NOVA IMS Information Management School

# MDSAA

Master Degree Program in
**Data Science and Advanced Analytics**

# Game Day

The role of analytics in sport organizations to help increase revenues and create a better fan experience on matchdays

Simão Eduardo Ramos Costa Pereira

Dissertation

presented as partial requirement for obtaining the Master Degree Program in Data Science and Advanced Analytics

**NOVA Information Management School**
**Instituto Superior de Estatística e Gestão de Informação**
Universidade Nova de Lisboa

**NOVA Information Management School**
**Instituto Superior de Estatística e Gestão de Informação**
Universidade Nova de Lisboa

# The role of analytics in sport organizations to help increase revenues and create a better fan experience on matchdays

by

Simão Eduardo Ramos Costa Pereira

Dissertation presented as partial requirement for obtaining the Master's degree in Data Science and Advanced Analytics, with a Specialization in Business Analytics

**Supervisor / Co Supervisor:** Miguel de Castro Neto

**Co Supervisor***:* Bruno Jardim

November 2023

## STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledge the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Simão Pereira

*Lisbon, 21-10-2023*

## ACKNOWLEDGEMENTS

I extend my heartfelt gratitude to those who made this thesis possible: I thank my dedicated advisors, Bruno Jardim (NOVA IMS) and Rui Valente (BI4ALL), for their guidance and expertise. I am grateful to my family for their unwavering support through this long journey. Lastly my closest friends, thank you for your encouragement and camaraderie.

# Abstract

The research gap identified in this study revolves around the limited utilization of analytics approaches by sports organizations in areas such as merchandising, food and beverages, marketing, user experience, and fan engagement. The research question seeks to understand how analytics can be employed in sports organizations to increase revenues and improve the matchday experience for fans.

The primary objective of this thesis is to describe and identify how the usage of big data and analytics by sports organizations can contribute to revenue growth and enhance the fan experience. This objective is further broken down into two sub-objectives. The first is to provide an overview of the current utilization of data by sports entities and assess its effectiveness. The second is to define a Business Intelligence approach and different technologies that can be employed by sports clubs to achieve their business objectives and also increase the fans' experience.

To achieve these objectives, a comprehensive analysis was conducted, utilizing a combination of qualitative and quantitative research methods. The study involved the development of Python scripts, azure cloud services such as Event Hubs, Stream Analytics and Azure SQL Database and also Power BI.

The key findings of the research demonstrate the efficacy of the implemented Business Intelligence solution in improving the fan experience and driving revenue growth. By leveraging real-time data and targeted interactions, sports organizations can provide personalized experiences, optimize ticket sales, and enhance fan engagement during matchdays.

This thesis contributes to the field by showcasing the potential of cloud-based BI solutions in the sports industry. It emphasizes the importance of analytics in improving revenue generation and enhancing the overall fan experience. The study demonstrates the feasibility and effectiveness of the proposed solution, laying the groundwork for further advancements in the application of real-time Business Intelligence in sports organizations.


Keywords **:** Analytics; Cloud-based BI solutions; Data; Fan experience; Business Intelligence; Revenue growth

INDEX

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS AND ACRONYMS

**BI -** Business intelligence

**IS -** Information systems

**SLB –** Sport Lisboa e Benfica

**ETL –** Extract, Transform, Load

**ELT –** Extract, Load, Transform

# 1.INTRODUCTION

## 1.1 MOTIVATION

Nowadays sport organizations are more than just the actual athlete's performance inside the field. They are big companies with huge economic and financial roles in society so it’s crucial that they keep being competitive sports wise but also keep growing in a business point of view. That’s when analytics and big data play an important role in today’s world. The sports entities already use lots of data to measure and improve the athlete’s performance but with the production of big data generated by the fans and stakeholders, it’s becoming crucial to use this information as an asset to create value for these organizations (Apostolou, K., & Tjortjis, C. ,2019, November 14). Many clubs are starting to put together data science/analytics departments because they see the usage of analytics and data as an opportunity to increase revenues, reduce costs and offer a better experience to the fans. The latter is highly important because the fans are without any doubt the biggest source of income for the sports industry and so, giving them a better experience and providing them value is a necessary approach in order to keep growing as a sport and business entity (Troilo, M., & Simkins, B. J. 2021).

## 1.2 OBJECTIVES AND METHODOLOGY

The purpose of this master thesis is to explore the role of analytics and big data in sports and build a solution to show how data-driven decision-making can enhance fan experience and engagement and increase revenue in sports organizations.

To accomplish this, two use cases will be presented in this thesis as two possible analytic solutions and approaches. It is important to address that only two use cases were developed given the academic purpose, available time and financial resources.

The two use cases simulate a situation that happens on a matchday and the final goal is to build BI solutions that can be used to make real time decisions and consequently increase the sport organizations revenue and the fans experience on the day of the match

From now on let´s consider that these use cases are applied in the football club Sport Lisboa e Benfica (SLB).

A lot of solutions could have been implemented and the possibilities are immense to show how data and analytics can have a positive impact in the fans experience and in the sports organization revenue but the following use cases were chosen because with these two approaches is possible to show how decisions can be made through cross information of the fans with the sports club database and with the club mobile application:

- Use case 1: Creation of a BI model that represents a possible SLB data warehouse and by accessing it in real-time we engage with fans live before and during the match by sending them a customized email.
- Use case 2: Azure cloud capabilities are used to ingest and process fans´ live location on a matchday in the Lisbon Area and engage with them by sending them an personalized email in real-time based on their location.

Overall, this work seeks to provide a comprehensive overview of the growing field of sports analytics and big data, highlighting its importance in the modern sports industry and its vast potential for continued growth, development and to serve as inspiration for many more potential BI approaches. Hence, our contributions are the following:

- Addressing the limited utilization of analytics in various areas of sports organizations, offering an opportunity to improve revenue and fan experiences.
- Set clear objectives, such as understanding the use of big data and analytics in sports, evaluating its effectiveness, and proposing Business Intelligence (BI) approaches and technologies for sports clubs.
- Employ a combination of qualitative and quantitative research methods, utilizing tools like Python scripts, Azure cloud services, and Power BI for analysis.
- Contributions to the field by showcasing the potential of cloud-based BI solutions in the sports industry. It emphasizes the importance of analytics in improving revenue generation and fan experiences, laying the groundwork for further advancements in real-time BI in sports organizations.

- Relevance to modern sports by highlighting the transformation of sports organizations into business entities that rely on data and analytics to improve not only athletic performance but also financial performance and fan experiences.
- Importance of Fan Engagement and the critical role of fan engagement, given that fans are a significant source of income for sports organizations.
- Two practical cases where we provide two situatuons that simulate real matchday situations and demonstrate how data-driven decision-making can be applied to enhance revenue and fan experiences.
- Inspiration for potential BI approaches in the field of sports analytics, encouraging further research and development in this area.

The previous points will be addressed in the remainder of this thesis: The "Background" section offers a review of the foundational concepts of the development of this research. In "Related Work," an analysis of previously implemented analytics approaches in the realm of sports is presented, shedding light on the existing landscape. "Methodology" elucidates the methodologies employed in both use cases, providing a clear roadmap of the research approach. The "Results and Discussion" section documents the experiments conducted and the resultant findings, facilitating a discussion and analysis of the outcomes. Serving as a synthesis of the research, the "Conclusions" section encapsulates and concludes the trajectory of work accomplished throughout this study. Lastly, the "Limitations and Recommendations for Future Works" section delves into the identified limitations of this research while also having suggestions and ideas for future investigations.

# 2. BACKGROUND

This section presents the fundamental BI concepts that are the basis for this project solutions. We review what is Business Intelligence and what are its stages (section 2.1) and discuss the differences of using cloud and on-premises solutions (section 2.2).

## 2.1. WHAT IS BUSINESS INTELLIGENCE AND ITS STAGES

Business intelligence is a vital area of study in the field of information systems. It involves the collection, integration, analysis, and presentation of data to support decision-making processes within organizations. The increasing availability of data and the need for data-driven decision-making have led to the growing importance of BI in modern organizations (Turban, E., Sharda, R., Delen, D., & Aronson, J. E., 2017).

The process of implementing a BI system can be complex and requires careful planning and execution. It is crucial to assure data quality and governance to ensure the success of BI initiatives: "Without clean and trustworthy data, BI is useless." (Turban, E., Sharda, R., Delen, D., & Aronson, J. E., 2017).

Overall, the importance of BI in modern organizations cannot be overstated. "BI is no longer a luxury, but a necessity for organizations seeking to survive and thrive in today's competitive business environment." By collecting, analyzing, and interpreting data, BI can help organizations gain insights into their operations, identify opportunities for growth, and make informed decisions to improve their performance (Sabherwal, R., & Becerra-Fernandez, I.,2010).

A BI solution involves several stages, from planning and design to implementation and maintenance. Here are the different stages of a BI solution (Watson, H. J., & Wixom, B. H., 2007):

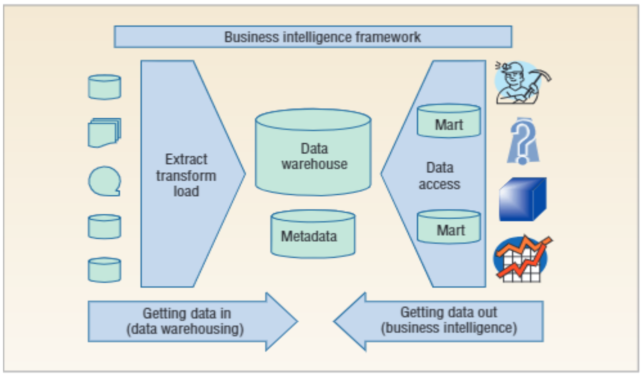


Figure 2.1- Business Intelligence Framework from Watson, H. J., & Wixom, B. H. (2007). The current state of business intelligence

1. Planning and requirements gathering: The first stage of a BI solution involves defining the objectives and scope of the project. This involves identifying the business requirements, data sources, and key performance indicators (KPIs) that will be used to measure success.

2. Data acquisition and integration: In this stage, data from various sources is acquired and integrated into a central data warehouse. This involves developing data extraction, transformation, and loading (ETL) processes to ensure data quality and consistency.

3. Data modeling and analysis: Once the data has been integrated, it needs to be modeled to support the reporting and analysis needs of the organization. This involves developing a multidimensional data model and creating data cubes for efficient querying and analysis.

4. Reporting and visualization: In this stage, reports and dashboards are developed to provide business users with easy-to-understand insights into the data. This involves selecting the appropriate reporting tools and designing reports and dashboards that meet the needs of the users.

5. Deployment and implementation: Once the BI solution has been developed, it needs to be deployed and implemented in the production environment. This involves testing the solution, training users, and ensuring that it is integrated with other systems.

6. Maintenance and support: After the solution has been deployed, it needs to be maintained and supported to ensure that it continues to meet the changing needs of

the organization. This involves monitoring performance, making updates and enhancements, and providing user support.

"Successful BI implementation requires careful attention to each of these stages, as well as effective communication and collaboration between the business and IT stakeholders involved in the project." (Watson, H. J., & Wixom, B. H., 2007). By following a structured approach to BI solution development, organizations can ensure that they are getting the most value from their data and making informed decisions based on accurate and timely information.

### 2.1.1 Modern Business Intelligence Architecture

Data architecture is the design of the systems to support the evolving data needs of an enterprise, achieved by flexible and reversible decisions reached through a careful evaluation of trade-offs (Reis, J., & Housley, M. ,2022).

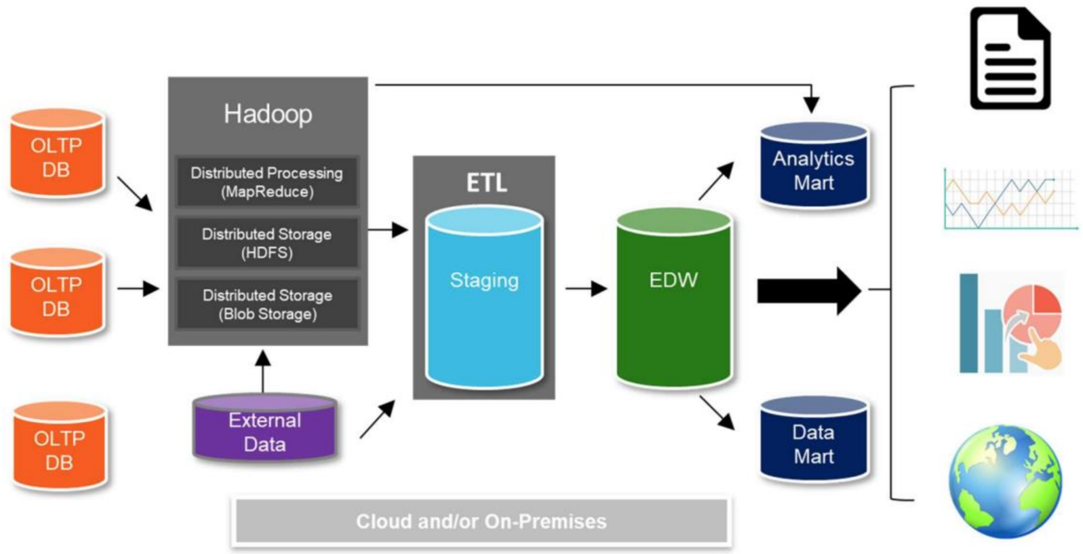


Figure 2.1 - Modern Business Intelligence Architecture - - Neto, M. C. (2022). Business Intelligence and Data-Driven Organizations. NOVA Information Management School, Lisbon.

The data architecture defines how the stages of the data lifecycle are implemented. These phases are responsible for converting raw data components into valuable end products, which are then ready for utilization by analysts, data scientists, ML engineers, and other professionals (Reis, J., & Housley, M. ,2022).

A good BI architecture should have into account the following 5 stages:

- Generation
- Storage
- Ingestion
- Transformation
- Serving data

The BI framework initiates its role in the data lifecycle by sourcing data from origin systems and securely storing it. Subsequently, the framework must ensure data undergoes transformation before fulfilling its primary objective: delivering accessible data to analysts, data scientists, ML engineers, and similar roles.

Generally, the intermediate stages – storage, ingestion, and transformation – might seem intricate and intermingled. This complexity is acceptable. While the authors delineate discrete segments within the data flow of a BI framework, the process isn't always a seamless, linear progression. Different stages of the cycle might loop back, transpire in non-sequential order, coincide, or interconnect in intriguing and unforeseen patterns (Reis, J., & Housley, M. ,2022).

### 2.1.2. ETL and ELT

Extract, transform, and load (ETL) is a data pipeline used to collect data from various sources. It then transforms the data according to business rules, and it loads the data into a destination data store. The transformation work in ETL takes place in a specialized engine, and it often involves using staging tables to temporarily hold data as it is being transformed and ultimately loaded to its destination (Jhawar, R., & Tejada, Z. ,2022, June 21).

