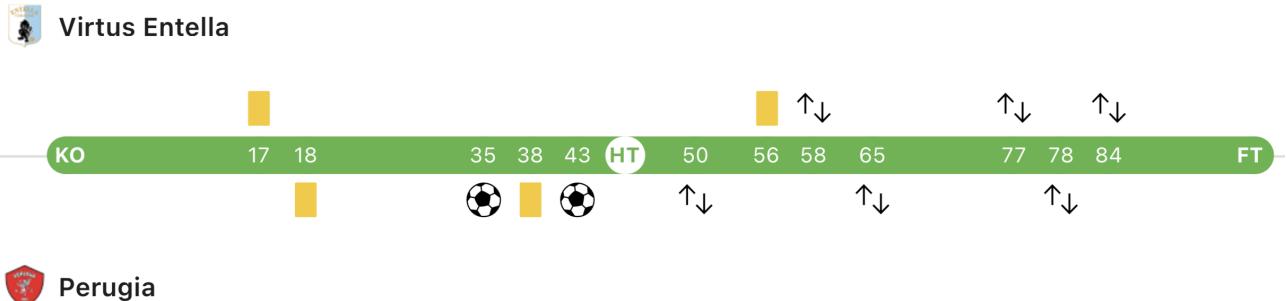


Figure 8: Match Timeline (Entella vs Perugia)

Even though it is at the beginning of the match, a 2414.08% increase between the 3rd and 4th minutes in the *DrawNoBet* market for the *Perugia* market option is excessively high.

In the *CorrectScore* market for the *0 - 2* market option, a 444.39% change in the 10 minutes, excluding the last 5 minutes, draws attention. Since this is the first-half score, it is unlikely that it will also be the final match score.

A 28.31% increase between the 30th and 40th minutes in the *HalfTimeScore* market for the *0 - 2* market option, considering the goals scored in the 35th and 43rd minutes, cannot be considered normal. The second goal was scored just minutes after the betting, and the bettor wins as a result!

In the *BothTeamsToScore* market for the *No* market option, the rise in odds over 5 minutes and the 69.65% decrease in money in the market between the 37th and 42nd minutes is noteworthy. One team had already scored a goal in the 38th minute. There was a subsequent decrease in odds and a slight increase in money coming into the market option.

In the *FirstHalfGoals1.5* market for the *Under 1.5 Goals* market option, there are 6 instances of disproportionate bets in the first 10 minutes. When the money percentage in the market option is between 60% and 70% and the odds are between 3.25 and 4.10. The *Under 1.5 Goals* option cannot be a winner given the score of 0-2 at the end of the first half. This indicates that bettors likely exited the market before the goals were scored.

Match Stats 31 show that the home team, Entella, who lost the match, possessed the ball approximately twice as much as Perugia. They also had more than twice the number of shot attempts compared to the away team and earned more corner kicks.

Criteria	Market	Market option	Details
4. criteria	DrawnoBet	Perugia	Percentage change between minute 3 and 4: 2414.08%
4. criteria	CorrectScore	0 - 2	Percentage change between row 570 and row 757: 444.39%
4. criteria	HalfTimeScore	0 - 2	Percentage change between minute 30 and 40: 28.31%
6. criteria	BothteamstoScore?	No	The odds raise in 10 minutes and 69.65% of the money left the market between minute 37 and 42
7. criteria	FirstHalfGoals1.5	Under1.5Goals	In first 10 minutes, there is a disproportion of bets (percentage: 69.91 and price: 3.25)
7. criteria	FirstHalfGoals1.5	Under1.5Goals	In first 10 minutes, there is a disproportion of bets (percentage: 69.91 and price: 3.50)
7. criteria	FirstHalfGoals1.5	Under1.5Goals	In first 10 minutes, there is a disproportion of bets (percentage: 69.76 and price: 3.75)
7. criteria	FirstHalfGoals1.5	Under1.5Goals	In first 10 minutes, there is a disproportion of bets (percentage: 69.76 and price: 3.85)
7. criteria	FirstHalfGoals1.5	Under1.5Goals	In first 10 minutes, there is a disproportion of bets (percentage: 69.55 and price: 4.00)
7. criteria	FirstHalfGoals1.5	Under1.5Goals	In first 10 minutes, there is a disproportion of bets (percentage: 69.46 and price: 4.10)

Table 9: Anomalies (Entella vs Perugia)

Haugesund vs Rosenborg (Norway Eliteserien)

Date: 26 Jul 2020 **Score:** 1-0

In the *MatchOdds* market for the *Haugesund* market option, just one minute before the goal, the percentage of money for the market option increased by 25.53% in one minute.

In the *FirstHalfGoals2.5* market for the *Under 2.5 Goals* market option, between the 3rd and 5th minutes, the percentage of money increased by 43.72%. If it were the *Under 1.5 Goals* option instead



Figure 9: Match Timeline (Haugesund vs Rosenborg)

of the *Under 2.5 Goals*, it could be more interesting. In fact, the increasing percentage of money in the market option is the case in the *FirstHalfGoals1.5* market for the *Under 1.5 Goals* market option. It is interesting to observe a 27.02% increase between the 7th and 8th minutes, before the goal scored at 9th minute.

Criteria	Market	Market option	Details
4. criteria	MatchOdds	Haugesund	Percentage change between minute 8 and 8: 25.53%
4. criteria	FirstHalfGoals2.5	Under 2.5 Goals	Percentage change between minute 3 and 5: 43.72%
4. criteria	HalfTime	Haugesund	Percentage change between minute 5 and 7: 336.29%
4. criteria	FirstHalfGoals1.5	Under 1.5 Goals	Percentage change between minute 7 and 8: 27.02%
6. criteria 7. criteria	Haugesund+1 AsianHandicap	Haugesund +1 Haugesund	The odds raise in 10 minutes and 96.63% of In first 10 minutes, there is a disproportion of bets (percentage: 68.72 and price: 5.29)

Table 10: Anomalies (Haugesund vs Rosenborg)

In the *HalfTime* market for the *Haugesund* market option, again before the goal, between the 5th and 7th minutes, the percentage of money in the market option increased significantly by 336.29%.

For the *Haugesund+1* market option, the odds rose by more than 2 points in 10 minutes, and 96.63% of the money left the market between the 88th and 93rd minutes, in the last 5 minutes of the match. This exit is not normal because, in this market, even the *Draw* option is a winner. When the +1 handicap is applied to Haugesund in the match against Rosenborg, the adjusted score for the bet becomes 1-1. Since the adjusted score results in a draw (1-1), the bet on Haugesund +1 is won. In handicap betting, if the adjusted score is a draw, the bet on the team with the handicap is considered a winner. Therefore, with a +1 handicap, Haugesund's adjusted score ties with Rosenborg, making the Haugesund +1 bet a winning bet. It is more usual to see an increasing percentage of money for this option. Because there is just one goal, with a score of 0-1, the goal scored at the beginning of the match, and nothing noteworthy observed during the match.

In the *AsianHandicap* market for the *Haugesund* market option, in the first 10 minutes, there is a disproportion of bets. The percentage of the money in the market option is 68.72% and the odds are 5.29. For Haugesund to be declared the winner of an Asian Handicap bet, they would need to have a handicap that adjusts the final score in their favor, despite losing the actual match 0-1. Typically, this would involve a handicap of +1.5 or +2, which would result in Haugesund's adjusted score being higher than Rosenborg's actual score.

Huddersfield vs Preston (English Football League)

Date : 4 July 2020 Score: 0-0

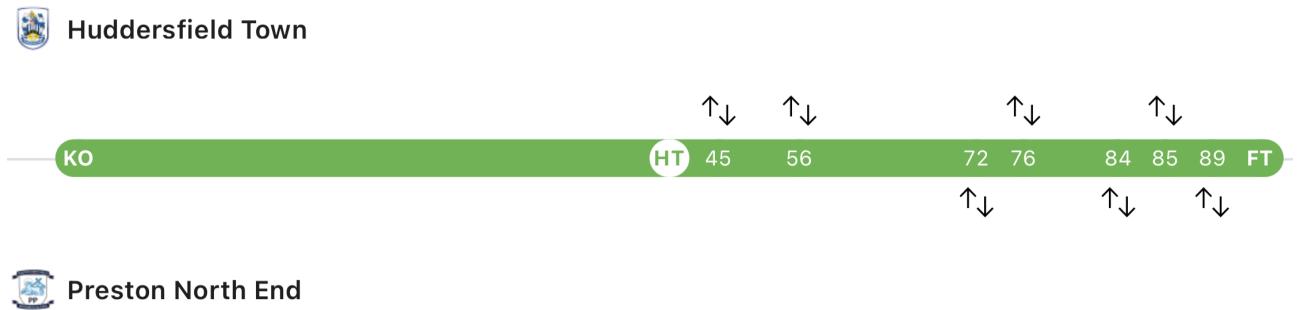


Figure 10: Match Timeline (Huddersfield vs Preston)

According to the four criteria defined, there is a noteworthy increase in the percentage of money in some market options: *OverUnder0.5Goals / Under 0.5 Goals*, *FirstHalfGoals0.5 / Under 0.5 Goals*, *HalfTimeScore / 0 - 0*, and *OverUnder1.5Goals / Under 1.5 Goals*. All these markets are related to either the half-time score or the match score.

Criteria	Market	Market option	Details
4. criteria	OverUnder0.5Goals	Under 0.5 Goals	Percentage change between minute 5 and 6: 176.35%
4. criteria	FirstHalfGoals0.5	Under 0.5 Goals	Percentage change between minute 2 and 4: 473.54%
4. criteria	FirstHalfGoals0.5	Under 0.5 Goals	Percentage change between minute 7 and 9: 29.02%
4. criteria	HalfTimeScore	0 - 0	Percentage change between minute 10 and 19: 29.92%
4. criteria	HalfTimeScore	0 - 0	Percentage change between minute 20 and 29: 28.88%
4. criteria	HalfTimeScore	0 - 0	Percentage change between minute 30 and 39: 37.23%
4. criteria	OverUnder1.5Goals	Under 1.5 Goals	Percentage change between minute 3 and 4: 359.25%
4. criteria	OverUnder1.5Goals	Under 1.5 Goals	Percentage change between minute 5 and 6: 31.54%
7. criteria	MatchOdds	Huddersfield	In first 10 minutes, there is a disproportionate of bets (percentage: between 64.80-69.01 and price: between 3.55-3.35) (7 Times)

Table 11: Anomalies (Huddersfield vs Preston)

In the *MatchOdds* market for the *Huddersfield* market option, there are 7 instances of disproportionate betting, with the percentage of money between 64% and 70% and odds between 3.35 and 3.55 in the first 10 minutes of the match. The final result for the *MatchOdds* market was a *Draw*, so it is likely that the bettor(s) exited the market before the bet was closed.

According to match Stats 32, the away team, Preston, is more dominant than the home team.

Hull City vs Middlesbrough (English Football League)

Date: 2 July 2020 Score: 2-1

In *HalfTime* market for *The Draw* market option, since one goal scored by away team (Middlesbrough),



Figure 11: Match Timeline (Hull vs Middlesbrough)

45.62% increase money in that option between minute 3 and 6 is not unusual. Similar logic in *HalfTimeScore* market for 1-1 market option the increasing even more than 30.97% money in 1-1 option is not unusual.

In the *HalfTime* market for the *The Draw* market option, a 45.62% increase in money between the 3rd and 6th minutes is not unusual, given that one goal was scored by the away team (Middlesbrough). Similarly, in the *HalfTimeScore* market for the 1-1 market option, an increase of more than 30.97% in money, occurring twice after the 11th minute, is also not unusual given that the score was 1-1 in the 8th minute.

Because a 56.01% increase in money percentage occurred between the beginning of the match, specifically between the 1st and 3rd minutes, in the *Hull+1* market for the *Hull +1* market option, the bet does not seem suspicious.

In *FirstHalfGoals2.5* market for Under 2.5 Goals market option, the odds rise more than 2 values between minute 8 and 13 minutes and 29.42% of the money left the market , considering in 8th minute, score is 1-1, leaving money is not suspicious because the probability to win decreased after 2 goals.

Criteria	Market	Market option	Details
4. criteria	HalfTime	The Draw	Percentage change between minute 3 and 6: 45.62%
4. criteria	HalfTimeScore	1-1	Percentage change between minute 11 and 20: 157.71%
4. criteria	HalfTimeScore	1-1	Percentage change between minute 21 and 30: 30.97%
4. criteria	Hull+1	Hull +1	Percentage change between minute 1 and 3: 56.01%
6. criteria	FirstHalfGoals2.5	Under 2.5 Goals	The odds raise in 10 minutes and 29.42% of the money left the market between minute 8 and 13
7. criteria	OverUnder3.5Goals	Under3.5Goals	In first 10 minutes, there is a disproportion of bets (percentage: 66.81 and price: 3.10)

Table 12: Anomalies (Hull vs Middlesbrough)

In the *FirstHalfGoals2.5* market for the *Under 2.5 Goals* market option, the odds rose by more than 2 values between the 8th and 13th minutes, and 29.42% of the money left the market. Considering that the score was 1-1 in the 8th minute, the departure of money is not suspicious, as the probability of winning decreased after the two goals were scored.

Having a disproportion of bets (percentage: 66.81% and price: 3.10) in the first 10 minutes could

be usual in the *OverUnder3.5Goals* market for the *Under 3.5 Goals* market option. After goals were scored in the 4th and 8th minutes, markets such as *Under 0.5 Goals* and *Under 1.5 Goals* would close, leading bettors to shift their focus to *Under 2.5 Goals*, *Under 3.5 Goals*, *Under 4.5 Goals*, and similar markets.

Match Stats 33 do not reveal any interesting points related to the match results and betting markets.

Even though this event matches the three criteria created, a detailed analysis reveals that it does not seem suspicious.

JuveStabia vs Entella (Italie Serie B)

Date: 10 July 2020 Score : 1-1

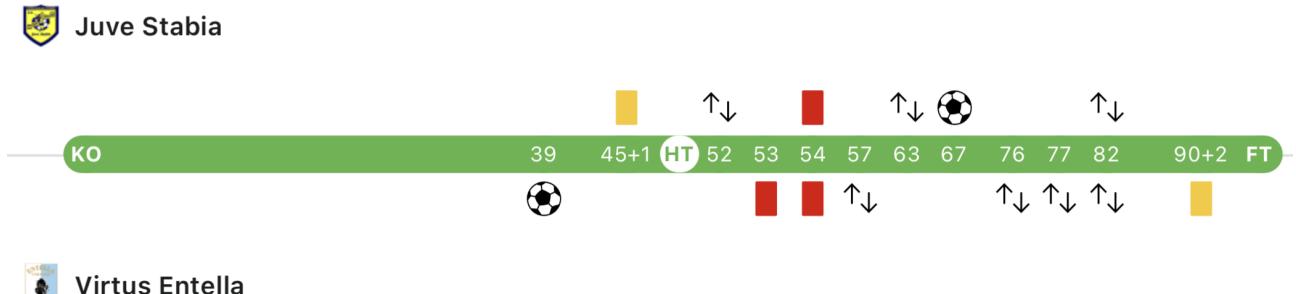


Figure 12: Match Timeline (JuveStabia vs Entella)

In HalfTime market for Entella market option, 65.04% increasing money percentage in market option between 4 and 6 minutes and 113.78% increasing money percentage in market option between 6 and 9 minutes seem suspicious considering Entella scored one goal and first half time score is 0-1, this bet seem suspicious. Same case, in FirstHalfGoals1.5 market for Under 1.5 Goals market option, 70.86% increasing money percentage in market option between 2 and 4 minutes.

Criteria	Market	Market option	Details
4. criteria	HalfTime	Entella	Percentage change between minute 4 and 6: 65.04%
4. criteria	HalfTime	Entella	Percentage change between minute 6 and 9: 113.78%
4. criteria	FirstHalfGoals1.5	Under 1.5 Goals	Percentage change between minute 2 and 4: 70.86%
7. criteria	MatchOdds	JuveStabia	In first 10 minutes, there is a disproportion of bets (percentage: 68.52 and price: 3.55)
7. criteria	MatchOdds	JuveStabia	In first 10 minutes, there is a disproportion of bets (percentage: 69.88 and price: 3.55)
7. criteria	MatchOdds	JuveStabia	In first 10 minutes, there is a disproportion of bets (percentage: 69.54 and price: 3.50)

Table 13: Anomalies (JuveStabia vs Entella)

In the *HalfTime* market for the *Entella* market option, a 65.04% increase in the money percentage between the 4th and 6th minutes, and a 113.78% increase between the 6th and 9th minutes, seem suspicious considering Entella scored one goal and the first-half score was 0-1. Similarly, in the *FirstHalfGoals1.5* market for the *Under 1.5 Goals* market option, a 70.86% increase in the money percentage between the 2nd and 4th minutes also seems suspicious.

In the first 10 minutes, there seem to be 3 instances of disproportionate bets in the *MatchOdds* market for the *Juve Stabia* market option (percentage: between 68.52-69.54 and price: 3.55-3.50). However,

this might be an error in the raw data. The Juve Stabia team never took the lead during the match, so there was no possibility to win, even if the bettor exited early with the *Juve Stabia* market option.

According to match Stats 34, it seems that Virtus Entella played a bit more dominantly than Juve Stabia, possessing the ball more and having more shots on goal. The match was intense, considering there were 3 red cards: one from Juve Stabia and two from Virtus Entella.

Kansas City vs Colorado (USA Major League Soccer)

Date: 18 July 2020 Score: 3-2

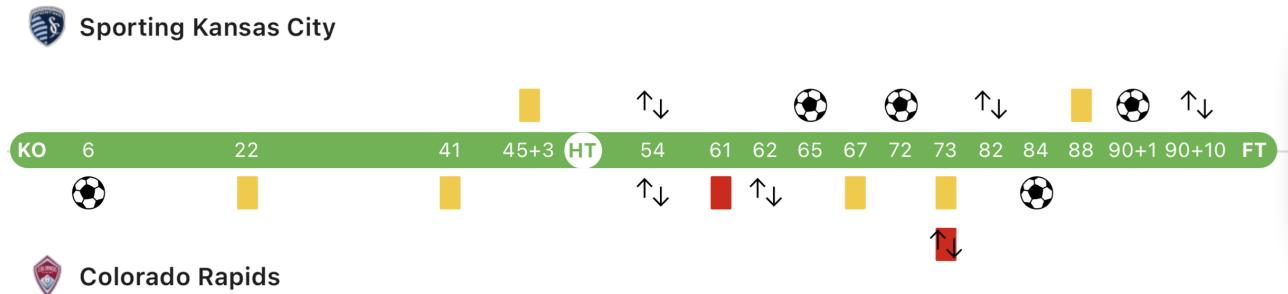


Figure 13: Match Timeline (Kansas City vs Colorado)

In the *HalfTimeScore* market for the *0 - 1* market option, there are 2 instances of more than a 25% increase in money percentage in the market option after the 6th-minute goal. This does not seem suspicious.

In the *OverUnder5.5Goals* market for the *Under 5.5 Goals* market option, there is a 52.03% increase in money percentage between the 1st and 3rd minutes. However, since this increase starts from the first minute, it cannot be considered an anomaly.

