NOVA IMS Information Management School

MGI

Master Degree Program in
Information Management

Business Intelligence in Football Scouting System

Dany Oliveira

Project Work

Presented as partial requirement for obtaining the Master Degree Program in Information Management

NOVA Information Management School

Universidade Nova de Lisboa

NOVA Information Management School
Universidade Nova de Lisboa

Business Intelligence in Football Scouting System

By
Dany Oliveira

Project work presented as a partial requirement for obtaining the Master's Degree in Information Management, with a specialization in Knowledge Management and Business Intelligence

Supervised by
Professor Nadine Côrte-Real, PhD, NOVA Information Management School
July, 2024

STATEMENT OF INTEGRITY

I declare having conducted this academic work with integrity, and I confirm that I have not used plagiarism or other illegal way of conducting this study. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School

Lisbon, July 2024

DEDICATION

To my parents that support me along of my live and a special dedication for the person that take care of me along this master, my lovely sister.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my mentor and thesis advisor, Professor Nadine Côrte-Real, PhD, for the support and attention along of the master theses process. Her insights, feedback, and group dynamics were crucial to achieve the final result.

Thank you to all the professors of the master's, from them, I can learn more about information management. Knowledge that I am trying to apply in my work and in this thesis.

Finally, to ReField, a company that I created years ago and keep the wild spirit that world needs.

ABSTRACT

Scouting departments play an important role in improve clubs and companies by being an important piece in the process of squad refinement. However, to maximize efficiency, a solid plan and decision-making processes are imperative. ReField was built to help scouting departments improving their strategies. By implementing a BI tool, ReField aims to provide invaluable insights and actionable information, empowering stakeholders to make informed decisions supported by real-time data. This innovative approach not only enhances business performance but also ensures that every decision is backed by tangible evidence. This project uses the design science research (DSR) methodology to develop a business intelligence (BI) dashboard using Microsoft Power BI tool. The value that this project wants to add for companies and clubs is the possibility to have a visual analytics solution that presents important information to sustained the decision-making process.

KEYWORDS

Business Intelligence; Dashboard; Football; Scouting; Decision-making; Power BI; Design Science Reasearch;

SUSTAINABLE DEVELOPMENT GOALS (SDG)

INDEX

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

BI	Business Intelligence
DSR	Design Science Research
DW	Data Warehouse
ETL	Extract, Transform, Load
IS	Information System
KPI	Key Performance Indicator

1. Introduction

1.1 Problem identification and motivation

In recent years, there has been a growing interest in exploring big data and analytics to enhance organizational performance (Manyika et al., 2011). However, adopting advanced data management systems remains a challenge for scouting departments within medium and smaller sports teams worldwide. This is exemplified by the case of first-division clubs in Portugal, where elementary tools in terms of scouting like Microsoft Excel are still used for managing scouting operations, which creates a big challenge in the maintenance of information.

Recognizing this gap, ReField emerges as a pioneering scouting platform designed to be a tool that helps clubs and companies daily (refield.app). Without the appropriate tools to do the job decision-making is often based on preconceived notions. For instance, when companies contemplate expansion, conventional wisdom might dictate selecting locations based on personal networks. However, data-driven analysis can reveal superior and more reliable options by considering various factors such as business frequency, networking opportunities, and costs (Erevelles et al., 2016).

According to LaValle et al. (2011), high- performing organizations demonstrate a propensity to make decisions based on rigorous data analysis at a rate exceeding double that of lower-performing counterparts. These organizations leverage analytic insights to inform both strategic planning and day-to-day operations, thereby maintaining a competitive edge in their respective industries.

1.2 Contextualization

Inside a scouting department, effective communication is one of the most important aspects, ensuring well-organized channels to facilitate operations. This necessity for direct and clear communication was the main goal behind the establishment of ReField in 2018.

ReField was built to be a comprehensive scouting platform designed to improve the management of scouting departments. It gives the possibility for users to generate reports, coordinate team tasks, schedule events, manage scouting teams, and oversee market operations. From its launch,

ReField has prioritized the development of a practical tool tailored for real-world scouting scenarios and focus on being easily understandable to be used for everyone. Usability is very important because many of ReField clients are predominantly former athletes, possess basic knowledge of information technology.

Through interactions with various stakeholders in the scouting community, ReField has identified significant potential in integrating business intelligence with scouting operations.

Analyses conducted across diverse sectors underscore the critical importance of two factors for achieving a competitive advantage: rapid access to accurate information and the ability to discern and utilize information selectively (Olszak & Ziemba, 2003).

1.3 Research objectives

The research objectives of ReField are based on the principles of business intelligence (BI), which is increasingly recognized as essential for informed decision-making in modern organizations (Power, 2008). The aim is to provide ReField customers with a BI experience that offers a comprehensive overview of their scouting team dynamics, enabling prompt analysis and response to management and decision-making requirements aligned with organizational strategies (Moss & Atre, 2003).

This feature seeks to answer the following research question: **who the clubs and companies are doing and if it agrees with their strategy?**

The main goal of this project is to offer a sophisticated dashboard interface to empower directors and owners with a powerful tool that helps them to lead scouting team performance, identify trends, and make strategic adjustments promptly.

Business intelligence (BI) should be used in a daily base of modern enterprises, embodying the systematic analysis of data to extract actionable insights crucial for decision-making. Without data, you are just another person with an opinion (Dresner, 2009). BI is a vital tool in transforming raw data into meaningful information to empower organizations. The goal is to turn data into information, and information into insight. Such insight-driven decision-making is instrumental in driving operational efficiency, fostering innovation, and gaining a competitive advantage in dynamic markets. (Marr, 2022)

1.4 Study Relevance

According to Kimball, R. & Ross, M. (2002), information builds the foundation of effective BI systems, enabling data collection, organization, and analysis to derive actionable insights. H. Dresner (2011) emphasizes that BI empowers businesses to make data-driven decisions, minimizing risks and capitalizing on opportunities. Through advanced analytics and visualization, BI tools enable organizations to interpret complex data sets, driving innovation and continuous improvement.