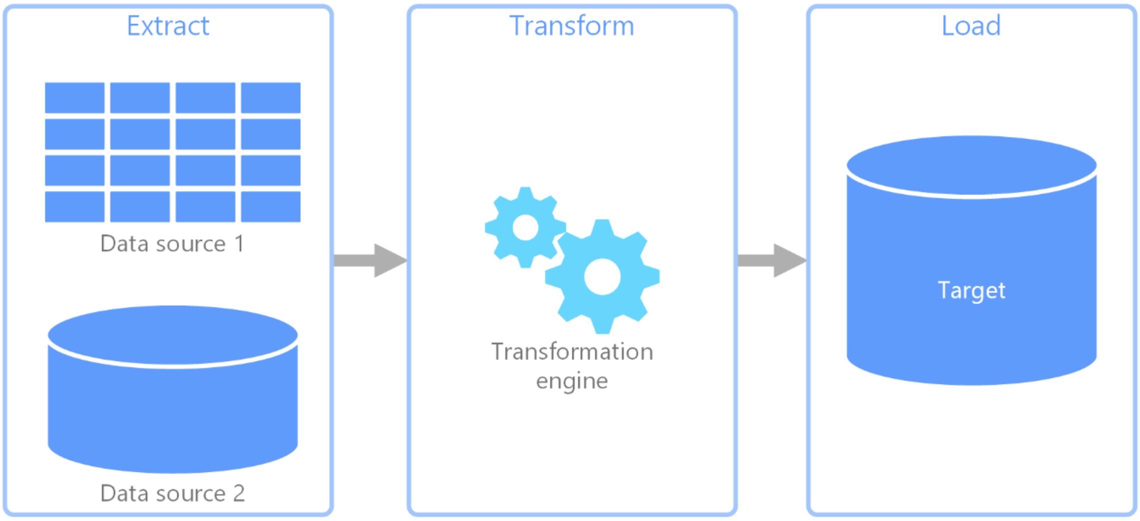


Figure 2.2 – ETL process from Jhawar, R., & Tejada, Z. (2022, June 21). **ETL process in Azure.** Mi*crosoft Azure Architecture Center.*

Extract, load, and transform (ELT) differs from ETL solely in where the transformation takes place. In the ELT pipeline, the transformation occurs in the target data store. Instead of using a separate transformation engine, the processing capabilities of the target data store are used to transform data. This simplifies the architecture by removing the transformation engine from the pipeline. Another benefit to this approach is that scaling the target data store also scales the ELT pipeline performance. However, ELT only works well when the target system is powerful enough to transform the data efficiently (Jhawar, R., & Tejada, Z. ,2022, June 21).

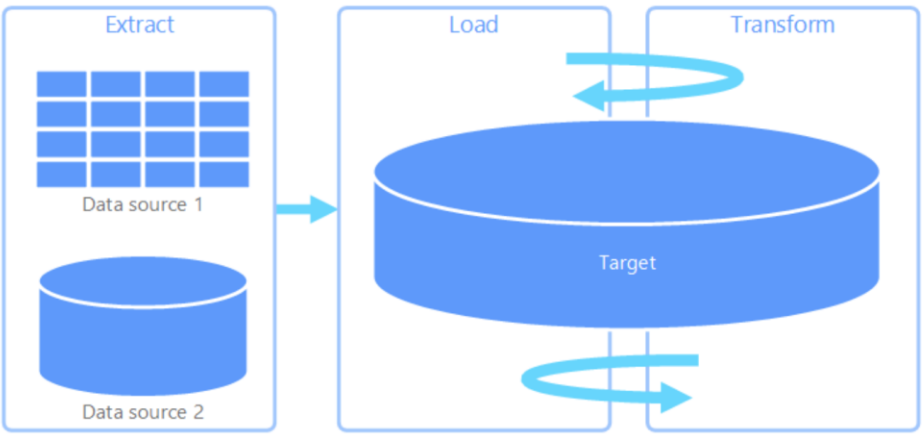


Figure 2.3 – ELT process from Jhawar, R., & Tejada, Z. (2022, June 21). **ETL process in Azure.** *Microsoft Azure Architecture Center.*

#### 2.1.2.1. Batch vs Streaming

Batch and streaming solutions are two ways of ingesting (or extract) data by organizations, but they have different strengths and weaknesses (Reis, J., & Housley, M. ,2022). Here is a brief overview of the differences between the two:

Batch solutions process data in batches, typically on a scheduled basis. They are often used for analysing large volumes of historical data to identify trends, patterns, and insights. Batch solutions can be effective for tasks such as customer segmentation, predictive modelling, and data warehousing. Some popular batch analytics tools used in sport organizations include Hadoop, Spark, and Amazon Redshift (Reis, J., & Housley, M. ,2022).

Advantages:

- Cost-effective: Batch solutions are typically less expensive than streaming solutions because they don't require real-time processing.
- High throughput: Batch solutions can handle large volumes of data and are designed for processing data in parallel.
- Scalability: Batch solutions can be easily scaled up or down to accommodate changing data volumes.

Disadvantages:

- Latency: Batch solutions may not be able to provide real-time insights due to the time required to process data in batches.
- Limited real-time decision-making: Batch solutions are not well-suited for real-time decision-making, as they are typically designed to analyze data after it has been collected.

Streaming solutions process data in (almost) real-time, as it is generated. They are often used for real-time analytics, such as monitoring social media sentiment or analyzing player performance during a game. Streaming solutions can be effective for tasks such as anomaly detection, fraud detection, and real-time personalization. Some popular streaming analytics tools used in sport organizations include Apache Kafka, AWS Kinesis, and Azure Stream Analytics (Reis, J., & Housley, M. ,2022).

Advantages:

- Real-time insights: Streaming solutions can provide real-time insights, which can be important for decision-making during a game or event.
- Real-time decision-making: Streaming solutions can enable real-time decision-making, such as adjusting ticket prices or promotions in real-time based on fan behavior.
- Better data freshness: Streaming solutions can provide more up-to-date data than batch solutions.

Disadvantages:

- Higher cost: Streaming solutions can be more expensive than batch solutions because they require real-time processing and data storage.
- Complexity: Streaming solutions can be more complex to set up and maintain than batch solutions, particularly if real-time processing is required.

Ultimately, the choice between batch and streaming solutions will depend on the specific needs and requirements of the sport organization. Batch solutions may be more suitable for historical analysis, while streaming solutions may be better suited for real-time decision-making and monitoring.

### 2.1.3. Dimensional Modelling in Data Warehousing

A data warehouse is a centralized repository of integrated data from one or more disparate sources. Data warehouses store current and historical data and are used for reporting and analysis of the data (Tejada, Z., 2022, July 25).

In the context of data warehousing, effective data modeling techniques play a pivotal role in enabling organizations to harness the power of business intelligence and analytics for informed decision-making. This chapter explores the intricacies of dimensional modeling, an approach designed to optimize data structures for efficient querying and reporting within the realm of decision support systems (Kimball, R., & Ross, M.,2013).

Dimensional modelling serves as the bedrock upon which data warehouses are built. It operates on the principle of organizing data into distinct dimensions and facts, providing a structured framework for comprehensive data representation (Kimball, R., & Ross, M.,2013).

Central to dimensional modelling is the concept of the star schema. This schema revolves around a central fact table that encapsulates quantitative measures, or facts, pertaining to various business events. Surrounding this fact table are dimension tables, each representing a distinct business entity such as time, product, customer, or location. These dimensions enrich the data by providing contextual attributes, enabling users to navigate and analyze the data from multiple perspectives (Kimball, R., & Ross, M.,2013).

An evolution of the star schema is the snowflake schema, where dimension tables are normalized to minimize data redundancy. While this normalization conserves storage space, it introduces additional complexity in querying due to increased table joins. Thus, the choice between star and snowflake schemas hinges on considerations of storage optimization versus query performance (Kimball, R., & Ross, M.,2013).

Hierarchies within dimension tables play a pivotal role in aiding users to traverse data in a structured manner. Hierarchies define relationships within dimensions, enabling users to drill down or roll up data to different levels of granularity. For instance, a time dimension might comprise hierarchies such as year > quarter > month > day, facilitating a comprehensive view of temporal data (Kimball, R., & Ross, M.,2013).

In the real world, dimension data is subject to change over time, necessitating strategies to address evolving attributes. Slowly Changing Dimensions (SCDs) categorize approaches to handle these changes. The choice of SCD strategy, whether Type 1 (overwrite), Type 2 (add new row), or Type 3 (add new attribute), hinges on the specific business requirements (Kimball, R., & Ross, M.,2013).

The granularity of a fact table, known as the grain, shapes the level of detail at which data is stored. Defining the grain has profound implications on how data is aggregated and subsequently analyzed. Clarity in specifying the grain ensures that analytical insights align with business objectives (Kimball, R., & Ross, M.,2013).

Dimensional modelling serves as a cornerstone in constructing robust data warehouses. By embracing the principles of star schemas, snowflake schemas, hierarchies, and carefully considered fact table grain, organizations can unlock the potential of their data for enhanced business intelligence and strategic decision-making. This Data warehousing and modelling is the foundation for the analytics applications that will be discussed next in this work (Kimball, R., & Ross, M.,2013).

### 2.1.4. Analytics

Business Analytics supplies the analysis models and the procedures for Business Intelligence. The ETL and Data warehousing/modelling stages are the foundation to the analytics stage in the BI architecture and are the ones that allow to take valuable insights from data obtained via analytics tools such as the ones below.

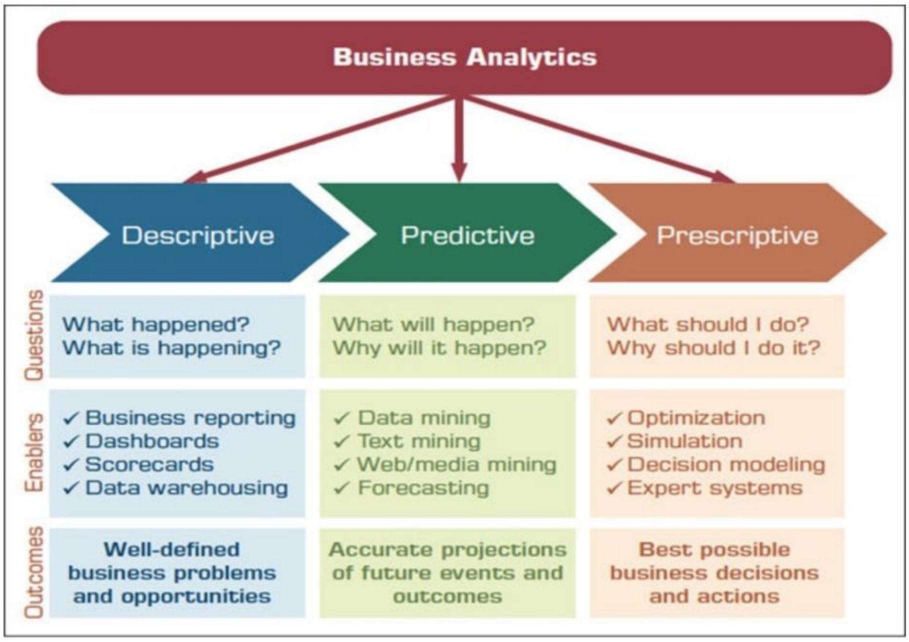


Figure 2.4 – Business Analytics approaches - Neto, M. C. (2022). Business Intelligence and Data-Driven Organizations. NOVA Information Management School, Lisbon.

**Descriptive Analytics**: Also known as reporting analytics, constitutes the retrospective analysis of historical data to answer the fundamental question of "what happened?" This framework leverages enablers such as Online Analytical Processing (OLAP) and Data Warehousing (DW), channelling data visualization, dashboards (e.g.: Power BI) and scorecards to succinctly portray trends and patterns. Furthermore, descriptive statistics play a pivotal role in distilling complex data into understandable summaries (Neto, M. C. (2022). Business Intelligence and Data-Driven Organizations. NOVA Information Management School, Lisbon ).

**Predictive Analytics**: Focus on foreseeing future events by scrutinizing past data. This framework's core objective lies in determining what is likely to happen in the future, empowering organizations with the capability to make informed decisions based on projected outcomes. Machine Learning, Data mining, text mining, web mining, and forecasting techniques, particularly time series analysis, form the bedrock of predictive analytics, propelling proactive decision-making through the anticipation of future trends (Neto, M. C. (2022). Business Intelligence and Data-Driven Organizations. NOVA Information Management School, Lisbon ).

**Prescriptive analytics:** Occupies the realm of determining the best course of action from a multitude of alternatives. This framework puts together the insights from both descriptive and predictive analytics to craft optimal solutions. Enablers encompass optimization methodologies, simulation techniques, multi-criteria decision modeling, and heuristic programming. By employing prescriptive analytics, organizations can navigate complexity and uncertainty, arriving at decisions that align seamlessly with strategic goals (Neto, M. C. (2022). Business Intelligence and Data-Driven Organizations. NOVA Information Management School, Lisbon ).

## 2.2. Cloud vs On Premises

Business intelligence (BI) solutions can be deployed either on the cloud or on-premises models, each with its own advantages and disadvantages.

According to two of the most well-known people in the Data Engineering field, Joe Reis and Matt Housley, there will be a dominance of cloud-based models over traditional on-premises approaches. Their insights are explained in the recent publication "Fundamentals of Data Engineering".

Cloud-based BI solutions are becoming increasingly popular due to their scalability, flexibility, and cost-effectiveness. With cloud-based BI, organizations can easily scale their infrastructure to meet changing needs, without having to invest in additional hardware or software. Cloud-based solutions also allow for easier collaboration and data sharing between teams, as data is stored in a centralized location accessible from anywhere with an internet connection. Additionally, cloud-based solutions often come with built-in security and compliance features, which can help ensure data privacy and protection (Reis, J., & Housley, M. ,2022).

Unlike on-premises approaches, cloud solutions offer the advantage of hardware rental, facilitating swift project commencement and experimentation devoid of protracted hardware planning concerns (Reis, J., & Housley, M. ,2022).

The allure of cloud services has been progressively augmenting among well-established enterprises. This resonance is primarily attributed to the dynamic and seamless scalability inherent in cloud environments. Notably, the outbreak of the COVID-19 pandemic in 2020 significantly catalyzed the traction gained by cloud adoption. This was underscored by the realization within companies of the pivotal role played by rapid data process scaling in elucidating insights amidst a markedly uncertain business landscape (Reis, J., & Housley, M. ,2022).

Sports organizations are no exception and therefore cloud tools have become increasingly popular in sport organizations for performing analytics due to their flexibility, scalability, and cost-effectiveness.

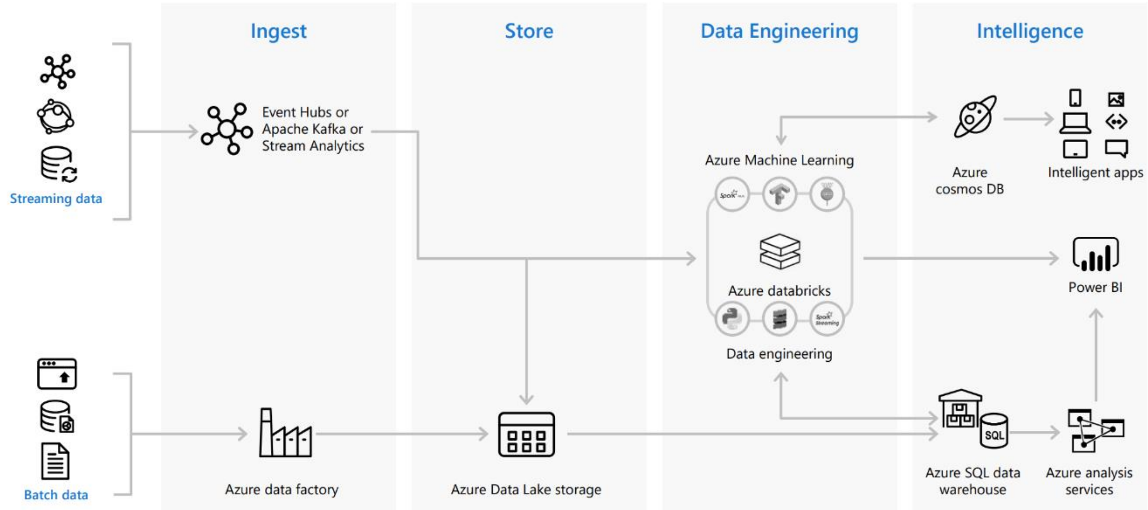


Figure 2.5 – Batch vs Streaming Azure cloud architecture (2022). *Microsoft Azure Architecture Center.*

In this figure we can see an example of a BI architecture solution using azure cloud services. The stages of BI do not change in a cloud environment, they just become easier to apply , safer and highly scalable. Here the ETL starts with the ingestion of data from different sources (batch or streaming) with the help of services such as Event Hubs, Stream Analytics and Azure data factory.

After being ingested and stored (in a data lake), the data suffers transformation/modelling in cloud services such Azure Machine Learning or Databricks. Finally the data is loaded to an analytics platform such as Azure analysis services or Power BI ready to be analysed.

# 3. RELATED WORK

In this chapter, we address some of the work done in the analytics sports community. A general idea of how the data and these analytics tools are being used in general in sports (section 3.1), how analytics is used regarding food , beverages and merchandising (section 3.2), fan engagement (section 3.3) and an example of cloud application in sports (section 3.3).