Criteria	Market	Market option	Details
4. criteria	HalfTimeScore	0 - 1	Percentage change between minute 3 and 11: 29.61%
4. criteria	HalfTimeScore	0 - 1	Percentage change between minute 22 and 31: 163.05%
4. criteria	OverUnder5.5Goals	Under 5.5 Goals	Percentage change between minute 1 and 3: 52.03%
4. criteria	DrawnoBet	Kansas City	Percentage change between minute 4 and 6: 32.93%
4. criteria	CorrectScore	3-2	Percentage change in 126th minute: 34.44%
7. criteria	MatchOdds	KansasCity	In first 10 minutes, there is a disproportion of bets (percentage: 68.52 and price: 3.45)
7. criteria	MatchOdds	KansasCity	In first 10 minutes, there is a disproportion of bets (percentage: 66.78 and price: 3.30)
7. criteria	MatchOdds	KansasCity	In first 10 minutes, there is a disproportion of bets (percentage: 65.76 and price: 3.50)

Table 14: Anomalies (Kansas City vs Colorado)

There is a 32.93% increase in money percentage between the 4th and 6th minutes in the *DrawNoBet* market for the *Kansas City* market option.

There is a 34.44% increase in money percentage in the 126th minute in the *CorrectScore* market for the *3-2* market option. This does not seem suspicious considering that the score was $3-2$ in the 90^{+1} st

minute.

In the first 10 minutes, there seem to be 3 instances of disproportionate bets in the *MatchOdds* market for the *Kansas City* market option (percentage: between 60-70 and price: 3.30-3.50). This attracted attention.

According to match Stats 35, Kansas City is more dominant than the away team, considering ball possession, shots on goal, shot attempts, and corner kicks. It was an intense match: There were 2 red cards, one of them resulting from a second yellow card and in the second half, there were 4 more goals.

Oita Trinita vs Vissel Kobe (Japan J1 league)

Date: 11 July 2020 Score: 1-1

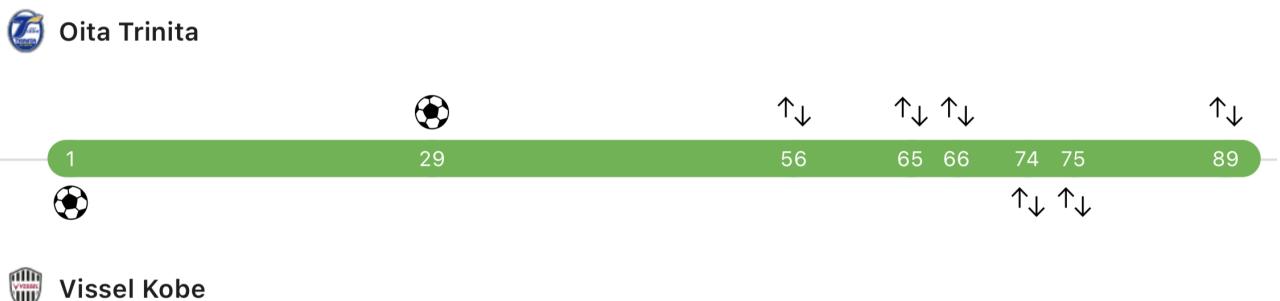


Figure 14: Match Timeline (Oita Trinita vs Vissel Kobe)

There is a 103.15% increase in money percentage between the 7th and 11th minutes in the *HalfTime* market for the *The Draw* market option, when one goal was scored by the away team and the score was 0-1.

Criteria	Market	Market option	Details
4. criteria	HalfTime	The Draw	Percentage change between minute 7 and 11: 103.15%
4. criteria	HalfTimeScore	1-1	Percentage change between minute 28 and 37: 1531.62%
4. criteria	CorrectScore	1-1	Percentage change between minute 15 and 33: 253.10%
6. criteria	OverUnder3.5Goals	Under 3.5 Goals	The odds raise in 10 minutes and 30.93% of the money left the market between minute 1 and 5
7. criteria	AsianHandicap	Oita	In first 10 minutes, the disproportion of bets (percentage: between 62.17-61.75 and price: between 3.67-3.33) (11 times)

Table 15: Anomalies (Oita vs Kobe)

There is a very high increase of 1531.62% in money percentage between the 28th and 37th minutes in the *HalfTimeScore* market for the *1-1* market option when the score was 1-1 in the 29th minute, which seems usual. On the other hand, there is a very high increase of 253.10% in money percentage between the 15th and 33rd minutes in the *HalfTimeScore* market for the *1-1* market option. Since this bet is related to the match score, this anomaly seems interesting to observe.

In the *OverUnder3.5Goals* market for the *Under 3.5 Goals* market option, the odds rose by more than 2 values in 10 minutes, and 30.93% of the money left the market between the 1st and 5th minutes. Since this occurred at the beginning of the event, it does not seem suspicious.

In the first 10 minutes, there seem to be 11 instances of disproportionate bets in the *AsianHandicap* market for the *Oita* market option, with the percentage of money in the market option between 61% and 63% and the odds between 3.33 and 3.67.

Pescara vs Frosinone (Italie Serie B)

Date: 17 July 2020 **Score:** 1-1

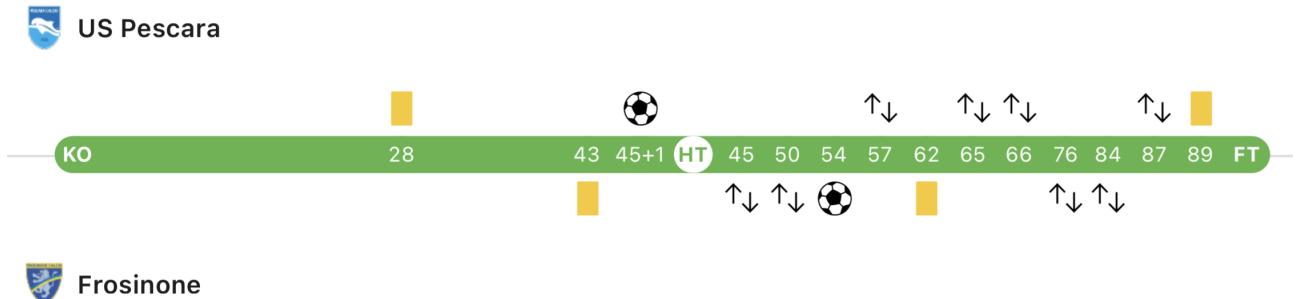


Figure 15: Match Timeline (Pescara vs Frosinone)

There is a 42.40% increase in money percentage between the 5th and 8th minutes in the *HalfTime* market for the *Pescara* market option and Pescara defined the 1-0 half time score at 45+1st minute.

In the *HalfTime* market for the *Pescara* option, there was a 42.40% increase in the percentage of money placed between the 5th and 8th minutes. This increase occurred before Pescara scored the opening goal, which was registered at the 45+1st minute. The significant rise in money percentage suggests that bettors accurately anticipated the goal before it happened.

In the first 10 minutes, there seem to be 2 instances of disproportionate bets in the *MatchOdds* market for the *Pescara* market option, with the percentage of money in the market option 60% and the odds between 3.15.

Criteria	Market	Market option	Details
4. criteria	HalfTime	Pescara	Percentage change between minute 5 and 8: 42.40%
7. criteria	MatchOdds	Pescara	In first 10 minutes, there is a disproportion of bets (percentage: 60.11 and price: 3.15)
7. criteria	MatchOdds	Pescara	In first 10 minutes, there is a disproportion of bets (percentage: 60.08 and price: 3.15)

Table 16: Anomalies (Pescara vs Frosinone)

Real Salt Lake vs Kansas City (USA Major League Soccer)

Date: 22 July 2020 **Score:** 0-2

There was a 29.91% increase in money percentage between the 2nd and 4th minutes in the *FirstHalfGoals1.5* market for the *Under 1.5 Goals* market option, despite one goal being scored in the 1st minute. This is unusual because the probability of winning this bet decreased with the goal, so one would expect a decrease in money percentage.

Similarly, there was a 26.51% increase in money percentage between the 2nd and 4th minutes in the *BothteamstoScore?* market for the *No* market option. Since this bet was placed at the beginning of the match, it cannot be considered an anomaly.

There were two increases in money percentage—169.52% and 33.45%—after the 1st-minute goal in the *FirstHalfGoals1.5* market for the *Under 1.5 Goals* market option. Since the bet was placed after the



Figure 16: Match Timeline (Real Salt Lake vs Kansas City)

goal, it cannot be considered an anomaly in terms of the actual score. However, the 169.52% increase at the beginning of the half time is noteworthy and worth paying attention to.

Criteria	Market	Market option	Details
4. criteria	FirstHalfGoals1.5	Under 1.5 Goals	Percentage change between minute 2 and 4: 29.91%
4. criteria	BothteamstoScore?	No	Percentage change between minute 1 and 2: 26.51%
4. criteria	HalfTimeScore	0 - 1	Percentage change between minute 2 and 11: 169.52%
4. criteria	HalfTimeScore	0 - 1	Percentage change between minute 12 and 21: 33.45%
6. criteria	FirstHalfGoals1.5	Under 1.5 Goals	The odds raise in 10 minutes and 33.09% of the money left the market between minute 3 and 9
7. criteria	OverUnder2.5Goals	Under2.5Goals	In first 10 minutes, there is a disproportionate of bets (percentage: between 66.94 and 61.10 and price: between 3.45 and 3.20) (8 times)

Table 17: Anomalies (Real Salt Lake vs Kansas City)

In the *FirstHalfGoals1.5* market for the *Under 1.5 Goals* market option, the odds rose by more than 2 values between the 3rd and 9th minutes, and 33.09% of the money left the market. There should have been an increase in money percentage in this market option because one goal had been scored before this betting period.

In the first 10 minutes, there seem to be 8 instances of disproportionate bets in the *OverUnder2.5Goals* market for the *Under2.5Goals* market option, with the percentage of money in the market option between 66.94% and 61.10% and the odds between 3.45 and 3.20.

According to match statistics 37, Real Salt Lake possessed the ball more, but the away team had more shots on goal, shot attempts, and corner kicks.

Rio Ave vs Portimonense (Portugal Primeira Liga)

Date: 9 July 2020 Score: 2-1

Between the 3rd and 4th minutes of the game, there was a 113.06% increase in the money percentage for the *Over 1.5 Goals* option in the *FirstHalfGoals1.5* market while the score remained 0-0.

In the *HalfTimeScore* market for the *1-1* option, there were two instances where the money percentage increased by more than 25%. The first increase occurred between the 20th and 29th minutes, following a goal scored in the 6th minute. The second increase occurred between the 30th and 39th minutes,



Figure 17: Match Timeline (Rio Ave vs Portimonense)

coinciding with a goal scored in the 35th minute. The first increase can be considered an anomaly, while the second increase aligns with the actual score progression, making it a typical betting pattern.

In the *HalfTime* market for the *Draw* option, there was a 53.53% increase in the money percentage between the 8th and 10th minutes, just after the first goal. Its suspicious level is similar to that of the *HalfTimeScore* market's first option.

Criteria	Market	Market option	Details
4. criteria	FirstHalfGoals1.5	Over 1.5 Goals	Percentage change between minute 3 and 4: 113.06%
4. criteria	HalfTimeScore	1-1	Percentage change between minute 20 and 29: 28.63%
4. criteria	HalfTimeScore	1-1	Percentage change between minute 30 and 39: 127.53%
4. criteria	HalfTime	The Draw	Percentage change between minute 8 and 10: 53.53%
7. criteria	MatchOdds	RioAve	In first 10 minutes, there is a disproportion of bets (percentage: 67.16 and the odds raise to 3.55)
7. criteria	MatchOdds	RioAve	In first 10 minutes, there is a disproportion of bets (percentage: 66.86 and the odds raise to 3.35)

Table 18: Anomalies (Rio Ave vs Portimonense)

In the first 10 minutes, there appear to be 2 instances of disproportionate bets in the *MatchOdds* market for the *RioAve* option, with the percentage of money in the market option between 67.16% and 66.86% and the odds between 3.55 and 3.35. These bets coincided with a goal scored by the Rio Ave team in the 6th minute.

According to match Stats 38, Rio Ave possessed the ball around 60%, whereas Portimonense possessed the ball around 39%. Other stats seem similar for both teams.

Rubin Kazan vs Rostov (Russia Premier League)

Date: 16 July 2020 **Score:** 0-0 There was a 49.24% increase in the money percentage between the 4th and 6th minutes in the *FirstHalfGoals0.5* market for the *Under 0.5 Goals* option. This could be an anomaly depending on the money placed before. For the same market option, the odds increased by more than 2 values between the 4th and 10th minutes, and 51.18% of the money left the market.

In the *DoubleChance* market for the *Home or Draw* option, there was a significant increase of 330.92% in the money percentage between the 2nd and 5th minutes.

In CorrectScore market for 0 - 0 market option, there was a increase of 33.36 in the money percentage

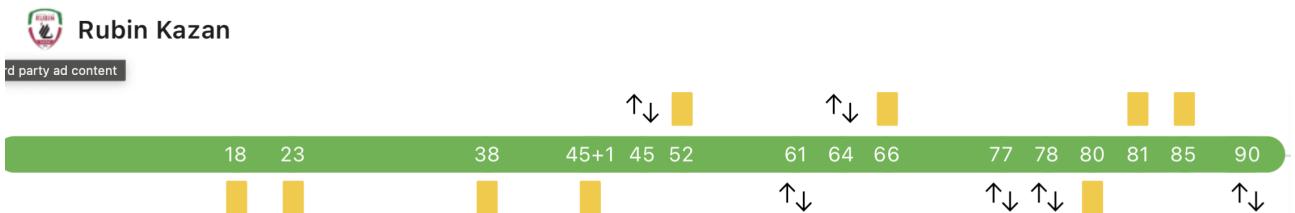


Figure 18: Match Timeline (Rubin Kazan vs Rostov)

between the 18th and 38th minutes. if it was firsthalf score it could be usual but this bet is related to match result.

Criteria	Market	Market option	Details
4. criteria	FirstHalfGoals0.5	Under 0.5 Goals	Percentage change between minute 4 and 6: 49.24%
4. criteria	DoubleChance	Home or Draw	Percentage change between minute 2 and 5: 330.92%
4. criteria	CorrectScore	0 - 0	Percentage change between minute 18 and 38: 33.36%
6. criteria	FirstHalfGoals0.5	Under 0.5 Goals	The odds raise in 10 minutes and 51.18% of the money left the market between minute 4 and 10
7. criteria	MatchOdds	RubinKazan	In first 10 minutes, there is a disproportionate of bets (percentage: between 68.98 and 61.27 and price: 3.25 and 3.20) (9 times)

Table 19: Anomalies (Rubin Kazan vs Rostov)

In the *CorrectScore* market for the *0-0* option, there was an increase of 33.36% in the money percentage between the 18th and 38th minutes. If this was for the *FirstHalfTime* score bets, it could be considered usual, but this bet is related to the match result.

In the first 10 minutes, there appear to be 9 instances of disproportionate bets in the *MatchOdds* market for the *RubinKazan* option, with the percentage of money in the market option ranging from 68.98% to 61.27% and the odds varying between 3.25 and 3.20. The match ended in a draw, but the bettor must have left before the closing of the market option.

According to match stats 39, there is nothing noteworthy to indicate.

Vikingur Reykjavik vs Valur (Iceland Premier League)

Date : 8 July 2020 Score : 1-5

There was more than a 50% increase in the money percentage in the 5th minute in the *OverUnder6.5Goals/Under 6.5 Goals*, *OverUnder5.5Goals/Over 5.5 Goals*, and *OverUnder7.5Goals/Under 7.5 Goals* markets/options.

The odds increased between the 12th and 19th minutes, and 60.31% of the money left the market in the *OverUnder6.5Goals* market for the *Under 6.5 Goals* option.

In the first 10 minutes, there appear to be 2 instances of disproportionate bets in the *MatchOdds* market for the *Watford* option, with the percentage of money in the market option being 69.45% at odds of 3.30 and 68.96% at odds of 3.35.



Figure 19: Match Timeline (Vikingur Reykjavik vs Valur)

Criteria	Market	Market option	Details
4. criteria	OverUnder6.5Goals	Under 6.5 Goals	Percentage change between minute 4 and 5: 92.21%
4. criteria	OverUnder5.5Goals	Over 5.5 Goals	Percentage change between minute 3 and 4: 56.41%
4. criteria	OverUnder7.5Goals	Under 7.5 Goals	Percentage change between minute 4 and 5: 77.48%
6. criteria	OverUnder6.5Goals	Under 6.5 Goals	The odds raise in 10 minutes and 60.31% of the money left the market between minute 12 and 19
7. criteria	MatchOdds	VikingurReykjavik	In first 10 minutes, there is a disproportion of bets (percentage: between 66.16% - 60.80% and price: 3.50 - 3.45) (4 times)

Table 20: Vikingur Reykjavik vs Valur (Iceland Premier League)

Watford vs Norwich (England Premier League)

Date: 7 July 2020 Score: 2-1

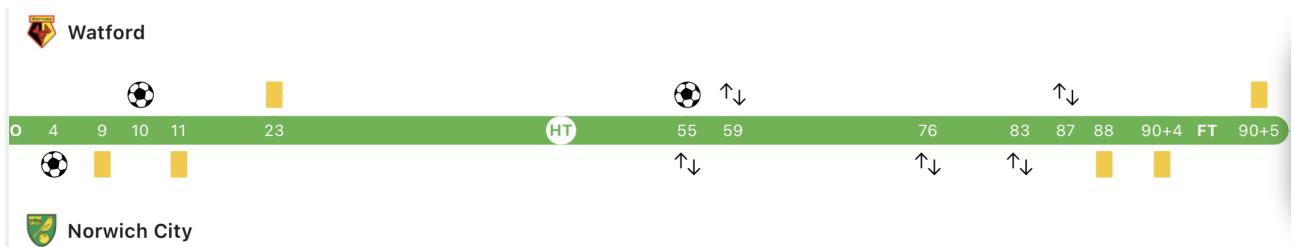


Figure 20: Match Timeline (Watford vs Norwich)

There is a 32.88% increase in money percentage between the 3rd and 4th minutes in the *DoubleChance* market for the *Home or Draw* market option and there is a 360.32% increase in money percentage between the 11th and 20th minutes in the *HalfTimeScore* market for 1-1 market option.

There is a 32.88% increase in money percentage between the 3rd and 4th minutes in the *DoubleChance* market for the *Home or Draw* option, and there is a 360.32% increase in money percentage between the 11th and 20th minutes in the *HalfTimeScore* market for the 1-1 option.

The odds increased more than 2 values between the 10th and 16th minutes, and 45.24% of the money left the market in the *FirstHalfGoals2.5* market for the *Under 2.5 Goals* option.

According to match stats ??, Norwich City possessed the ball more, around 56%, and had more shots

on goal and shot attempts.