Scouting departments are made up of different visions and missions, but everyone should be converging in the same goal. This project will be an essential part to help in the creation of the concept ‘single version of truth', ensuring that decision-makers have access to consistent, reliable information for effective strategic planning and execution (Inmon, 2002). This iterative process is fundamental for adding long-term value and driving sustained performance enhancements within sports organizations (Power, 2008).

2. Literature Review

2.1 Business Intelligence

Business Intelligence (BI) is a multifaceted process involving the collection, processing, and validation of data from diverse sources that impact an organization's operations, followed by the analysis and transformation of this data into strategic knowledge (Massoud et al., 2004).

Business Intelligence (BI) has evolved significantly over the years, transforming from basic reporting tools to sophisticated analytics platforms. The term "business intelligence" was created by Hans Peter Luhn, an IBM researcher, in 1958, to describe the process of transforming data into actionable insights for business decision-making (Luhn, 1958). After 20 years, there was the rise of data warehousing (Watson & Wixom, 2001), while the 1990s witnessed the advent of reporting tools like Crystal Reports and Business Objects. The 2000s marked a shift towards analytics with the growth of the internet and e-commerce (R. Sabherwal & I. Becerra-Fernandez, 2011). In the 2010s, in-memory computing and real-time analytics became prevalent, along with the rise of self-service BI tools like Tableau and Power BI (Watson & Wixom, 2001).

In a comprehensive review by Wanda and Stian (2015), the primary benefits derived from BI include making better decisions, providing decision support, and enhancing business processes. It plays a critical role in enabling organizations to gain insights from their data, leading to improved operational efficiency and competitive advantage (R. Sabherwal & I. Becerra-Fernandez, 2011). According to Howard Dresner (2009), BI is about using data to make better decisions. L. Moss and S. Atre (2003) also emphasize how BI supports decisions across different parts of a company. BI helps companies understand what works and what does not, making them more successful in the long term.

Developing a BI strategy that adds significant business value requires careful consideration and acceptance by the company's board of directors. In order to achieve it is fundamental to identify the critical business activities that must be effectively executed to drive organizational success.

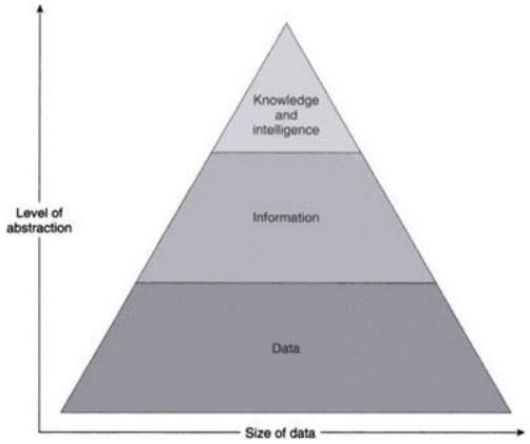

Figure 1- BI level of abstraction

2.2 Project Architecture

Business Intelligence comprises a structured architecture consisting of five components: Data Sources, ETL (Extract, Transform, Load), Data Warehousing, Reporting and Dashboarding, and End Users. R. Kimball and M. Ross (2002) talk about this setup, which helps organizations collect, process, store, and present data in a useful way for making decisions. Each part has a role in the

Business Intelligence system, from gathering data to making it understandable for people who need it.

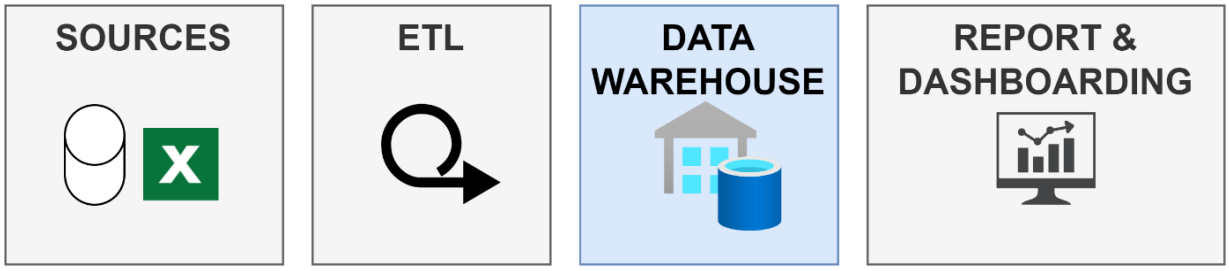

Figure 2- Project Architecture

2.2.1 Data Sources

This component refers to the various systems, databases, and applications where data originates. These sources may include transactional databases, CRM systems, ERP systems, spreadsheets, and more (Inmon, 2002).