## 3.1. ANALYTICS APPLICATIONS IN SPORTS

Analytics can play a crucial role in helping sport organizations increase revenues and create a better fan experience on matchdays. By collecting and analysing data on fan behaviour and preferences, teams and venues can make informed decisions on how to optimize revenue streams and enhance the overall fan experience (Troilo, M., & Simkins, B. J., 2021).

Here are some examples in which analytics are being used:

1. Fan behaviour analysis: By analysing data on fan behaviour, such as social media activity, ticket purchases, concession sales, and merchandise sales, teams and venues can identify patterns and trends in fan spending. This information can be used to optimize pricing strategies and sales promotions, as well as to personalize marketing messages and improve the overall fan experience (Gong, X., & Wang, Y. ,2021).

2. Fan engagement: Analytics can be used to personalize the fan experience and increase engagement. By analysing data on fan preferences and behavior, teams, venues, weather, location and traffic it is possible to engage with the fan in a personalized way (Arkenberg, C., Giorgio, P., & Deweese, C) (Huettermann, M).

3. Predictive modelling: By analysing data sports organization can make informed decisions on investments and predict how much a transfer of a specific athlete might cost based on his performance to make better and safer financial decisions (McHale, I. G., & Holmes, B. ,2023).

To perform these analysis some technical tools have been used such as :

1. Customer relationship management (CRM) systems: These systems are used to collect and manage data on fan behaviour, such as ticket purchases, concession sales, and merchandise sales. Popular CRM systems used in sport organizations include Salesforce, Microsoft Dynamics, and SAP (Apostolou, K., & Tjortjis, C. 2019, November 14) ( Singh, N., 2020).

2. Business intelligence (BI) tools: These tools (e.g: Tableau or Power BI ) are used to analyse data and create visualizations and reports that help teams and venues make informed decisions (Venugopal, S., 2023).

3. Social media monitoring tools: These tools are used to monitor and analyse social media activity related to the team. This can include sentiment analysis, which helps teams understand how fans are feeling about the team or the game. Popular social media monitoring tools used in sport organizations include Hootsuite, Sprout Social, and Brandwatch (Gong, X., & Wang, Y. ,2021).

4. Machine learning, Predictive analytics/ predictive modelling, Multivariate regression (Singh, N., 2020).

5. Cloud Capabilities (Bouquet, P., Molinari, A., Bortolan, L., Dal Pra, A., Bezzi, G., & Vettorato, S., 2022).

Sports are swiftly embracing the realm of data analytics, recognizing its profound impact on organizational performance. To retain a competitive edge, sports entities are innovating approaches to harness the potential of data. Across the spectrum, sports organizations, athletes, and fans are immersed in this data revolution. This chapter delves into the pivotal role of data in the sports domain, particularly through the lens of analytics (Ratten, V., & Dickson, G.,2020).

Although there is a lot of scientific research done about this subject without any doubts that the usage of analytic tools to evaluate and increase athletes performance is way more established by now in the sports industry than the usage of tools as BI or Machine Learning to improve the fans experience and increase the organization´s revenues (Ratten, V., & Dickson, G., 2020).

Real time interaction with the fans for these purposes is also something that just started and so there is not a lot of literature regarding that area but next are presented some examples.

### 3.2. Analytics in Food, beverage, and merchandising

The incorporation of analytics within the domains of food and beverage as well as merchandising delves into additional dimensions of professional sports franchises (Sutton, W. A., Urban, T. L., Troilo, M., Bouchet, A., & Mondello, M., 2020). This phenomenon can be attributed, in part, to its relative contribution to an organization's revenue stream and the intricacies involved in collaborating with external service providers. The authors highlight that during the development of analytical capabilities, many organizations undergo a 'prove-it phase,' undertaking localized projects that illustrate the potential merits of business analytics to management.

Given that a mere five percent of revenue for professional sports entities stems from concessions and merchandise sales (Hoover’s Inc., 2016), it becomes evident that other analytic endeavours might yield more measurable advantages. Moreover, as numerous teams outsource their food, beverage, and merchandising operations, internal projects could offer analytic exposure and sponsorship before progressing to partnerships with external concessionaires. Further research that delves into the role played by external service providers in the service profit chain (Heskett and Schlesinger, 1994) could be illuminating. It is noteworthy that while food and beverage's role has historically been minor, this dynamic might be evolving. Presently, various ownership groups hold equity stakes in concession companies that serve their stadiums. Collectively, the realms of food, beverage, and merchandising offer considerable potential for analytic initiatives, possibly serving as a strategic resource for competitive advantage for managers. Identifying optimal marketing and pricing strategies can enhance existing revenue streams and open doors to new sources. For instance, NBA teams typically generate $20 million in revenue from premium-seating customers (Lawrence et al., 2013). Understanding how additional amenities, such as upscale food and beverage offerings, influence this customer segment could greatly benefit market expansion. As a starting point, Harrison et al. (2016) examine the preferences of female spectators concerning food, beverage, apparel, and merchandise at NFL games. Further insights into these aspects could contribute to retaining and attracting another crucial market segment. Additionally, as highlighted earlier, analytics can address inventory and supply-chain challenges, such as identifying which products sell during games and why. Analytics also unravels trends, encompassing factors like the day of the week, opponent, etc., influencing food, beverage, and merchandise sales throughout the season. This landscape encompasses opportunities for both cost reduction and revenue generation.

A noteworthy facet emerges regarding the interaction between analytics departments supporting on-field success and those contributing to revenue generation. An interviewee underscores a "church and state relationship" between these domains. When prompted for clarification, he expounds, "there has always been a separation between the team (players, coaches, and scouts) and the rest of the business. Many teams – including ours – this separation is physical, meaning our team operations are located in a different area of the city than our business operations" (personal communication, 2016). Nonetheless, knowledge exchange and the mobilization of knowledge between firms stand as pivotal elements in optimizing organizational resources. Interestingly, this contrast with the application of analytics in more conventional industries. For instance, one of the early adopters of revenue management through analytics is American Airlines, which established a dedicated department named American Airlines Decision Technologies (later renamed SABRE). This department aided employees across business domains with analytics, occasionally surpassing the airline's core passenger air travel business in profitability (Cross, 1997). Similarly, Proctor & Gamble streamlined their analytics teams into a global analytics department under the company's IT division (Harris, 2007). Prior to this, analytics personnel were dispersed throughout the organization. These firms perceive business analytics as a means to attain a competitive edge. In essence, they view investments in analytics as ventures into tacit knowledge that can confer an advantage over competitors. Given the structure of North American professional sports leagues, where organizations compete on the field while respecting territorial rights, cultivating a core competency in revenue-oriented business analytics can indeed confer a lasting competitive edge.

The integration of analytics within the domains of merchandise, food, and beverage offers untapped potential for boosting revenues and curbing expenses. Despite not yet being universally perceived as a source of competitive advantage, the significance of enhancing business analytic capabilities remains widely acknowledged.

### 3.3. Analytics in Fan Engagement

Fan involvement cultivates an ambiance that seeks to extend the duration of patrons' presence within the venue, leading to instances where attendees arrive earlier and depart later than the actual match duration. This elongated engagement window often translates into increased spending. Notably, the interplay between fan engagement and operational efficiency holds sway over the decision of whether fans will revisit the stadium in the future. Within this context, technology emerges as a potent catalyst, adept at yielding optimal outcomes. Existing research within these realms serves as the foundation for hypotheses that have undergone empirical testing to ascertain the segments of consumers who exhibit heightened awareness and a willingness to engage with and financially support digitalization initiatives (Caulfield, J., & Jha, A. K., 2022). The related literature has hypothisized two types of fans:

$H_1$: Technology will be most welcomed by flâneurs (measured by low frequency attending), implying they attend for the experience more than loyalty to the team.

$H_2$: Die-hard supporters categorized as supporters, who are in attendance purely for the enjoyment of the match (measured by high frequency attending), will be indifferent to new technology.

Illustrative instances of fan engagement encompass applications tailored for fans, enabling actions like food ordering directly to seats or fostering interactions with fans on both match-days and non-match days through social media platforms. When facilitated through technological channels, fan engagement not only confers a competitive edge in the realm of winning but also furnishes invaluable data for customer segmentation and targeted outreach. Moreover, it places the fan at the core of activities, fostering enhanced transparency. In the battle to capture consumers' leisure time, organizations equipped with tools to precisely segment and target their customers on an individual basis gain a distinct advantage in cultivating customer loyalty and fostering enduring relationships. Given that the younger demographic constitutes the most actively engaged fan base, comprehending social media's potential as a conduit to connect with fans stands as an imperative for stadiums. The crux of effective strategies to satisfy and exhilarate fans lies in the mastery of identifying the target market and understanding associated costs. Hence, another hypothesis can be formulated:

$H_3$: Spectators in younger age demographics will be more likely to pay a higher ticket price for technology and therefore represent a primary target market.

While some stadiums have introduced customer-facing technological advancements, others have directed their efforts towards backend enhancements, aiming to streamline operations or leverage analytics to curtail expenses. An exemplar is Croke Park in Dublin, Ireland, which leverages Microsoft's Azure IoT suite. This technology assimilates data into machine learning algorithms, thereby discerning patterns that predict optimal times for illuminating the pitch with heat lamps. Such strategic illumination facilitates optimal growth and tangible cost savings for the stadium proprietors (Caulfield, J., & Jha, A. K., 2022).

This aligns with our endeavour to ascertain whether fan preferences lean towards internal or external initiatives. The physical stadium environment significantly influences the extent to which spectators desire to remain within and revisit the stadium. However, it's worth noting the limited mention of technology, which nowadays constitutes a substantial portion of the physical environment. This influence manifests either directly in the hands of consumers through tools like stadium apps, previously mentioned fan engagement endeavors, and social media interactions, or from an operational standpoint, such as mitigating overcrowding or preemptively resolving inventory challenges (Caulfield, J., & Jha, A. K., 2022). This brings us to the next hypothesis.

$H_4$: Fan-facing (external) initiatives will be more demanded and profitable to consumers than stadium-facing (internal) initiatives.

The authors executed a survey to gauge these hypotheses, revealing that emerging technologies contribute significantly to heightened operational efficiency in stadiums, culminating in enhanced benefits for fans. Remarkably, technological interventions foster fan engagement, generating a mutually reinforcing cycle that enriches the spectator experience, amplifies customer retention, and bolsters ticket sales. This reciprocal dynamic empowers stadiums and teams to channel these advantages into digital advancements or other assets, thus propelling their progress. Seizing the intersection of sports and technology, these entities occupy a vantage point that capitalizes on the convergence of these domains. As stadium technology advances and the experiential economy gains momentum, the demand for their services naturally surges, further solidifying their importance.

The paper introduces a noteworthy mechanism aimed at fostering informed decision-making concerning the selection of technology within stadiums. Aligned with existing research advocating for innovative yet sustainable strategies, this mechanism serves as a guiding beacon. It furnishes stadium managers with a tool to navigate the delicate equilibrium between innovation and sustainability. By optimizing decision-making, the paper presents a framework that not only elevates the fan experience but also curbs superfluous investments in non-essential innovations.

### 3.4. Cloud application in sports

The NBA has introduced the NBA CourtOptix, an AI-driven data analytics system designed to enhance fan experience by providing valuable insights into player and ball positioning during games. This system, already visible in the official NBA Twitter feed and game broadcasts, capitalizes on data related to geographical positioning to provide meaningful information about in-game actions (Microsoft., 2021, October 1).

The process involves cameras tracking player movements and the ball 25 times per second, generating around 1.5 million geographical coordinates per game. For the NBA, this translates to nearly 2 billion data points throughout the season. The organization collaborates with a tracking partner to acquire this data, and Microsoft Azure serves as the pivotal tool for processing and transforming this massive data pool into actionable insights.

Key Azure components utilized include Azure Data Lake Storage, Azure Machine Learning, Azure Databricks, Apache Airflow, MLflow, and Delta Lake, along with Azure Functions, Azure Cosmos DB, Azure Kubernetes Service (AKS), and Azure Event Hubs. These tools consolidate tracking and other data in the cloud, while historical records remain stored in an on-premises system powered by Microsoft SQL Server. Azure Databricks proves to be a versatile tool, enabling various skilled employees to work collaboratively.

Machine learning and modelling are fundamental to the effectiveness of NBA CourtOptix. The system learns basketball concepts through machine learning, enabling it to interpret plays, dribbles, and shots. This learning process results in more sophisticated insights about the game as the system evolves.

The NBA aims to personalize fan engagement by tailoring content based on individual preferences. The Azure-based platform facilitates this customization, with services like Personalizer being leveraged to present fans with tailored experiences that align with their behaviors and preferences.

This strategic shift to Azure empowers the NBA to comprehend fan consumption patterns, which devices they use, and their content preferences, all contributing to a customized fan experience. Azure's capabilities offer the NBA opportunities for expansion and innovation, ensuring it stays ahead in a competitive sports and entertainment landscape.

# 4.METHODOLOGY

In order to address the study objectives mentioned in section 1.2 and based on the concepts and work addressed in sections 2 and 3, this section describes the methodologies used for the BI solutions in both use cases.

In both use cases we consider the football club Sport Lisboa e Benfica and matches that occur in its stadium, Estádio da Luz in Lisbon.

## 4.1. USE CASE 1: REAL TIME ACCESS TO A DATA WAREHOUSE AND TAKING LIVE ACTIONS

A comprehensive football club data warehouse was constructed to explore additional real-time fan interaction methods. The objective was to replicate a plausible Benfica database and develop a BI model (figure 4.1) capable of conducting analysis and live actions. Although the following model is more used in this case, it also plays a role in the second use case as we will see.

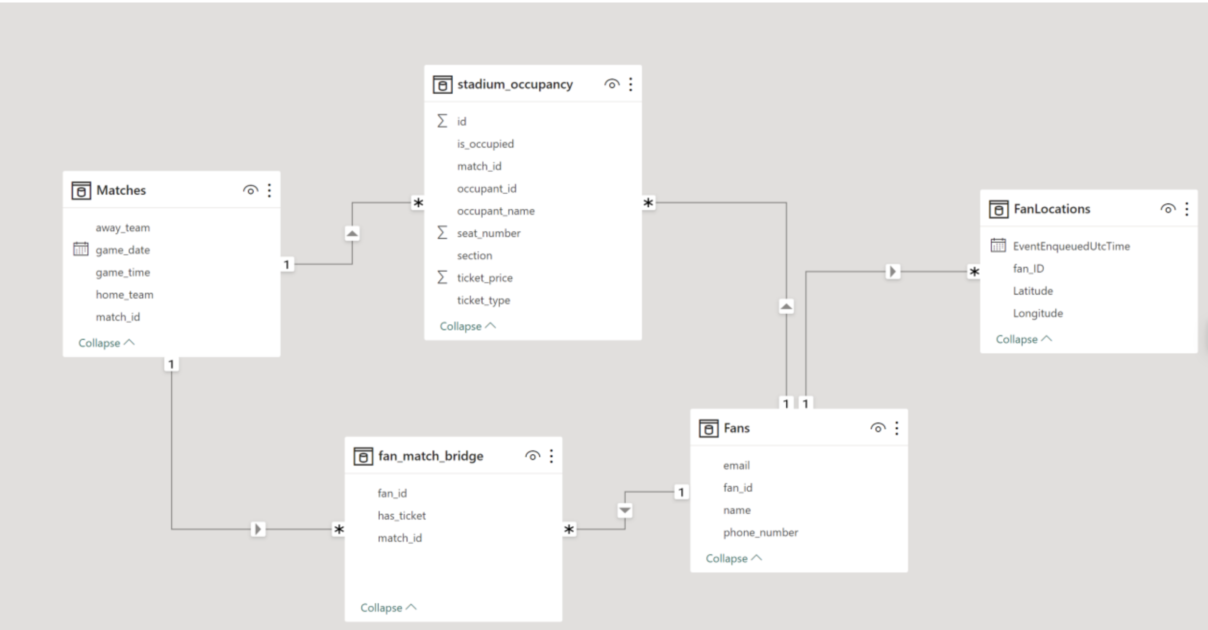


Figure 4.1 – BI Data warehouse/Model

This well-structured model harnesses the power of fan_id and match_id fields, along with their interconnected relationships. These key elements serve as the catalyst for conducting a multitude of insightful analyses, unlocking valuable insights and optimizing decision-making processes for the organization.