Criteria	Market	Market option	Details
4. criteria	DoubleChance	Home or Draw	Percentage change between minute 3 and 4: 32.88%
4. criteria	HalfTimeScore	1-1	Percentage change between minute 11 and 20: 360.32%
6. criteria	FirstHalfGoals2.5	Under 2.5 Goals	The odds raise in 10 minutes and 45.24% of the money left the market between minute 10 and 16
7. criteria	MatchOdds	Watford	In first 10 minutes, there is a disproportion of bets (percentage: 69.45 and price: 3.30)
7. criteria	MatchOdds	Watford	In first 10 minutes, there is a disproportion of bets (percentage: 68.96 and price: 3.35)

Table 21: Anomalies (Watford vs Norwich)

Yokohama FM vs FC Tokyo (Japan J1 league)

Date: 12 July 2020 Score: 1-3

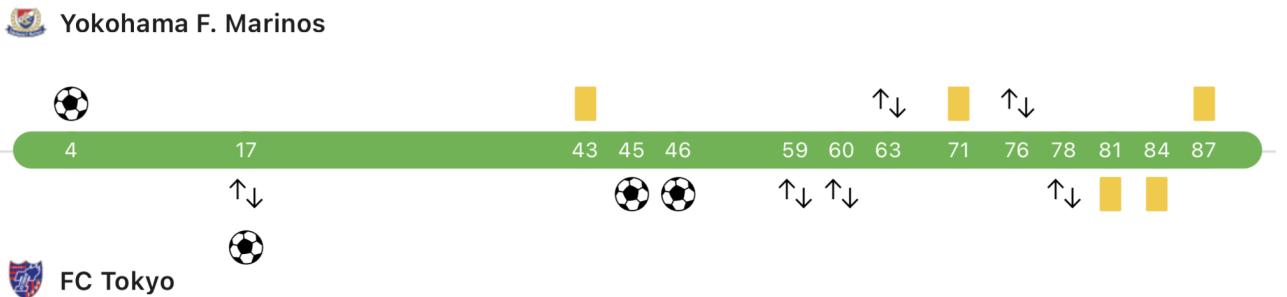


Figure 21: Match Timeline (Yokohama FM vs FC Tokyo)

In the *HalfTime* market for *FC Tokyo*, there is a 76.70% increase in money percentage between the 3rd and 5th minutes, and a substantial 320.22% increase between the 5th and 9th minutes, which coincides with FC Tokyo's goal scored in the 17th minute. In the *OverUnder3.5Goals* market for the *Over 3.5 Goals* option, money percentage rises by 51.35% from the 1st to the 4th minute, the time when Yokohama scored a goal, and by 68.88% from the 4th to the 5th minute. Additionally, in the *FirstHalfGoals2.5* market for the *Over 2.5 Goals* option, there is a 46.15% increase in money percentage between the 6th and 8th minutes.

In the *HalfTime* market for *FC Tokyo*, there is a 76.70% increase in money percentage between the 3rd and 5th minutes, and a substantial 320.22% increase between the 5th and 9th minutes. The latter increase coincides with FC Tokyo's goal scored in the 17th minute. In the *OverUnder3.5Goals* market for the *Over 3.5 Goals* option, money percentage rises by 51.35% from the 1st to the 4th minute, when Yokohama scored a goal, and by 68.88% from the 4th to the 5th minute. Additionally, in the *FirstHalfGoals2.5* market for the *Over 2.5 Goals* option, there is a 46.15% increase in money percentage between the 6th and 8th minutes. Notably, in the *HalfTimeScore* market for the *1-2* option, there is an incredible 38000.00% increase in money percentage between the 28th and 37th minutes, when the score is 1-1 just before the 45th minute goal.

Criteria	Market	Market option	Details
4. criteria	HalfTime	FC Tokyo	Percentage change between minute 3 and 5: 76.70%
4. criteria	HalfTime	FC Tokyo	Percentage change between minute 5 and 9: 320.22%
4. criteria	OverUnder3.5Goals	Over 3.5 Goals	Percentage change between minute 1 and 4: 51.35%
4. criteria	OverUnder3.5Goals	Over 3.5 Goals	Percentage change between minute 4 and 5: 68.88%
4. criteria	FirstHalfGoals2.5	Over 2.5 Goals	Percentage change between minute 6 and 8: 46.15%
4. criteria	HalfTimeScore	1-2	Percentage change between minute 28 and 37: 380.00%
7. criteria	AsianHandicap	YokohamaFM	In the first 10 minutes, there is a disproportion of bets (percentage between 63.03% and 61.70%, and price between 3.05 and 3.13) (5 times)

Table 22: Anomalies (Yokohama FM vs FC Tokyo)

7 Discussion

As a match progresses, various events (like goals, cards, time remaining) can significantly influence betting markets. These events lead to rapid shifts in betting odds and volumes as bettors react to the changing circumstances.

Popular matches with high liquidity can see more frequent updates in the price and volume graphs. This is because there is a continuous flow of bets being placed and matched, leading to constant recalibration of odds and volumes.

Odds are updated based on pre-match factors like team news, weather, and historical performance. Updates are less frequent but can be substantial if significant news breaks (e.g., a key player gets injured). Odds are updated in real-time based on match events. This results in frequent and often rapid updates to the graphs as the betting market reacts instantly to on-field events.

Main Markets include bets on the overall outcome of the match (win, lose, draw). They tend to have more stable and gradual changes compared to more specific markets (e.g. first half goals 2.5). Bets like “over/under 7.5 goals” can have more volatile updates as they are more directly influenced by specific in-game events.

Some bet types appeared in certain parts of the graphs, which meant that as soon as the match started, bets might not have been placed immediately. It could also have been that bets were placed before the match.

For a few matches, we could not see the graphs for some bet types because no bets were placed for that bet type, no information was provided, or there was an error during data extraction.

Markets related to halftime, such as *HalfTimeScore* or *FirstHalfGoals*, deserve attention. These markets focus on outcomes that do not affect the final score, reducing risk for manipulated player performances, potentially making them easier to manipulate.

Towards the last seconds before the market closes, there is typically stability in the current market option, with fewer chances for unexpected surprises.

In *CorrectScore* markets, towards the end of the match, more money is placed on the current score. For instance, if the score is 0-2 in the 85th minute, more money and bets will be placed on this score.

In such cases, *AnyOtherHome*, *AnyOtherDraw*, and *AnyOtherAway* market options have a very low probability of winning. If these outcomes occur, it would be a significant surprise.

For *FirstHalfGoals* bets (0.5, 1.5, and 2.5), bets on the current score tend to increase sharply in the last few minutes of the first half. For example, if the score is 0-0, bets on *FirstHalfGoals0.5* increase sharply in the last five minutes, which is expected. For *FirstHalfGoals2.5*, bets on the under 2.5 goals option start to increase around the 30th minute. Similarly, for *FirstHalfGoals1.5*, an increase in bets around the 40th minute is typical.

Market Options While the match odds market typically offers three options, the *correct score* market can have significantly more, depending on the goals scored in the match. For instance, in a match ending 0-2, there can be up to 19 different betting options. (Discussion)

Components of graphs Graphs provided various insights into the dynamics of bets and odds related to betting markets :

On the *Price_and_Volume_graphs*, when the odds decrease and the volume increases, it indicated that more bettors believe the event is likely to happen, hence they are willing to accept lower returns (lower odds). Sharp movements in price indicated significant events or new information affecting the perception of the outcome's probability. Additionally, large spikes in volume often occur after significant match events (e.g., goals) as bettors react to the new information.

On the *Percent.money_graph*, high percentage for one bet type option indicates a strong consensus or significant betting interest in that option. Fluctuating Percentages indicate changing opinions and/or potentially live reactions to match events.

On the *Total.money_graph*, a steady increase in total money indicates consistent betting interest. Sudden increases can signal significant betting activity, possibly due to major match events or new information. Periods of little to no increase may indicate a lull in betting activity, potentially during uneventful periods of a match.

Frequency of changing The frequency of changing plots or bars on graphs in the context of a football match betting market, particularly on a *Price_and_Volume_graph*, is influenced by several factors as popularity of the match, bet type or bet type option:

High-profile matches (events) (e.g., matches involving top teams) attract more betting activity, leading to more frequent updates in price and volume data. Matches, especially during critical moments (e.g., goals, red cards), see rapid changes in betting odds and volume. without having cards information, it cannot be observed clearly if yellow or red card provide rapid changes. But when some matches were analyzed manually it was observed that the goals and cards particularly red cards influence the betting frequency.

Different markets (bet types) (e.g., match odds, half time score) might have varying levels of activity and thus different update frequencies.

Within a bet type, specific options (e.g., the draw) might experience less frequent changes if there is high profile team or news related to that option. Options with higher volumes of bets will likely have more frequent updates due to the constant flow of new information and transactions.

Markets Generally, the amount of money placed on the winner bet type option is higher than on the loser's options, but this is not always the case. For example, in the *Both Teams to Score* market, even if one team scores a goal, supporters often wait until the end of the match hoping for the other team to score a goal as well.

In some bets, particularly in *Asian handicap* or *Double Chance* markets, there can be multiple winners, whereas in other types of bets such as *teamHome+1* market, there may be no clear winner at all.

In the *Double Chance* market, there is a higher probability of a favorable outcome compared to the

Asian Handicap market, with a likelihood of 2 out of 3.

The *Correct Score* market type is popular for betting on the least probable outcome. There were 19 different bet types available, but only one of them resulted in a win. The outcome of these bets depends on the final score of the match.

In matches where one team is strongly favored to win, in the *Draw No Bet market*, the odds can vary widely, with bets sometimes approaching or reaching close to 100% for one team.

In markets related to goals scored, such as *First Half Goals ?* and *Over Under ? Goals*, when there are no goals or fewer goals scored, there is a high probability of winning in almost all these markets. For instance, if the score is 0-0 at half time, all bets under 0.5 goals would be winners, including bets like *FirstHalfGoals0.5*, *FirstHalfGoals1.5*, and *FirstHalfGoals2.5*. Similarly, in match score markets like *OverUnder0.5Goals* and *OverUnder1.5Goals* up to *OverUnder8.5Goals*, if the final match score is 0-0, all bets placed under those thresholds would be winners.

In the *Goal Lines* market, where winning depends on the total number of goals scored by both teams, bets placed in the last minutes/seconds, often with large amounts of money, can significantly impact outcomes. This is due to the high probability of achieving a win in such scenarios.

When there are no goals scored by around the 20th minute, bets related to *under 5.5*, *6.5*, *7.5*, and *8.5* goals often see a very high percentage of money placed on them.

There is generally an inverse relationship between price and volume. When the price increases, the volume tends to decrease, and vice versa. However, if both the price and volume increase in the same direction, particularly by a significant margin (e.g., a 10% increase in price accompanied by an equal or greater increase in volume), this could indicate an unusual market condition. For example, an increase in volume from 100 to 120 may not be logical if it coincides with a price increase, as typically one would expect volume to decrease when prices rise. —

8 Conclusion

This study used historical data from Betfair to ultimately detect and analyse potential anomalies in football match betting markets. The analysis used automated processes to convert, process and visualise betting data, providing insights into betting patterns.

Automated tools and libraries such as Pandas and Matplotlib have made it easier to process large datasets and create meaningful visualizations. This approach has minimized manual intervention and saved time, making the anomaly detection process more systematic and repeatable.

There were a number of limitations that affected the study. Accessing Betfair data required a VPN due to geographic restrictions, and account verification posed additional hurdles. Data processor file size restrictions and the presence of erroneous rows in some files meant that not all historical data could be analysed. Additionally, data management issues and format incompatibilities occasionally hampered the efficiency of the analysis. The dataset was limited to July 2020 and only covered select match events, which may not be comprehensively representative of the wider betting market.

Anomalies were detected based on predefined criteria related to betting patterns and market behavior. The criteria included sharp changes in odds, significant increases in betting volume, and unusual distribution of the money placed. The study identified 19 suspicious match events that showed signs of irregular betting patterns, such as the Henan v Jiangsu Suning, Vikingur Reykjavik vs Stjarnan and the Sampdoria v AC Milan events. These anomalies suggest potential areas for further investigation into match fixing or unusual betting activity.

The visualizations created provided valuable insights into betting dynamics. Price and volume charts

reflected bettors' reactions to match events by showing how odds and volumes changed over time. Percentage money and total money charts revealed changes in betting interest and the distribution of money across different types of bets. These visualizations are crucial for understanding betting trends and identifying potential anomalies.

The study observed that betting markets, particularly in high-profile matches, showed frequent updates and rapid changes in response to match events. Anomalies were often associated with significant market option movements, prices or unusual betting volumes towards the end of both half times , indicating possible manipulation or unexpected events.

While the results are compelling, however, it does not necessarily indicate that match-fixing has been detected based on finding anomalies in this study. Rather, it can be said that anomalous patterns consistent with possible match-fixing situations have been identified in Betfair exchange historical data.

This study successfully demonstrated the effectiveness of automated data conversion, processing and visualization in detecting betting anomalies, while also highlighting the challenges related to data accessibility and limitations. The study can be further improved by expanding the dataset, including a wider range of sources, and further improving the anomaly detection criteria to increase the accuracy and comprehensiveness of the analysis. The information obtained from this study provides a basis for understanding betting market patterns and identifying potential irregularities.

References

- [1] Shah Imran Alam, Syed Imtiyaz Hassan, Ihtiram Raza Khan, M Afshar Alam, and Anil Kumar Mahto. Potential of open data in sustainable and open governance: A search for imperceptible barriers. In *Proceedings of the 2nd International Conference on ICT for Digital, Smart, and Sustainable Development, ICIDSSD 2020, 27-28 February 2020, Jamia Hamdard, New Delhi, India*. EAI, 3 2021.
- [2] Remote Gambling Association. *Sports Betting: Legal, Commercial and Integrity Issues*. London, 2010. Online; accessed: 2024-07-05.
- [3] Ronald J. Bacigal. Making the Right Gamble: The Odds on Probable Cause. *Mississippi Law Journal*, 74:340, 2004.
- [4] Nicole Beebe. Digital forensics research: the good, the bad, and the unaddressed. In *Fifth annual IFIP WG 11.9 international conference on digital forensics*. Springer, January 2009.
- [5] Libor Benes. OSINT, New Technologies, Education: Expanding Opportunities and Threats. A New Paradigm. *Journal of Strategic Security*, 6(3 Suppl.):22–37, 2013.
- [6] Ludo Block. The long history of OSINT. *Journal of Intelligence History*, 23(2):95–109, 2024.
- [7] Kevin Carpenter. Match-Fixing—The Biggest Threat to Sport in the 21st Century? *International Sports Law Review*, 2:13–23, 2012.
- [8] Pompeu Casanovas. Conceptualisation of rights and meta-rule of law for the web of data. *Journal of Governance and Regulation*, 4(4):116–129, 09 2015.
- [9] Sudhanshu Chauhan and Nutan Kumar Panda. Chapter 2 - Open Source Intelligence and Advanced Social Media Search. In Sudhanshu Chauhan and Nutan Kumar Panda, editors, *Hacking Web Intelligence*, pages 15–32. Syngress, Boston, 2015.
- [10] Collective work of all DFRWS attendees. A Road Map for Digital Forensic Research. In *Proceedings of The Digital Forensic Research Conference (DFRWS 2001 USA)*, Utica, NY, August 7-8 2001.
- [11] Manuel Delgado, Manuela Aparicio, and Carlos Costa. Using Open Source for Forensic Purposes. In *Proceedings of the Workshop on Open Source and Design of Communication*, OSDOC '12, pages 31–37, New York, NY, USA, 2012. Association for Computing Machinery.
- [12] Christian Deutscher, Eugen Dimant, and Brad Humphreys. Match Fixing and Sports Betting in Football: Empirical Evidence from the German Bundesliga, 01 2017.
- [13] Helmut Dietl and Christian Weingärtner. Betting scandals and attenuated property rights - How betting related match fixing can be prevented in future. Working Papers 0154, University of Zurich, Institute for Strategy and Business Economics (ISU), Feb 2012. Cited from: Sharpe, G. (1997). *Gambling on Goals*. Edinburgh: Mainstream Publishing Company (Edinburgh) LTD.
- [14] Helmut Dietl and Christian Weingärtner. Betting scandals and attenuated property rights - How betting related match fixing can be prevented in future. Working Papers 0154, University of Zurich, Institute for Strategy and Business Economics (ISU), Feb 2012.
- [15] European Commission. Communication from the Commission to the European Parliament, the Council, the Economic and Social Committee and the Committee of the Regions: Developing the European Dimension in Sport, 2011. Online; accessed: 2024-07-05.
- [16] European Commission. Communication from the Commission to the European Parliament, the Council, the Economic and Social Committee and the Committee of the Regions: Towards a comprehensive European framework for online gambling, 2012. Online; accessed: 2024-07-05.

- [17] European Commission. Match-fixing in sport: A mapping of criminal law provisions in eu 27. Technical report, European Commission, March 2012.
- [18] European Commission. Study on Sports Fraud, 2018. Online; accessed: 2024-07-05.
- [19] European Commission. Fighting Match-Fixing, 2024. Online; accessed: 2024-07-05.
- [20] European Data Portal. What is open data?, 2024. Online; accessed: 20-Feb-2024.
- [21] European Gaming & Betting Association. European Online Gambling – Key Figures 2021 Edition, 14.01.2022. Online; accessed: 2024-07-06.
- [22] European Gaming & Betting Association. Sports betting: Commercial and integrity issues, 2014. Online; accessed: 2024-07-07.
- [23] European Gaming & Betting Association. European Online Gambling – Key Figures 2022 Edition, 22.12.2022. Online; accessed: 2024-07-06.
- [24] Thomas Feltes. Match Fixing in Western Europe, 2013. Online; accessed: 2024-07-07.
- [25] David Forrest. The threat to football from betting-related corruption. *International Journal of Sport Finance*, 7(2):99–116, May 2012.
- [26] David Forrest, I.G. Mchale, and Kevin McAuley. “Say It Ain’t So”: Betting-Related Malpractice in Sport. *International Journal of Sport Finance*, 3:156–166, 08 2008.
- [27] David Forrest and Robert Simmons. Sport and Gambling. *Oxford Review of Economic Policy*, 19(4):598–611, 12 2003.
- [28] Sally M. Gainsbury and Alex Russell. Betting Patterns for Sports and Races: A Longitudinal Analysis of Online Wagering in Australia. *J Gambl Stud*, 31:17–32, 2015.
- [29] Simson Garfinkel, Paul Farrell, Vassil Roussev, and George Dinolt. Bringing science to digital forensics with standardized forensic corpora. *Digital Investigation*, 6:S2–S11, 2009. The Proceedings of the Ninth Annual DFRWS Conference.
- [30] Simson L. Garfinkel. Digital forensics research: The next 10 years. *Digital Investigation*, 7:S64–S73, 2010. The Proceedings of the Tenth Annual DFRWS Conference.
- [31] GESPA. Competition manipulation – national platform annual review 2023, 2023. Online; accessed: 2024-07-07.
- [32] Owen Gibson. Europe hit by 'biggest-ever' match-fixing scandal. *Guardian*, November 20 2009. Online; accessed: 2024-07-07.
- [33] Stevyn D Gibson. *Examining the Role of Intelligence in Policing*. PhD thesis, Cranfield University, 2011.
- [34] Declan Hill. How Gambling Corruptors Fix Football Matches. *European Sport Management Quarterly*, 9:411–432, 11 2009.
- [35] John T. Holden, Christopher M. McLeod, and Marc Edelman. Regulatory Categorization and Arbitrage: How Daily Fantasy Sports Companies Navigated Regulatory Categories Before and After Legalized Gambling. *American Business Law Journal*, 57(1):113–167, 2020.
- [36] Arthur S. Hulnick. The Dilemma of Open Sources Intelligence: Is OSINT Really Intelligence? In Loch K. Johnson, editor, *The Oxford Handbook of National Security Intelligence*, pages 229–241. Oxford University Press, U.S.A., March 2010. Cited from: Intelligence Community Directive Number 301, National Open Source Enterprise.
- [37] IBM. Types of Historical Data, 2024. Online; accessed: 2024-07-05.