2.2.2 ETL (Extract, Transform, Load)

ETL is a fundamental process in data management and analytics. ETL is very important in the data warehousing process, it has significance in integrating diverse data sources and ensuring data quality for effective decision-making and strategic planning within organizations (Kimball and Caserta, 2004). The ETL workflow typically involves the following stages:

- **Extract**: It involves identifying and accessing relevant data sets from disparate sources, ensuring completeness and accuracy in data retrieval. (Inmon, 2002)
- **Transform**: It is essential for converting raw data into a structured format that aligns with the requirements of the target system or data warehouse. (Kimball and Caserta, 2004)
- **Load**: It ensures that the transformed data is organized and accessible for reporting, querying, and decision-making activities within the organization. (Inmon, 2002)

2.2.3 Data Warehousing

According to Kimball, R. & Ross, M. (2002), a data warehouse is a copy of transaction data specifically structured for query and analysis. Kimball also emphasizes the importance of

designing a data warehouse to support the analytical needs of the organization efficiently. Bill Inmon, defines a data warehouse as a subject-oriented, integrated, time-variant, and nonvolatile collection of data in support of management's decision-making process. Inmon's emphasis on integration highlights the importance of consolidating data from multiple sources into a unified format within the data warehouse. There are key characteristics in a data warehouse such as:

- **Subject-Oriented**: Data is organized around key subjects or business areas to support analytical queries specific to those domains. (R. Kimball and M. Ross, 2002)
- **Integrated**: Data from various operational systems across the organization is consolidated and standardized within the data warehouse, ensuring consistency and reliability in reporting and analysis. (Ponniah, 2001)
- **Time-Variant**: The data warehouse retains historical information, allowing users to analyze trends and patterns over time to identify long-term insights and make forecasts. (Inmon, 2002)
- **Nonvolatile**: Once data is loaded into the data warehouse, it is rarely modified or deleted. This preserves a reliable historical record and ensures data consistency for decision-making processes. (Golfarelli & Rizzi, 2010)

A well-defined data warehouse architecture becomes indispensable to provide users with prepared, structured, and transformed data for BI reporting, dashboards, and other analytics tools (Orenstein et al., 2017). Part of this well-defined data warehouse is the existence of a staging area. Ralph Kimball, describes the staging area as a crucial component in the ETL process. It works as a landing zone where data extracted from various source systems is initially deposited before undergoing cleansing, validation, and transformation to align with the data warehouse schema.

Kimball and Inmon give a critical role of data warehouses and staging areas in data management. While a data warehouse stores structured historical data for analysis, a staging area facilitates data cleansing and integration, ensuring quality and reliability for decision-making processes.

2.2.4 Reporting and Dashboarding

A business intelligence (BI) dashboard provides users with a comprehensive overview of the process situation using data from various sources. This interdependent relationship enables faster and more informed decision-making (Hansoti, 2010).

According to a survey, nearly 95 percent of IT executives recommend using a dashboard as a tool to ensure consistency, reliability, and accuracy, all of which are essential for enhancing decision-making capabilities (Hurwitz, 2005). The advantages of using a dashboard include the representation of data in interactive output, which is vital for enhancing user productivity and efficiency (Few, 2006). Furthermore, analyzing performance information within the context of goals and objectives allows users to reduce the time spent on data analysis and decision-making processes. Dashboards help users improve their decisions and reduce costs over time (Few, 2006).

During the development process of a dashboard, certain precautions should be taken, especially regarding design. It is crucial to avoid overwhelming the dashboard with too much information (Few, 2006). Many users may not be highly computer-literate, but the interface should be as simple as possible (Few, 2006). Additionally, audience targeting and role-based security measures should be implemented to ensure data integrity and confidentiality (Dyche, 2006).

2.2.4.1 Microsoft Power BI

Microsoft Power BI, as described by Microsoft, is a unified and scalable platform designed for self-service and enterprise business intelligence (BI) needs. With Power BI, users can seamlessly connect and visualize data from various sources, and integrate these visualizations into their applications (Microsoft, 2023).

Through Power BI projects, companies can make better decisions with confidence, seeing the insights derived from their data (Microsoft, 2023).

2.2.5 End Users

End users are the individuals or groups within an organization who consume and utilize BI reports and dashboards to make informed decisions. These users may include executives, managers, analysts, and operational staff (Turban, 2019).

2.3 Football Scouting

Scouting in football has a multifaceted approach, serving both as a means to analyze opponents and to unearth promising talents for recruitment (dos Santos, P., 2012). It embodies a philosophy of understanding the nuances to improve team strengths and address weaknesses (dos Santos, P., 2012).

Historically, football clubs relied on managers and chairmen to scout and acquire players. However, as the sport has evolved into a business (dos Santos, P., 2012), clubs have embraced structural changes, leading to the establishment of specialized departments such as the scouting department. This evolution reflects the recognition of football as more than just a sport, but also a lucrative enterprise (Smith, J., 2020).

The scouting department operates as an intermediary between a football club's specific needs and

talent, assuming the role of an impartial evaluator within the organization (Wilson, 2008). It is fundamental to be neutral in the discernment and evaluation of potential value and talent across both domestic and international markets (dos Santos, P., 2012). In practice, scouting adopts a methodical approach that encompasses needs assessment, strategic planning, meticulous player observation, and the dissemination of outcomes (Kuper, S., & Szymanski, S., 2019). These outcomes are typically encapsulated in exhaustive reports and shadow teams, providing actionable insights for the club's decision-makers (Cox, 2017).

3. Methodology

The methodology that was chosen to follow in the ReField Business Intelligence project was the Design Science Research (DSR). This methodology is fundamentally a problem-solving paradigm that puts focus on human knowledge with the creation of innovative artifacts and the generation

of design knowledge via innovative solutions to real-world problems (Hevner et al., 2004).

The goal of a DSR research project is to design new and innovative artifacts represented by constructs, models, methods, and instantiations. DSR results in information systems or other domains such as engineering, architecture, business, economics, and sports that shown to create significant economic and societal impact (Hevner, 2004; Winter & Hevner, 2013).