### 4.1.1. Tables

To achieve the referred model, multiple SQL tables were created to emulate the data warehouse of a football club such as Benfica. The data is included in the Fans table (Table 4.1), the Matches table (4.2), Fan_match_bridge table (Table 4.3) and Stadium_Occupancy table (Table 4.4).

| Field | Data Type |
|---|---|
| Fan_id | int (Identity, Not Null) |
| Name | varchar(100) (Not Null) |
| Phone_number | varchar(20) (Not Null) |
| Email | varchar(100) (Not Null) |

Table 4.1 - Metadata of Fans SQL table

| Field | Data Type |
|---|---|
| Match_id | int (Identity, Not Null) |
| Home_team | varchar(100) (Not Null) |
| Away_team | varchar(100) (Not Null) |
| Game_date | date (Not Null) |
| Game_time | time(7) (Not Null) |

Table 4.2 - Metadata of Matches SQL table

| Field | Data Type |
|---|---|
| Fan_id | int (Identity, Not Null) |
| Match_id | int (Identity, Not Null) |
| Has_ticket | bit (Null) |

Table 4.3 - Metadata of Fan_match_bridge SQL table

| Field | Data Type |
|---|---|
| ID | int (Identity, Not Null) |
| Match_id | int (Null) |
| Occupant_id | int (Null) |
| Section | varchar(50) (Not Null) |
| Seat_number | int (Null) |
| Is_occupied | bit(Null) |
| Occupant_name | varchar(100) (Null) |
| Ticket_type | varchar(50) (Null) |
| Ticket_price | decimal(10,2) (Null) |

Table 4.4 - Metadata Stadium_Occupancy SQL table

### 4.1.2. Database

With the intention of simulating the SLB database, the previous tables were randomly populated with the following data:

| | fan_id | name | phone_number | email |
|---|---|---|---|---|
| 1 | 1 | John Smith | 1234567890 | simao.er.costa.pereira@gmail.com |
| 2 | 2 | Jane Doe | 9876543210 | simao.er.costa.pereira@gmail.com |
| 3 | 3 | Michael Johnson | 5551234567 | simao.er.costa.pereira@gmail.com |
| 4 | 4 | Sarah Wilson | 9998887777 | simao.er.costa.pereira@gmail.com |
| 5 | 5 | David Lee | 1112223333 | simao.er.costa.pereira@gmail.com |
| 6 | 6 | Emily Davis | 4445556666 | simao.er.costa.pereira@gmail.com |
| 7 | 7 | Robert Thompson | 2223334444 | simao.er.costa.pereira@gmail.com |
| 8 | 8 | Jennifer Brown | 7778889999 | simao.er.costa.pereira@gmail.com |
| 9 | 9 | Christopher Martinez | 6665554444 | simao.er.costa.pereira@gmail.com |
| 10 | 10 | Jessica Taylor | 3332221111 | simao.er.costa.pereira@gmail.com |

Table 4.5 - Fans SQL table

| | match_id | home_team | away_team | game_date | game_time |
|---|---|---|---|---|---|
| 1 | 1 | Benfica | Sporting CP | 2023-07-15 | 18:00:00.0000000 |
| 2 | 2 | Benfica | FC Porto | 2023-07-16 | 18:00:00.0000000 |
| 3 | 3 | Benfica | SC Braga | 2023-07-17 | 18:00:00.0000000 |
| 4 | 4 | Benfica | Arouca | 2023-07-18 | 18:00:00.0000000 |
| 5 | 5 | FC Porto | Benfica | 2023-07-21 | 18:00:00.0000000 |
| 6 | 6 | Sporting CP | Benfica | 2023-07-22 | 18:00:00.0000000 |
| 7 | 7 | SC Braga | Benfica | 2023-07-23 | 18:00:00.0000000 |
| 8 | 8 | Arouca | Benfica | 2023-07-24 | 18:00:00.0000000 |

Table 4.6 - Matches SQL table

| | fan_id | name | email | match_id | home_team | away_team | game_date | game_time | has_ticket |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | John Smith | simao.er.costa.pereira@gmail.com | 1 | Benfica | Sporting CP | 2023-07-15 | 18:00:00.0000000 | 1 |
| 2 | 1 | John Smith | simao.er.costa.pereira@gmail.com | 2 | Benfica | FC Porto | 2023-07-16 | 18:00:00.0000000 | 1 |
| 3 | 1 | John Smith | simao.er.costa.pereira@gmail.com | 3 | Benfica | SC Braga | 2023-07-17 | 18:00:00.0000000 | 1 |
| 4 | 1 | John Smith | simao.er.costa.pereira@gmail.com | 4 | Benfica | Arouca | 2023-07-18 | 18:00:00.0000000 | 1 |
| 5 | 1 | John Smith | simao.er.costa.pereira@gmail.com | 5 | FC Porto | Benfica | 2023-07-21 | 18:00:00.0000000 | 1 |
| 6 | 1 | John Smith | simao.er.costa.pereira@gmail.com | 6 | Sporting CP | Benfica | 2023-07-22 | 18:00:00.0000000 | 1 |
| 7 | 1 | John Smith | simao.er.costa.pereira@gmail.com | 7 | SC Braga | Benfica | 2023-07-23 | 18:00:00.0000000 | 0 |
| 8 | 1 | John Smith | simao.er.costa.pereira@gmail.com | 8 | Arouca | Benfica | 2023-07-24 | 18:00:00.0000000 | 0 |
| 9 | 2 | Jane Doe | simao.er.costa.pereira@gmail.com | 1 | Benfica | Sporting CP | 2023-07-15 | 18:00:00.0000000 | 0 |
| 10 | 2 | Jane Doe | simao.er.costa.pereira@gmail.com | 2 | Benfica | FC Porto | 2023-07-16 | 18:00:00.0000000 | 1 |
| 11 | 2 | Jane Doe | simao.er.costa.pereira@gmail.com | 3 | Benfica | SC Braga | 2023-07-17 | 18:00:00.0000000 | 1 |
| 12 | 2 | Jane Doe | simao.er.costa.pereira@gmail.com | 4 | Benfica | Arouca | 2023-07-18 | 18:00:00.0000000 | 0 |
| 13 | 2 | Jane Doe | simao.er.costa.pereira@gmail.com | 5 | FC Porto | Benfica | 2023-07-21 | 18:00:00.0000000 | 0 |
| 14 | 2 | Jane Doe | simao.er.costa.pereira@gmail.com | 6 | Sporting CP | Benfica | 2023-07-22 | 18:00:00.0000000 | 1 |
| 15 | 2 | Jane Doe | simao.er.costa.pereira@gmail.com | 7 | SC Braga | Benfica | 2023-07-23 | 18:00:00.0000000 | 0 |
| 16 | 2 | Jane Doe | simao.er.costa.pereira@gmail.com | 8 | Arouca | Benfica | 2023-07-24 | 18:00:00.0000000 | 1 |
| 17 | 3 | Michael Johnson | simao.er.costa.pereira@gmail.com | 1 | Benfica | Sporting CP | 2023-07-15 | 18:00:00.0000000 | 0 |
| 18 | 3 | Michael Johnson | simao.er.costa.pereira@gmail.com | 2 | Benfica | FC Porto | 2023-07-16 | 18:00:00.0000000 | 0 |
| 19 | 3 | Michael Johnson | simao.er.costa.pereira@gmail.com | 3 | Benfica | SC Braga | 2023-07-17 | 18:00:00.0000000 | 1 |
| 20 | 3 | Michael Johnson | simao.er.costa.pereira@gmail.com | 4 | Benfica | Arouca | 2023-07-18 | 18:00:00.0000000 | 0 |
| 21 | 3 | Michael Johnson | simao.er.costa.pereira@gmail.com | 5 | FC Porto | Benfica | 2023-07-21 | 18:00:00.0000000 | 1 |
| 22 | 3 | Michael Johnson | simao.er.costa.pereira@gmail.com | 6 | Sporting CP | Benfica | 2023-07-22 | 18:00:00.0000000 | 0 |
| 23 | 3 | Michael Johnson | simao.er.costa.pereira@gmail.com | 7 | SC Braga | Benfica | 2023-07-23 | 18:00:00.0000000 | 1 |
| 24 | 3 | Michael Johnson | simao.er.costa.pereira@gmail.com | 8 | Arouca | Benfica | 2023-07-24 | 18:00:00.0000000 | 1 |
| 25 | 4 | Sarah Wilson | simao.er.costa.pereira@gmail.com | 1 | Benfica | Sporting CP | 2023-07-15 | 18:00:00.0000000 | 0 |
| 26 | 4 | Sarah Wilson | simao.er.costa.pereira@gmail.com | 2 | Benfica | FC Porto | 2023-07-16 | 18:00:00.0000000 | 0 |
| 27 | 4 | Sarah Wilson | simao.er.costa.pereira@gmail.com | 3 | Benfica | SC Braga | 2023-07-17 | 18:00:00.0000000 | 1 |
| 28 | 4 | Sarah Wilson | simao.er.costa.pereira@gmail.com | 4 | Benfica | Arouca | 2023-07-18 | 18:00:00.0000000 | 0 |
| 29 | 4 | Sarah Wilson | simao.er.costa.pereira@gmail.com | 5 | FC Porto | Benfica | 2023-07-21 | 18:00:00.0000000 | 0 |
| 30 | 4 | Sarah Wilson | simao.er.costa.pereira@gmail.com | 6 | Sporting CP | Benfica | 2023-07-22 | 18:00:00.0000000 | 1 |
| 31 | 4 | Sarah Wilson | simao.er.costa.pereira@gmail.com | 7 | SC Braga | Benfica | 2023-07-23 | 18:00:00.0000000 | 1 |
| 32 | 4 | Sarah Wilson | simao.er.costa.pereira@gmail.com | 8 | Arouca | Benfica | 2023-07-24 | 18:00:00.0000000 | 1 |

Table 4.7 - Fan_details SQL table

| | id | match_id | occupant_id | section | seat_number | is_occupied | occupant_name | ticket_type | ticket_price |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | East Stand | 101 | 1 | John Smith | Premium | 50.00 |
| 2 | 2 | 1 | 7 | East Stand | 102 | 1 | Robert Thompson | Standard | 35.00 |
| 3 | 3 | 1 | NULL | East Stand | 103 | 0 | NULL | Standard | 35.00 |
| 4 | 4 | 1 | 8 | West Stand | 201 | 1 | Jennifer Brown | Premium | 50.00 |
| 5 | 5 | 1 | 9 | West Stand | 202 | 1 | Christopher Martinez | Standard | 35.00 |
| 6 | 6 | 1 | 10 | West Stand | 203 | 1 | Jessica Taylor | Standard | 35.00 |
| 7 | 7 | 1 | NULL | West Stand | 203 | 0 | NULL | Standard | 35.00 |

Table 4.8 - Stadium occupancy (for match_id =1) SQL table

### 4.1.3. Actions

Within this first use case, two types of actions were executed: one prior to the start of the match and another during the match. To deploy these actions Python and its connection to Azure SQL Database were used.

Potential actions include:

1. Before the match begins: If the club's database indicates that a specific fan does not possess a ticket five hours before the match, an email can be sent as a reminder to purchase tickets.
2. During halftime: If the club's database confirms that a specific fan holds a ticket for the game, a notification can be sent at halftime featuring promotional offers.

#### 4.1.3.1. Action before the match

By leveraging the created data warehouse that includes matches table, fans table and stadium occupancy data, we can determine the availability of tickets in real-time. If there are still tickets available before the match starts, we can send personalized emails to fans, reminding them to purchase the remaining tickets.

Using python capabilities to connect with the SQL Database and by performing SQL queries it is possible to send a personalized email to a specific fan based if he has a ticket or not. For instance, the following code allows to get the needed conditions and information to send the email:

```python
try:
    # Get the current datetime
    current_time = datetime.now()

    # Query the game start time from the matches table
    with conn.cursor() as cursor:
        query = "SELECT game_date, game_time FROM Matches WHERE match_id = 1"
        cursor.execute(query)
        result = cursor.fetchone()
        game_date, game_time = result[0], result[1]

    # Combine the date and time to create the game start time
    game_start_time = datetime.combine(game_date, game_time)

    # Calculate the time difference in hours
    time_difference_hours = (game_start_time - current_time).total_seconds() / 3600

# except Exception as e:
#     print("An error occurred:", str(e))



# Check if the game starts in 5 or fewer hours --- CHECK THE START TIME IN THE SQL TABLE. IT NEEDS TO BE CHANGED WITH TIME FO
    if time_difference_hours <= 5:
        # Check if the stadium has at least one unoccupied seat
        with conn.cursor() as cursor:
            query = "SELECT COUNT(*) FROM stadium_occupancy WHERE is_occupied = 0"
            cursor.execute(query)
            result = cursor.fetchone()
            unoccupied_count = result[0]

        if unoccupied_count > 0:
            # Create a cursor object
            with conn.cursor() as cursor:
                # Execute the query
                query = """

                SELECT name, email FROM fan_details WHERE has_ticket = 0 and match_id = 1

                """

                cursor.execute(query)

                # Fetch all the rows
                rows = cursor.fetchall()

                # Iterate over the rows
                for row in rows:
                    name, email = row

                    # Send an email to fans without a ticket
                    send_email(email)
                    print(f"Email sent to {name} at {email}")
        else:
            print("No unoccupied seats in the stadium.")
    else:
        print("Current time is not within the specified cutoff period.")



finally:
     # Close the database connection
        conn.close()
```

Figure 4.2 – Python code used to produce the action before the match

The provided code snippet demonstrates a Python script designed to automate the process of sending email notifications to fans who do not have tickets for an upcoming Benfica match. The script leverages the previous referred database to retrieve relevant information and make informed decisions.

The script begins by obtaining the current datetime using the datetime.now() function. It then queries the database to retrieve the start time of a specific game from the "Matches" table (in this case match 1). The retrieved game date and time are combined to create the actual game start time. By calculating the time difference in hours between the current time and the game start time, the script determines how soon the game is scheduled to begin.

If the game is set to start within a specific time frame (configured as 5 hours in this example), the script proceeds to check whether the stadium has unoccupied seats available. It queries the "stadium_occupancy" table to count the number of unoccupied seats (is_occupied = 0).

If there are unoccupied seats, the script further queries the database to retrieve the names and email addresses of fans who have not yet obtained a ticket for the specified match. For each fan without a ticket, the script simulates sending an email notification using the send_email() function and prints a message indicating that an email has been sent to the respective fan.

If there are no unoccupied seats in the stadium, the script notifies that there are no available seats. Similarly, if the current time falls outside the specified cutoff period, the script indicates that the current time does not align with the scheduled game.

The script utilizes exception handling (try and finally) to ensure proper closure of the database connection (conn.close()) even in case of errors.

Overall, this script showcases how technology can be employed to automate fan engagement and ticket-related notifications in the context of sports events.

#### 4.1.3.2. Action during the match

With access to the fan's information and by knowing those who are attending the game, we can engage with them during the match. For example, during halftime, we can interact with fans, providing marketing actions such as promotions, merchandise offers, or concession discounts based on their preferences and purchasing history.