- [38] IBM Global Business Services. The quantifiable value of cloud and cognitive services, 2017. Online; accessed: 28-May-2024.
- [39] Golden G. Richard III and Vassil Roussev. Digital Forensic Tools: The Next Generation. *Digital Crime and Forensic Science in Cyberspace*, pages 75–90, 01 2006.
- [40] Iris, University of Salford, Cabinet PRAXES-Avocats and CCLS (Université de Pékin). Sports Betting and Corruption: How to Preserve the Integrity of Sport, 2012. Online; accessed: 2024-07-05.
- [41] Kathleen Janssen. The influence of the PSI directive on open government data: An overview of recent developments. *Government Information Quarterly*, 28(4):446–456, 2011.
- [42] Marijn Janssen, Yannis Charalabidis, and Anneke Zuiderwijk. Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, 29(4):258–268, 2012.
- [43] Michael C. Johannesmeyer, Ashish Singhal, and Dale E. Seborg. Pattern Matching in Historical Data. *AIChE Journal*, 48(9):2125–2135, April 2004.
- [44] Chul Kim, Jae-Hyung Park, and Jae-Young Lee. AI-based betting anomaly detection system to ensure fairness in sports and prevent illegal gambling. *Scientific Reports*, 14:6470, 2024.
- [45] Richard LaBrie, Debi LaPlante, Sarah Nelson, Anja Broda, and Howard Shaffer. Assessing the Playing Field: A Prospective Longitudinal Study of Internet Sports Gambling Behavior. *Journal of Gambling Studies*, 23:347–362, 06 2007.
- [46] Office of the Director of National Intelligence. Intelligence Community Directive Number 301, National Open Source Enterprise, 2006.
- [47] Open Knowledge Foundation. The Open Definition, 2024. Online; accessed: 20-Feb-2024.
- [48] Dimitra Papatsaroucha, Yannis Nikoloudakis, Ioannis Kefaloukos, Evangelos Pallis, and Evangelos K. Markakis. A Survey on Human and Personality Vulnerability Assessment in Cyber-security: Challenges, Approaches, and Open Issues. *ArXiv*, abs/2106.09986, 2021.
- [49] Aino Pietikäinen. The Future of the British Remote Betting and Gaming Industry, 2014. Online; accessed: 2024-07-05.
- [50] Darren Quick and Kim-Kwang Raymond Choo. Impacts of increasing volume of digital forensic data: A survey and future research challenges. *Digital Investigation*, 11(4):273–294, 2014.
- [51] Darren Quick and Kim-Kwang Raymond Choo. Digital forensic intelligence: Data subsets and Open Source Intelligence (DFINT+OSINT): A timely and cohesive mix. *Future Generation Computer Systems*, 78:558–567, 2018.
- [52] Darren Quick, Ben Martini, and Kim-Kwang Raymond Choo. Chapter 1 - Introduction. In Darren Quick, Ben Martini, and Kim-Kwang Raymond Choo, editors, *Cloud Storage Forensics*, pages 1–11. Syngress, Boston, 2014.
- [53] Olivier Ribaux. *Police scientifique: Le renseignement par la trace*. Université de Lausanne, 2013. Online; accessed 2024-07-09.
- [54] Olivier Ribaux. Notes from 'Forensic Intelligence' course in Ecole des Sciences Criminelles, UNIL, Spring semester 2022. Course notes, Spring 2022.
- [55] Olivier Ribaux, Simon J. Walsh, and Pierre Margot. The contribution of forensic science to crime analysis and investigation: Forensic intelligence. *Forensic Science International*, 156(2):171–181, 2006.

- [56] Chris Sanders and Jason Smith. Chapter 14 - Friendly and Threat Intelligence. In Chris Sanders and Jason Smith, editors, *Applied Network Security Monitoring*, pages 385–420. Syngress, Boston, 2014.
- [57] Raymond D. Sauer. The Economics of Wagering Markets. *Journal of Economic Literature*, 36:2021–2064, December 1998.
- [58] Inge Sebyan Black and Lawrence J. Fennelly. Chapter 20 - Investigations using open source intelligence (OSINT). In Inge Sebyan Black and Lawrence J. Fennelly, editors, *Investigations and the Art of the Interview (Fourth Edition)*, pages 179–189. Butterworth-Heinemann, Boston, fourth edition edition, 2021.
- [59] Uthayasanckar Sivarajah, Muhammad Mustafa Kamal, Zahir Irani, and Vishanth Weerakkody. Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70:263–286, 2017.
- [60] Michael A. Smith and Leighton Vaughan Williams. Chapter 19 - Betting Exchanges : A Technological Revolution in Sports Betting. In Donald B. Hausch and William T. Ziemba, editors, *Handbook of Sports and Lottery Markets*, Handbooks in Finance, pages 403–418. Elsevier, San Diego, 2008.
- [61] Sportradar Group AG. ANTI-MATCH-FIXING: Comprehensive approach to combatting betting-related fraud and the manipulation of sporting events, 2004. Online; accessed: 2024-07-05.
- [62] Statista. Online Sports Betting - Worldwide, Highlights, 2024. Online; accessed: 2024-07-09.
- [63] Swiss Federal Office for the Environment (BAFU). Procedure: Historical Investigation, 2024. Online; accessed: 2024-07-05.
- [64] Fahimeh Tabatabaei and Douglas Wells. *OSINT in the Context of Cyber-Security*, pages 213–231. 01 2016.
- [65] Katie Terrell and Hanna Ivy Wigmore. Historical Data. Online; accessed: 2024-07-05.
- [66] UEFA. Sports betting and integrity workshop in Nyon, 2009. Online; accessed: 2024-07-07.
- [67] United Nations Office on Drugs and Crime. Criminal Intelligence for Analysts, n.d.
- [68] Fons Wijnhoven, Michel Ehrenhard, and Johannes Kuhn. Open government objectives and participation motivations. *Government Information Quarterly*, 32(1):30–42, 2015.
- [69] Wikipedia contributors. Acun Ilcali, 2024. Online; accessed: 2024-07-07.
- [70] Wikipedia contributors. Betfair — Wikipedia, The Free Encyclopedia, 2024. [Online; accessed 2024-06-10].
- [71] Wikipedia contributors. Locard’s exchange principle — Wikipedia, The Free Encyclopedia, 2024. Online; accessed 2024-07-09.
- [72] U. Yogish Pai and K. Krishna Prasad. Open Source Intelligence and its Applications in Next Generation Cyber Security - A Literature Review. *International Journal of Applied Engineering and Management Letters (IJAEML)*, 5(2):1–25, 2021.

A Ethics

The study was conducted within legal limits and was used free data that is publicly available and can be obtained by anyone for a fee from *betfair.com* website. Deleting Historical Data, which is the subject of the study, is automated during the data processing phase. Furthermore, the entirety of the data collected will be deleted after the end of the study.

The datasets obtained in question do not contain any personal or corporate private information.

The methodology of the study is completely personal. The existence of collaboration with UEFA consists only of providing free supply of data and an overview of the subject.

Since the data obtained as a result of the study were created by determining completely relative criteria, and as a result, the results are relative, it is said that the anomaly detection results are not binding on any person or institution or are not the subject of any legal investigation.

B Appendix

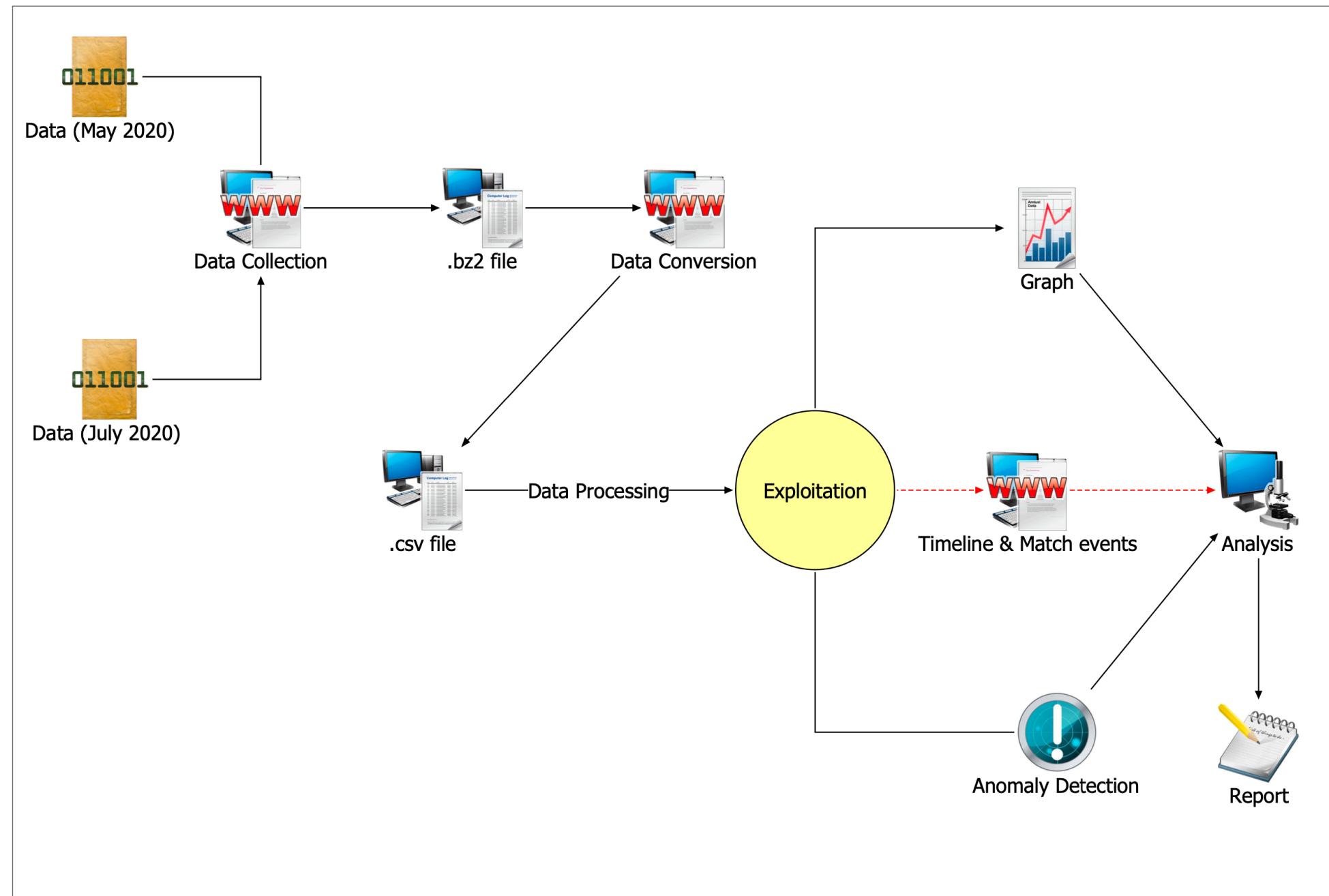


Figure 22. Visualization

Table 23: Football Betting Market Types on Betfair

No	Market Type	Explanation	Scenario
1	teamHome+1	This means teamHome starts with a one-goal advantage. This handicap levels the playing field between teams of different strengths.	If you bet on <i>teamHome+1</i> and teamHome wins, draws, or loses by exactly one goal, you win. If teamHome loses by more than one goal, you lose the bet.
2	Both Teams to Score	This market is based on both teams scoring at least one goal each. The outcome hinges on both teams finding the net.	If both teams score, the bet wins. If either team does not score, the bet loses.
3	Correct Score	The exact final score of the match is predicted.	If you bet on a specific final score, such as 2-1, you win if the match ends with that exact score. Any other result means a loss.
4	Double Chance	Allows betting on two out of three possible outcomes: a win for one team or a draw.	If you bet on a team to win or draw, you win if the team wins or the match ends in a draw. A loss occurs only if the opposing team wins.
5	Draw No Bet	Bets are placed on a team to win, with the draw outcome resulting in a refund.	If you bet on a team to win, and they win, you win the bet. If the match ends in a draw, your stake is refunded. If the opposing team wins, you lose the bet.
6	FirstHalfGoals 0.5	Bets are placed on whether there will be at least 0.5 goals scored in the first half.	If betting on <i>Over 0.5</i> , you win if at least one goal is scored in the first half. If no goals are scored, you lose. Betting on <i>Under 0.5</i> means you win if no goals are scored in the first half. If one or more goals are scored, you lose.
7	FirstHalfGoals 1.5	The total number of goals in the first half is wagered on, with options for over or under 1.5 goals.	- Over 1.5: Wins if two or more goals are scored in the first half. Loses if one or no goals are scored. - Under 1.5: Wins if one or no goals are scored in the first half. Loses if two or more goals are scored.
8	FirstHalfGoals 2.5	Bets are placed on whether there will be at least 2.5 goals scored in the first half.	If betting on <i>Over 2.5</i> , you win if 3 or more goals are scored in the first half. If 2 or fewer goals are scored, you lose. Betting on <i>Under 2.5</i> means you win if 2 or fewer goals are scored in the first half. If 3 or more goals are scored, you lose.
9	Goal Lines	Bets are placed on the total number of goals scored by both teams combined. A line is set, and bets are made on whether the total will be over or under this line.	If betting on <i>Over X</i> , you win if the total number of goals scored is greater than X. If it is X or fewer, you lose. Betting on <i>Under X</i> means you win if the total goals are X or fewer. If more than X goals are scored, you lose.

Continued on next page

Table 23 – continued from previous page

No	Market Type	Explanation	Scenario
10	Half Time	Bets are placed on the outcome of the match at halftime. This can include predictions for which team is leading, the exact score, or whether the match is drawn.	If betting on <i>Half Time Home Win</i> , you win if the home team is leading at halftime. Betting on <i>Half Time Draw</i> means you win if the score is tied at halftime. Bets on <i>Half Time Away Win</i> win if the away team is leading at halftime.
11	Half Time Full Time	Bets are placed on the outcome of both halftime and fulltime. This includes predicting the result at halftime and the final result of the match.	If betting on <i>HalfTimeHomeWin / FullTimeHomeWin</i> , you win if the home team is leading at halftime and also wins by the end of the match. Betting on <i>HalfTimeDraw / FullTimeAwayWin</i> wins if the match is tied at halftime and the away team wins by fulltime.
12	Half Time Score	Bets are placed on the exact score at halftime. This market predicts the precise score-line of the match at halftime.	If betting on <i>HalfTimeScore 1-0</i> , you win if the exact score at halftime is 1-0 in favor of the chosen team. Any other score at halftime results in a loss.
13	Match Odds	Bets are placed on the outcome of the entire match. This market predicts which team will win, or if the match will end in a draw.	If betting on <i>teamHome</i> to win and <i>teamHome</i> wins the match, you win the bet. If the match ends in a draw or the away team wins, you lose the bet.
14	Over Under 0.5 Goals	Bets are placed on whether the total number of goals scored in the match will be over or under 0.5.	If betting on <i>Over 0.5 Goals</i> and at least one goal is scored, you win. If no goals are scored, you lose the bet. For <i>Under 0.5 Goals</i> , if no goals are scored, you win. If at least one goal is scored, you lose the bet.
15	Over Under 1.5 Goals	Bets are placed on whether the total number of goals scored in the match will be over or under 1.5 goals.	If betting on <i>Over 1.5 Goals</i> and the match ends with 2 or more goals, you win. If the match ends with 1 or no goals, you lose the bet.
16	Match Odds and Both Teams to Score	Bets are placed on the match outcome (Home win, Draw, Away win) and whether both teams will score at least one goal each.	If betting on <i>Home Win</i> and the home team wins, you win the bet. If betting on <i>Both Teams to Score</i> and both teams score, you win the bet. If either or both conditions are not met, you lose the bet.
17	Asian Handicap	A form of handicap betting where one team is given a virtual advantage or disadvantage to balance the odds.	If betting on <i>teamHome-1</i> , <i>teamHome</i> must win by more than one goal for a win. If <i>teamHome</i> wins by exactly one goal, the bet is a push (stake returned). If <i>teamHome</i> draws or loses, the bet loses.