The performance of SDR projects has been based on several process models, such as Nunamaker, Chen & Purdin (1990), Walls, Widmeyer & El Sawy (1992), Hevner (2004) and Kuechler & Vaishnavi (2008). The most widely referenced model is one proposed by Peffers, Tuuanen, Rothenberguer & Chatterjee (2007) where the DSR process includes six steps:

- **Problem Identification and Motivation**: This initial step defines the specific research problem and justifies the necessity of a solution. It involves thorough analysis to understand the underlying issues and challenges.
- **Definition of Objectives**: Objectives of the solution are delineated, which can be quantitative or qualitative. Quantitative objectives may specify improvements over existing solutions, while qualitative objectives might focus on understanding the design process itself. These objectives must align closely with the identified problem.
- **Design and Development**: Activities such as scenario analysis, ideation, and prototyping are crucial for exploring potential market scenarios and refining the design to meet user needs effectively.
- **Demonstration**: This step is essential for illustrating how the solution functions in practical contexts and how it addresses the identified problem.
- **Evaluation**: A thorough comparison between the achieved results and the initial objectives is conducted. Positive outcomes indicate readiness to proceed to the next phase, while discrepancies may necessitate revisiting the objective definition phase to refine the solution.
- **Communication**: The final step is where the conclusions are communicated to relevant stakeholders and the broader audience. Effective communication ensures that insights

gained from the DSR process are shared and understood, facilitating knowledge dissemination and potential adoption of the solution.

These steps emphasize the need for researchers to continually refine their solutions and methodologies to address evolving market dynamics and consumer demands. (Winter, R., & Hevner, A. R. 2013)

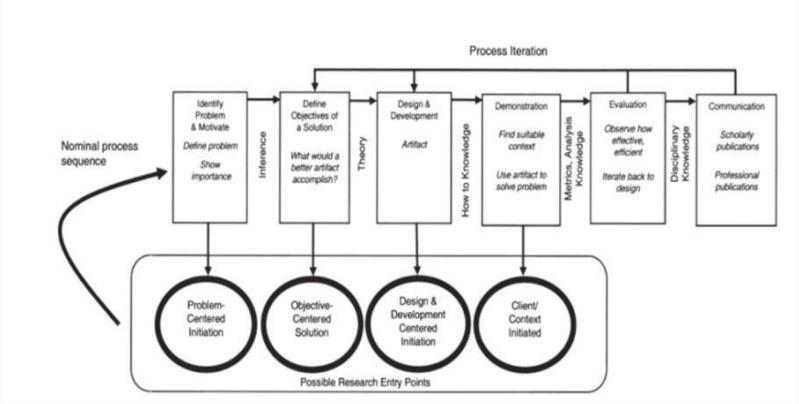

Figure 3- Design Science Research Architecture

4. Project Management

4.1 Project Management Methodology

The Manifesto for Agile Software Development, often referred to as the Agile Manifesto, is a foundational document outlining the core values and principles of Agile methodology. Created by a group of software developers in 2001, the manifesto emphasizes individuals and interactions over processes and tools, working software over comprehensive documentation, customer collaboration over contract negotiation, and responding to change over following a plan. (agilemanifesto.org)

Due to the level of uncertainty on this project, it was decided that the best approach is to use Kanban Methodology. Kanban is a visual management method that promotes continuous flow and limits work in progress (Anderson, 2010).

Every major task can be broken down into smaller tasks and sub-tasks, detailing the specific work needed to accomplish the objectives of the study. As one iteration concludes, a new one starts, and this iterative approach persists until the product is fully developed, and its value is realized.

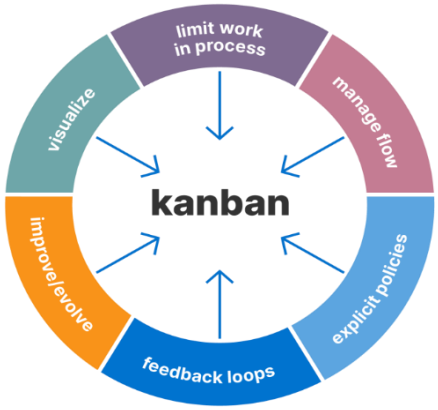

Figure 4 – Kanban cycle

4.2 Project Management Tool

JIRA, developed by Atlassian, is a leading project management software renowned for its robust issue-tracking and agile project management capabilities. It enables teams to organize tasks, track progress, and customize workflows to suit their project needs efficiently. JIRA's intuitive interface and extensive features empower teams to collaborate effectively, ensuring transparency and productivity throughout the project lifecycle (Atlassian, n.d.).

The reason the author chose to use JIRA is related to the fact that it is a versatile project management tool offering robust issue tracking, customizable workflows, and support for agile methodologies. Its integration capabilities with various tools enhance collaboration, while reporting features provide valuable insights for data-driven decision-making (Atlassian, n.d.).

4.3 Requirement Elicitation

In any project, clients must have a clear understanding of their needs and desires. This understanding forms the foundation for creating a product that aligns with their expectations and is truly valuable to them. Stakeholders' needs and desires translate into specific

requirements, outlining what they aim to build or deliver (Glinz, 2022). This initial clarity sets the stage for successful project outcomes and effective collaboration between all parties involved.

To achieve the stakeholder's necessities, it is necessary to create a product backlog that expresses their needs and for it to happen user stories and job stories is an important step in agile development. It starts by gathering requirements, prioritizing them based on business value, and writing to capture needs succinctly. It is necessary to do the effort estimation, and the backlog is managed using tools like JIRA. Regular refinement sessions ensure alignment with project goals and stakeholder feedback. This iterative process guides in delivering value to customers efficiently.

According to Atlassian a user story is an informal explanation of a software feature written from the perspective of the end user. Its purpose is to articulate how a software feature will provide value to the customer. For instance, a user story can be formatted as follows: "As a, I need to do that." or "In order to, as a, I need to." (Breno, 2024). On the other hand, a job story is focused less on the user performing some function than on the job to be done by that story (Klement). For instance, a job story can be formatted as follows: "When I want to so I can." or "When someone, actor(s) so that." (Breno, 2024).

5. Project

5.1 Business Requirements

To attain success, it is essential to identify the key components of business intelligence systems within an organization. In pursuit of understanding clients' perspectives and needs, ReField conducted interviews with two of their clients: Luis Trigueiro representing Team of Future, and Nuno Barroco representing SC Farense.