The code below shows how to access the database and how to make sure we are only sending emails to fans that are attending the match and that happens at halftime.

```python
try:
    # Get the current datetime
    current_time = datetime.now()

    # Query the game start time from the matches table
    with conn.cursor() as cursor:
        query = "SELECT game_date, game_time FROM Matches WHERE match_id = 1"
        cursor.execute(query)
        result = cursor.fetchone()
        game_date, game_time = result[0], result[1]

    # Combine the date and time to create the game start time
    #game_start_time = datetime.combine(game_date, game_time)
    game_start_time = datetime(2023, 7, 14, 17, 0)

    # Calculate halftime
    halftime = (game_start_time + timedelta(minutes=45))

# except Exception as e:
#     print("An error occurred:", str(e))


# Check if the game is at halftime
    if current_time >= halftime:
        # Check if the stadium has at least one unoccupied seat
        with conn.cursor() as cursor:
            query = "SELECT COUNT(*) FROM stadium_occupancy WHERE is_occupied = 1"
            cursor.execute(query)
            result = cursor.fetchone()
            unoccupied_count = result[0]

        if unoccupied_count > 0:
            # Create a cursor object
            with conn.cursor() as cursor:
                # Execute the query
                query = """

                SELECT name, email FROM fan_details WHERE has_ticket = 1 and match_id = 1

                """

                cursor.execute(query)

                # Fetch all the rows
                rows = cursor.fetchall()

                # Iterate over the rows
                for row in rows:
                    name, email = row

                    # Send an email to fans without a ticket
                    send_email(email)
                    print(f"Email sent to {name} at {email}")
        else:
            print("No unoccupied seats in the stadium.")
    else:
        print("Current time is not within the specified cutoff period.")



finally:
     # Close the database connection
        conn.close()
```

Figure 4.3 – Python code used to produce the action during the match

The provided Python script showcases the automation of sending email notifications to fans who already have tickets for Benfica matches, specifically targeting the halftime of the match. Similar to the previous script, this script also relies on the database to fetch relevant information and make informed decisions.

The script starts by obtaining the current datetime using the **datetime.now()** function. It then queries the database to retrieve the start time of a specific game from the "Matches" table.

The retrieved game date and time are utilized to set a predefined game start time (in this example, July 14, 2023, at 17:00).

The script calculates the halftime of the match by adding 45 minutes to the game start time using the **timedelta()** function.

Following this, the script compares the current time with the calculated halftime. If the current time is greater than or equal to halftime, the script proceeds to check whether the stadium has occupied seats by querying the "stadium_occupancy" table.

If there are occupied seats, the script further queries the database to retrieve the names and email addresses of fans who have tickets for the specified match. For each fan with a ticket, the script simulates sending an email notification using the **send_email()** function and prints a message indicating that an email has been sent to the respective fan.

If there are no occupied seats in the stadium, the script notifies that there are no available seats. Additionally, if the current time falls outside the halftime period, the script indicates that the current time is not within the specified halftime cutoff.

The script utilizes exception handling (**try** and **finally**) to ensure proper closure of the database connection (**conn.close()**) even in case of errors.

Overall, this script demonstrates how technology can be employed to engage fans and provide timely notifications during a sports match, specifically targeting halftime.

### 4.2. Use case 2 - Using real-time location on a matchday to interact with the fans

As a brief overview of the application's functionality, SLB would have a mobile application where they could access live information about the fans and engage with them via that same app but in this work the focus is primarily on the BI solution and its architecture, rather than the application itself, leading to certain assumptions and adjustments being made.

The methodology employed in this study aimed to simulate the functionalities that could be achieved through the mobile app, including the retrieval of fan locations and sending notifications. Various types of data can be processed using this application, encompassing live weather updates, real-time traffic information, current fan locations, ticket status, entrance timestamps at the stadium, and historical purchase data of fan members.

The deployed BI solution leverages Azure's capabilities in real-time data ingestion, processing, and streaming analytics. This choice is driven by Azure's robustness in collecting live data, its streamlined architecture, robust security protocols, scalability, and a variety of user-friendly services for capturing, converting, and examining real-time information.

The potential outcomes from the utilization of this BI solution are extensive. These include the ability to conduct live marketing campaigns, enhance security measures within the stadium, administer satisfaction questionnaires, provide fans with weather and traffic tips, and increase ticket sales on match days.

By implementing this BI solution and exploring the diverse array of data types, sports organizations can unlock numerous possibilities to enhance fan experiences, drive revenue growth, and optimize operational efficiency.

### 4.2.1. Solution Architecture

The following image represent the inspirational architecture used to build the real-time BI solution.

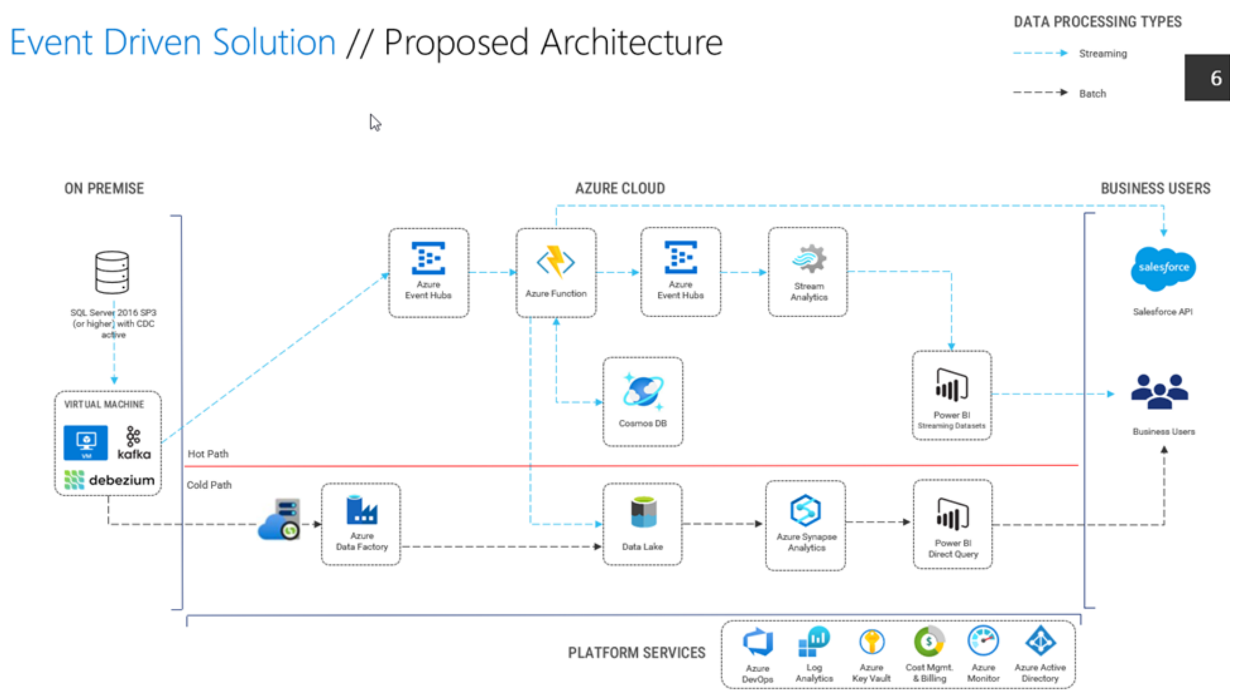


Figure 4.4 – Azure Cloud inspirational architecture

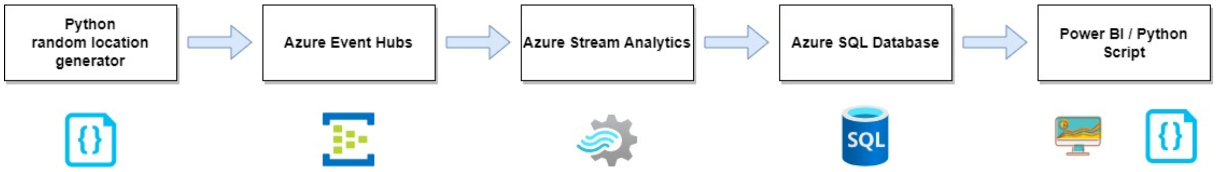


Figure 4.5 – Solution architecture

This diagram above represents the final architecture of the solution that enables the ingestion of real-time fan data and facilitates immediate actions, while also storing the data in an SQL database for future in-depth analysis and real-time visualization with Power BI.

### 4.2.2 Software

Now it will be explained all the cloud services and BI tools used in this research, more specifically what which software does and what was used for.

#### 4.2.2.1 Python – Location generator

Python played a pivotal role in two crucial aspects of the research: generating fan locations and facilitating seamless interaction with an Azure SQL database to send personalized notifications. **Python was used in this work as a replacement for what would be the mobile application software in a real-life scenario**.

In [25]:

```python
# Replace these with your actual Azure Event Hub credentials
EVENT_HUB_CONNECTION_STRING = 'Endpoint=sb://thesis-eventhub.servicebus.windows.net/;SharedAccessKeyName=spereira-python;Shar
EVENT_HUB_NAME  = 'python-location'


def generate_random_location(fan_id):
    # Lisbon latitude and longitude range
    lisbon_latitude_range = (38.700045, 38.805514)
    lisbon_longitude_range = (-9.211691, -9.124352)

    latitude = random.uniform(*lisbon_latitude_range)
    longitude = random.uniform(*lisbon_longitude_range)

    return {"fan_id": fan_id, "latitude": latitude, "longitude": longitude}


async def send_event_to_eventhub(producer, location):
    fan_id = location["fan_id"]
    latitude = location["latitude"]
    longitude = location["longitude"]

    # Format the event data as JSON
    event_data = json.dumps({
        "fan_id": fan_id,
        "latitude": latitude,
        "longitude": longitude
    })

    # Send the event data as bytes
    await producer.send_batch([EventData(body=event_data.encode("utf-8"))])
    print(f"Sent event for Fan ID {fan_id}")


async def simulate_fan_locations(duration=60, num_fans=10):
    producer = EventHubProducerClient.from_connection_string(EVENT_HUB_CONNECTION_STRING)

    fan_ids = random.sample(range(1, num_fans + 1), num_fans)

    try:
        async with producer:
            tasks = []
            for fan_id in fan_ids:
                location = generate_random_location(fan_id)
                tasks.append(send_event_to_eventhub(producer, location))

            # Send location updates for the specified duration
            end_time = asyncio.get_event_loop().time() + duration
            while asyncio.get_event_loop().time() < end_time:
                await asyncio.gather(*tasks)
                await asyncio.sleep(1)
    except Exception as e:
        print(f"An error occurred: {e}")
    finally:
        await producer.close()

def main():
    loop = asyncio.get_event_loop()
    if loop.is_running():
        loop.create_task(simulate_fan_locations(duration=60, num_fans=10))
    else:
        loop.run_until_complete(simulate_fan_locations(duration=60, num_fans=10))
        loop.run_until_complete(loop.shutdown_asyncgens())
        loop.close()

if __name__ == "__main__":
    main()
```

Figure 4.6 – Python code used to generate random locations

This code snippet demonstrates how to use Python to send event data at regular intervals. It establishes a connection to an azure event hub using the provided connection string and event hub name. The code generates random location coordinates of each fan in the Lisbon area , converts them to a dictionary, and sends them as events to the event hub. The process continues for a specified duration, with events being sent at specified intervals. Finally, the connection to the event hub is closed.

The code generates only random location inside the Lisbon area and sends randomly for each one of the 10 fans one location within that range at a specific time.

#### 4.2.2.2. Azure Event Hub

Azure Event Hubs is a cloud-based data streaming platform provided by Microsoft Azure. It is a fully-managed event ingestion service that can receive and process millions of events per second from a wide variety of sources. Event Hubs is designed to be scalable, reliable, and highly available, making it well-suited for processing large volumes of streaming data in real-time.

Event Hubs can be used for a variety of use cases, including telemetry data ingestion, log and event aggregation, and real-time analytics. Some of the key features of Azure Event Hubs include high throughput, scalability, durability, security, and integration with other Azure services.

In the context of analytics in sport organizations, Azure Event Hubs can be used to ingest and process real-time data from various sources, such as cameras, sensors, and social media, to provide real-time insights into fan behavior, player performance, and other relevant factors. By leveraging the scalability and real-time processing capabilities of Azure Event Hubs, sport organizations can improve their ability to analyze and act upon streaming data to enhance fan experience, increase revenues, and optimize performance.

In this study azure event hub was used to capture and process data from the python script that simulated live location from the fans hours before the match. Due to the service being a highly scalable data streaming platform that enables real-time ingestion of massive amounts of data, this process happens almost instantly.

#### 4.2.2.3. Azure Stream Analytics

Azure Stream Analytics is a cloud-based real-time data analytics service provided by Microsoft Azure. It allows organizations to analyze and gain insights from streaming data in real-time, with minimal setup and maintenance.

Azure Stream Analytics enables organizations to perform various tasks, including:

1. Data ingestion: Stream Analytics can ingest data from various sources, including event hubs, IoT hubs, and other streaming platforms.
2. Real-time data processing: Stream Analytics provides real-time data processing capabilities, allowing organizations to analyze streaming data and act upon it in real-time.

3. Data transformation: Stream Analytics allows organizations to transform streaming data into different formats, such as JSON or CSV, making it easier to work with.

4. Integration with other Azure services: Stream Analytics can be integrated with other Azure services, such as Azure Machine Learning, Azure Event Grid, and Azure Functions, allowing for a complete end-to-end streaming solution.

5. Visualizations: Stream Analytics can be used to create real-time dashboards and visualizations of streaming data, enabling organizations to monitor and track data trends and patterns.

By leveraging the real-time processing capabilities of Azure Stream Analytics, sport organizations can improve their ability to analyze and act upon streaming data to enhance fan experience, increase revenues, and optimize performance.

Using Azure Stream Analytics, the incoming data from Event Hubs is processed and transformed in real-time. Stream Analytics allows to check the data that is retrieved from the azure event hub and then transformations are applied through SQL query and finally sent to an output service, in this case, an azure SQL Database.

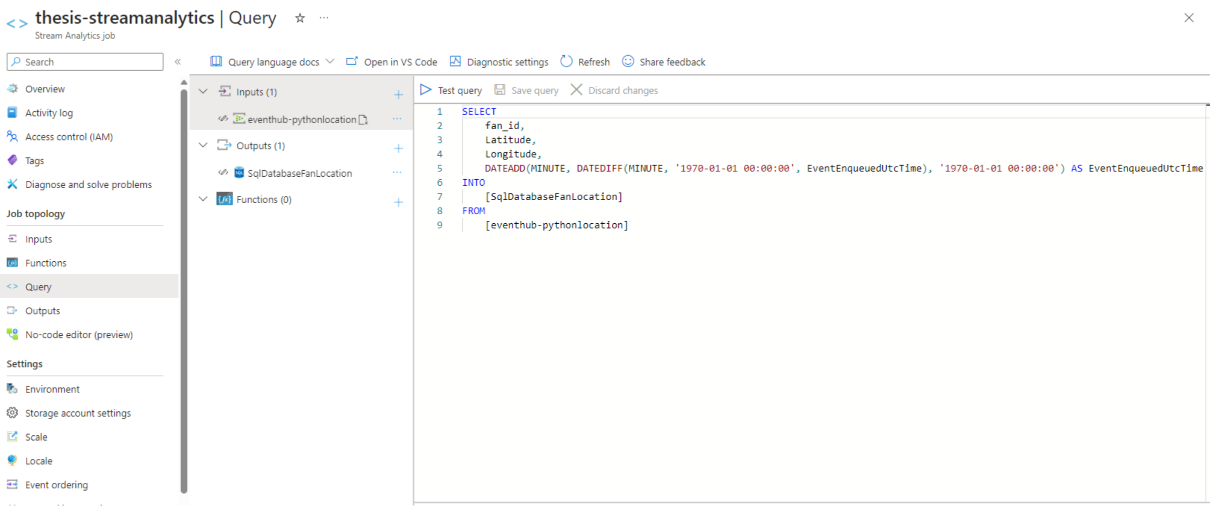


Figure 4.7 – Azure Stream Analytics interface

The Azure Stream Analytics interface, as depicted in Figure 4.7, showcases the workflow where data is inputted and processed before being sent to an output destination via an intermediary SQL query.

As presented the input source in this case is an event hub containing real-time location data of fans. In the provided illustration, a query has been implemented to fetch relevant information from the event hub. This query is designed to select essential fields such as fan_id, Latitude, and Longitude. Additionally, the default event hub field that captures the time of each event's entry into the event hub is also included in the query.

To enhance data manipulation and streamline the data format, a transformation step has been directly integrated into the Azure Stream Analytics query. Specifically, the datetime format has been modified to adopt the pattern '1970-01-01 00:00:00'. This transformation enables more efficient data processing and prepares the information for subsequent stages of analysis and engagement.