**25.bz2 (marketID)
(include all 24 files)**

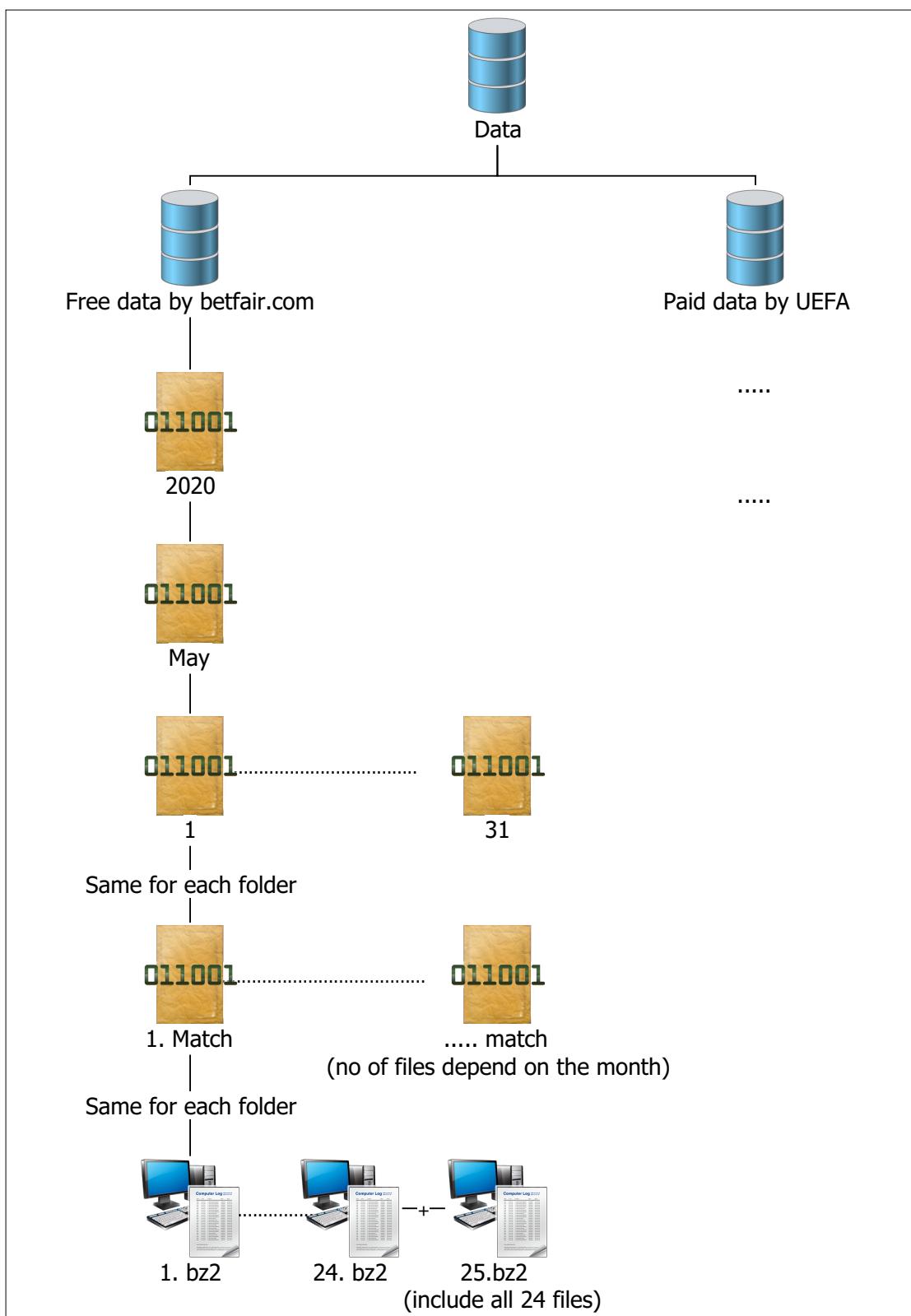


1.Market_type.csv



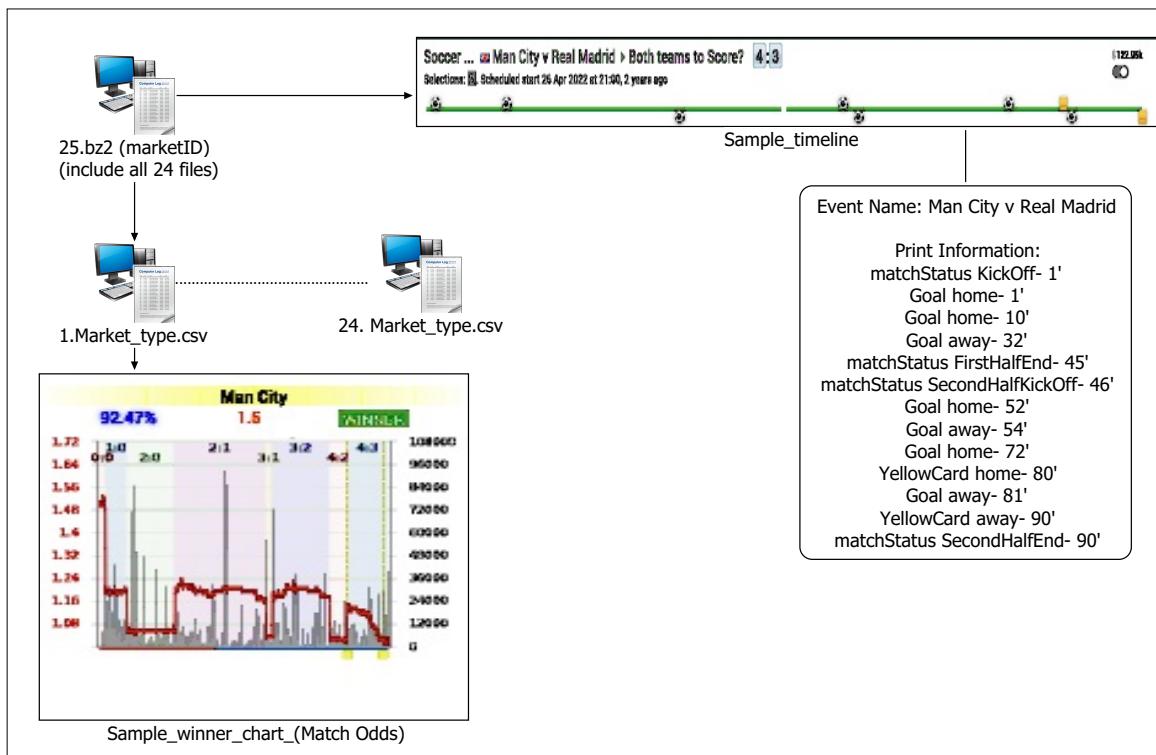
24. Market_type.csv

Figure 23: Visualisation



Page 1 sur 1

Figure 24: Data Organization



Page 1 sur 1

Figure 25: Visualisation

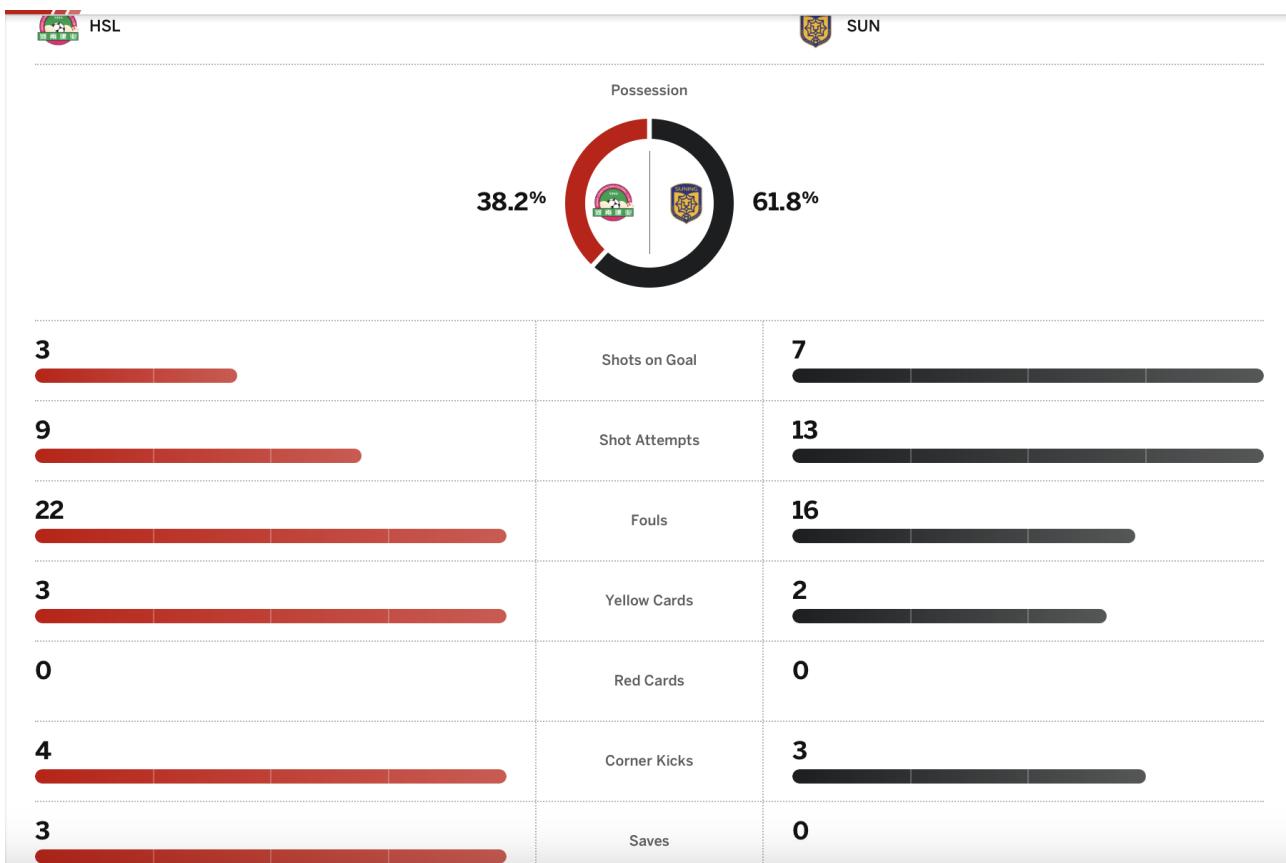


Figure 26: Henan vs Jiangsu (Match Stats)

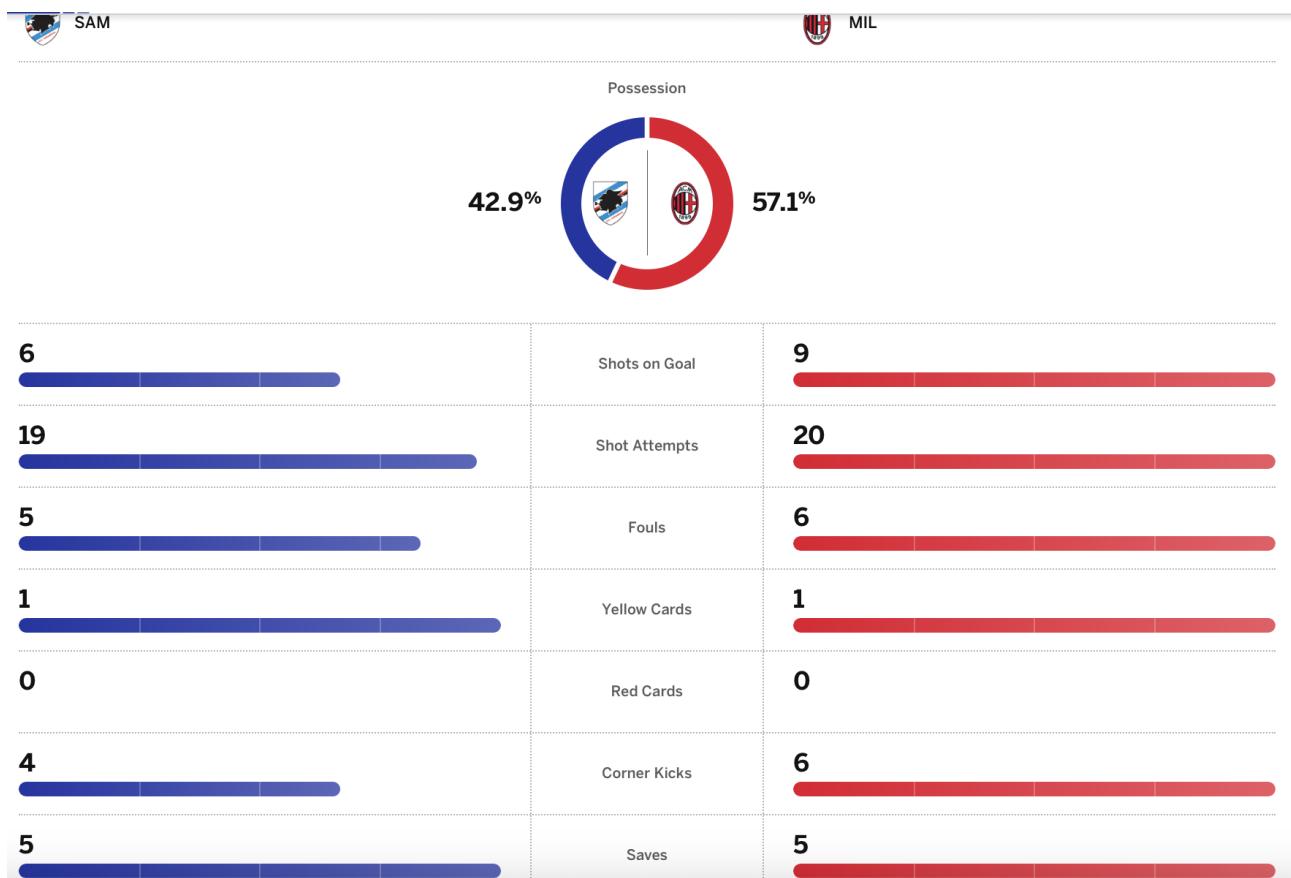


Figure 27: Sampdoria vs AC Milan (Match Stats)

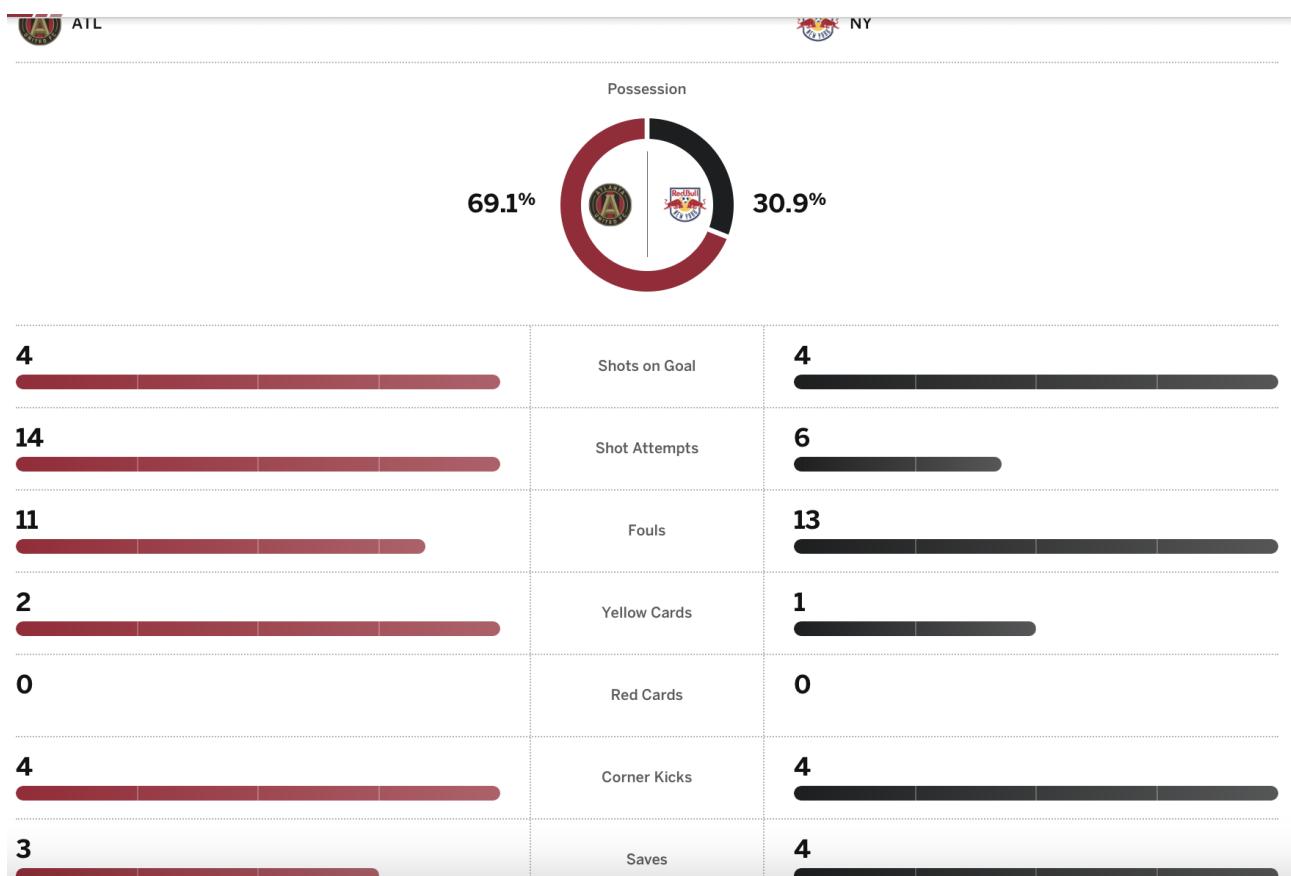


Figure 28: Atlanta Utd vs New York Red Bulls (Match Stats)

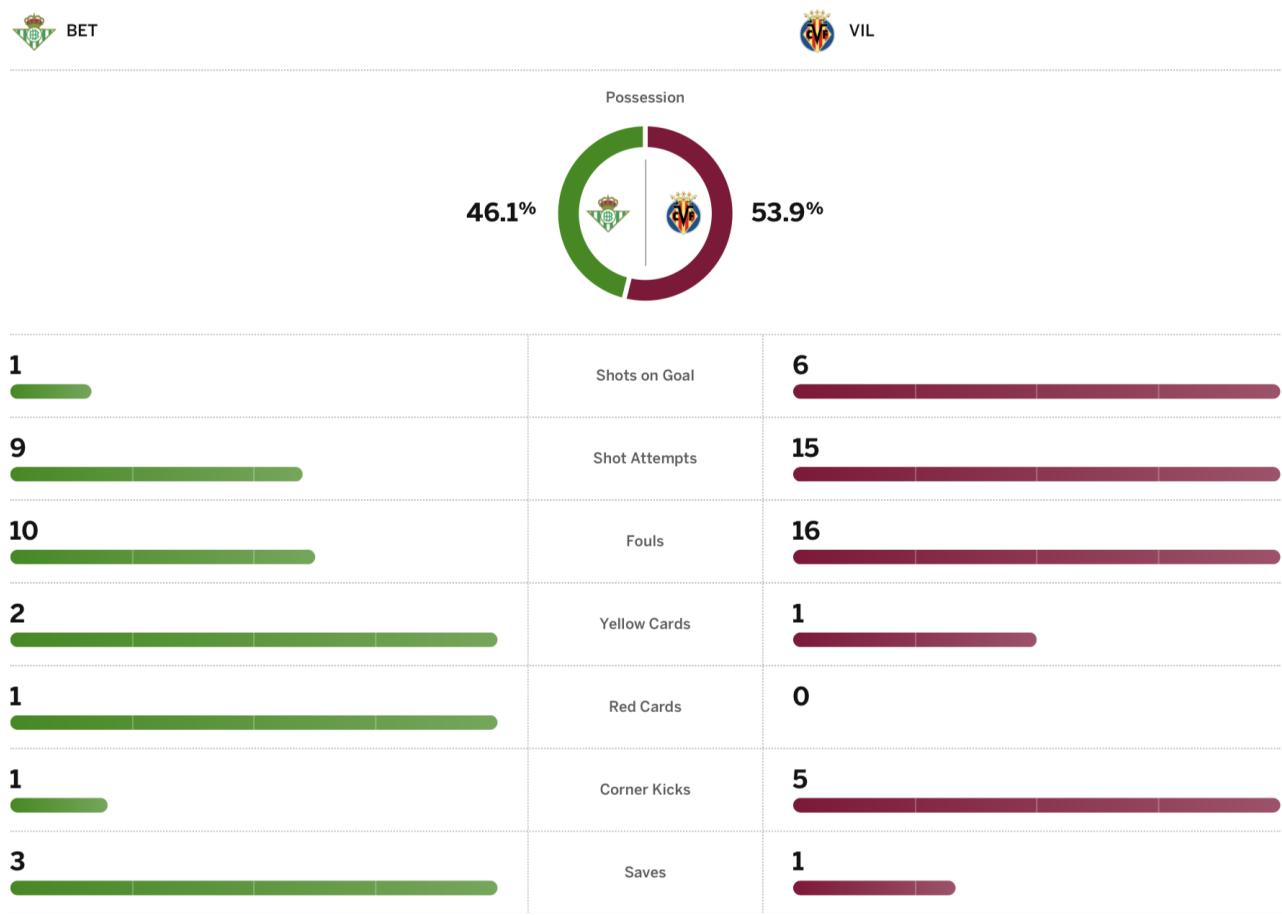


Figure 29: Betis vs Villarreal (Match Stats)