The outcome of this work can be analyzed in the section Business Requirements List, as user stories/job stories.

5.2 Data Flow

The data flow begins with two primary sources: a Microsoft Excel file and a SQL Server Database. In the subsequent stage, the Extract, Transform, Load (ETL) process takes place. Here, data from

the sources undergoes processing and transformation to a format suitable for consumption by Power BI Reports.

5.3 Conceptual Data Model

The model used in this project is based on the constellation schema. The constellation schema, also known as a galaxy schema, is a more complex design that involves multiple fact tables sharing dimension tables. It is used when there are multiple fact tables with different measures and each fact table is related to several common dimension tables. (Goel, 2023). The ReField data warehouse comprises 15 tables, categorized into two fact tables: reports and report games. These tables work as the destination for importing data from the SQL Server database.

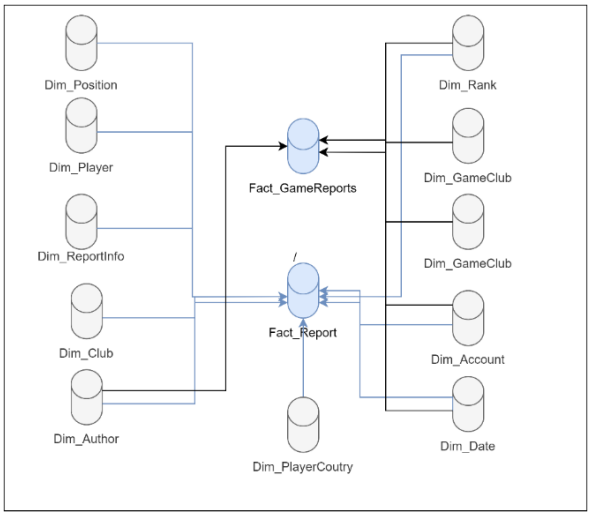

Figure 5 - Constellation Schema, adapted from Microsoft Learn (2023)

5.4 Metrics and KPIs

Stakeholders can collaboratively define the essential components and metrics that drive the functionality and utility of the dashboard, thereby fostering a data-driven culture and enabling stakeholders to control the power of actionable insights for strategic and tactical purposes (Gorman & Bakker, 2009).

Metrics and KPIs provide the context and descriptive attributes surrounding the facts in the dataset, enabling meaningful analysis and reporting (Kimball, 2002).

Identifying Metrics and KPIs in a data warehouse involves understanding the key aspects of the business and its data. One effective approach is to use user stories to determine what dimensions are relevant and necessary for analysis and reporting.

The outcome of this work can be found in the section Business Requirements List.

5.5 Conceptual Dashboard

ReField emphasizes the importance of having three fundamental pillars when determining the content and information that should be present in the dashboard. These pillars serve as guiding principles, ensuring that the dashboard effectively conveys relevant data and insights to users. By anchoring their dashboard design and content strategy on these pillars, ReField and its clients try to enhance usability, facilitate informed decision-making, and optimize operational efficiency within their respective domains.

ReField dashboard has 3 pages:

- **Players**: They are the core business, without them, there is no business or team on the field. So, it is very important to have information about them. This information will help clubs and companies to find the best approach for the market and where they should be.
- **Reports**: Football structures are complex, and there are a lot of reports being produced every day. This fact complicates good tracking for everyone, so information about who did the report, the period of more workflow, or which player has more observations is fundamental to analyzing how the scouting team is working and if the goals will be completed.
- **Game Reports**: Clubs are the provider and they are the ones that decide their players' future. So, ReField clients should be watchful for what clubs have interested players, and from where they are this information is vital to grow the networking in a way to be close to them.

5.5.1 Mockups

The dashboard mockup was designed following the literature review recommendations (Few & Dyche, 2006), where it is explained that it is very important to avoid overwhelming the dashboard with too much information.

Through clear and concise presentation, essential information is showcased via graphs and key Performance Indicators (KPIs), enabling users to quickly understand the information.

Moreover, the inclusion of intuitive filters enhances usability by empowering users to efficiently navigate through data relevant to their needs and across various dimensions.