#### 4.2.2.4. Azure SQL Database

Azure SQL Database is a cloud-based relational database management system provided by Microsoft Azure. It provides a fully-managed database service that offers high performance, scalability, and security for applications that require a relational database.

In the context of analytics in sport organizations, Azure SQL Database can be used to store and manage data from various sources, such as transactional data, customer data, and other relevant data. By leveraging the high performance and scalability of Azure SQL Database, sport organizations can ensure that their applications can handle increased workload demands and provide fast access to data for analysis and reporting.

In the purposed solution the processed data coming out of stream analytics is stored in a Azure SQL database (table 4.9). This database serves as a central repository for fan information, including their current location and other relevant details. The data is stored in a matter of seconds in a cloud database and ready to be used for live actions or future historical analysis.

| Field | Data Type |
|---|---|
| Fan_id | bigint (Not Null) |
| Latitude | float(Null) |
| Longitude | float(Null) |
| EventEnqueuedUtcTime | datetime(Null) |

Table 4.9 – Metadata Fan Locations SQL table

This is was the table created that includes the fields fan_id, Latitude, Longitude and EventEnqueuedUtcTime field representing the datetime of each event. The data coming out of the stream analytics will populate this table in real-time.

#### 4.2.2.5. Power BI

Power BI is a cloud-based business intelligence (BI) and data visualization platform provided by Microsoft. It enables organizations to connect to various data sources, transform and model data, and create interactive reports and dashboards.

In the context of analytics in sport organizations, Power BI can be used to create interactive reports and dashboards that provide insights into various aspects of the organization, such as fan behavior, player performance, and revenue streams. By leveraging the easy-to-use interface and data connectivity capabilities of Power BI, sport organizations can quickly gain insights into their data and make data-driven decisions to enhance fan experience, increase revenues, and optimize performance.

Using the live connection capabilities of Power BI is possible to read in real time data from the Azure SQL database and visualize the current location of the fans by using the geographical capabilities of this service.

#### 4.2.2.6. Python – Sending Email

Python's integration with Azure services such the process of connecting to an Azure SQL database, turned out to be an essential component of the BI solution. Through the utilization of the pyodbc library, Python successfully established a connection to the database and it was possible to query from the SQL databases, retrieving essential fan information and location data in real-time.

By implementing the Haversine formula, Python efficiently calculated the distances between various fan coordinates (in the Lisbon area) and Estádio da Luz.

It was used the location of the Benfica stadium (Estádio da Luz) as benchmark which coordinates were obtained with the google maps API: Latitude = 38.752711 and Longitude = -9.1847739.

```python
import math
import pyodbc
from datetime import datetime, timedelta
import smtplib


# Haversine formula to calculate distance between two coordinates
def calculate_distance(lat1, lon1, lat2, lon2):
    R = 6371  # Radius of the Earth in kilometers

    # Convert coordinates to radians
    lat1_rad = math.radians(lat1)
    lon1_rad = math.radians(lon1)
    lat2_rad = math.radians(lat2)
    lon2_rad = math.radians(lon2)

    # Calculate differences in coordinates
    dlat = lat2_rad - lat1_rad
    dlon = lon2_rad - lon1_rad

    # Calculate distance using Haversine formula
    a = math.sin(dlat / 2) ** 2 + math.cos(lat1_rad) * math.cos(lat2_rad) * math.sin(dlon / 2) ** 2
    c = 2 * math.atan2(math.sqrt(a), math.sqrt(1 - a))
    distance = R * c

    return distance

# Database connection details
server = 'thesis-sql-server.database.windows.net'
database = 'Thesis-asql'
username = 'spereira'
password = '[redacted]'

# Connect to the SQL Server database
conn = pyodbc.connect(
    f'DRIVER=ODBC Driver 17 for SQL Server;SERVER=thesis-sql-server.database.windows.net;DATABASE=Thesis-asql;UID=spereira;P
)
cursor = conn.cursor()

# Execute an SQL query to retrieve data
cursor.execute("""
    SELECT TOP 1 a.fan_id, a.name, a.email, b.Latitude, b.Longitude, b.EventEnqueuedUtcTime
    FROM fans a
    LEFT JOIN [FanLocations] b ON a.fan_id = b.fan_ID
    WHERE a.fan_id = 1
    ORDER BY b.EventEnqueuedUtcTime DESC
""")

data = cursor.fetchone()

# Close the database connection
conn.close()
```

Figure 4.8 – Part 1 of Python Code used to send emails based on fans live location

Python's exceptional capabilities further came to the fore when enabling the sending of personalized notifications to fans. Leveraging the smtplib library, Python effectively configured the email parameters, facilitating the dispatch of customized notifications to individual fans based on their locations in proximity to the stadium and based on the timestamp. The powerful combination of Python and Azure's real-time data handling capabilities made it possible to send timely, targeted messages, enriching the fans' experience and engagement.

```python
# Reference point coordinates (Estádio da Luz)
reference_latitude = 38.752711
reference_longitude = -9.184773999999999

# Check distances and send personalized messages
fan_id, name, email, latitude, longitude, event_time = data
distance = calculate_distance(latitude, longitude, reference_latitude, reference_longitude)

message = ""
if distance > 5:
    message = f"Dear {name},\n\nWe noticed that you are still far away from the stadium, and the kick-off is near. We advise
else:
    message = f"Dear {name},\n\nSince you are nearby the stadium, come visit our official store and enjoy 20% off on every ar


# Email configuration
smtp_server = 'smtp.office365.com'
smtp_port = 587
smtp_username = 'simao.costa.pereira@hotmail.com'
smtp_password = '[redacted]'
sender_email = 'simao.costa.pereira@hotmail.com'


current_time = datetime.now()
match_time = datetime(current_time.year, current_time.month, 21, 18, 0)  # Set the desired match time
time_difference = match_time - current_time

if time_difference <= timedelta(hours=3.0):
    # Compose the email body with personalized messages
    subject = 'Information ahead of the game'
    body = message

    # Compose the email
    email_message = f'Subject: {subject}\n\n{body}'


    # Send the personalized email to the fan
    if email:
        receiver_email = email

        # Compose the email body with personalized messages
        subject = 'Information ahead of the game'
        body = message

        # Compose the email
        email_message = f'Subject: {subject}\n\n{body}'

        # Send the email
        try:
            with smtplib.SMTP(smtp_server, smtp_port) as server:
                server.starttls()
                server.login(smtp_username, smtp_password)
                server.sendmail(sender_email, receiver_email, email_message)
            print(f'Email sent successfully to {name}!')
        except Exception as e:
            print(f'An error occurred while sending the email: {str(e)}')
    else:
        print('Fan email not available. Cannot send personalized email.')
```

```
Email sent successfully to John Smith!
```

Figure 4.9 – Part 2 of Python Code used to send emails based on fans live location

The code means that if the distance of the fan to the stadium is more than 5km and it is less than 3 hours to the kick-off, then the following email is sent:

“We noticed that you are still far away from the stadium, and the kick-off is near. We advise you to start heading to the stadium as soon as possible due to intense traffic."

Note that, in this research, we do not have access to live traffic, therefore this email is just a suggestion of what could be done given the location of the fan.

As a second possible outcome of the code, if the fan is less then 5km away from the stadium and the game starts within 3 hours, the following email is sent:

“Since you are nearby the stadium, come visit our official store and enjoy 20% off on every article before the game starts."

To be able to get the live location , the right fan name and also the email the before mentioned BI model comes into play. For instance using the Fans (Table 4.5) and FanLocation (Figure 5.6 ) SQL tables and connecting these two tables through the following SQL code we get the following output :

```
SELECT TOP 1 a.fan_id, a.name, a.email, b.Latitude, b.Longitude, b.EventEnqueuedUtcTime
   FROM fans a
   LEFT JOIN [FanLocations] b ON a.fan_id = b.fan_ID
   WHERE a.fan_id = 1
   ORDER BY b.EventEnqueuedUtcTime DESC
```

Results Messages

| fan_id | name | email | Latitude | Longitude | EventEnqueuedUtcTime |
|---|---|---|---|---|---|
| 1 | John Smith | simao.er.costa.pereira@gmail.com | 38,7717372058991 | -9,18790291248247 | 2023-07-21 15:16:00.000 |

Figure 4.10 – SQL code and output used to personalize the emails

So, thanks to the BI model we can retrieve information from different tables and then send a personalized email via python.

In summary, Python's versatility and integration with Azure services were instrumental in both the data generation and database interaction aspects of my thesis. By harnessing Python's capabilities, it was possible to successfully simulated real-time fan movements, established efficient connections with the Azure Event Hubs and Azure SQL database, and facilitated personalized notifications, culminating in an enriched and innovative Business Intelligence model for sports organizations.

# 5.RESULTS AND DISCUSSION

This chapter presents the final actions taken in each Use Case and discuss the effectiveness and potential of these BI solutions.

## 5.1. FIRST USE CASE: REAL-TIME ACCESS TO THE DATA WAREHOUSE AND INTERACTION WITH THE FANS

Using the explained methodology in section 4.1 and 4.2 a personalized email is sent to a specific fan giving certain conditions and the figures 5.1 and 5.2 show how that email would look like.

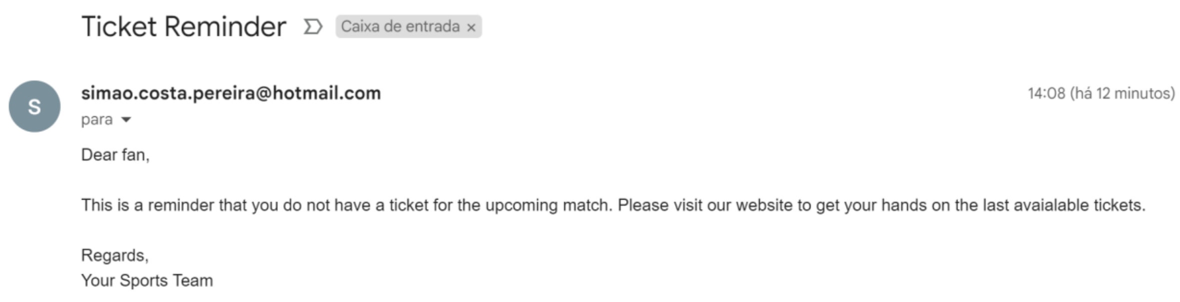


Figure 5.1 – Email sent before the match

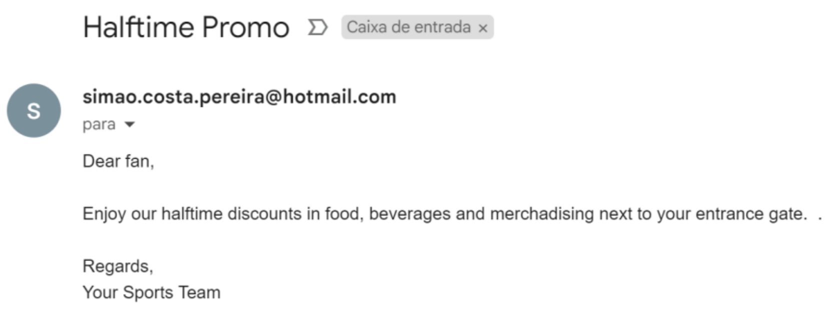


Figure 5.2 – Email sent during the match

Here we observe a range of automated email communication strategies that can be employed to engage with fans prior to and during sporting events. This is made feasible through the utilization of the BI model, which we have previously explained, and the accompanying data pipeline architecture. What sets these communication initiatives apart is their ability to deliver emails to fans in a matter of seconds, effectively in real time. Furthermore, the email content and structure are highly adaptable, allowing for a variety of formats to align with the organization's objectives.

While these outcomes may appear straightforward at first glance, they underscore the profound impact that data and analytical capabilities can have in facilitating real-time interactions with fans. This, in turn, carries the potential to significantly enhance both the overall fan experience and revenue generation.

## 5.2. Second use case: Real-time Location and interaction with the fans

In this section, we present the results throughout each phase of the data pipeline.

### 5.2.1. Azure Event Hub

The code showed in section 4.2.2.1 sends the location of 10 fans to the azure event hub. This what happens if the event hub successfully gets the data from python (Figure 5.3).

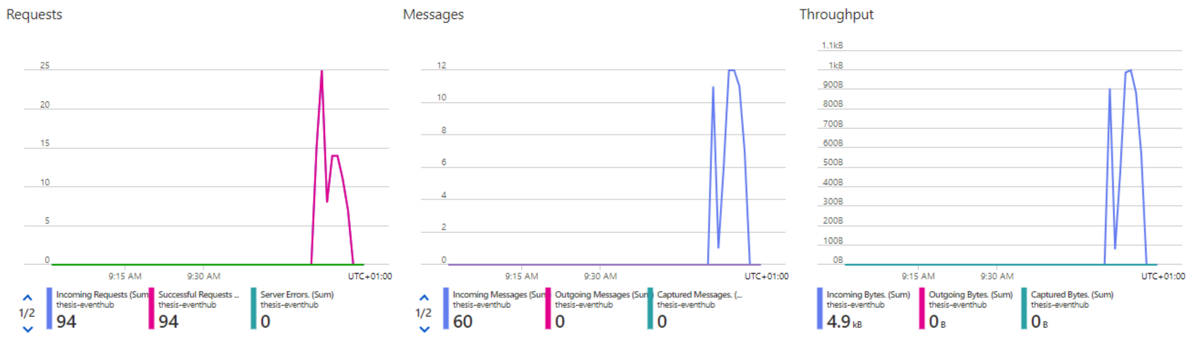


Figure 5.3 – Azure Event Hub incoming graphics overview

The image illustrates the successful output of the Azure Event Hub after running the code. The lines in the graphics confirm that the data from Python was successfully transmitted to the event hub. The image also displays the number of requests made and the corresponding arrival time of the messages. It is noteworthy that the data transmission from Python to the event hub is exceptionally quick, highlighting the effectiveness of event hubs for real-time data ingestion.

### 5.2.2. Azure Stream Analytics

But it is only with the help of stream analytics that we can actually see the data that comes from python and that is ingested by the event hub.

#### 5.2.2.1. Input preview

The following figure shows the input data that comes from the event hub. The data that comes from python is the fan_id, latitude and longitude. The rest of the fields are automatically created by azure event hub such as the EventEnqueuedUtcTime field.

The input data shows that each fan got an id and a random location (in Lisbon region) associated like it was required (Figure 5.4).

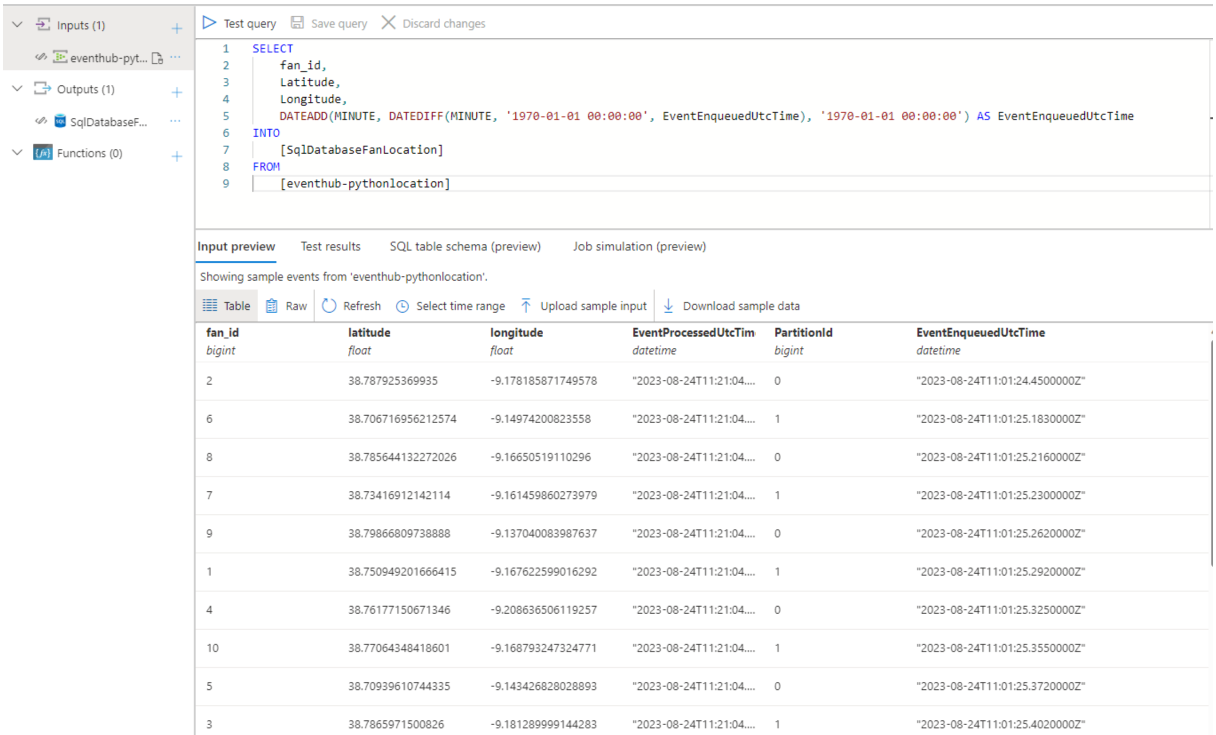


Figure 5.4 – Azure Stream Analytcis Input view

#### 5.2.2.2. Test results

With stream analytics and using the sql query functionality we can test which result will be generated to the output software we choose. In this case as its possible to see on upper left part of the image (Figure 5.4), the azure sql database is our output.