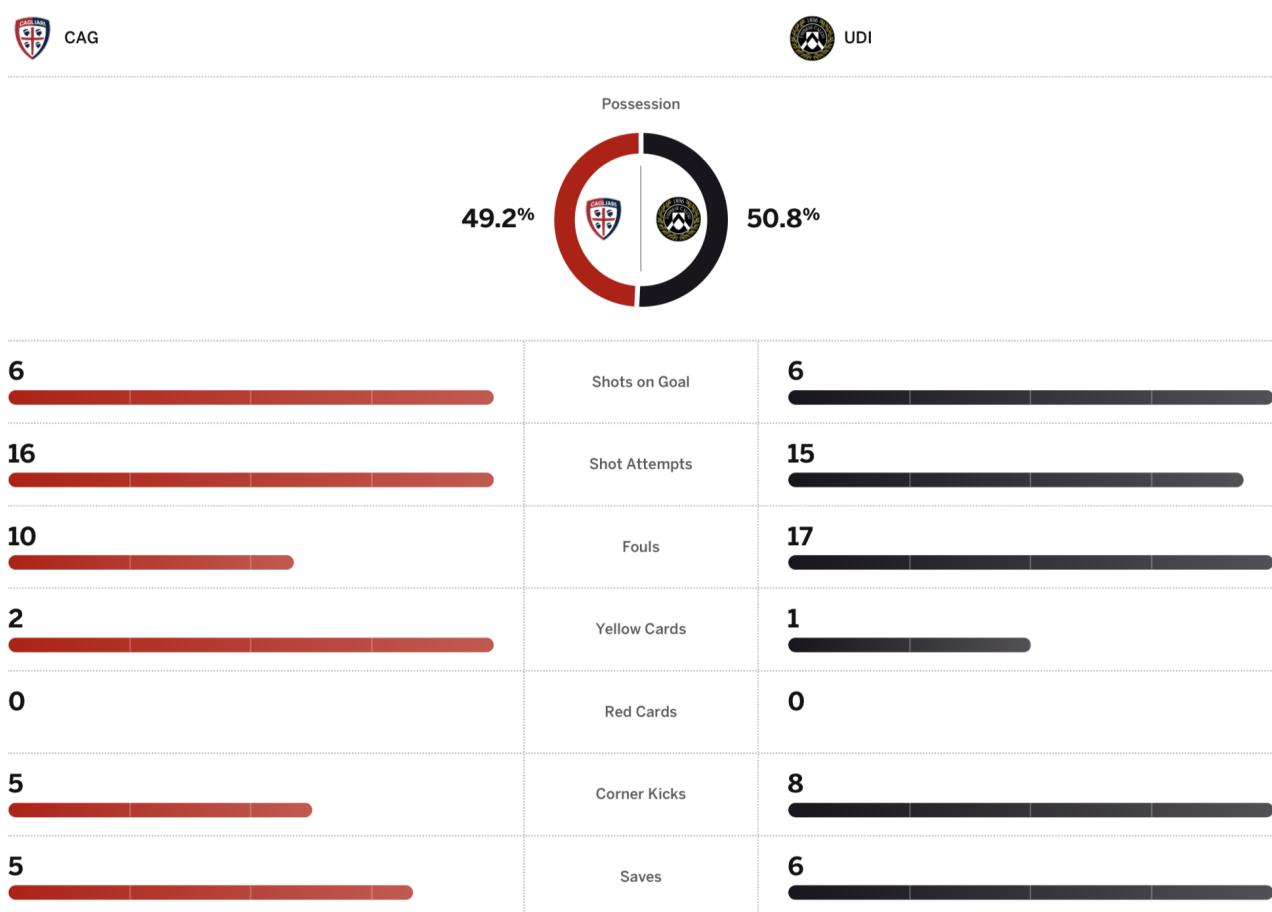


Figure 30: Cagliari vs Udinese (Match Stats)

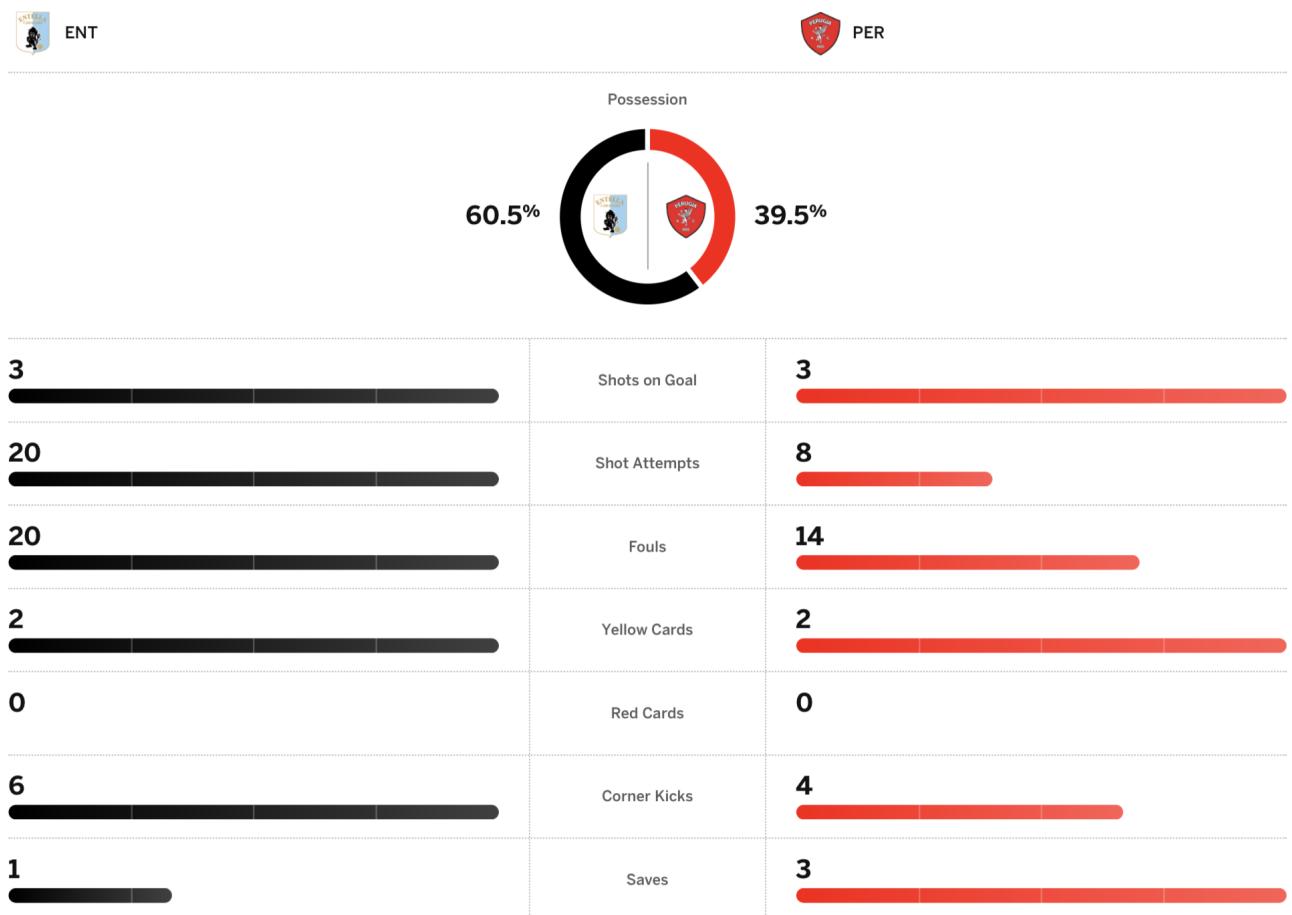


Figure 31: Entella vs Perugia (Match Stats)

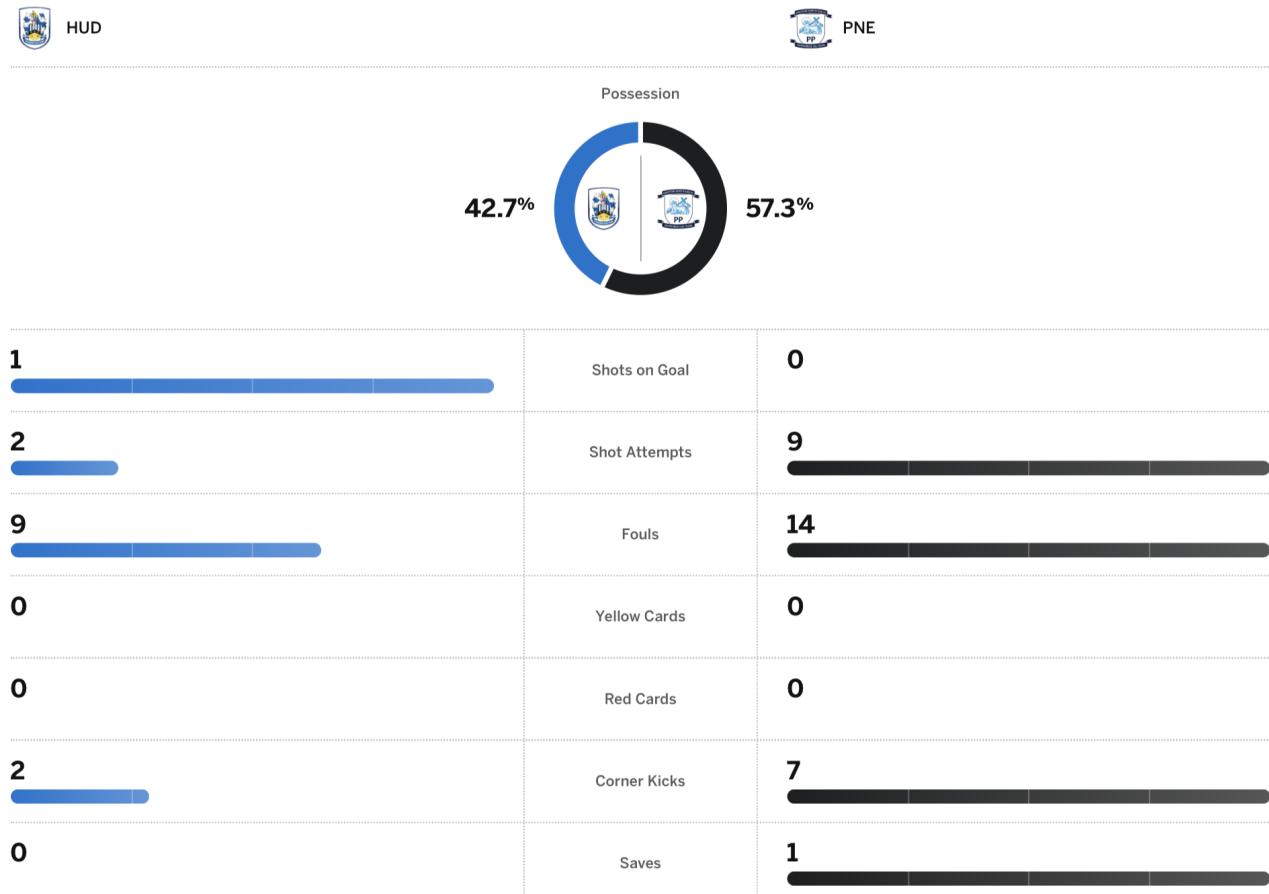


Figure 32: Huddersfield vs Preston (Match Stats)

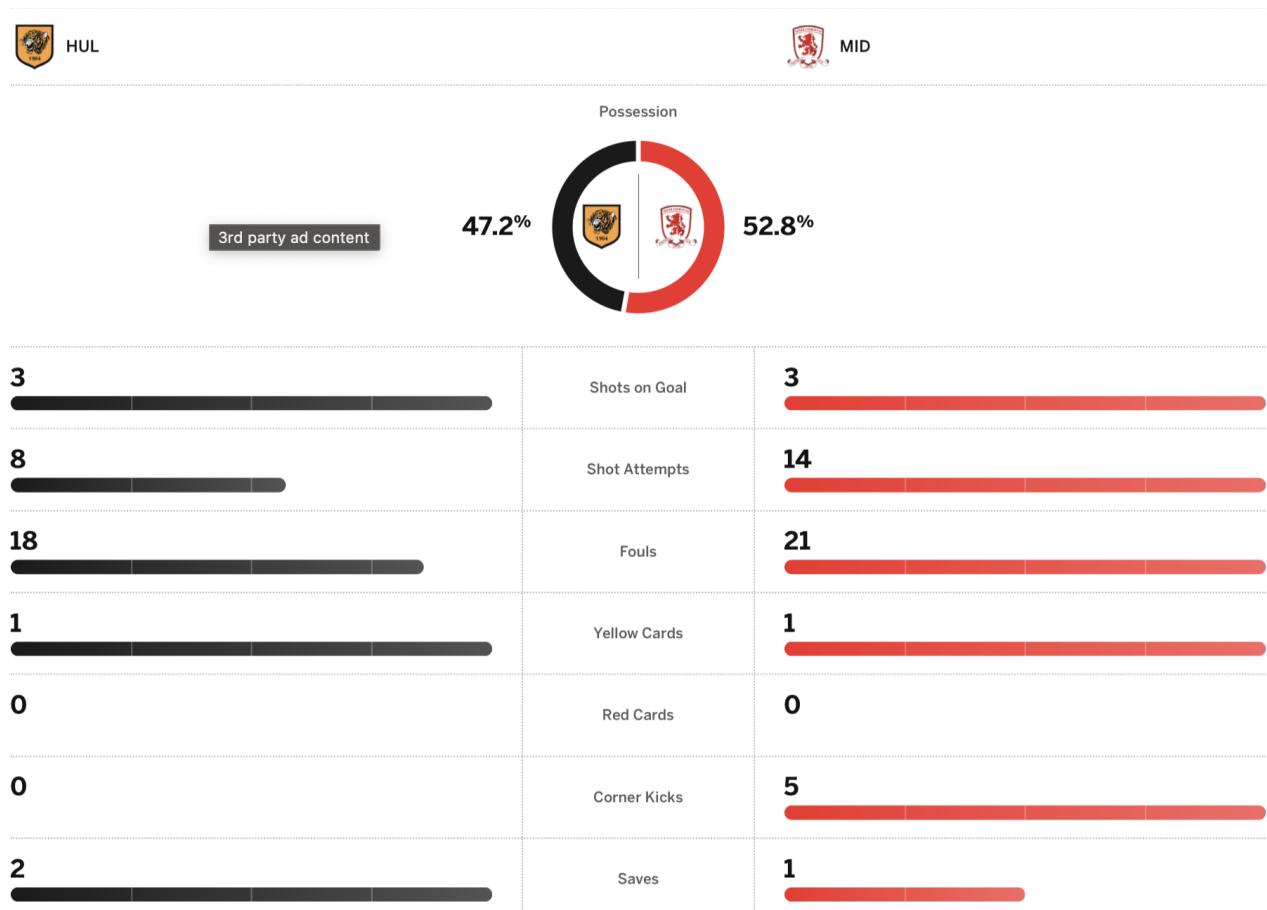


Figure 33: Hull vs Middlesbrough (Match Stats)



JVS



ENT

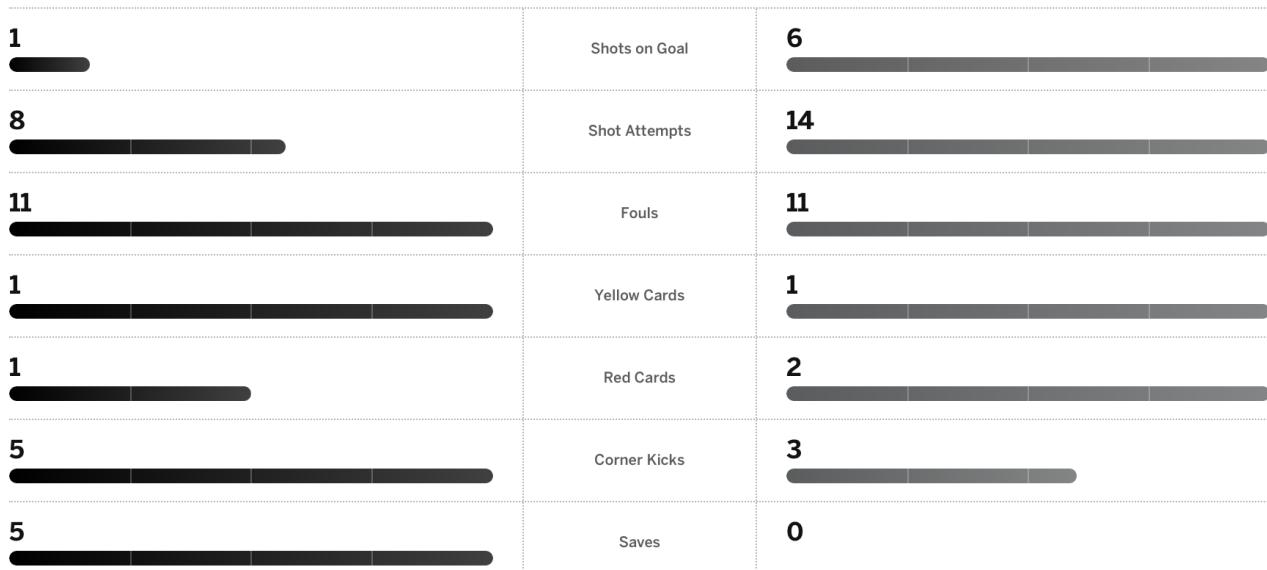
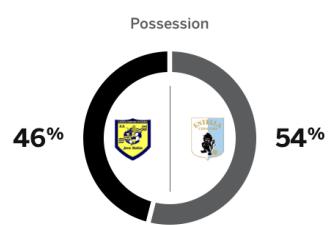


Figure 34: JuveStabia vs Entella (Match Stats)