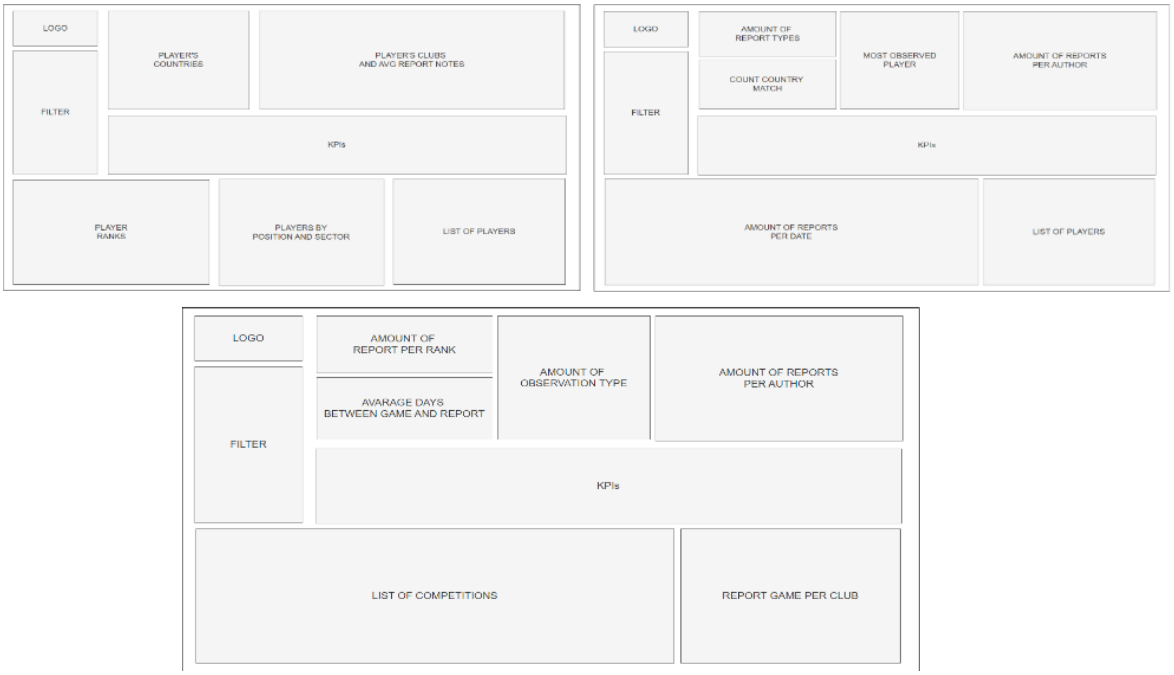

Figure 6 – Dashboards mockups

6. Development

6.1 Data Source

ReField database is stored in SQL Server, housing information important to its operations. All stakeholders within the platform unanimously support providing the requisite data for this thesis and the ReField board also agreed.

There is a Microsoft Excel File that contains information about dates from 2017 until 2024 that will be very helpful to populate tables related to date.

6.2 ETL

The ReField ETL Project comprises three primary SSIS packages: the MASTER Control of ALL ETL, responsible for orchestrating Staging Area and Data Warehouse operations. The Staging Area ETL

package handles staging processes, while the Data Warehouse ETL package manages data warehouse transformations and loading.

In terms of connection managers, the ReField ETL Project employs various connections depending on the specific SSIS package being executed. The Staging Area package utilizes connections to facilitate data movement and transformation, while the Data Warehouse package handles connections for extracting, transforming, and loading data into the target warehouse.

Specifically, within the Staging Area SSIS package, three distinct connections are utilized. Firstly, there is a connection to Excel, which serves as the source of information crucial for populating the Dim_Date table. Secondly, a connection to the ReField Database is established, enabling data extraction from this source to populate tables within the data flow. Lastly, a connection with the ReField Staging Database is employed as the destination, where data from various sources is staged for further processing.

Conversely, the Data Warehouse SSIS Package operates with two primary connections. Initially, a connection is established with the ReField Staging Area Database, serving as the source of information for the transformation process. Subsequently, a connection is made with the ReField Data Warehouse Database, designated as the final destination where transformed and structured data is loaded to populate the warehouse tables.

6.3 Data Model

A dimensional model is a data modeling technique used in data warehousing to organize and structure data for efficient querying, reporting, and analysis (Inmon, 2002; Kimball & Ross, 2002). It is designed to provide a simple and intuitive representation of business data, focusing on business processes and user requirements. The dimensional model typically consists of two types of tables: fact tables and dimension tables.

Fact tables contain quantitative, numeric data that represent business facts or measures. Each row in a fact table represents a specific event or transaction within the analyzed business process.

Dimension tables contain descriptive attributes or context for the data in the fact tables. These attributes provide the necessary context for analyzing and interpreting the quantitative data in the fact tables.

Analyzing the goals of the ReField BI project we can conclude there are two facts tables on the project. The first fact is reports which incorporate players, the player's clubs, report evaluation, date of reports, and authors. Secondly, the fact report games include the teams that played the games, the date of the reports game, authors, and rank of the game.

6.3.1 Data Tables

Table Name	Type	Overview
Fact_Report	Fact	Information about players' reports.
Fact_GameReport	Fact	Information about games reports.
Dim_Account	Dimension	ReField accounts are presented in this table, in this table it is possible to get information about accounts like name or country.
Dim_Author	Dimension	ReField users are stored in Dim_Author, information such as name and userId can be searched here.
Dim_Club	Dimension	This table contains the list of clubs that have one or more players observed.
Dim_Date	Dimension	This table contains calendar dates, including the full date, day, month and year.
Dim_Game	Dimension	This table has storage for game information, such as competition and scouting type.
Dim_GameClub	Dimension	This table contains the list of clubs that have one or more games where they played.
Dim_Player	Dimension	This table has the player data that were observed, information such as name and foot can be found here.
Dim_PlayerCountry	Dimension	This table has the countries list.
Dim_Position	Dimension	This table has the positions list, where it is possible to find the sector of which position.
Dim_Rank	Dimension	This table has the rank list.
Dim_ReportInfo	Dimension	This table stores information about reports like report type.
Log_Errors	Log	It plays an important role in error handling and troubleshooting, enabling teams to identify, analyze, and rectify issues that may arise during data extraction, transformation, or loading stages.

Log_ETL	Log	Conversely, the Log_Stg_ETL table is designed to record key actions and events related to the ETL process.

Figure 7- Data Source Overview

The complete map of attributes from each table is present in section Fact and Dimension Attributes List .

7. Dashboard

This chapter was used as the base for the conclusions taken in the Conceptual Dashboard. Thus, there are three pages: players, reports, and game reports.

7.1 Players

This page has information about players' data that were observed by the scouting team, here clients can have metrics and KPIs that will be game changers in the decision-making process. This page was created to meet the needs of stakeholders, and it allows them to adjust using filters as necessary.

There are four KPIs present in the players' dashboard, that provide essential insights into player performance and effectiveness:

- **Number of Players:** Total number of players observed.
- **Average of Age:** Average Age of players that were observed for the scouting team.
- **Average of Height:** Average of height of players that were observed for the scouting team.
- **Number of Left/Both/Right Foot:** Total amount by preferential foot.