The tests results are shown below and this is what we can expect to have in the azure sql database once the stream analytics job is running.

With the presented sql query the goal was to retrieve data from fan_id, latitude, longitude and also get the time of the event rounded to minutes which is was achieved as shown in figure 5.5.


| fan_id<br>*bigint* | Latitude<br>*float* | Longitude<br>*float* | EventEnqueuedUtcTime<br>*datetime* |
|---|---|---|---|
| 2 | 38.787925369935 | -9.178185871749578 | "2023-08-24T11:01:00.0000000Z" |
| 6 | 38.706716956212574 | -9.14974200823558 | "2023-08-24T11:01:00.0000000Z" |
| 8 | 38.785644132272026 | -9.16650519110296 | "2023-08-24T11:01:00.0000000Z" |
| 7 | 38.73416912142114 | -9.161459860273979 | "2023-08-24T11:01:00.0000000Z" |
| 9 | 38.79866809738888 | -9.137040083987637 | "2023-08-24T11:01:00.0000000Z" |
| 1 | 38.750949201666415 | -9.167622599016292 | "2023-08-24T11:01:00.0000000Z" |
| 4 | 38.76177150671346 | -9.208636506119257 | "2023-08-24T11:01:00.0000000Z" |
| 10 | 38.77064348418601 | -9.168793247324771 | "2023-08-24T11:01:00.0000000Z" |
| 5 | 38.70939610744335 | -9.143426828028893 | "2023-08-24T11:01:00.0000000Z" |
| 3 | 38.7865971500826 | -9.181289999144283 | "2023-08-24T11:01:00.0000000Z" |

Figure 5.5 – Azure Stream Analytcis output view

### 5.2.3. Azure SQL Database

After running the stream analytics job, the data shown in the test results pass in a matter of seconds to the chosen output. And here it is the results stored in an azure sql database (Figure 5.6).

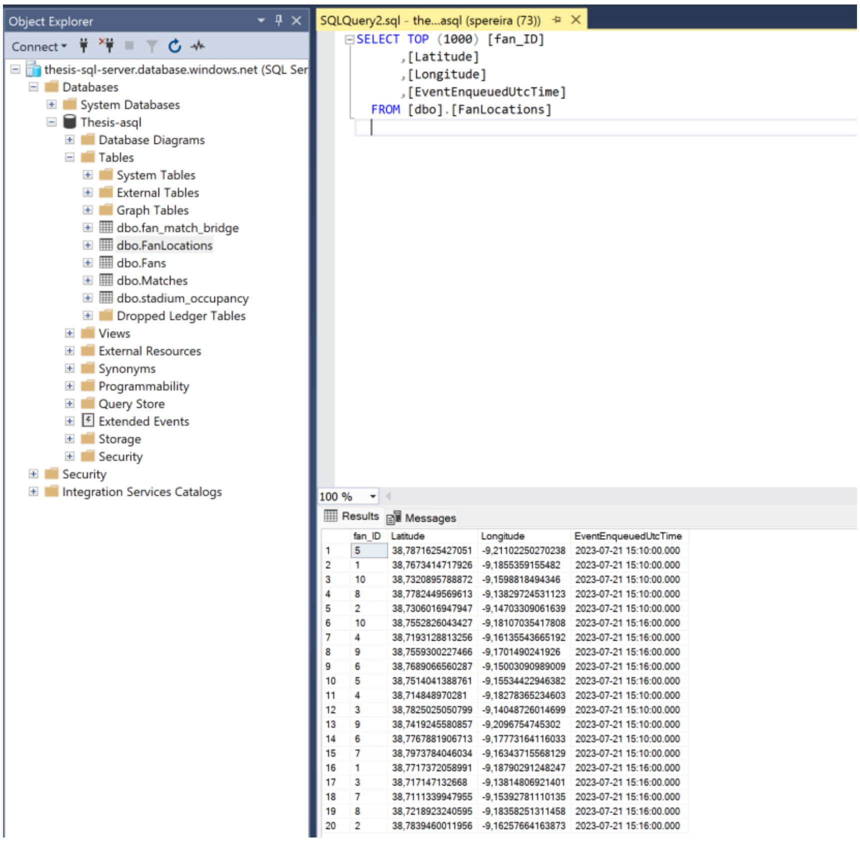

Figure 5.6 – FanLocations data in Azure SQL Database

The image above displays the output in the SQL Database (FanLocations table), showcasing that only the selected columns from the Stream Analytics query are populated in the table. Each row in the table represents the current location of a fan at a specific time.

Here it is possible to check the change of the current position of the fan inside a 6 minute interval. There is data from 3:10pm and another batch from 3:16pm.

For example, the photo above shows all the historical data of the current location of the fans but of course it is possible to filter this information per fan_ID, per specific time range, etc.

### 5.2.4. Power BI

The Power BI interface presents a compelling visualization of real-time data. By establishing a live connection with the previous SQL table, the location of the 10 fans is dynamically displayed as events occur. Notably, the data can be conveniently filtered based on specific criteria, such as viewing the visuals for 3 pm. Furthermore, the flexibility of Power BI enables interactive filtering by fan and/or Datetime, empowering users to derive actionable insights from the data with remarkable ease and precision.

### 5.2.4.1. Location of all the fans at 3:10pm

In this photo we can see where all the 10 fans were located on that day at 3:10pm (Figure 5.7).

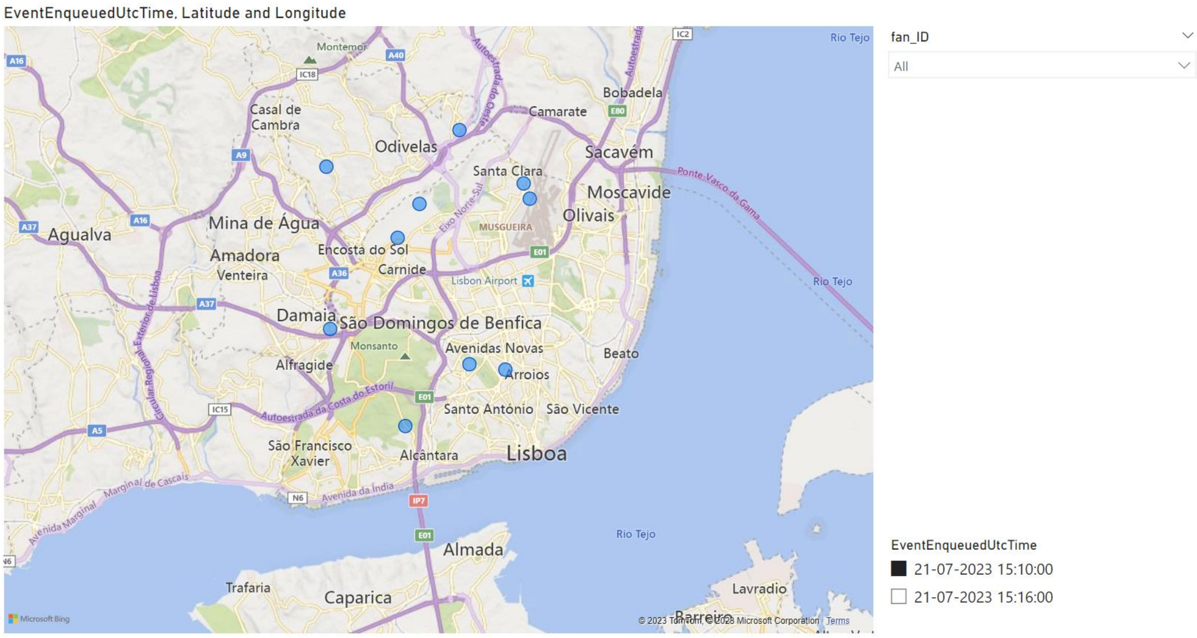


Figure 5.7 – Power BI visualization of fans’ location at 3:10pm

### 5.2.4.2. Location of all the fans at 3:16pm

Here the circles represent the fans’ location at 3:16pm so it is possible to have a idea of the movement of the fans throughout the time

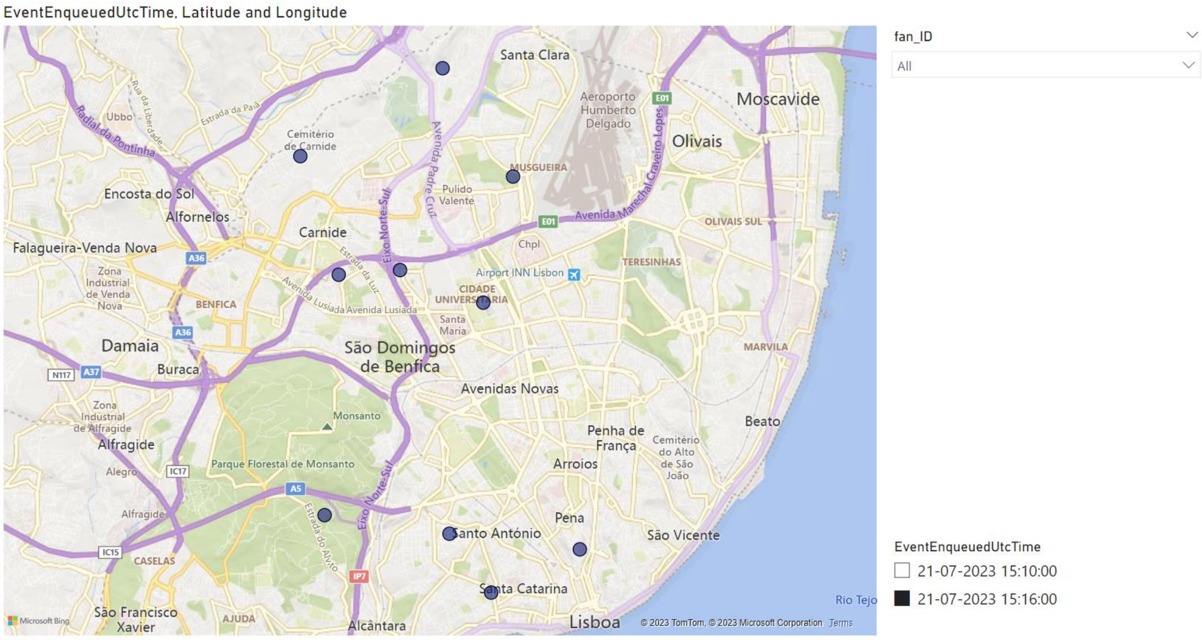


Figure 5.8 – Power BI visualization of fans location at 3:16pm

### 5.2.4.3. Location of fan with ID=1 at 3:10pm and 3:16pm

The next visual shows how we can use the Power BI capabilities, having the BI model as foundation, to make even deeper analysis. Here we are selecting the location at 3:10pm and 3:16pm and also filtering by a specific fan (in this case fan with id =1) and we can understand better his movements in just a space of 6 minutes.

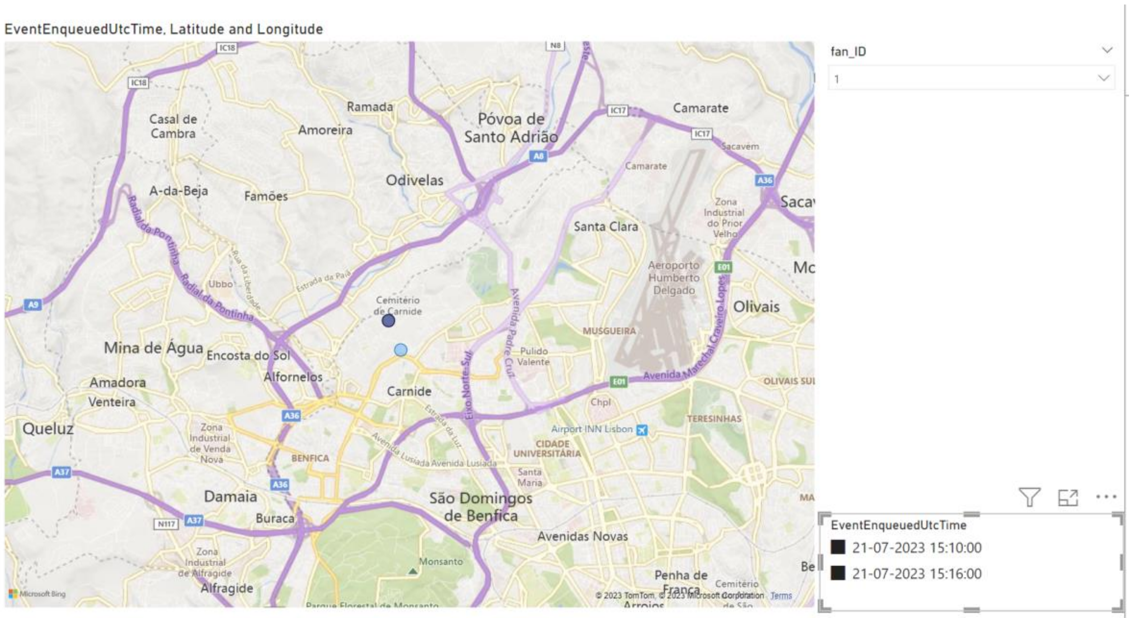


Figure 5.9 – Power BI visualization of fan(with id=1) location

### 5.2.5. Action with Python

After the data goes through this pipeline it is time to take meaningful actions to engage with the fans and to try to increase the revenues and create a better fan experience. Several actions can take place using this information but for the study purpose the two examples mentioned in section 4.2.2.6 will be shown.

For example, when using the previous data flow if a fan is close to the stadium (in this study we consider close, less than 5km away from Estádio da Luz stadium) and the game starts within three hours, the following targeted email (figure 5.10) is sent to alert for some promotion campaigns going on in the stadium.

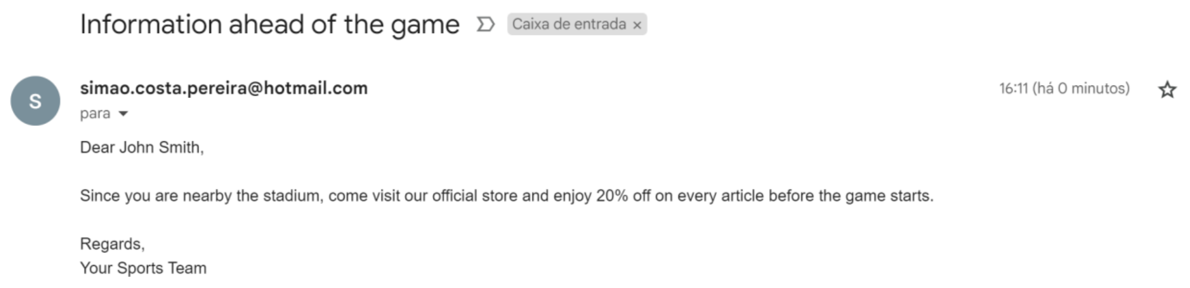


Figure 5.10 - Email sent in case the fan is close to the stadium

But if the fan is far away and the game starts within 3 hours we can increase his experience by reminding him of the intense traffic and advise him to go to the stadium as soon as possible (figure 5.11).