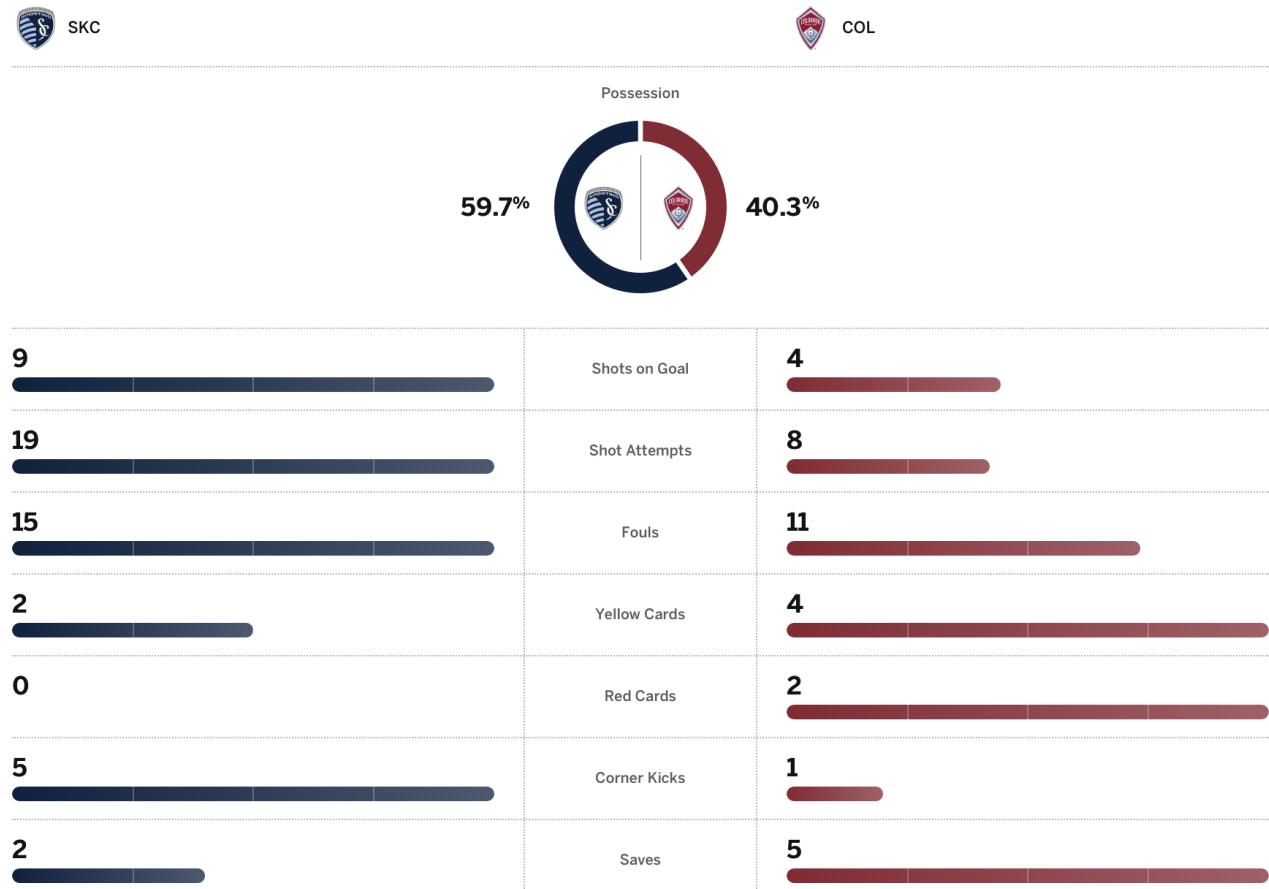


Figure 35: Kansas City vs Colorado (Match Stats)

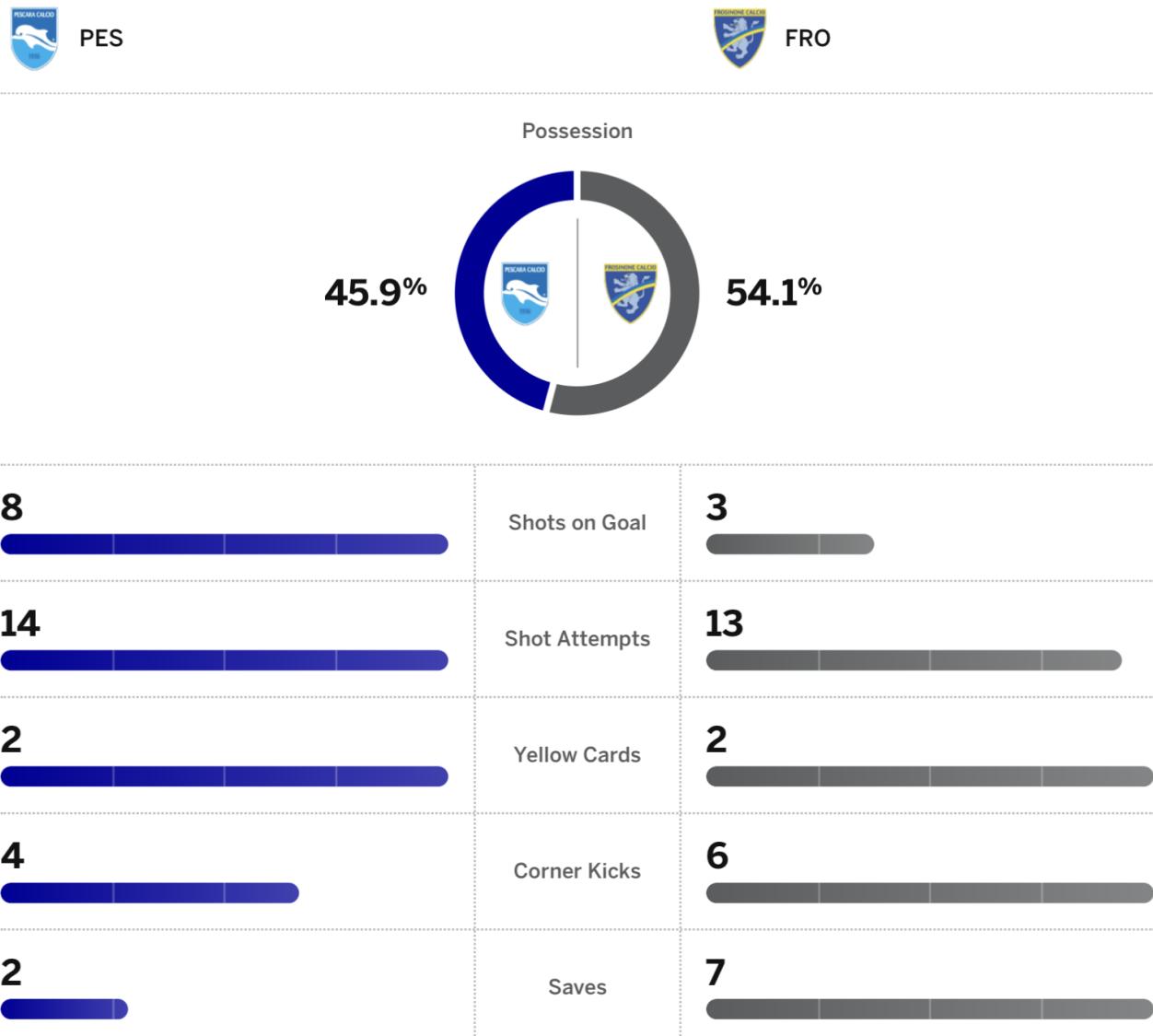


Figure 36: Pescara vs Frosinone (Match Stats)

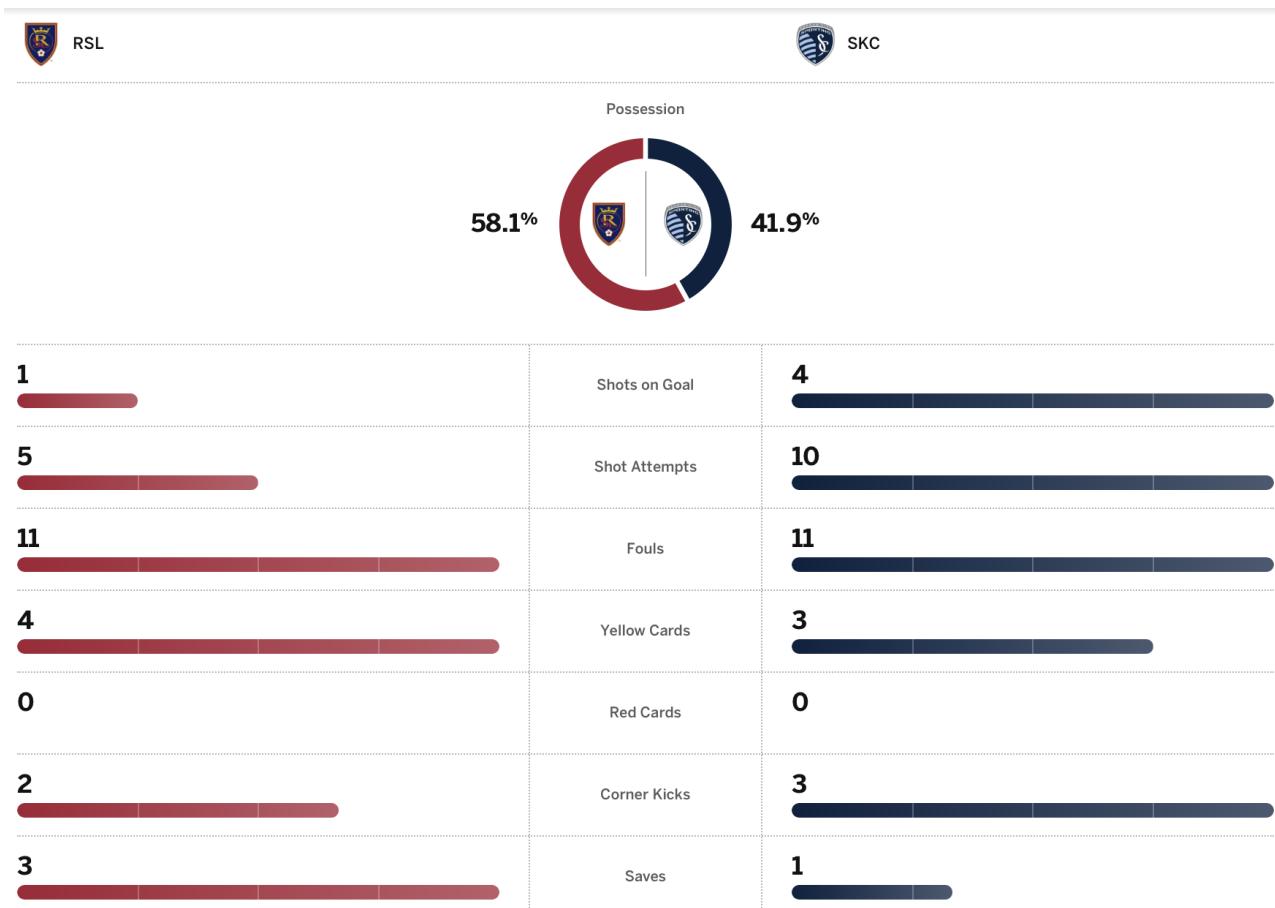


Figure 37: Real Salt Lake vs Kansas City (Match Stats)



RIO



PTMO

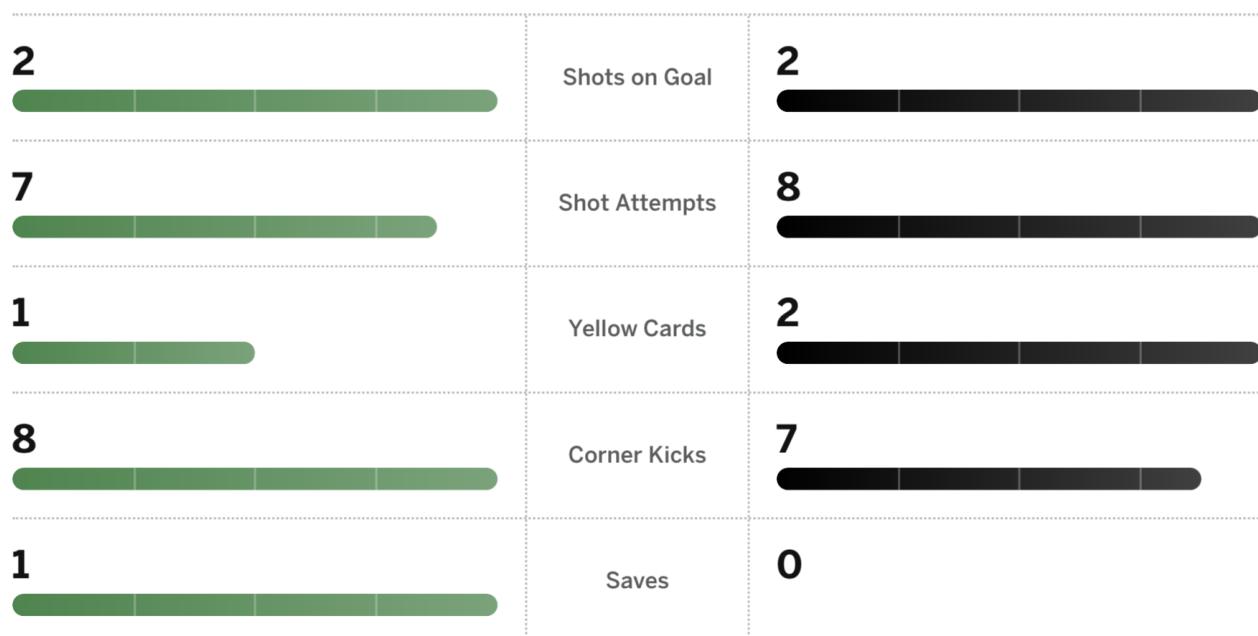
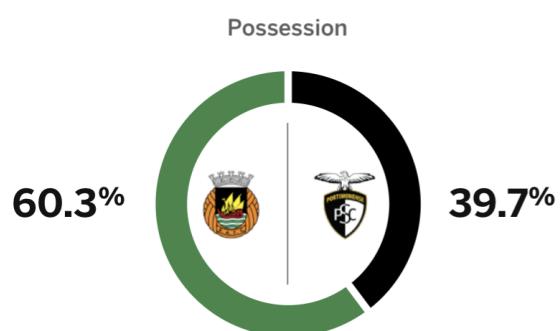


Figure 38: Rio Ave vs Portimonense (Match Stats)

TEAM STATS		
	Shots	11
14	Shots on target	3
6	Possession	55%
45%	Passes	469
390	Pass accuracy	81%
78%	Fouls	8
12	Yellow cards	5
4	Red cards	0
0	Offsides	3
3	Corners	5
4		

Figure 39: Rubin Kazan vs Rostov (Match Stats)



WAT



NOR

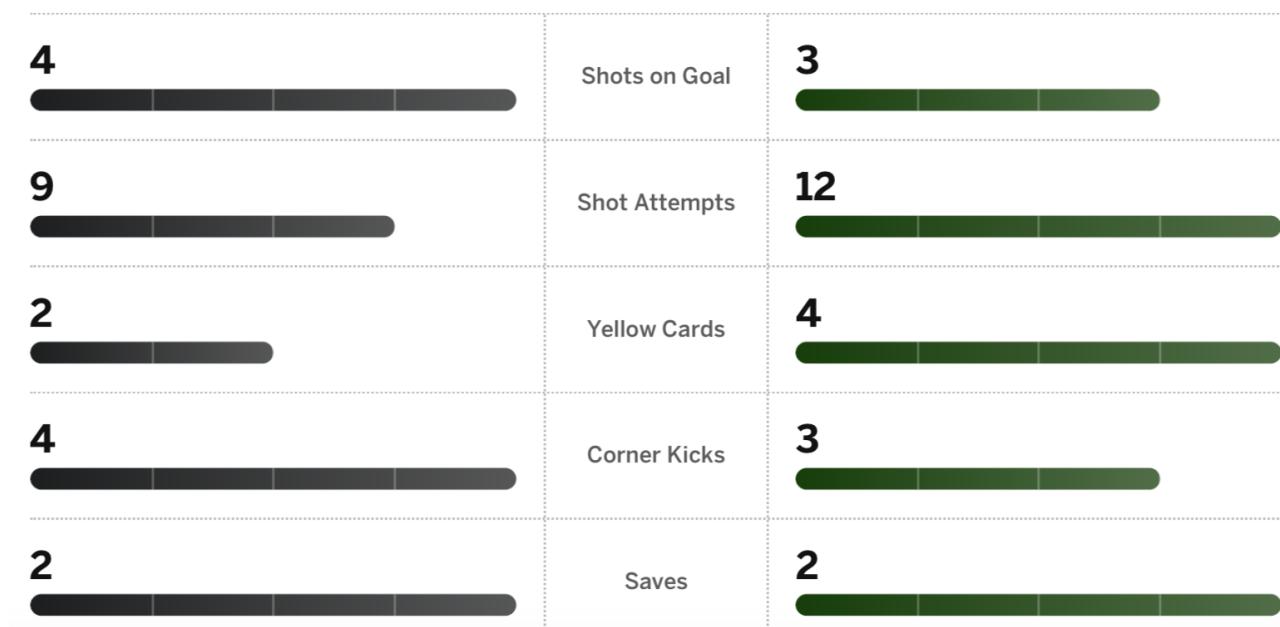
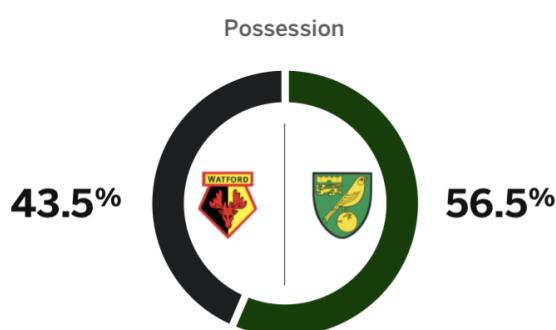


Figure 40: Watford vs Norwich City (Match Stats)

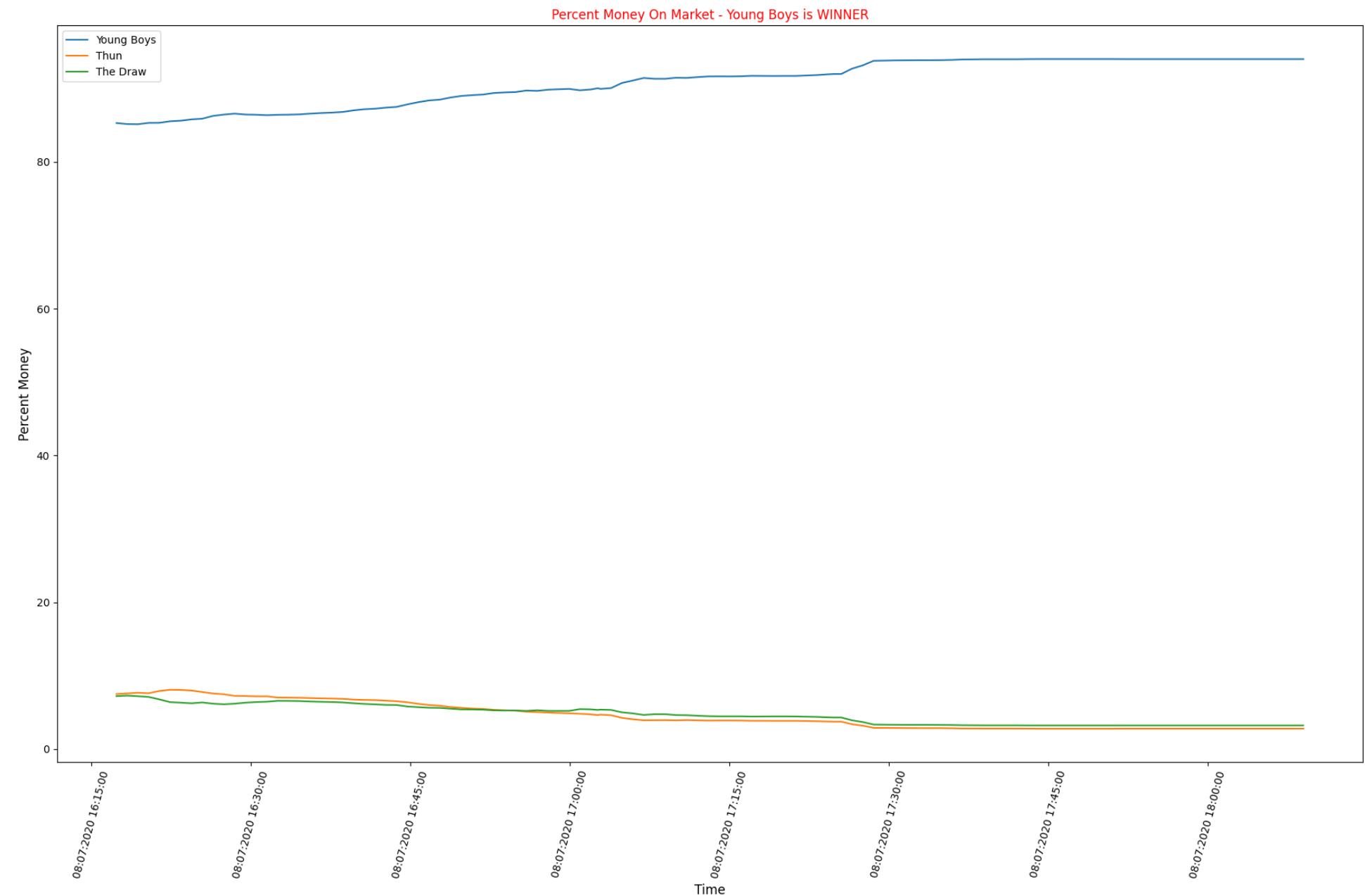


Figure 41: Sample (Percentage of money) YoungBoys vs Thun (MatchOdds)

Total Money On Market - Any Other Home Win is WINNER

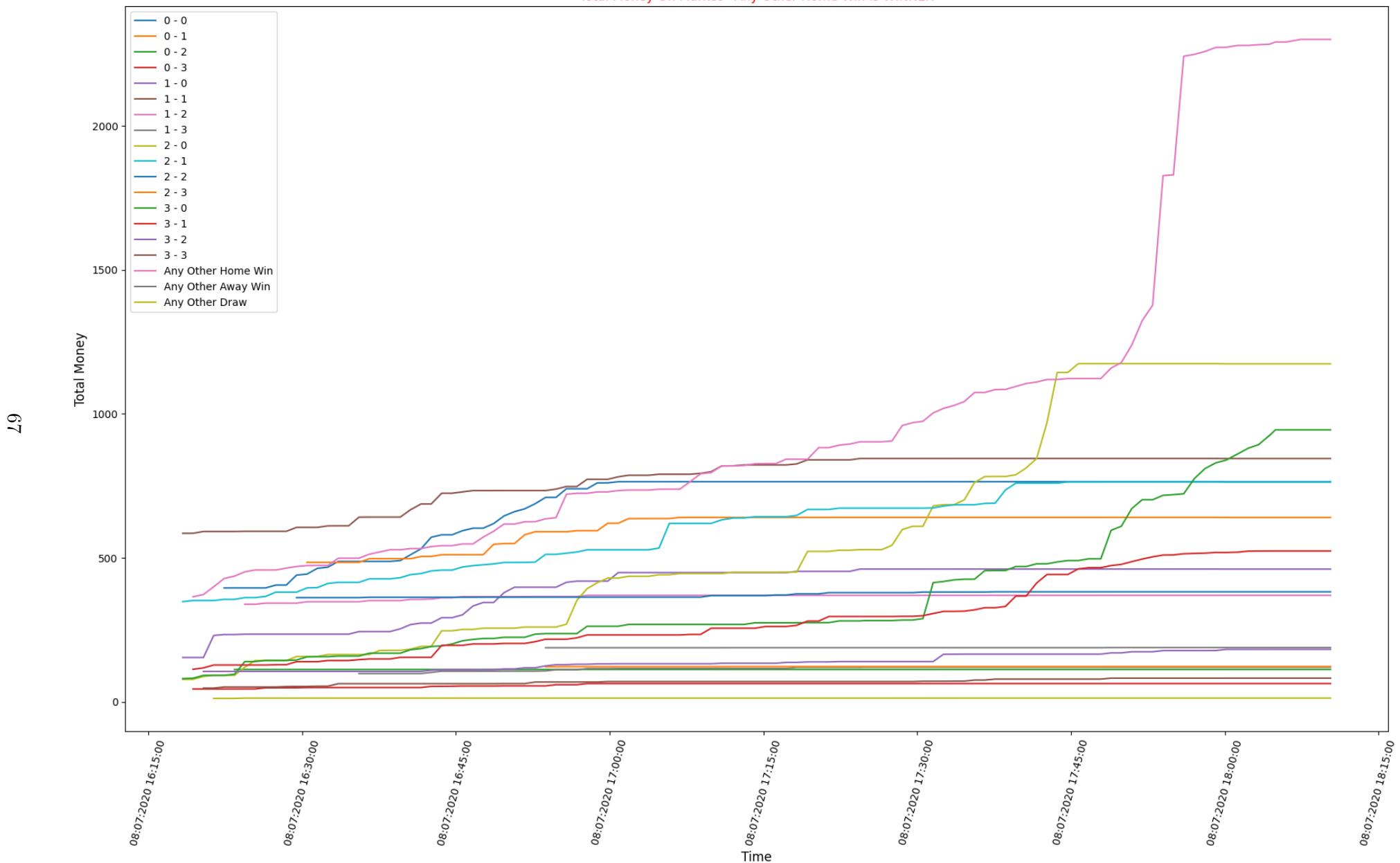


Figure 43: Sample (Total Money) YoungBoys vs Thun (CorrectScore)

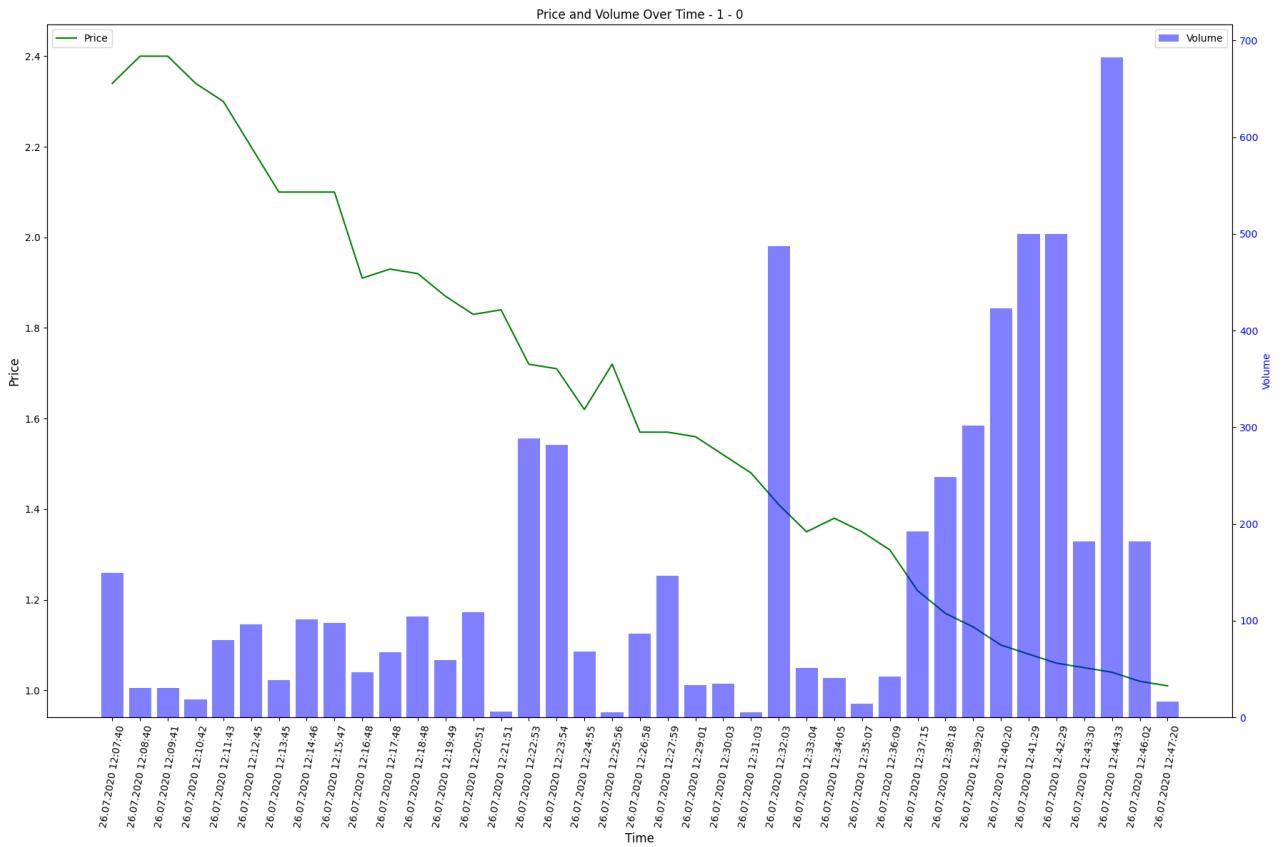


Figure 46: 1-0_(Henan)

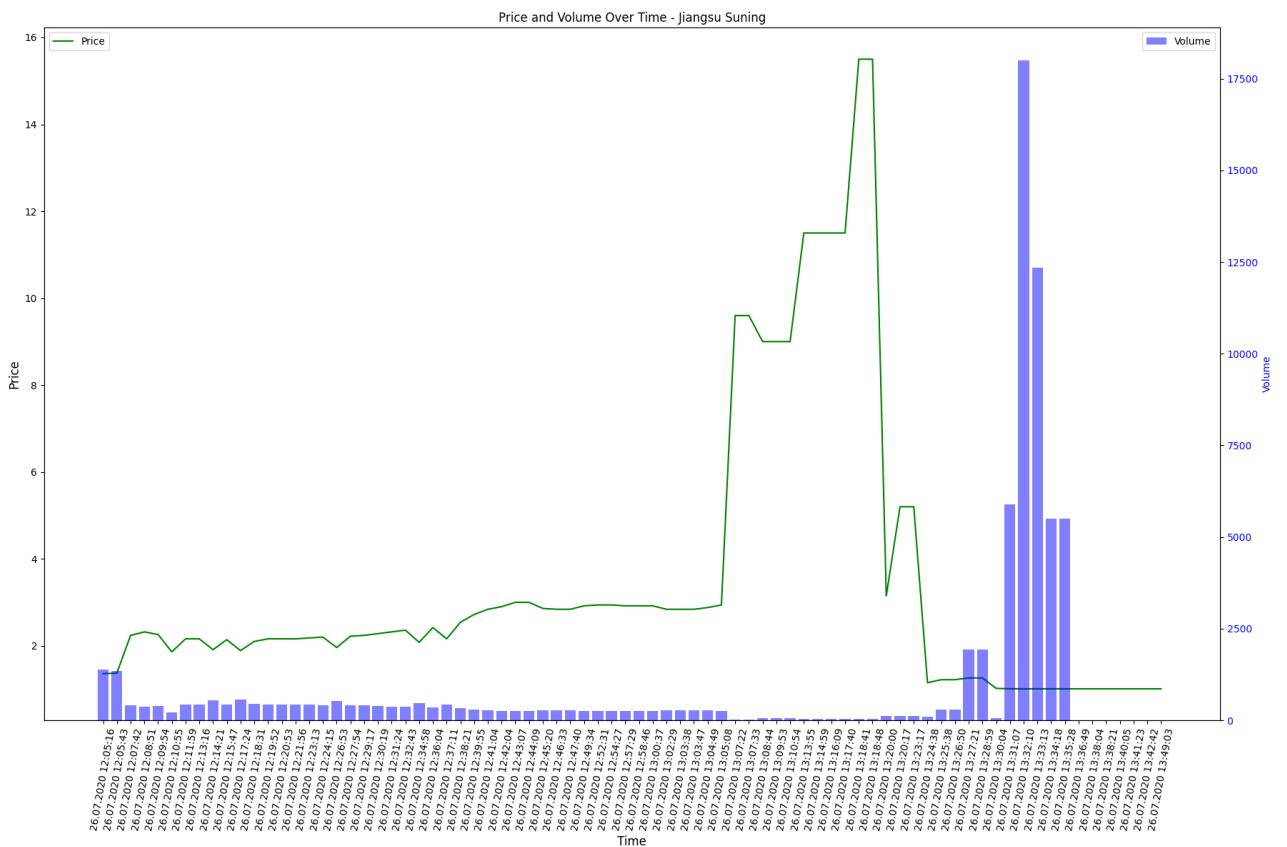


Figure 47: Jiangsu Suning_ (Henan)

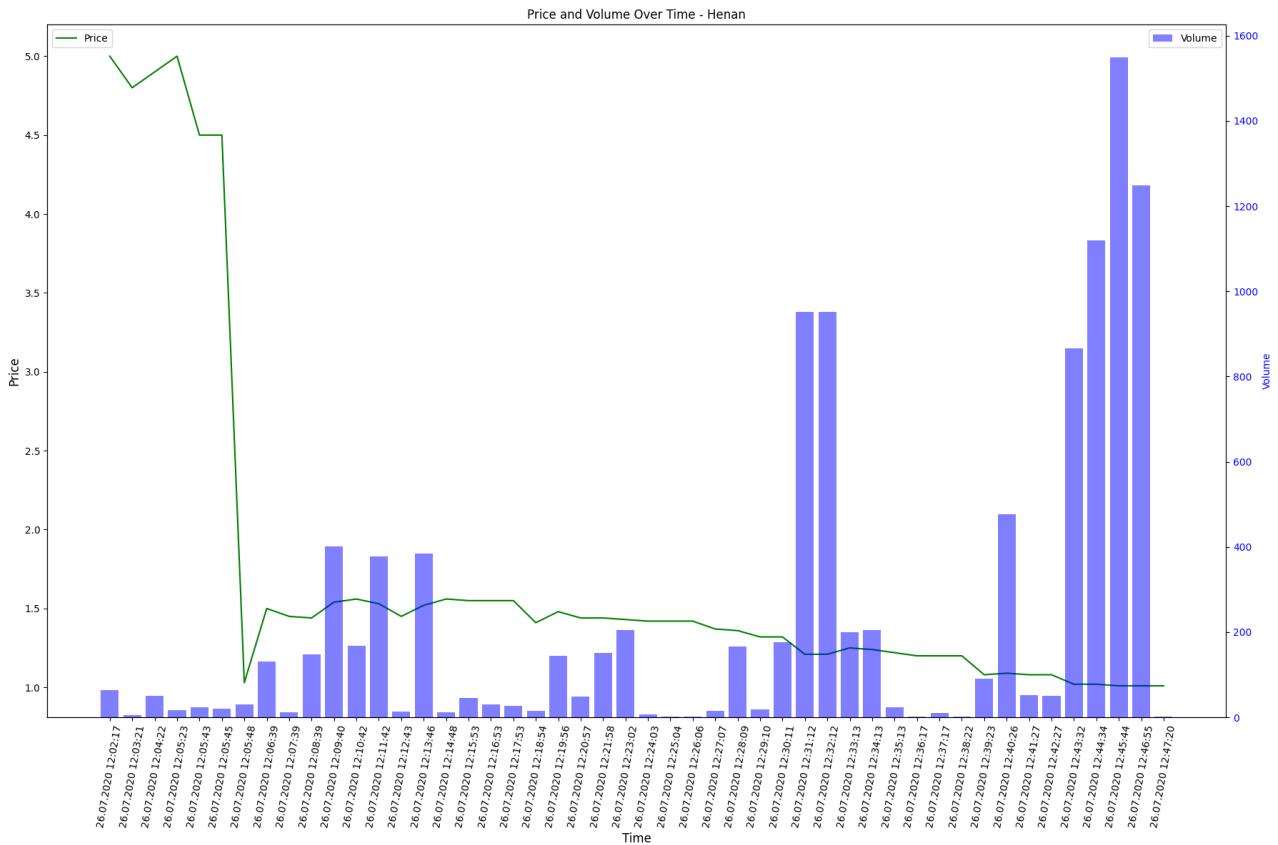


Figure 48: Henan_option_(Henan)

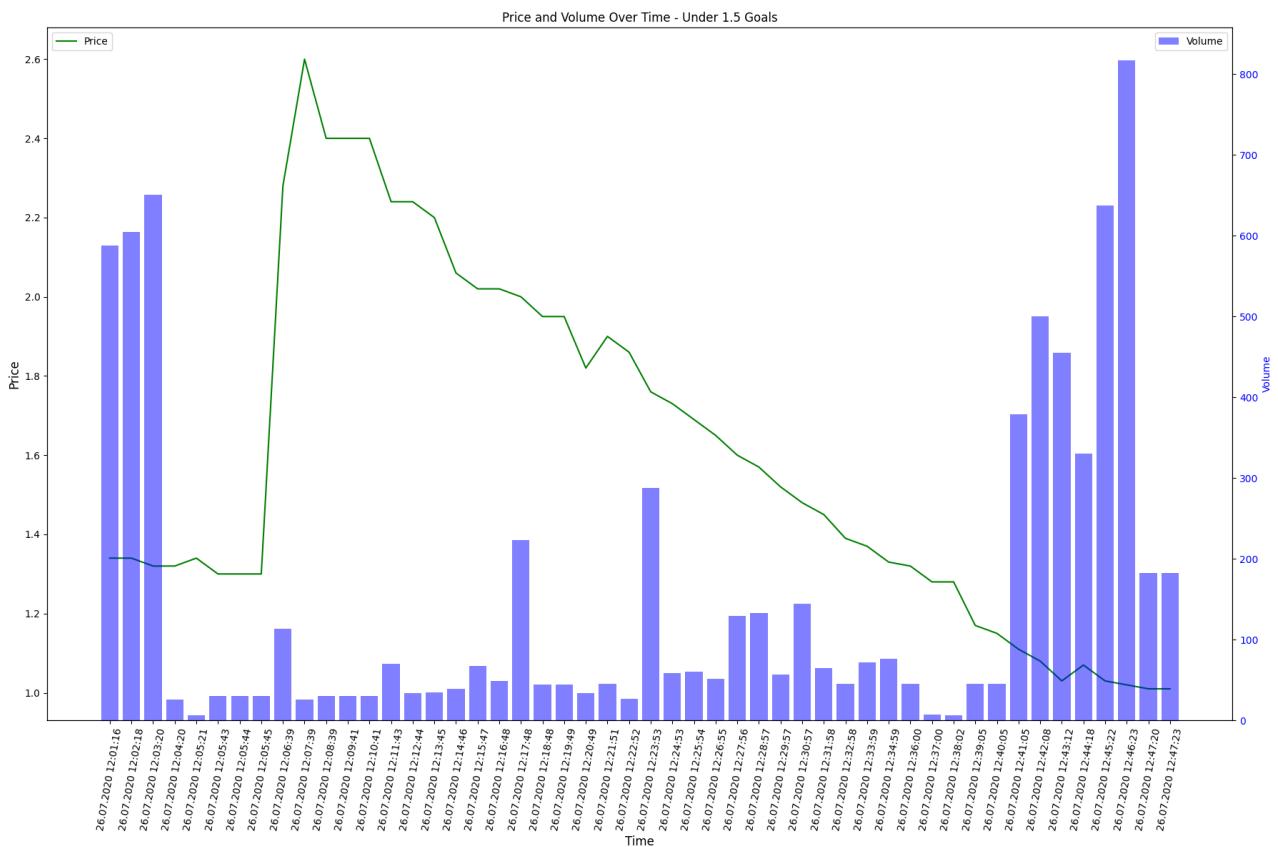


Figure 49: Under 1.5 Goals_ (Henan)