The graphs displayed on the dashboard show various insights and trends. They help us understand different aspects of the data we have collected. In terms of graphs presented in the dashboard are about:

- **Player's countries:** allows the user to study the provenance of players. For example, this can be useful for the establishment of the strongest network in a country.
- **Player's Clubs and Avg Report Notes:** allows the user to know which club has more interested players, and the association with an average note of report is possible if that club has a good level of players.
- **Player's Ranks:** presents the number of players distributed by rank.

- **Number of players by position and sector:** shows the number of players ordered by positions and which position has a color of its sector. It helps the chief scout to study if the necessities in terms of position are covered.
- **Table of players:** It is possible to see where the players fit in the research. Useful to get players' names and if necessary try to negotiate with them.

Figure 8 – Player Scouting Dashboard

7.2 Reports

The reports page has information about reports done by the scouting team. Here it is possible to have an overview about the reports produced throughout metrics and KPIs. This page was created to meet the needs of stakeholders, and it allows them to adjust using filters as necessary. There are four KPIs present in reports dashboard, that provide essential insights:

- **Number of Reports**: Amount of reports done, this value is variable depending on filters used.
- **Number of Players**: Number of players observed, it is not necessary for the number of reports and players to be the same because one player could be observed multiple times. This value is variable depending on the filters used.
- **Number of Account Members**: Count the number of members registered in the account.
- **Average of Report Notes**: Average of notes given by scouts to the reports done.

The dashboard graphs offer valuable insights and trends, adding to stakeholder's comprehension of diverse data aspects. They provide a visual representation of crucial information related to our objectives.

- **General report**: Graph that represents the amount of general and non-general reports, important to know which is the provenance of reports.
- **Country match count**: Here it is possible to know if the club country is the same as the players' country.
- **Most observed player**: present the top 5 of most observed players by the scouting team. The number is the player's ID and the text is the player's name.
- **Amount of reports per author**: allows the user to know which team member has more players observed, and the association with an average note of report is possible if that team member has a good level of players.
- **Amount of reports per date**: present the number of players that were observed along time, there is the possibility to navigate throughout years, months, and days.
- **Players table**: It is possible to see where the players fit in the research. Useful to get players' names and if necessary try to negotiate with them.

Figure 9 – Reports Dashboard

7.3 Game Reports

The game Reports page has information about game reports done by the scouting team. On this page, it is possible to have an overview of the game reports done through metrics and KPIs.

There are four KPIs present in the reports dashboard, that provide essential insights:

- **Number of game reports**: presents the amount of game reports created for the scouting team.
- **Average notes per game**: shows the average of notes given to every player observed in the context of the game.
- **Number of competitions**: Total amount of observed competitions.
- **Number of ranks**: Amount of distinct ranks observed.

In this page there are six different graphs in which it is possible to analyze information about games that were observed.

- **Game report ranks**: Graph that represents the amount of reports done for each rank.
- **Observation type:** Shows the amount of reports created per observation type, if it is live or video scout.
- **Amount of reports per author**: allows the user to know which team member has more games observed, and the association with an average note of players allows to check if the scout observed a good level of players.
- **List of competitions**: Table with the name of the competition, the number of games observed in each competition, and the average note.
- **Report game per club**: allows to know how many games each club was observed.

Figure 10 – Report Games Dashboard

8. Conclusions

The BI ReField project was presented to the clients and they agreed that it can be a game changer in their structures. Luis Trigueiro representing the Team of Future said “This dashboard will help our organization a lot and it can be fundamental to improve decision-making.”.

Following the identification of the opportunity and defining the main business goals, the DSR methodology was chosen to support the scientific aspects of the project. With the support of the existing literature the project was successfully achieved and it was created a dashboard using Microsoft Power BI, which presented the KPIs and metrics desired by clients. The management was done by JIRA Software through a Kanban board to ensure smooth and timely completion of tasks.

The value proposition of the final product lies in its provision of a comprehensive tool, empowering clients with access to information to improve decision-making and in-depth analysis of performance indicators.

9. Bibliography

dos Santos, P. (2012). *O modus operandi de um Departamento de Scouting de Futebol*. (Link)

Wilson, J. (2008). *Inverting the pyramid: The history of football tactics*. Orion Publishing Group. 275-290. (Link)

Dresner, H. (2009). *The Performance Management Revolution: Business Results Through Insight and Action.* John Wiley & Sons. 7-19. (Link)

Dresner, H. (2011). *Howard Dresner's Wisdom of Crowds Business Intelligence Market Study*. Dresner Advisory Services. 10-35 (Link)

Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research, 69*(2), 898-903. (Link)

Golfarelli, M., & Rizzi, S. (2010). *Data Warehouse Design: Modern Principles and Methodologies*. McGraw-Hill Education. 4-8. (Link).

Hansoti, B. (2010). *Business Intelligence Dashboard in Decision Making*. (Link).

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). *Design Science in Information Systems Research*. In Design Science in IS Research MIS Quarterly. 77-105 (Link)

Inmon, W. H. (2002). *Building the data warehouse*. John Wiley & Sons. 3-17, 100-125. (Link)

Kimball, R., Caserta, J. (2004*). The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. 69-80 (Link).

Kimball, R. & Ross, M. (2002). *The data warehouse toolkit: The definitive guide to dimensional modeling*. John Wiley & Sons. 32-36, 380-390 (Link)

Kuechler, W., & Vaishnavi, V. (2008). On theory development in design science research: Anatomy of a research project. European Journal of Information Systems, 17(5), 492-505. (Link)

Kuper, S., & Szymanski, S. (2019). *Soccernomics: Why England loses, why Germany and Brazil win, and why the U.S., Japan, Australia, Turkey—and even Iraq—are destined to become the kings of the world's most popular sport* (3rd ed.). Bold Type Books.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). *Big data, analytics, and the path from insights to value*. *MIT Sloan Management Review.* 22-24. (Link)

Luhn H. P. (1958), *A Business Intelligence System*, in IBM Journal of Research and Development, 314-319. (Link)

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011*). Big data: the next frontier for innovation, competition, and productivity*. McKinsey & Company. (Link)

Marr, B. (2022). "How Business Intelligence Delivers Insights to Drive Business Success." Forbes. Unavailable

Massoud, Y., Osterfield, S. R., King, Y., & White, J. K. (2004). *Improving the accuracy of interconnect delay estimation for RC interconnect trees*. In *Proceedings of the 2004 IEEE/ACM International Conference on Computer-Aided Design.* 460-464. (Link)

Microsoft. (2023). *What is Power BI*. (Link)

Moss, L. T., & Atre, S. (2003). *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications.* Addison-Wesley Professional. 29-48 (Link)

Nunamaker, J. F., Chen, M., & Purdin, T. D. M. (1990*). Systems development in information systems research*. Journal of Management Information Systems, 7(3), 89-106. (Link)

Olszak, C. M., & Ziemba, E. (2003). *Approaches to business intelligence systems: Increasing the efficiency of business decision making*. *Intelligent information systems*. 140-142. (Link)

Orenstein, G., Doherty, C., Boyarski, M., & Boutin, E. (2017). *Data warehousing in the age of artificial intelligence* (First edition). O'Reilly Media. (Link)

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of Management Information Systems, 24(3), 45-77. (Link)

Ponniah, P. (2001). *Data warehousing fundamentals*. John Wiley & Sons

Sabherwal, R., & Becerra-Fernandez, I. (2011). *Business Intelligence: Practices, Technologies, and Management.* John Wiley & Sons. 4-23. (Link)

Silva, J. P. (2019). *Scouting - Tudo sobre a observação e análise de jogadores*. Editora Futebol.

Smith, J. (2020). *Anatomy of a football scout: An in-depth look at player recruitment*. Sports Publishing House.

Walls, J. G., Widmeyer, G. R., & El Sawy, O. A. (1992). Building an information system design theory for vigilant EIS. Information Systems Research, 3(1), 36-59. (Link)

Wanda, P., & Stian, S. (2015). *The Secret of my Success: An exploratory study of Business Intelligence management in the Norwegian Industry*. *Procedia Computer Science*, 240–247. (Link)

Wixom, B. H., & Watson, H. J. (2001). *The current state of business intelligence*. *Proceedings of the 2001 Conference on Information Systems*. (Link)

Winter, R., & Hevner, A. R. (2013). *Positioning and presenting design science research for maximum impact*. MIS Quarterly, 37(2), 337-355. (Link)

10. Annexes

10.1 Business Requirements List

ID	User Story	Fact / Measure	Dimension
1	As chief scout, I need to know where the players under observation are.	N/A	Country, Reports
2	As chief scout, I want to know where the players under observation are playing at the moment and where the AVG of reports is higher	Report notes scenario	Clubs, Report Notes
3	As an administrator, I need to have real stats about players, such as the number of players that were observed, Average Age, Average height, and foot of players	Player scenario reports	Players
4	As a chief scout, the rank is important to define our strategy for defining tasks	N/A	Ranks
5	As a chief scout, I need to know what are the top positions and sectors most observed	N/A	Position
6	As a chief scout, I need to know what the top players that were observed and what the average report notes	Player scenario reports	Players, Report Notes
7	As a chief scout, I want to know what kind of reports we are doing mostly	N/A	Reports
8	As a chief scout, I want to know what the time period during which we generate the most and least reports throughout the year	Reports date scenario	Reports, Date
9	As a chief scout, I want to know which competitions we observe mostly.	N/A	Competitions
10	As a chief scout, I need to know which members of my scout team have the highest and lowest performance in terms of observation (quantity, average ratings)	Scouting team scenario	Users, Accounts, Report Notes

Figure 11- Business requirements

10.2 Fact and Dimension Attributes List

HIERACHY /ATTIBUTE	FACT TABLE		DIMENSION TABLE										
	Fact_Reports	Fact_ReportGames	Dim_Account	Dim_Author	Dim_Club	Dim_Date	Dim_Game	Dim_GameClub	Dim_Player	Dim_PlayerCountry	Dim_Position	Dim_Rank	Dim_ReportInfo
FK_IdReport (BK,SK)	X												X
FK_Account (BK,SK)	X	X	X										
FK_Author (BK,SK)	X			X									
FK_Date (BK,SK)	X	X				X							
FK_Player (BK,SK)	X								X				
FK_Club (BK,SK)	X				X								
FK_Rank (BK,SK)	X	X										X	
FK_Position (BK,SK)	X										X		
FK_PlayerCountry (BK,SK)	X									X			
IsAccountMainRole	X												
IsReportFromAARank	X												
IsGeneralReport	X												
ReportNote	X												
FK_IdGame		X					X						
FK_HomeClub		X						X					
FK_AwayClub		X						X					
Competition	X						X						
ScoutType	X												
HowManyDays BetweenGameAndReport	X												
IsAccountMainRole	X												
IsReportFromAARank	X												
AVGPlayersNotes	X												
HowManyReports CreatedFromGame	X												
Name				X	X			X	X	X	X	X	
Logo					X			X					
Country					X			X					
Flag					X			X		X			
FullDate						X							
MonthdayNumber						X							

WeekdayNumber						X							
WeekdayName						X							
MonthNumber						X							
MonthName						X							
SemesterNumber						X							
Year						X							
ProperDate						X							
ScoutType							X						
Photo									X				
Born									X				
Height									X				
Foot									X				
Link									X				
Transfer									X				
ShortName										X	X		
Sector											X		
Year												X	
Age												X	
ReportType													X

Figure 12- Fact and Dimension Attributes