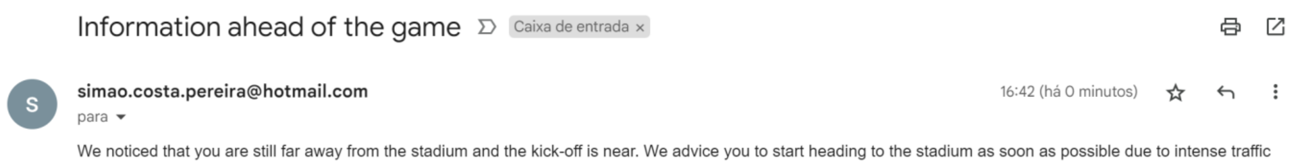


Figure 5.11 - Email sent in case the fan is far away from the stadium

### 5.3. Discussion

When considering the implementation of real-time data and interaction with fans, the efficiency of the solutions employed becomes crucial. In this context, both the Python random location generator and the integration of Event Hubs, Stream Analytics, SQL Database, and the final Python script offer exceptional performance. The flow of data through these platforms occurs within a matter of seconds, ensuring that organizations have access to up-to-date information in a timely manner. This real-time capability is of utmost importance when dealing with streaming data, as it enables organizations to quickly identify and respond to relevant events or trends. Furthermore, the ability to take live actions based on this data is equally significant. By ensuring that the information reaches the fans swiftly, organizations can have the desired effects, whether it be sending personalized emails regarding ticket availability or engaging fans during the match. The speed and efficiency of these solutions enable organizations to create a seamless and immediate connection with fans, enhancing their overall experience and optimizing revenue opportunities.

These actions were tested several times and it is possible to say that these processes are fast and close to real-time. The emails are sent and received in a matter of seconds making these BI solutions appropriate for the challenge. Of course, it is worth mentioning that the data used for this experience is way smaller than what would happen in real-life.

Both BI solutions offer several potential actions in real time. For instance, if we identify a fan located in close proximity to the stadium one hour before the game, we can utilize this solution to send them a targeted marketing message, reminding them to explore ongoing merchandising promotions at the stadium's store on that matchday. Another example would be to notify a fan who is still far from the stadium's location, advising them to take the subway instead of driving due to heavy traffic congestion in the stadium's vicinity.

Furthermore, through future analysis of the stored real-time data, significant insights can be gained. For instance, by analyzing the fan location data during the two hours prior to the game, patterns may emerge. It may be observed that most fans only approach the stadium approximately 45 minutes before the game. Consequently, this information allows for strategic decisions, such as launching marketing campaigns one hour prior to the game or adjusting security measures three hours before the match based on the anticipated fan presence.

Emails were just one of the actions that could have been made to interact with fans. Also, the content of email and the ways to personalize even more the content is enormous. For instance, instead of email, a text message could have been sent because we have the information of the fans phone number.

In terms of content and with this data warehouse that was created the email/text message could also contain information such as which kind of tickets are still available before the match or for what price the fans can still get them.

By utilizing real-time data, we can create a dynamic and personalized fan experience, enhance ticket sales, and increase fan engagement. The possibilities for interacting with fans and tailoring marketing efforts are vast, ultimately leading to improved fan satisfaction and revenue generation for the organization. These examples illustrate a fraction of the multitude of possibilities that can be realized through the implementation of analytics within sports organizations. While this case study may not cover the full extent of these possibilities, it provides valuable insights into the potential benefits that analytics can offer within the sports industry.

The research undertaken in this master thesis has successfully addressed the identified research gap surrounding the limited integration of analytics approaches within diverse domains of sports organizations. By offering insights into areas such as merchandising, food and beverages, marketing, user experience, and fan engagement, the study showcases the potential of analytics to revolutionize revenue growth and enhance the matchday experience for fans. The effectiveness of the implemented analytics solution resonates strongly with the research question and contributes to the advancement of the sports industry's operational and engagement paradigms.

The initial research gap pinpointed the lack of comprehensive analytics utilization in non-traditional facets of sports organizations. Through meticulous investigation and empirical application of analytics techniques, the thesis has effectively bridged this gap. The showcased Business Intelligence approach utilizing Python scripts, Azure cloud services (Event Hubs, Stream Analytics, Azure SQL Database), and Power BI exemplifies the adaptability of modern technologies in revolutionizing operations and interactions within sports entities.

One of the central aspects addressed by the research question pertains to the potential of analytics in increasing revenues for sports organizations. The findings reveal that the integration of real-time data and targeted interactions leads to tangible benefits. By personalizing fan experiences and optimizing ticket sales strategies, sports organizations can maximize revenue generation. The presented analytics solution serves as empirical evidence of this potential, demonstrating that data-driven insights can facilitate smarter business decisions that directly impact financial growth.

Enhancing the matchday experience for fans emerged as another significant dimension of the research question. The implemented analytics solution effectively showcases how real-time data analysis can elevate the fan experience. Through personalized interactions, tailored recommendations, and improved engagement, the study highlights the power of analytics in fostering a more connected and satisfied fan base. This resonates with the research gap, as the thesis not only addresses this aspect but also substantiates it through practical implementation.

# 6. CONCLUSION

With this research, a comprehensive exploration of analytics solutions and real-time interactions within the realm of sports organizations was undertaken, addressing critical research gaps and significantly advancing the understanding of how data-driven strategies can reshape revenue growth and fan experiences. Through the development and examination of two distinct use cases, the thesis showcased the transformative potential of analytics within the sports industry.

It is worth noting that the emphasis of this work was not on the software or app development aspect. The data generation employed in this thesis was purely illustrative, and the presented BI solution can be applied to any context, regardless of how the data is collected from the fans.

The first use case focused on harnessing analytics to drive revenue growth within a football club. By creating, accessing, and reading from a real-time data warehouse, the study demonstrated the tangible benefits of leveraging data insights to optimize various aspects of club operations. The practical implementation emphasized how analytics can fine-tune merchandising strategies, enhance marketing approaches, and improve the overall fan experience. This use case effectively translated theoretical concepts into real-world application, highlighting the direct impact of data-driven decisions on revenue generation.

The second use case delved into the live interaction aspect, showcasing how Azure cloud services can revolutionize fan engagement. Through real-time location tracking and personalized interactions, the study illustrated the potential for technology to create seamless and immersive experiences for fans during matchdays. The implementation of Azure cloud services such as Event Hubs, Stream Analytics, and Power BI provided a robust framework for understanding how real-time data can be harnessed to foster stronger connections between sports organizations and their fan base.

Both use cases converge to demonstrate the power of analytics and real-time interactions in driving revenue growth and fan engagement. The strategic utilization of data insights and technological solutions offers sports organizations the opportunity to make informed decisions, optimize operations, and enhance the overall experience for fans. By addressing research gaps and providing tangible examples, this work serves as a contribution to the sports industry's digital transformation journey.

Regarding the research topic, which delves into the role of analytics in sport organizations to increase revenues and enhance the fan experience on matchdays, this study has shed light on the tremendous opportunities available in this area. It is evident that analytics plays a critical role in the present and future of sports organizations, and there is substantial room for improvement and innovation in this field.

In conclusion, this master thesis serves as a testament to the transformative potential of analytics solutions and real-time interactions within the sports industry. By building and analyzing two comprehensive use cases, the study not only answered the research question but also showcased how data-driven strategies can drive revenue growth and enhance fan experiences. As sport organizations continue to evolve in the digital age, the insights from this thesis offer valuable guidance and inspiration for leveraging technology and data to create a new era of sports engagement and operations.

# 7.LIMITIONS AND RECOMMENDATIONS FOR FUTURE WORKS

In this thesis, certain limitations were encountered that provide avenues for future work. One significant limitation relates to the software engineering aspect, which could have added substantial value and completeness to the study. Developing an actual mobile app for a sports organization would have allowed for the extraction of fan data and the testing of the BI solution in a real-world scenario. While this was not the primary objective of the research, integrating the solution with a functioning app would enhance its practical applicability.

Additionally, the limitation of having only two use cases must be acknowledged. Due to time constraints and academic considerations, the exploration of fan data was limited, although there were numerous opportunities to delve into different aspects of fan engagement. Expanding the scope of use cases would enrich the research and provide a more comprehensive understanding of the potential applications of the real-time BI solution.

Budget constraints and the expenses associated with cloud services imposed limitations on the exploration of additional functionalities within the Azure cloud platform. Future research could address this limitation by allocating sufficient resources to explore more advanced features and capabilities of cloud services. Additionally, considering different cloud service providers and alternative streaming data architectures could provide further insights and enhance the solution's power and efficiency.

Moving forward, future research should aim to incorporate various types and larger quantities of data to enrich the analysis. For instance, integrating live traffic data, ticket holder information, or real-time weather data could enhance the precision and contextual relevance of the solution. This would allow for a more comprehensive understanding of fan behavior and enable personalized and targeted interactions.

Lastly, to validate the effectiveness of the solution, it would be valuable to develop or utilize an existing app where the proposed BI solution can be applied and tested on a bigger dataset. This real-world implementation would provide empirical evidence of its practicality and efficacy in enhancing fan experiences and driving revenue.

By addressing these limitations and pursuing future research along these lines, a more comprehensive understanding of real-time BI solutions for sports organizations can be achieved, paving the way for enhanced fan experiences and revenue generation in the industry.

# 8.BIBLIOGRAPHY


Apostolou, K., & Tjortjis, C. (2019, November 14). Sports Analytics Algorithms for Performance Prediction. In Proceedings of the IEEE Conference. Publisher: IEEE. ISBN: 978-1-7281-4960-8.

Troilo, M., & Simkins, B. J. (2021). Business analytics, revenue management, and sport: Evidence from the field. Journal of Revenue and Pricing Management. https://www.researchgate.net/profile/Michael-Troilo/publication/346167787_Business_analytics_revenue_management_and_sport_evidence_from_the_field/links/5fc68fc4a6fdccfea3f0830f/Business-analytics-revenue-management-and-sport-evidence-from-the-field.pdf

Gong, X., & Wang, Y. (2021). Exploring dynamics of sports fan behavior using social media big data - A case study of the 2019 National Basketball Association Finals. *Applied Geography*, *129*, 102438. https://doi.org/10.1016/J.APGEOG.2021.102438

McHale, I. G., & Holmes, B. (2023). Estimating transfer fees of professional footballers using advanced performance metrics and machine learning. *European Journal of Operational Research*, *306*(1), 389–399. https://doi.org/10.1016/J.EJOR.2022.06.033

Arkenberg, C., Giorgio, P., & Deweese, C. (2019). Redesigning stadiums for a better fan experience Putting sports fans at the center of smart stadium transformation.

Huettermann, M. (2021). Fan engagement in professional team sports Digital Marketing in Switzerland View project. https://www.researchgate.net/publication/357396428

Ratten, V., & Dickson, G. (2020). Big data and business intelligence in sport.

Singh, N. (2020). Sport Analytics: A Review. The International Technology Management Review, 9(1), 64. https://doi.org/10.2991/itmr.k.200831.001´

IEEE Staff. (2013). 2013 3rd International Conference on Consumer Electronics, Communications and Networks (CECNet). IEEE.

Bouquet, P., Molinari, A., Bortolan, L., Dal Pra, A., Bezzi, G., & Vettorato, S. (2022). A cloud-based management system for a data infrastructure in four winter sport facilities. 2022 IEEE

International Workshop on Sport, Technology and Research, STAR 2022 - Proceedings, 94–99. https://doi.org/10.1109/STAR53492.2022.9859651

Venugopal, S. (2023). Big Data - Enhancement in Sports : Exploring the Impact of Technology Integration. 2023 International Conference on Computer Communication and Informatics, ICCCI 2023. https://doi.org/10.1109/ICCCI56745.2023.10128427

Reis, J., & Housley, M. (2022). Fundamentals of Data Engineering. O'Reilly Media, Inc.

Kimball, R., & Ross, M. (2013). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. John Wiley & Sons.

Sabherwal, R., & Becerra-Fernandez, I. (2010). Business intelligence: Practices, technologies, and management. John Wiley & Sons.

Turban, E., Sharda, R., Delen, D., & Aronson, J. E. (2017). Business intelligence and analytics: systems for decision support. Pearson

Watson, H. J., & Wixom, B. H. (2007). The current state of business intelligence. Computer, 40(9), 96-99.

Jhawar, R., & Tejada, Z. (2022, June 21). ETL process in Azure**.** *Microsoft Azure Architecture Center.* https://learn.microsoft.com/en-us/azure/architecture/data-guide/relational-data/etl

Tejada, Z. (2022, July 25). Data Warehousing**.** *Microsoft Azure Architecture Center.* https://learn.microsoft.com/en-us/azure/architecture/data-guide/relational-data/data-warehousing

Neto, M. C. (2022). Business Intelligence and Data-Driven Organizations. NOVA Information Management School, Lisbon.

Sutton, W. A., Urban, T. L., Troilo, M., Bouchet, A., & Mondello, M. (2020). Business analytics, revenue management, and sport: evidence from the field. *International Journal of Revenue Management*, *11*(1), 1. https://doi.org/10.1504/ijrm.2020.10032774

Caulfield, J., & Jha, A. K. (2022). Stadiums and Digitalization: An Exploratory Study of Digitalization in Sports Stadiums. *Journal of Decision Systems*, *31*(S1), 331–340. https://doi.org/10.1080/12460125.2022.2073629

Ratten, V., & Dickson, G. (2020). *Big data and business intelligence in sport*

Microsoft. (2021, October 11). NBA uses Microsoft Azure to create new content that increases fans engagement. *Microsoft Customers.* Retrieved from https://customers.microsoft.com/en-us/story/1423195401562064416-nba-media-entertainment-azure-pt-br

Microsoft. Get started with Event Hubs using Python. Azure Event Hubs documentation. https://learn.microsoft.com/en-us/azure/event-hubs/event-hubs-python-get-started-send?tabs=passwordless%2Croles-azure-portal. Accessed on December 10, 2022

Temporal Data (AALTD). https://ieeexplore.ieee.org/abstract/document/8029342

Microsoft. Azure Event Hubs - Overview. Microsoft Azure. https://learn.microsoft.com/en-us/azure/event-hubs/event-hubs-about. Accessed on March 15, 2023

Microsoft. Process data with Stream Analytics. Microsoft Azure. https://learn.microsoft.com/en-us/azure/event-hubs/process-data-azure-stream-analytics. Accessed on April 10, 2023

Microsoft. Azure Stream Analytics - Real-time data stream processing from millions of devices. Microsoft Azure. https://learn.microsoft.com/en-us/azure/stream-analytics/. Accessed on April 18, 2023

Microsoft. Azure Functions – Serverless computing. Microsoft Azure. https://learn.microsoft.com/en-us/azure/azure-functions/functions-overview. Accessed on April 18, 2023

Microsoft. Azure Databricks – Fast, easy, and collaborative Apache Spark-based analytics platform. Microsoft Azure. https://learn.microsoft.com/en-us/azure/databricks/. Accessed on April 14, 2023

Microsoft. (2023, March 30). Stream data from Azure SQL Database to Azure Stream Analytics. Microsoft Azure. https://learn.microsoft.com/en-us/azure/azure-sql/database/stream-data-stream-analytics-integration?view=azuresql. Accessed on March 30, 2023

Microsoft. Connect to Azure data services from Power BI. https://learn.microsoft.com/en-us/power-bi/connect-data/service-azure-and-power-bi. Accessed on May 10, 2023

Microsoft. Develop Azure Functions by using Visual Studio Code. Azure Functions. https://learn.microsoft.com/en-us/azure/azure-functions/functions-develop-vs-code?tabs=csharp. Accessed on May 10, 2023

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa