NOVA IMS Information Management School

MDSAA

Master Degree Program in
Data Science and Advanced Analytics

Data Driven Sports Organisations

Empowering Small Sports Organizations with Data-Driven Training Strategies

Tiago Filipe Ramalho dos Santos

Project Work

presented as partial requirement for obtaining a Master's Degree 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

Data Driven Sports Organisations

Empowering Small Sports Organizations with data-Driven Training Strategies

by

Tiago Filipe Ramalho dos Santos

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

Co-Supervisor: Miguel Neto

Co-Supervisor: Bruno Jardim

February, 2024

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 acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Tiago Filipe Ramalho dos Santos

Belas, 11/02/2024

DEDICATION

I dedicate this project to my parents António and Manuela, and sister Andreia, whose support, motivation and love have been invaluable for my success.

I want to also dedicate it to my friends, namely Sarra Jbeli, Catarina Teixeira, Nina Urbančič, Renan Stoffel and Simão Pereira. Thank you for your words, encouragement and solidarity.

Having you all with me on this journey made it all the more rewarding.

ACKNOWLEDGEMENTS

I wish to express my gratitude to several people for their roles in the completion of this project.

To my professor and advisor Bruno Jardim, for the captivating classes of Business Intelligence and Power BI, and his coaching throughout this project.

To my professional supervisor Rui Valente and my team, namely Alice Vale, Beatriz Crispim, Carlos Lima, Catarina Rocha, Cátia Reis, Jorge Pessoa, Maria Ana Teixeira and Pedro Ribeiro. Their support and teaching were invaluable.

ABSTRACT

In modern days, the use of technology has reached unprecedented levels, spreading to almost every industry, including the industries of Health, Wellness and Sports. In this sense, Business Intelligence can be leveraged to help athletes maximize their potential, basing their strategies on knowledge-based decisions. However, the use of such techniques in small clubs, neighborhood organizations and unofficial groups of athletes is very limited. In this project, with the aim of benefiting small organizations and amateur athletes, Data Preparation and Visualization techniques were applied to create a Power BI Report that transforms raw data into applicable business intelligence. Through its analysis, the users can draw conclusions on what exercises are more relevant to their performances and adjust their training schedule, basing their decisions on knowledge.

KEYWORDS

Data-Driven; Sports; Efficiency; Business Intelligence; Power BI

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS AND ACRONYMS

BI	Business Intelligence
BMI	Body Mass Index
DAX	Data Analysis Expression
DSS	Decision Support Systems
EI	Effort Index(es)
EIS	Executive Information Systems
MIS	Management Information Systems
NBA	National Basketball Association
UI	User Interface
UX	User Experience
VBA	Visual Basic

1. INTRODUCTION

1.1. MOTIVATION

Now-a-days, technology use is at an all-time high, and it's being applied on health and wellness world. Technology allows us to keep track of health and fitness data like heart rate, body temperature, level of stress, calories burnt, blood oxygen levels and other indicators, through smart watches and even rings.

This information is being used for: keeping track of diseases by, for example, letting the user know when their temperature is dropping or increasing to unusual levels; helping with mental health, by alerting the user to when their stress is increasing, which might help them detect situations that trigger stress and anxiety; simpler tasks as keeping track of the quality in sleep which allows the user to work on a better pre-sleeping routine.

The data that technology enables us to gather can and should also be used in sports-related matters. Business Intelligence can be leveraged to help athletes maximize their potential and improve their performances.

The use of business intelligence in small clubs, neighborhood organizations and unofficial groups of athletes is very limited. According to Wright (2017), analytics methods like BI are rarely applied in amateur sports organizations. This leaves these sports organizations in a situation of basing their decisions on intuition and "traditional wisdom". This last pillar of their decisions is the wisdom that has been passed from generation to generation and is valuable to a certain level, but has its flaws and doesn't take into consideration all the factors available that can be better applied when using technology.

1.2. OBJECTIVES

With the goal of helping smaller sports organizations take advantage of technology and bringing Business Intelligence to these organizations, we envisioned a simple and intuitive tool that can be used to keep track of athletes' performances and trainings, and through its analysis take conclusions on what exercises are more relevant to their performances.

This solution would guarantee a level of ease of comprehension, utilization and maintenance that would satisfy organizations with very limited resources. This requirement led us to the use of some indexes that are examples of measures of effort, and to allow users to replace them with others that they consider more enlightening and valuable. Also, on the subject of preparing the solution to make it easy to be used by a smaller organization, we prepared the report to load data that is stored in a local excel, and athletes' images that are stored in a local folder.

Through the use of our solution, a small sports organization will have a better understanding of its athletes' performance and effort put into trainings: the evolution of ratings in events, the level of dedication the athlete invested in each training, and the correlation between both indicators, allowing the coaches and athletes to better plan and strategize their trainings. This report allows practices to become more efficient, helps athletes achieve better results and boosts their confidence in the ability to reach their full potential. Throughout this paper we will provide a detailed explanation of how these goals are achievable.

The process of using our solution can be seen in Figure 1.

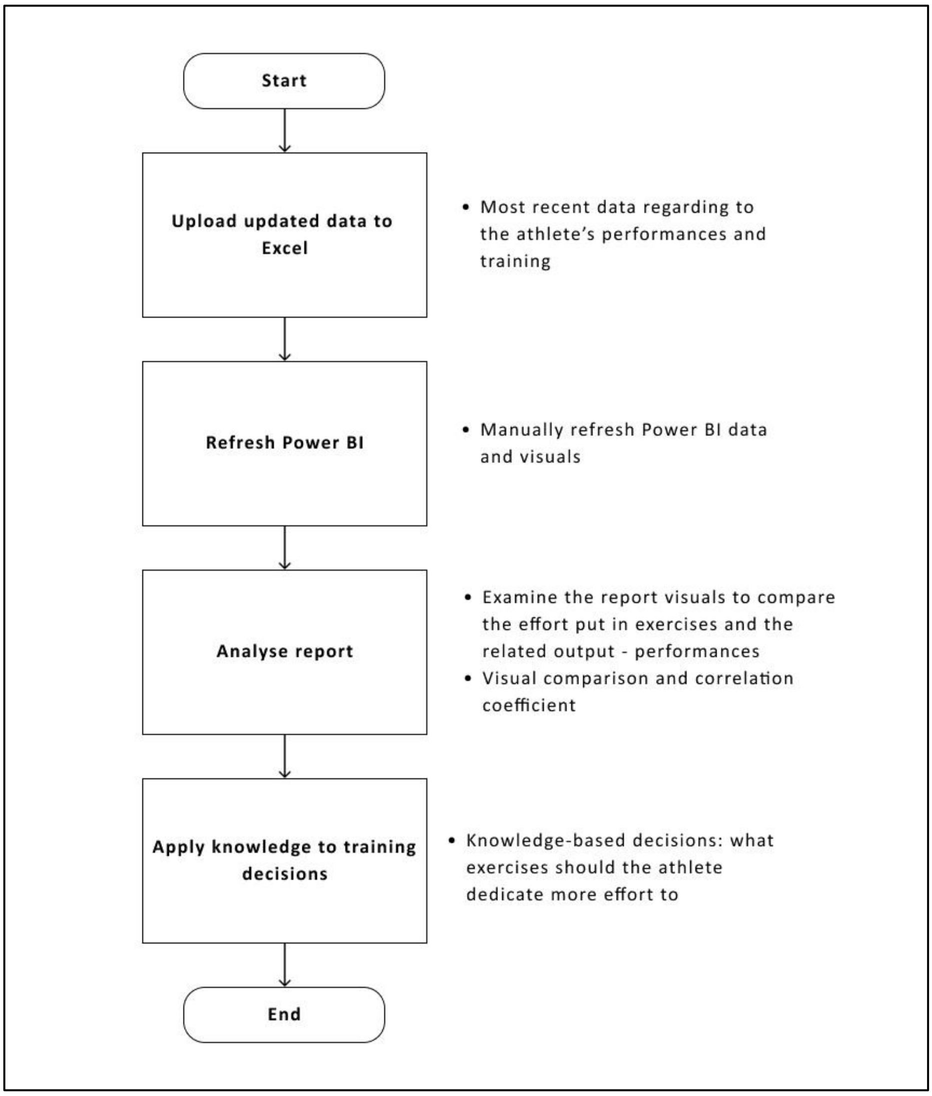

Figure 1 – User Interaction Process

It is important to note that the solution is a tool meant to facilitate the analysis of the data and bring value to the athlete and organization, but this analysis should be accompanied by business intuition. Alone, neither of them is enough to take the athletes to the next level.

This document describes the steps taken to reach the solution envisioned and its structure will consist of the following sections. The first section is “Introduction”, in which we expose the Motivation and Objectives of this project. The second section “Literature Review” covers the examination of relevant literature for the project. The third is “Methodology”, through which we present and describe the generation, preparation and modelling of our data, and the analysis techniques applied to it. After this, follows “Results and Discussion”, in which we expose the dashboards developed, the objective in each visual created and their interactions. Section 5 is “Limitations and Future Work”, exposing the obstacles found throughout the project and future developments in our report that might create more value to the user. The last section is number 6 “Conclusion”, which consists of the summary of the whole project.

2. LITERATURE REVIEW

2.1. BUSINESS INTELLIGENCE

Business Intelligence (BI) is a series of techniques, strategies and technologies that provides wisdom to businesses. The wisdom is derived from significant data, turned into information and, later, knowledge, that is created with the purpose of enabling the user to ground their decision-making processes with knowledge that reflects “historic operational data” (Sun et al., 2018). According to Trieu (2017) even though Business Intelligence is predominantly used for decision-making activities, in recent years there’s been a growing application on institutional learning and transformation, refining efficiency and enhancing “organizational intelligence”. Moreover, Lea et al. (2018) defined BI as the process that, through the use of raw historical data from the organization’s internal sources, identifies trends and patterns, makes knowledge-based strategic decisions, and provides real-time information to support the operational decision-making.

Having established an understanding of what Business Intelligence means, it's important to acknowledge its history and evolution, assess what benefits BI can bring to an organization and what are Business Intelligence tools.

2.1.1. The Evolution of Business Intelligence

Business intelligence emerged in response to the necessity to address the challenges associated with data collection and transforming the data into knowledge for decision-making. Its early development was a result of the progressive evolution and reduced costs of technology in data gathering, analysis, interactivity, and personal computing. The transformative journey of BI began with the introduction of spreadsheets in the 1980s and has since evolved into one of the top investment areas of IT organizations. Marjamäki (2017) outlines the evolution of Business Intelligence into two main phases: the first spanning from 1960 to 1989 and the second from 1990 to the present day.

During the first phase, BI emerged and evolved from other managerial decision support systems, namely Decision Support Systems (DSS), Management Information Systems (MIS), and Executive Information Systems (EIS). However, what sets BI apart from these systems is its interactive features along with the inclusion of Online Analytical Processing (OLAP) features.

By the late 1980s, BI gained more popularity after the development of data warehousing, database management, and analytical methods. Eventually, it became a huge interest after it was introduced by consultants as the new “organizational innovation” (Gilad & Gilad, 1986) and it became an established part of organizational functions.

By the early to mid-1990s, BI became and remained a mainstream tool supporting informed decision-making. Organizations now continue to recognize the value of BI-based technology making it one of the top investments in IT departments for many years (Pekka Marjamäki, 2017).

2.1.2. The Benefits of Business Intelligence

According to Choi (2022), surveys show that the most significant benefits of BI are: acquiring reliable data to help decision-making, enhancing the ability to assess opportunities and threats, enhancing organizational knowledge, and improving information exchange.

Watini et al. (2022) discloses the main benefits as: cost savings from market data consolidation, time saving for suppliers, cost saving for industry, best choices in decision making and objective and profitable strategies.

In 2020, Yu Yin Lim and Ai Ping Teoh presents the benefits of BI in 3 big advantages:

- Enhancing Economic Performance: There is a positive and significant correlation between the application of BI techniques and an impact on a company's economic performance. BI allows the company to monitor "sale trends, customer requests and complaints, anticipate customer behavior and market demand" leading to the company reaching economic goals, including improvement in "cost reduction, productivity (…), product development, customer service development, and income".
- Strengthening Environmental Performance: A big challenge that every company should concern themselves with is to keep some balance between their profit and the effect that their operations have on the environment. To keep up with this challenge, the availability of relevant and live knowledge of the business is essential. As a result, the use of BI methods is advised, given that companies can "observe the strategic impact in terms of diminishing the corporate impact on the natural environment, global warming, and respond strategically to the ongoing climate change".
- Reinforcing Social Performance: BI enhances the availability of knowledge and innovation, which leads to the development of, not only the operational side but also, human capital and relations with the community. These tools and techniques provide the company with knowledge regarding customers, suppliers, competitors and employees, and insights regarding "potential areas for contributing to society via effective corporate social programs that address the needs of the community".

In conclusion, BI can offer support in financial monitoring, keeping track of the operational performance and in the process of identifying key matters that require the attention of the organization (Sorour et al., 2020).

2.2. BUSINESS INTELLIGENCE IN SPORTS

While BI is, traditionally, associated with financial and operational challenges, its versatility results in the use of its tools in many fields and for different types of challenges.
Given its goal of improving the success of decision-making, the application of BI tools has captured the attention and interest of a broad range of industries. In particular, studies have proved its value for financial institutions, entertainment businesses, retail companies, and healthcare (Caya, O., & Bourdon, A. (2016). The Sport industry also understood the benefit of Business Intelligence techniques, specifically competitive and professional sports.

Alamar (2013) defined the usage of these techniques in the sports industry as "the management of structured historical data, the application of predictive analytic models that utilize that data of information systems, in order to inform decision makers and enable them to assist their organizations in gaining a competitive advantage on the field of play". BI tools enable the decision makers, i.e. coaches, athletes, and other staff, to improve their decision-making strategies and methodologies, by providing a "reliable and systematic analysis of the data".
The analysis provided by BI tools, enable the coaches and managers to objectively evaluate performance levels in competitive contexts in order to adapt and improve it (Morgulev et al., 2018).
Caya and Bourdon (2016) emphasize the value that comes from the analysis of historic data, to both individual athletes, who for example compete in running, cycling or athletics, and organizations such as football clubs. Business Intelligence and Analytics provide techniques and strategies that can be of great interest for these industries, in situations such as strategically organize sports events, monitor games and their statistics, they can enhance the protection of players throughout events, optimizing performances, and even allow better management of the team and staff, both in day-to-day situations and big decisions as deciding what coach to hire.

2.2.1. Sports Analytics Framework

Caya and Bourdon (2016) illustrated the process through which a sports analytics project would go, from raw data to "knowledge-based strategic decisions" (Figure 2).

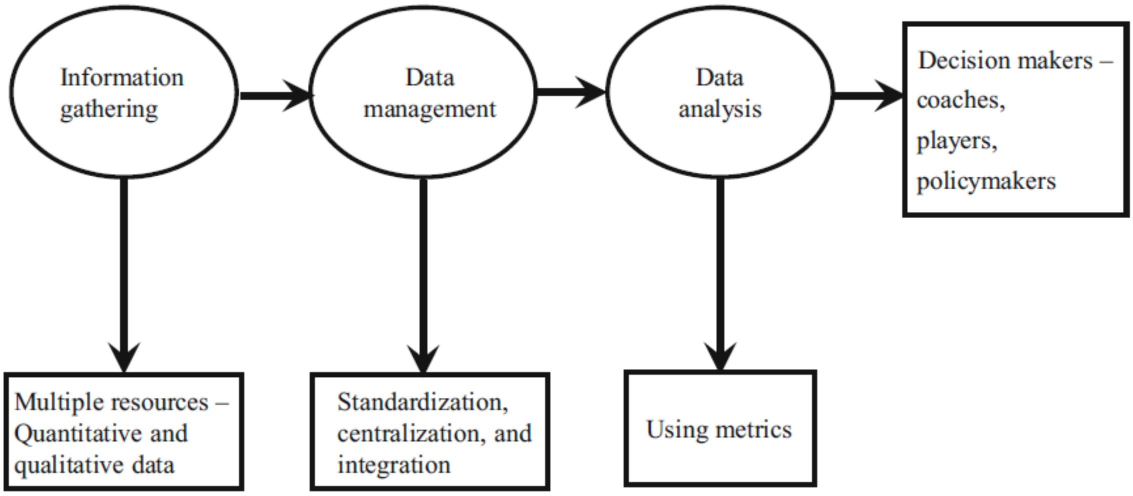

Figure 2 - A sports analytics framework - Caya and Bourdon (2016)

Quantitative and qualitative data is collected from multiple sources, after which is handled and transformed with techniques of standardization, centralization, and integration. This allows the data to be better understood and to easier to be used on visualizations and consequent analysis. Finally, the data is analysed through metrics allowing the production and sharing of specialised knowledge. This step is what creates value, with the use of knowledge in decision making.

2.2.2. Applications in Sports

The use of analytical techniques as BI in sports is anything but recent. Although today one of the bases of BI is the use of technology, it is merely a way to make the process more efficient and to lead to data-driven decisions faster (Foote, K., 2023). There are many examples of individuals and organizations using Business Intelligence before the introduction of the BI tools used in the present days.

One of the first scientific publications in sports analytics is the work of Reep and Benjamin (1968), who studied the number of passes that lead to a goal in a football match, and analysed the positions in the field, from which they come from. According to Sarmento et al. (2014), this study led to the "Long Ball", which was the English trademark playstyle for decades, by analysing how the ball possession would influence the final score, comparing the probability of goal-scoring and goal-conceding.

Another example of the use of these techniques in sports is the real-life events that lead to the book “Moneyball – The Art of Winning an Unfair Game”, by Michael Lewis. According to Morgulev et al. (2018) the newly appointed general manager of an American Baseball team used, in 1997, an approach similar to Business Intelligence to build his new roster with underperformers and underachievers. Through the use of techniques that allowed him to gather knowledge from statistical data, he created a very efficient team by buying players that would complement each other.

Today, the use of Business Intelligence methods to enhance the performance of Sports Organizations is extremely widespread. Professional Clubs understand the contribution and importance of data to succeed on the field. Recently, America has seen a growth in success of teams who follow a more analytical path - like the Boston Red Sox, 49ers and Dallas Mavericks -, and an unexpected level of competitiveness on smaller teams like Oakland’s A and Green Bay Packers. Despite this, most Clubs don’t employ large analytical staff. The analytical team does not perform the first step of the Sports Analytical Framework (Fig. 1) in-house but is provided by a third party. Some examples of third-party data providers and the data it provides are Bloomberg Sports and Player performance data, ShamSports and NBA salaries, Sports Reference and data of most professional sports, Opta and English Premier League status (Davenport and Harris, 2014)

Morgulev et al. (2018) mentions another recent and successful use of Business Intelligence techniques in NBA (National Basketball Association). The Supersonics were searching for a player that would pass the ball rather than shoot (point guard). When Russel Westbrook’s name was brought-up it was dismissed because his position was of one who would shoot rather than pass. Despite this, and because of the use of BI techniques and measures the team signed him for the position of point guard The measures brought up to the staff’s attention that despite his original position, there was a direct effect on Westbrook’s passes and his teammate’s ability to shoot. This means that his teammates had more success in scoring when the pass came from Westbrook. His numbers, regarding this measure, were on the same level of the best point guards in NBA.

In 2020, Patel et al. commented on the other aspects of sports affected by Business Intelligence:

1. Avoiding over-training and its consequences.
2. Maximizing the potential of changing the players’ positions.
3. Creating new and more efficient training exercises/activities.
4. Detection and control of doping.

They also mention the possibility of replacing personal trainers by drawing conclusions from the previous training data and the athlete’s current biological factors and indexes.

2.2.3. Business Intelligence in Athlete Training

One final opportunity in which Patel et al. (2020) address the application of business intelligence in Sports, is the usage of biological factors to provide information on the athlete's training and a way to classify them. In order to reach these goals, it is a very effective practice to use effort indexes as a way to measure the effort incurred and gain insights into the optimal training (David Rodríguez-Rosell et al., 2018). The authors defined Effort Index (EI) as a variable to "monitor the level of effort" during exercises. These are mostly biological indicators whose levels change depending on the effort put in an exercise and, consequently, whose variance reveals the effort that an athlete dedicated to their exercise.

According to this definition of EI, we can consider the following physiological parameters as Effort Indexes: Heart Rate and Blood Lactate Concentration (Rodríguez-Rosell et al., 2018), Oxygen Consumption (Osgnach et al. 2016), Heart Rate Recovery (Michael et al., 2017), Creatine Kinase (Magal et al., 2010), Tidal Volume (Nicolò and Sacchetti, 2023), Muscle Glycogen (Ivy, 2012), Muscle Temperature (Cochrane et al., 2008) and Blood Glucose (Adams, 2013). Through the use of BI, these indexes can be better evaluated and analyzed, delivering a lot of value to the sports industry.

3. METHODOLOGY

In this chapter, the methodology followed throughout the project will be explained. First a description of the tools used and how they were applied to each part of the project, then on how the Data was retrieved and why, its transformation and challenges, and its Modeling. Lastly, we will explain our approach to Analysis Techniques through the creation of DAX Measures.

In this project, we used two tools to generate and store data, one tool to create the report and transform the data, and a tool to create the backgrounds for the report:

- Microsoft Excel, to generate the data concerning Athletes, Game Events and to store the Dimensions that would feed the Facts table.
- Jupyter Notebook, to generate data referring to Trainings.
- Microsoft Power BI to load the data from an Excel Workbook and a folder, transform and model the data, and to create the necessary measures and the final Report.
- Figma, to design the backgrounds we'd use in the Report.

Due to the unavailability of public data to feed our dataset and the unwillingness of companies to share their data, we had the need to generate our data, using Excel and Python. For this we had to carefully think of what data would be necessary to achieve our goals. Once the data was generated, we needed to explore, process and model it. For the first two steps, we used Power BI's Power Query. For the third, we used the functionalities of Power BI's Model View.

Besides the data that would fit our model and that would create a report to be analyzed, we gathered pictures of athletes in a folder. The pictures were retrieved from unsplash.com, a free stock photo website. The website adds a disclaimer that the photos are free and can be downloaded for commercial and non-commercial purposes, without need for permission.

3.1. DATA GENERATION

Regarding the generation of our data, its premises are that we have 25 athletes, each athlete performs at a Game Event each week and during the week that precedes that event, the athlete practices 5 types of exercise. Right after each exercise, the Effort Index are measured. For the data generation we used a period of 15 weeks, each referred to by the first day (Sunday), e.g. "Week of 01/01/2023".

As Effort Indexes, we used Blood Glucose, Blood Lactate, Creatine Kinase, Heart Rate, Heart Rate Recovery, Muscle Glycogen, Muscle Temperature, Oxygen Consumption, Tidal Volume.

These are merely examples of physiological parameters that fit into the definition of Effort Indexes, and that can be used to achieve our goal.

After carefully thinking about the data and characteristics that we would need to achieve our goals, we came up with the following needs:

- Athletes' information and characteristics.
- Game Events information to qualify the athlete's performance.
- Description of the Effort Indexes used and their Units of Measure.

In excel we defined the "Athletes" data, by randomly generating the athlete's Age, Weight, and Height. These variables were created using the function *Randbetween*, which produces a number within the specified range. In addition to these variables, we randomly assigned each athlete a sport, and we also calculated each athlete's Body Mass Index (BMI), using its formula (Equation 1).

$$\frac{\boldsymbol{Weight}\ (\boldsymbol{kilograms})}{\boldsymbol{Height}\ (\boldsymbol{meters})^{2}}$$

Equation 1- Body Mass Index

The "Game Events" table was also generated on excel. We created this table of events and overall rating with two starting columns: the athlete, and the weeks in which he performed. Based on these two columns, we randomly generated for each athlete a performance rating, by using the function Randbetween.

Besides generating this data, excel was also used to store information that would be loaded as Dimensions in our Dataset: "Indexes" and their Unit of Measure, "Sports" and "Exercise Type".

Following, we had the need to use Python to create the data for the "Training" table for this table is used to store the most complex information of our dataset. It provides details of each training: Who was the athlete who performed it, in what week did this happen, the exercise practiced, and the recorded values of each EI.

The complexity of this information is that we needed to replicate the distribution of real people. Considering this, we had to take into account two main factors:

- Each index has a unique minimum and maximum value, between which it needed to be randomized.
- In a real scenario, almost every athlete would put more effort into a certain type of exercise, as opposed to others. As a consequence, the randomization of effort couldn't be completely left to chance.

Figure 3 is an illustration of the processes taken to generate the training data.

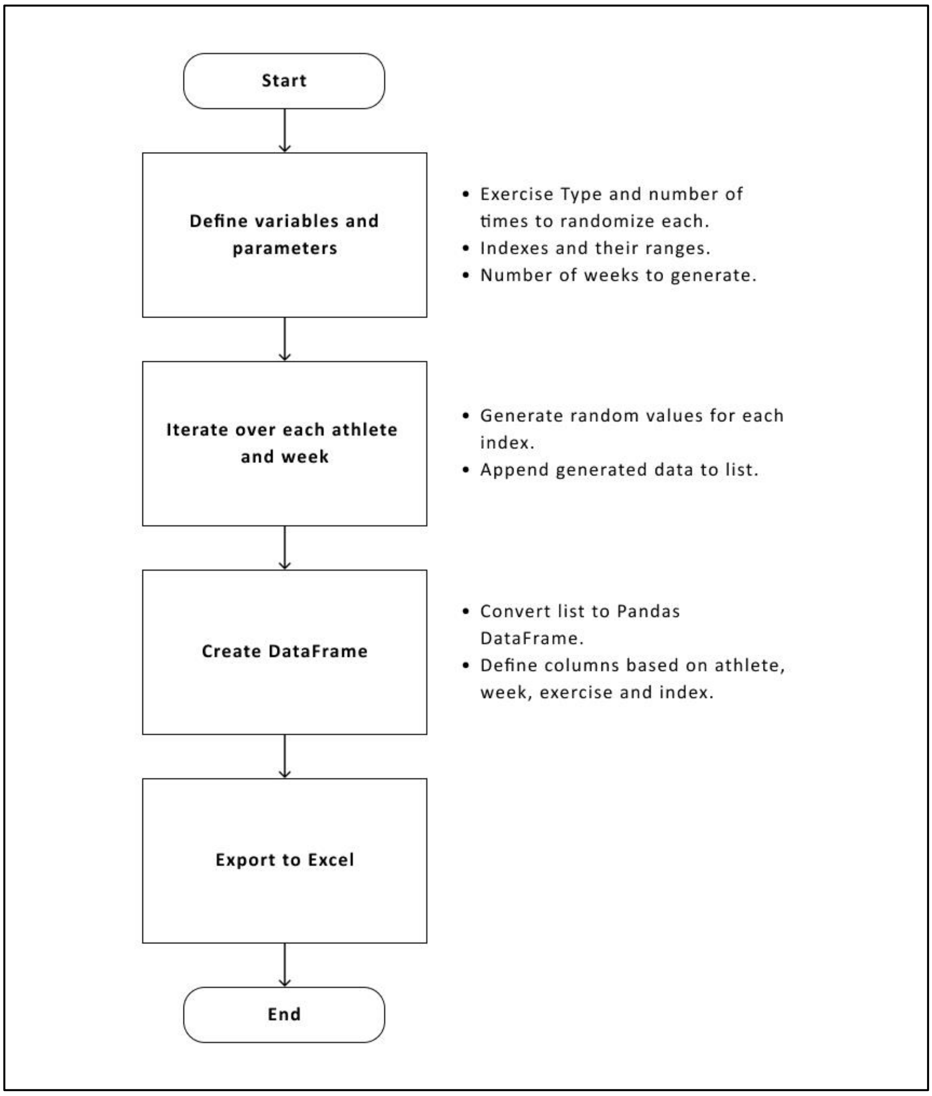

Figure 3 – Data Generation Process in Python

The code generates a table with 12 columns: "Athlete", "Week of", "Exercise Type" and 9 columns, named after each Effort Index.

The first step was to define what exercise types each athlete would have generate for them and the number of times to randomize each exercise's indexes, the indexes that would be generated for each exercise type, the ranges between each varies, and the number of weeks to generate data for.

Having all these defined, we had the algorithm iterate over each athlete and week, generating random values for each index, corresponding to all the exercises defined earlier, and to append those values to a list.

Lastly, that list was converted into a Pandas DataFrame, under the defined columns and exported to an excel spreadsheet, named after the date and time it was created on.

According to the second main complexity factor mentioned above, we had to make sure that the data was not completely and had some variations. To address this issue and create data as close to reality as possible, we generated data for 11 weeks without interfering with the parameters and the data of 4 weeks while forcing the algorithm to get slightly better results at certain exercises, depending on the athlete.

In order to do so, we added the variable that indicates how many times each exercise's indexes would be generated for the defined athlete. When this variable's value is, for example 6, the algorithm will randomize the related values 6 times and store the maximum value. For example, if we define that, when generating Andy's data, that the indexes for "Strength" will have be randomized 1 time, those values will be as random as possible. On the other hand, if we determine that it should be randomized 6 times and the value to be stored should be the maximum, we are increasing the chances of Andy dedicating more effort into the Strength exercises.

The output of these steps of Data Generation and Storage was compiled in a single excel workbook. That workbook was provided as source to Power BI, and that Dataset's structure is exposed on Table 1.

Table 1 - Dataset Structure

Table	Description	Columns	Rows
Athlete	Athletes' physical attributes and the Sports they practice	Name, Age, Weight, Height, BMI, Gender, Sports.	25
Game Event	Athlete's ratings on events occurred in specified week	Week of, Athlete, Performance Rating.	375
Index	Effort Indexes used for analysis and their Units of Measure	Index, Unit.	9
Sports	List of sports practiced by the athletes	ID, Sports.	4
Exercise Type	List of exercises practiced by the athletes	ID, Exercise Type.	5
Training Data	Information regarding each athlete's training as the week the training happened, the exercises practiced, and the values of effort indexes registered for each occurrence	Training, Athlete, Blood Glucose, Blood Lactate, Creatine Kinase, Heart Rate, Heart Rate Recovery, Muscle Glycogen, Muscle Temperature, Oxygen Consumption, Tidal Volume, Exercise type and Week of.	16.875

3.2. Data Preparation

Having the whole data in an excel workbook, divided into sheets, we loaded it on Power BI and proceeded with the Data Preparation using Power BI's Power Query. To store the location of the workbook, we created a parameter called "Path", and the same technique was used regarding the location of the Folder containing the athlete pictures, under "Path Images". This way, if the workbook or folder's location is changed, we don't need to change the source of every table manually, just the value of the parameter.

In each table sourced from Path, we used the parameter as Source and then added a Navigation step to identify the sheet. After this, we followed the basic steps of Promoting Headers and Change Type for all the tables.

As it was referred to before, we stored pictures of our athletes in a folder, and named each of them after the athlete it concerned. The parameter Path Images was used as the source to retrieve information from the Folder, and then we used Remove Other Columns to eliminate unnecessary columns and keep only Name and Content. From content, we created a custom column using M language code that converts the image data into a text-based format: "data:image;base64, " & Binary.ToText([Content], BinaryEncoding.Base64).

Encoding to Base64 divides the image into groups of 6 bits and represents each in a set of 64 characters. The output represents the whole image, in a text format, allowing us to use the images in the report without actually storing them. For the proper use of this field, we then had to categorize the column as "Image URL".

After this complex step, it was time to remove the Content column, extract the name of the picture except for the extension (".jpg"), and to rename the columns and change the type of the new column, Image, to text.

Given that this table is auxiliary to the table "Athlete", the column "Image" was merged into "Athlete" and the option Enable Load for this table was disabled.

For the tables "Exercise Type", "Index" and "Sports", there was no need for further Data Preparation, considering that we confirmed that there were no errors or missing values using the column quality function on Power BI and explored the data in each table. Additionally, it was confirmed that for these tables, there were no duplicates nor outliers, given that they are supporting tables.

For the "Athlete" table, there was a need to remove blank rows and columns. After doing so, we merged this table with "Athlete Images" and retrieved the "Image" column.

For the "Game Events" table, other than the referred initial steps, we removed blank columns and merged the "Athlete" and "Week of" columns to create a key to connect to the "Training Data".

The "Training Data" sheet was more complex. The table was structured with each EI as a column which was necessary for the data generation, but this sort of structure is inadequate to the analysis and creates an unnecessary level of complexity.

To fix this issue, the first steps we used were to Unpivot Other Columns, to create two new columns: Index, with the index name, and Value, for the registered value of that index for that occurrence. This way, instead of having a record for each combination of athlete, exercise type and week, we'll have 9, one for each Index recorded in that combination.

This approach doesn't have a big impact on performance, given the size of this dataset, and the expected size of real users' datasets. These steps were applied to the "Training Data" and were used for the creation of 2 additional tables: "Training_max" and "Training_min".

The values recorded for each index vary between very distinct orders of magnitude. For example, the values of Creatine Kinase can be between 2.000 and 7.000 U/L, while the Muscle Glycogen levels vary from 1.2 and 1.7 g/min. This makes it very difficult to compare and join information from both Indexes to create knowledge. To solve this problem, we used "Minmax" to normalize the data, meaning that we transformed the EI values to fit a range between 0 and 1. This process is done comparing the maximum and minimum values of each Index. The "Minmax" formula can be observed on Equation 2.

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Equation 2 – Minmax Normalization formula

To help with this normalization process, "Training_max" and "Training_min" were created. Each table featured the steps mentioned before, regarding "Training Data". After these, we merged the columns "Athlete" and "Index" and used the function Table.Group to group this table by "Merged" value, retrieving the minimum or maximum value of each occurrence. This step allowed us to get the minimum and maximum value of each Index for the corresponding athlete, making it possible to normalize the Index values in the "Training Data" table. After we gathered this information, the Enable Load option for "Training_max" and "Training_min" was disabled, for these tables are auxiliary and there is no need to load them into the report.

Returning to the "Training Data" table on Power BI, after the initial steps of data transformation, we merged the columns "Athlete" and "Index" to be able to connect this table with "Training_max" and "Training_min". This step enabled us to merged the tables and retrieve the maximum and minimum values regarding each "Athlete" and "Index", with which the "Normalised" column was created, using the "Minmax" formula. The last steps were to remove the unnecessary columns min and max, and to and merged the "Athlete" and "Week of" columns to create a key to connect to the "Game Events" table.

To be able to properly calculate the correlation between Effort applied in a training and the performance rating in events, we had the need to merge the tables "Training Data" and "Game Events". To do so, our source was "Training Data" after all the transformations applied, to which we merged the "Game Events" table.

We kept the columns that would be necessary to calculate the correlation coefficients: "Athlete", "Index", "Exercise type", "Week of", "Normalised Value" and "Overall Rating".

After these steps 7 tables were loaded on Power BI report view. The columns and data types can be observed on Table 2.

Table 2 – Tables Metadata

Table	Columns	Data Type
Exercise Type	ID	Whole Number
Exercise Type	Exercise Type	Text
Index	Index	Text
Index	Unit	Text
Sports	ID	Whole Number
Sports	Sports	Text
Athlete	Name	Text
Athlete	Age	Whole Number
Athlete	Weight	Whole Number
Athlete	Height	Decimal Number
Athlete	BMI	Decimal Number
Athlete	Gender	Text
Athlete	Sports	Whole Number
Athlete	Image	Text
Game Events	Week of	Date
Game Events	Athlete	Text
Game Events	Overall Rating	Decimal Number
Game Events	Sports	Text
Training	Training	Whole Number
Training	Athlete	Text
Training	Index	Text
Training	Exercise Type	Whole Number
Training	Week of	Date
Training	Original Value	Decimal Number
Training	Normalised (MinMax)	Decimal Number
Training	AthleteWeek id	Text
Merged Training_Game	Athlete	Text
Merged Training_Game	Index	Text
Merged Training_Game	Exercise Type	Whole Number
Merged Training_Game	Week of	Date
Merged Training_Game	Normalised MinMax	Decimal Number
Merged Training_Game	Overall Rating	Decimal Number

3.3. DATA MODELING

After the steps of Data Preparation we had 7 tables to model, and needed to make sure that all would be necessary and to consider if we needed any other.

After some deliberation regarding the tables “Training Data”, “Game Events”, and “Merged Training_Game”, we decided to keep all of them. The merged table was created with the specific objective of enabling the calculation of correlation coefficients between effort put in an exercise type and the performance at the event. For other analysis, specifically related to performances or training, the model will have a better performance using the original tables.

In order to use dates as a filter and feature in the model, there was a necessity to create a “Date” table. On Table View, we generated a calculated date table using the minimum and max values of the column “Week Of” from “Training Data”, for these are the period within which our dataset has data. In the same table, we included a column for “Year”, “Month” (number) and “Month Name”. We wanted to add two more columns, but couldn’t include those in the original table formula, because it took in consideration the column “Date”, which is created only after the original formula is applied. For this reason, we used DAX to create the calculated columns “Week of Month” and “WMY Conc”, which is the concatenation of the values Week, Month and Year, to be used as in a card. In order to enable all of Power BI’s functions regarding date, we later defined this as the Date Table of our model. The DAX code used for these calculations can be observed in Appendix Table 1.

To store the measures later created, we also added a “Measures” table. To do so, we selected Enter Data, and added an empty column. Once we created our first measure, that column was deleted, and Power BI automatically identified it as a Measures table.

After these steps we had 3 Facts tables – “Training Data”, “Game Events” and “Merged Training_Game”, 5 Dimension tables – “Athlete”, “Index”, “Exercise Type”, “Date” and “Sports” and an auxiliary table “Measures_”.

When it comes to relationships between them, we had to create the following, displayed on Table 3.

Table 3 - Relationship between tables

From Table	To Table(s)	Columns Used
Athlete	Training Data, Game Events, Merged Training_Game	Name - Athlete
Date	Training Data, Game Events, Merged Training_Game	Date – Week Of
Index	Training Data, Merged Training_Game	Index – Index
Exercise Type	Training Data, Merged Training_Game	ID – Exercise Type
Sports	Athlete	Id - Sports

All of these relationships have a "One to Many" cardinality, and a "Single" cross filter direction. The first five have a filter direction from the Dimension to the Facts table, and the last from "Sports" to "Athlete".

This transformations and modeling techniques led us to a Snowflake model, which means that not all the dimensions are connected to the Facts tables, but there's one ("Sports") that is connected to another Dimension table, as we can see on Figure 4.

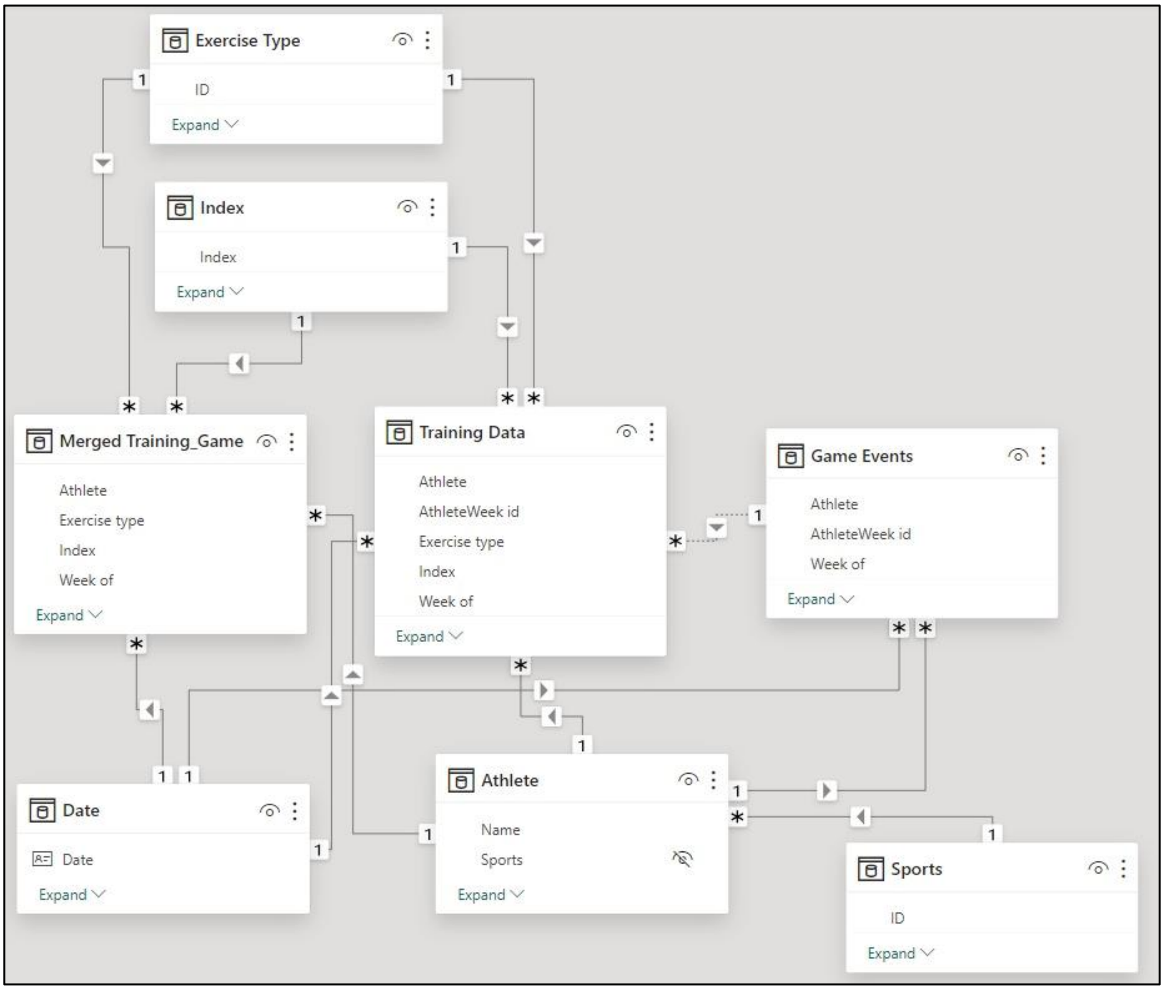

Figure 4 – Data Model

The last topic to mention, regarding Data Modeling, is the Hierarchies created. For this model we only had the need to create the Date Hierarchy: Year, Month and Week of Month, in order. This Hierarchy would allow us to better filter the periods we want to analyze in the report.

3.4. Analysis Techniques

With the Data Modeling process completed, it was time for us to address the report and its needs in terms of Measures, Visualizations and Dashboard Interactions. With that purpose we used Power BI's coding languages: DAX (Data Analysis Expressions), and M. "M" is used for Data Transformation purposes on the Power Query module, while DAX is used on the Data Modeling and Analysis, to create calculated columns, calculated tables, and measures to perform advanced data analysis. Besides using DAX to calculate the Date table, we also used it to create Measures that allowed us to perform dynamic calculations on the data model, and aggregate values based on the analysis' needs and Report's interactions.

Consequently, our next step was to create DAX measures in compliance with the calculation needs of our report. For organization purposes, we divided the created measures in 3 folders:

Athlete, Correlation and Distribution. On Appendix Table 2, we share the Athlete measures created and explain their purpose.

Measures regarding age, height and weight were created to facilitate the display of these characteristics with their unit of measure. The measure Sport was necessary to display the Sport that the selected athlete practices. This is necessary because the cross filter direction is from the type 2 dimension table Sport to the Dimension table Athlete, which makes it impossible to filter Sports by selecting an athlete.

Regarding the “Athlete count”, it was necessary because we would use it in a card and if, instead of this measure, we used, for example, an automatic count of athletes by their name and there was no athlete in the selection, the output would be “blank” instead of “0”.

On Appendix Table 3, we have the measures created that concern the correlation between the athlete’s effort on each specific sport and the performance they had.

“S. Correlation all” is the basis for all the analysis related to the correlation. It is a replica of the formula of Spearman’s Rank Correlation Coefficient (Spearman, 1904), which we perceive in Equation 3.

$$\boldsymbol{p = 1 - \frac{6 \, \Sigma \, d_i^2}{n \, (\, n^2 - 1 \,)}}$$

Equation 3 - Spearman’s Rank Correlation Coefficient

In this formula, *p* represents the correlation coefficient, *d* represents the difference between ranks of the two coordinates in each data pair, and n represents the number of data pairs in the dataset. The replication of this formula meant that we needed to get the two columns “Normalised MinMax” and “Overall Rating”, from table “Merged Training_Game”. We then had to calculate the rank of each value of “Normalised” and “Rating” between that same column, calculate the difference between those two ranks and square them.

To make it clearer, we display an example on Table 4.

Table 4 - Spearman’s Calculation Example

#	Normalised MinMax	Overall Rating	“MinMax” Rank	“Performance” Rank	Rank Difference (d)	Squared Difference (d^2)
1	0,5	2	3	4	-1	1
2	0,1	5	4	2	2	4
3	0,9	9	1	1	0	0
4	0.6	3	2	3	-1	1

With this example, the formula would then become what we can observe on Equation 4.

$$p = 1 - \frac{6 * (1 + 4 + 1)}{4\ (\ 4^2 - 1\ )}$$

Equation 4 - Spearman's Calculation Example

According to Spearman (1904), the coefficient varies between 1 and -1, and the closer the coefficient is to 0, the weaker is the correlation between the variables. This way, a coefficient of 1 expresses a perfect association of ranks, a coefficient of 0 expresses no association, and a coefficient of -1 expresses a perfect negative association.

The last measure of this group, S. Correlation abs, calculates the absolute value of Spearman's coefficient. It was created to be used in a visual from which the user can confirm the "strength" of association between the exercise types and the athlete's performance, regardless of if the association is positive or negative.

On Appendix Table 4 we have the measures created regarding the distribution of our data. These are simple measures that were calculated to display the average, maximum and minimum values of Overall Rating, taking into consideration context like slicers. This means that if we use a slicer to filter the Sport practiced by the athletes, the values calculated by these measures will change. Besides these measures we also created others to calculate the maximum, minimum and average of the selected dimensions on Field Parameters "SwitchDimensions1" and "SwitchDimensions2".

These field parameters allow the user to select which dimensions they want displayed in a visual. "SwitchDimensions2" as options Athlete, Age, Gender, Week of Year, Sport and Exercise Type, while "SwitchDimensions2" has Athlete Age, Height, Weight and BMI. The code to create the field parameters can be found on Appendix Table 5.

With all the necessary measures created, we were ready to start developing the report.

4. RESULTS AND DISCUSSION

4.1. RESULTS

In this section, we will analyze the report created (which can be accessed through this link), dashboards developed, explain the objective in each visual created, their interactions and how the user should use it in their analysis. In order to achieve our goals, the following dashboards were developed:

- Overview.
- High Level Summary.
- Correlation Details.

The landing page “Overview” has a button at the bottom that takes the user to the “High Level Summary” page, by selecting the athlete they want to analyse. The other dashboards have page navigation buttons on the left side. “High Level Summary” has buttons to either take the user back to the “Overview” or to the “Correlation Details”, while “Correlation Details” has page navigation buttons to “Overview” or “High Level Summary”.

The first dashboard, Overview allows us to have an overall understanding of the player’s characteristics and the relative best and worst performances. Figure 5 presents the layout of dashboard “Overview”.

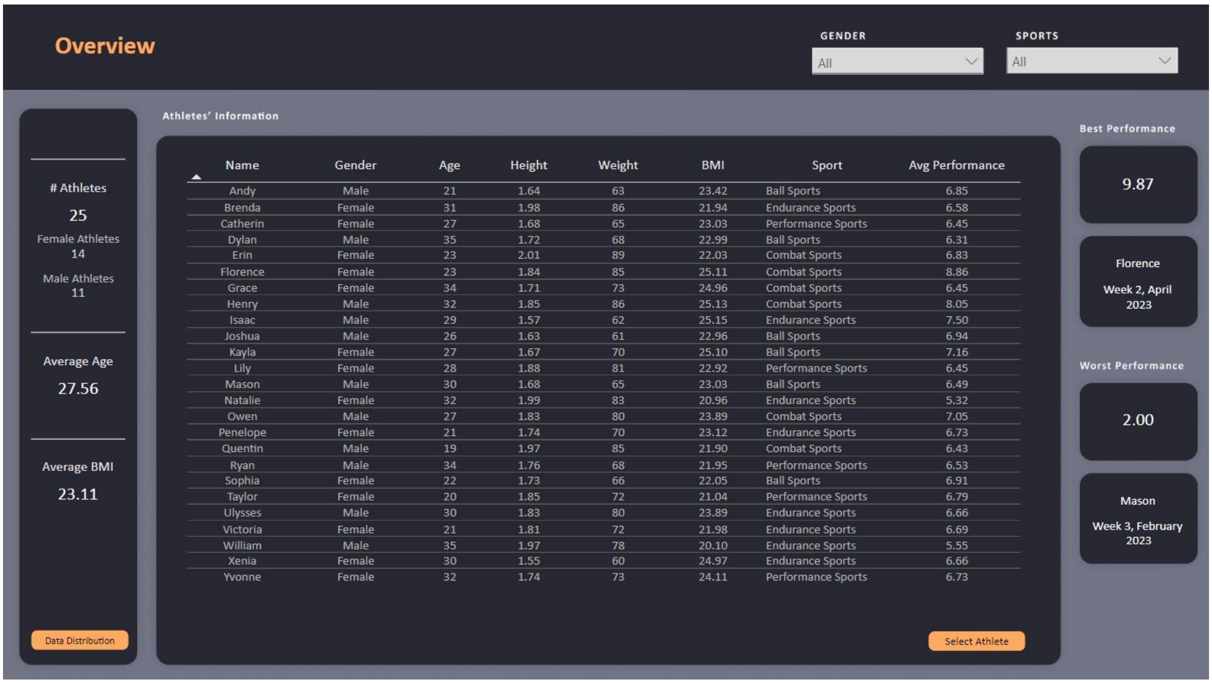

Figure 5 – Overview dashboard

At the top we have two slicers that enable the user to reduce the athletes in the table: Gender and Sports. On the left we have some information regarding the population: overall number of athletes and number of athletes by gender, average age, and BMI. On the right side of the dashboard, we have information regarding the best and worst performance of the selected group of athletes: Overall Rating, the performing athlete, and the period when it happened (Week of Month, Month and Year). At the centre of the dashboard, we have a table displayed with all the information related to the athletes: Name, Gender, Age, Height, Weight, BMI, Sport practiced, and Average Performance. Using the slicers at the top will affect the whole dashboard, recalculating both the cards on the left and on the right side of the page, and the table in the middle.

Pressing the button “Select Athlete” will trigger a pop-up to display. In this pop-up we can select the athlete of whose data we wish to see in more detail, by opening the slicer, choosing an athlete and pressing the button “More info”. To leave this pop-up, the user can select the “X” on the top right corner, or the greyed-out area of the table behind it. We can see Overview with the pop-up showing on Figure 6.

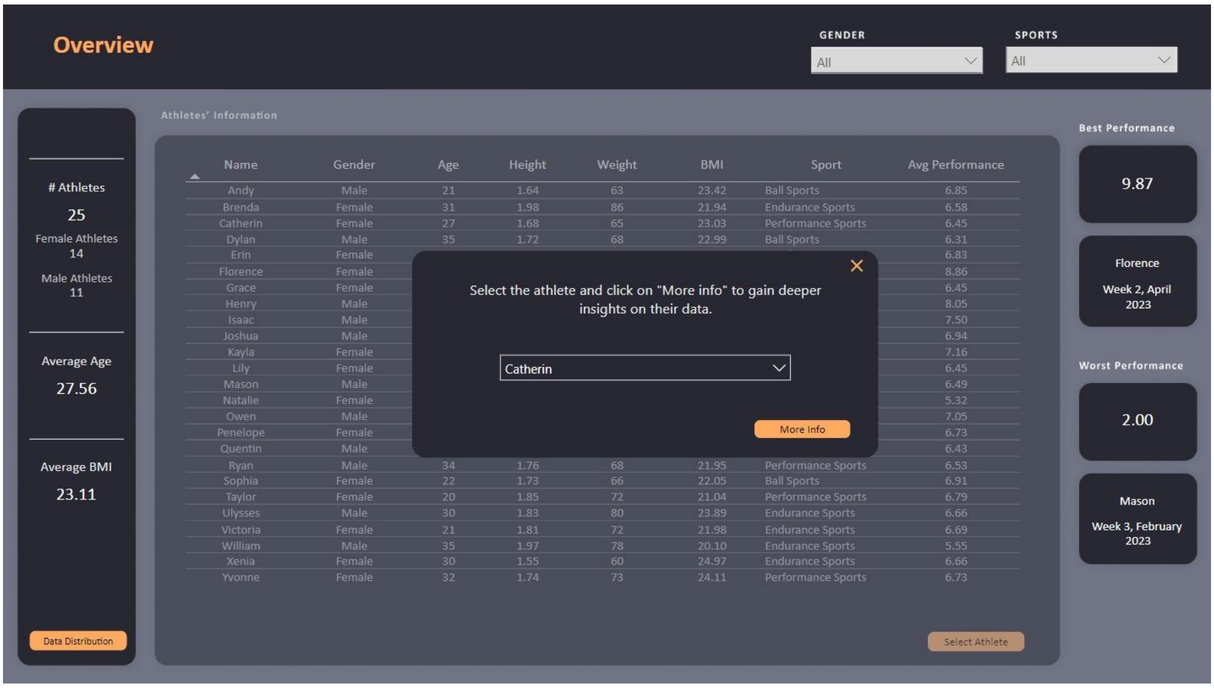

Figure 6 – Overview dashboard with pop-up

In the second dashboard, we have details regarding the selected athlete. On the top, there are 4 slicers, “Athlete”, “Period”, “Index” and “Exercise Type”. At the left we have information regarding the selected Athlete: Name, picture, age, weight, height, and Sport practiced. On the right side we have information regarding their performances: Average, Best and Worst. In the middle we have two line charts: Effort Evolution and Performance Evolution.

Both line charts display the evolution throughout the selected period of time. This layout and information can be perceived on Figure 7.

Figure 7 – High Level Summary dashboard

Effort Evolution exhibits the development of effort levels that the athlete applies in their trainings. The level of effort in each week and exercise type is the result of the sum of every index recorded. This sum is an example of why we found the need to previously normalise the Indexes data. Without it, we would be summing, for example, 5.200 U/L of Creatine Kinase and 1.4 g/min of Muscle Glycogen. It is important to note that this visual is where the Index slicer can make a difference on the analysis: we have summed the normalised values of every index for this "effort" value, but the user might consider that, for example, the Oxygen consumption and Heart rate recovery are not useful indexes in this analysis. The slicer enables them to select the indexes considered upon the calculation. The slicer "Exercise Type" regulates the Exercises displayed on the visual, by eliminating the unselected exercise's line from it. Regarding the chart "Performance Evolution", it displays the performance of the selected athlete throughout time.

By selecting the top right corner button "Combine graphs", we get a pop-up with information regarding both visuals, in a line-chart: the evolution of performances and the evolution of effort. A big difference between this aggregated visual and the two displayed before is that the effort is no longer separated in different lines, according to the exercise type. The user must select the exercise type(s) whose evolution they want to compare to the evolution to performance. This pop-up can be seen in Figure 8.

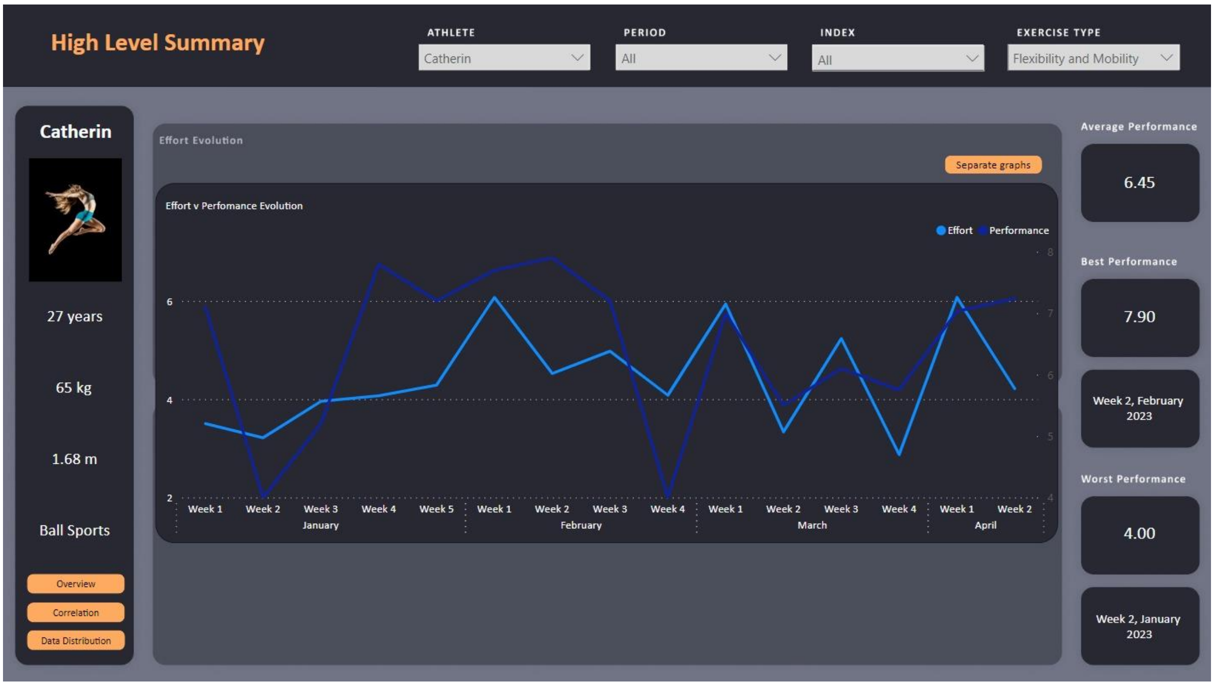

Figure 8 – High Level Summary dashboard with pop-up

This visualization makes it much easier for the user to compare the effort applied to the training, regarding one exercise type, and how successful the athlete was on the event, in terms of time period. To close the pop-up the user must press “Separate graphs” on the top right corner, or the greyed-out area of the visuals behind it.

The last dashboard, “Correlation Details” can be seen on Figure 9, and it displays for the user the calculated Spearman’s Correlation coefficient.

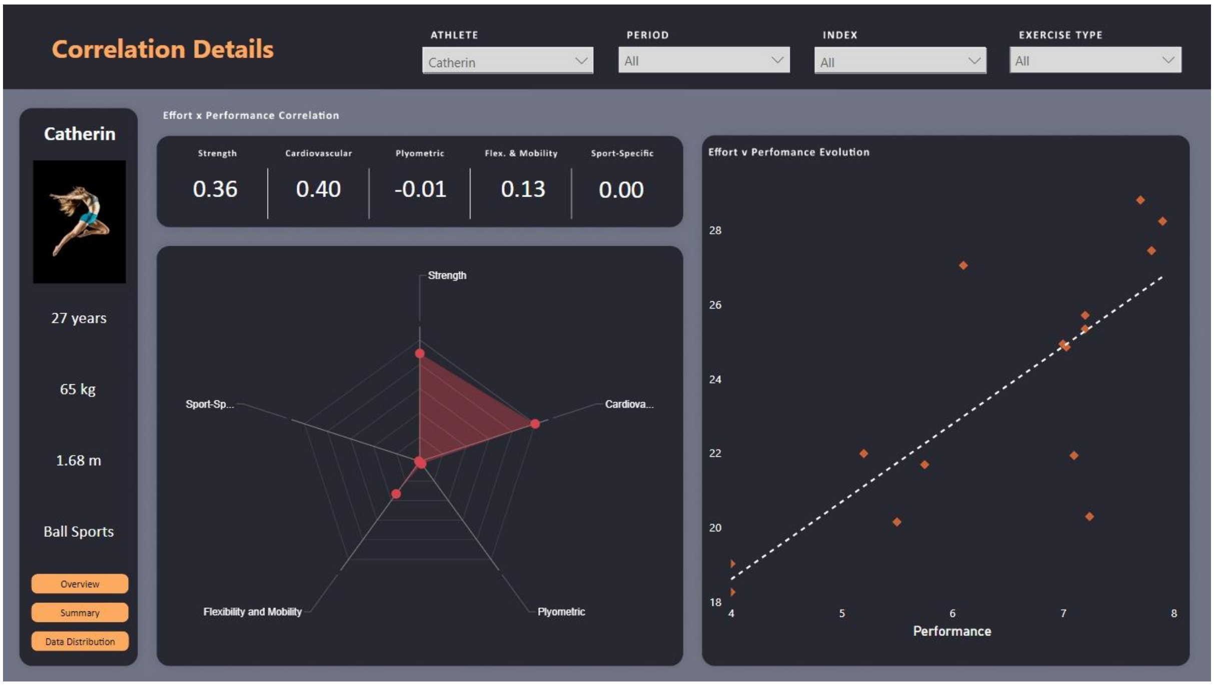

Figure 9 – Correlation Details dashboard

On the left side of the dashboard, we kept the athlete's information, and on the top the same four slicers the user can use on "High Level Summary". All of these slicers are synced, which means that if we filter a specific index on "High Level Summary", it will be applied to the "Correlation Details" dashboard when the user navigates from it, and vice-versa. In the middle of the page, we have five cards exhibiting the correlation coefficient concerning each exercise type. The values presented in these cards represent the influence that the effort put in this exercise has on the performance of the athlete. These values will change upon different selections of the slicers "Period", "Index", and "Athlete", allowing more control and choices for each user. Underneath these cards we have the same information displayed in a radar chart, with the difference that these values are in an absolute form. This means that the chart visually represents the influence of effort in performance levels, regardless of if that influence is negative or positive. On the right side of the dashboard, we have a Scatter chart displaying the relationship between effort and performance, considering every record of the selected athlete. This is the only visual in this dashboard that is affected by the slicer "Exercise Type".

Every page has, at the bottom left, a button to "Data Distribution". This button takes the user to the dashboard presented in Figure 10.

Figure 10 – Data Distribution dashboard

On this dashboard, the user can get some insights regarding the data displayed through the whole report. On the left, the number of Athletes, Weeks, Exercise Types and Effort Indexes that are present in the database. In the middle, there are two bar and line charts, displaying the minimum, maximum and average values of Effort and Performance, regarding the dimension on the x-axis. This dimension can be changed using the slicer "X Axis Dimension", which includes Athlete Name, Age, Gender, Week of Year, Sport and Exercise Type.

The values of the visual Effort Distribution consider the selection on the 'Index' slicer and include a card that displays the corresponding Unit of Measure for the selected Index. On the right side, there are 3 cards displaying the average, maximum and minimum values of the selected Athlete Dimension. This dimension can be changed using the “Athletes’ Dimension” slicer at the top, which contains the options Age, Height, Weight, and BMI.

4.2. DISCUSSION

For the purpose of better explaining what type of analysis can be performed by the user, we will illustrate an example, using the dashboards High-Level and Correlation Details dashboards previously displayed.

Considering Figure 7, the user is analyzing Catherin’s evolution of effort and performance. Using the cards on the right, it is possible to observe the weeks with better and worse performance in the visuals at the center. Catherin’s best performance was on the second week of February, when she got a 7.9 performance. In that week, the Exercises she put more effort into were Strength and Cardiovascular. Observing the week in which she had her worse performance, on January’s 2^{nd} week, the effort put into those exercises were way lower. This might indicate that the exercises Strength and Cardiovascular have some relation with the performance level achieved. On the same page, using the “Combine graph” option, we can confirm the relationship between exercises and effort, throughout the selected period, in a single visual and disregarding the other exercise types.

To have a more clear and scientific confirmation of these relationships, the user should access the Correlation Details dashboard. Considering Figure 9, at the top the user can confirm that the exercises with a stronger correlation to the performance are Strength and Cardiovascular. As mentioned in section 3.4 – Analysis Techniques, Spearman’s coefficient varies between 1 and -1, and the closer the coefficient is to 0, the weaker is the correlation between the variables. In this example, the user can conclude that, for Catherin, Cardiovascular is the Exercise type with the most influence in a good performance, followed closely by Strength exercises.

These two exercises and Flexibility and Mobility have a positive influence on the athlete’ performance, while Sports-specific exercises have no impact, and the Plyometric exercise has an insignificant negative impact.

With these insights, the user can make the decision of reducing or canceling Plyometric, and Sports-specific exercises for the Catherin, while increasing number of Strength, Cardiovascular and Flexibility and Mobility exercises. The level of increase of effort put into these exercises can also be studied in this dashboard, given that with (new) data that reflects these changes, the user can examine if the level of increase was too much or could be raised even more. The more data provided by the user to the report, the more accurate the insights will be.

We consider that this report is very intuitive to use, well-structured, and very easy to draw conclusions from. Through its use, the user should achieve a level of knowledge at which they can comfortably ground their decision-making and increase the efficiency of their training leading to more successful performances.

5. CONCLUSIONS

This paper explains a Business Intelligence approach to a project in which the goal was to create a solution to leverage data into better results for small sports organizations. This project was developed using tools such as Excel, Jupyter Notebook and Power BI, and programming languages including Python, M and DAX.

The process was complex and difficult from the beginning as the data we were looking for was not publicly available and companies who had it couldn't share it for privacy matters or had reservations regarding the value of sharing their data. Despite this obstacle in our path, we managed to generate the needed data and dedicated some time to assuring it was as similar to real data as possible. After the data generation, we focused on Data Preparation, Modeling, creation of DAX Measures and creating meaningful visualizations to allow the user to analyze the available information and retrieve knowledge.

The solution created provides details on what training exercises each player is getting better results from. In other words, the user can conclude what is the impact of a player applying more effort into a particular training exercise, on their performance in the following competitions. This tool aims to be a contributing factor in the process of decision-making regarding athletes' training strategies, helping organizations motivate their athletes through better results and by demonstrating that decisions are based on knowledge, not just intuition.

6. LIMITATIONS AND FUTURE WORK

After the description of the project, its results, and conclusions, it is important to reflect on its limitations and possible future work.

6.1. LIMITATIONS

Throughout this project, our main difficulties were regarding data collection, not so much with the creation of the dashboard and general use of Power BI as a tool.

There was a lot of time invested in the data generation and attempts to ensure it is as similar as possible to real data collected from real athletes. The creation of the algorithm was a big challenge, not only when it comes to coding, but also determining the variables that should be taken into account and how they would interact with one another, and then merging the output of this code with the rest of the data. It was an interesting challenge, but it consumed time that could have been better allocated to enhancing our report or to further develop it.

We also consider that the use of real data on our project might offer us some insights that would take us on a slightly different path in terms of data preparation and perhaps even dashboard design.

6.2. FUTURE WORK

Although we are pleased with the results and believe that our project reached its goals and has the capability of leveraging BI in sports' decisions, the project has a lot of potential for further development. The main points for future work are its basis - generated data, feedback from users, and its deployment on a website or any other platform.

Regarding the use of generated data to base this project on, as mentioned above, despite our efforts, the data might not be reflective of reality and through the use of real data and user feedback, there certainly are possibilities we didn't consider in this project. Additionally, any organization that considers the embedding of our project in a website or an organizational platform opens the door for future developments, being in the layout or landing page of our report, or other features they want to include. The final aspect to consider for future work is automatic data refresh. These would take into consideration how often new data is loaded in the database, and guarantee that the report would detail information based on the most updated data available.

BIBLIOGRAPHICAL REFERENCES

Adams, O. P. (2013). The impact of brief high-intensity exercise on blood glucose levels. Diabetes, Metabolic Syndrome and Obesity, 6, 113–122. https://doi.org/10.2147/dmso.s29222

Alamar, B. C. (2013). Sports Analytics: A Guide for Coaches, Managers, and Other Decision Makers. Columbia University Press.

Arai, Y., Saul, J. P., Albrecht, P., Hartley, L. H., Lilly, L. S., Cohen, R. J., & Colucci, W. S. (1989). Modulation of cardiac autonomic activity during and immediately after exercise. American Journal of Physiology-Heart and Circulatory Physiology, 256(1), H132–H141. https://doi.org/10.1152/ajpheart.1989.256.1.H132

Caya, O., & Bourdon, A. (2016). A Framework of Value Creation from Business Intelligence and Analytics in Competitive Sports. 2016 49th Hawaii International Conference on System Sciences (HICSS), 1061–1071. https://doi.org/10.1109/HICSS.2016.136

Choi, L. K., Panjaitan, A. S., & Apriliasari, D. (2022). The Effectiveness of Business Intelligence Management Implementation in Industry 4.0. Startupreneur Business Digital (SABDA Journal), 1(2). https://doi.org/10.33050/sabda.v1i2.106

Cload, P., Mallow, P., Mako, J., & Gunnarsson, C. (2011). PMD57 Data Visualization for Business Intelligence: Assessing an Online Tool Used for Benchmarking Hospital Procedure Costs to Reimbursment in Cardiac Catheterization and Electophysiology Procedures. Value in Health, 14(7), A254. https://doi.org/10.1016/j.jval.2011.08.128

Cochrane, D. J., Stannard, S. R., Sargeant, A. J., & Rittweger, J. (2008). The rate of muscle temperature increase during acute whole-body vibration exercise. European Journal of Applied Physiology, 103(4), 441–448. https://doi.org/10.1007/s00421-008-0736-4

Coco, M., Di Corrado, D., Calogero, R. A., Perciavalle, V., Maci, T., & Perciavalle, V. (2009). Attentional processes and blood lactate levels. Brain Research, 1302, 205–211. https://doi.org/10.1016/j.brainres.2009.09.032

Davenport, T., & Harris, J. (2014). Competing on Analytics: The New Science of Winning.

Foote, K. D. (2023, April 6). A Brief History of Business Intelligence. DATAVERSITY. https://www.dataversity.net/brief-history-business-intelligence/

Gilad, T., & Gilad, B. (1986). SMR Forum: Business Intelligence - The Quiet Revolution - ProQuest. 27 (4), 53–61. https://www.proquest.com/docview/211487983?sourcetype=Scholarly%20Journals&parentSessionId=t%2B8aAao6EVHYZuNQw2IWZNyiW34bUS8EAfI1giQ3r2A%3D

Ivy, J. L. (2012). Muscle Glycogen Synthesis Before and After Exercise. Sports Medicine, 11(1), 6–19. https://doi.org/10.2165/00007256-199111010-00002

Lea, B.-R., Yu, W.-B., & Min, H. (2018). Data visualization for assessing the biofuel commercialization potential within the business intelligence framework. Journal of Cleaner Production, 188, 921–941. https://doi.org/10.1016/j.jclepro.2018.02.288

Lim, Y. Y., & Teoh, A. P. (2020). Realizing the strategic impact of business intelligence utilization. Strategic Direction, 36(4), 7–9. https://doi.org/10.1108/SD-09-2019-0184

Magal, M., Dumke, C. L., Urbiztondo, Z. G., Cavill, M. J., Triplett, N. T., Quindry, J. C., McBride, J. M., & Epstein, Y. (2010). Relationship between serum creatine kinase activity following exercise-induced muscle damage and muscle fibre composition. Journal of Sports Sciences, 28(3), 257–266. https://doi.org/10.1080/02640410903440892

Marjamäki, P. (2017). Evolution and trends of business intelligence systems: A systematic mapping study. https://www.semanticscholar.org/paper/Evolution-and-trends-of-business-intelligence-%3A-a-Marjam%C3%A4ki/cbe86d72795cd546aadd319620419471b080c7f8

Michael, S., Graham, K. S., & Davis, G. M. (2017). Cardiac Autonomic Responses during Exercise and Post-exercise Recovery Using Heart Rate Variability and Systolic Time Intervals—A Review. Frontiers in Physiology, 8. https://www.frontiersin.org/articles/10.3389/fphys.2017.00301

Morgulev, E., Azar, O. H., & Lidor, R. (2018). Sports analytics and the big-data era. International Journal of Data Science and Analytics, 5(4), 213–222. https://doi.org/10.1007/s41060-017-0093-7

Nicolò, A., & Sacchetti, M. (2023). Differential control of respiratory frequency and tidal volume during exercise. European Journal of Applied Physiology, 123(2), 215–242. https://doi.org/10.1007/s00421-022-05077-0

Osgnach, C., Paolini, E., Roberti, V., Vettor, M., & Prampero, P. E. di. (2016). Metabolic Power and Oxygen Consumption in Team Sports: A Brief Response to Buchheit et al. International Journal of Sports Medicine, 37(1), 77–81. https://doi.org/10.1055/s-0035-1569321

Passlick, J., Guhr, N., Lebek, B., & Breitner, M. H. (2020). Encouraging the use of self-service business intelligence – an examination of employee-related influencing factors. Journal of Decision Systems, 29(1), 1–26. https://doi.org/10.1080/12460125.2020.1739884

Patel, D., Shah, D., & Shah, M. (2020). The Intertwine of Brain and Body: A Quantitative Analysis on How Big Data Influences the System of Sports. Annals of Data Science, 7(1), 1–16.

Prevedello, L. M., Andriole, K. P., & Khorasani, R. (2008). Business Intelligence Tools and Performance Improvement in Your Practice. Journal of the American College of Radiology, 5(12), 1210–1211. https://doi.org/10.1016/j.jacr.2008.08.018

Reep, C., & Benjamin, B. (1968). Skill and Chance in Association Football. Journal of the Royal Statistical Society. Series A (General), 131(4), 581–585. https://doi.org/10.2307/2343726

Riley, L. (2016). Indicator Metadata Registry Details. https://www.who.int/data/gho/indicator-metadata-registry/imr-details/2380

Rodríguez-Rosell, D., Yáñez-García, J. M., Torres-Torrelo, J., Mora-Custodio, R., Marques, M. C., & González-Badillo, J. J. (2018). Effort Index as a Novel Variable for Monitoring the Level of Effort During Resistance Exercises. The Journal of Strength & Conditioning Research, 32(8), 2139. https://doi.org/10.1519/JSC.0000000000002629

Sarmento, H., Marcelino, R., Anguera, M. T., CampaniÇo, J., Matos, N., & LeitÃo, J. C. (2014). Match analysis in football: A systematic review. Journal of Sports Sciences, 32(20), 1831–1843. https://doi.org/10.1080/02640414.2014.898852

Spearman, C. (1904). The Proof and Measurement of Association between Two Things. The American Journal of Psychology, 15(1), 72–101. https://doi.org/10.2307/1412159

Sorour, A., Atkins, A. S., Stanier, C. F., & Alharbi, F. D. (2020). The Role of Business Intelligence and Analytics in Higher Education Quality: A Proposed Architecture. 2019 International Conference on Advances in the Emerging Computing Technologies (AECT), 1–6. https://doi.org/10.1109/AECT47998.2020.9194157

Sun, Z., Sun, L., & Strang, K. (2018). Big Data Analytics Services for Enhancing Business Intelligence. Journal of Computer Information Systems, 58(2), 162–169. https://doi.org/10.1080/08874417.2016.1220239

Tavera Romero, C. A., Ortiz, J. H., Khalaf, O. I., & Ríos Prado, A. (2021). Business Intelligence: Business Evolution after Industry 4.0. Sustainability, 13(18), Article 18. https://doi.org/10.3390/su131810026

Trieu, V.-H. (2017). Getting value from Business Intelligence systems: A review and research agenda. Decision Support Systems, 93, 111–124. https://doi.org/10.1016/j.dss.2016.09.019

Wangoo, D. P. (2020). Intelligent Software Mining with Business Intelligence Tools for Automation of Micro services in SOA: A Use Case for Analytics. 2020 7th International Conference on Computing for Sustainable Global Development (INDIACom), 98–101. https://doi.org/10.23919/INDIACom49435.2020.9083682

Watini, S., Latifah, H., Rudianto, D., & Santoso, N. A. (2022). Adaptation of Digital Marketing of Coffee MSME Products to Digital Transformation in the Era of the Covid-19 Pandemic.

Wright, I. (2017). Analytics for Decision-Making in Amateur Sport Organizations—ProQuest

APPENDIX

Appendix Table 1– Date table and Columns

Variables	Calculus
Date (Table) Date, Year, Month, and Month Name (Columns)	Date = VAR date_from = DATE(YEAR(MIN('Training Data'[Week of])),MONTH(MIN('Training Data'[Week of])),01) VAR date_to = DATE(YEAR(MAX('Training Data'[Week of])),MONTH(max('Training Data'[Week of])),31) RETURN ADDCOLUMNS(CALENDAR(date_from, date_to), "Year", YEAR([Date]), "Month", MONTH([Date]), "Month Name", FORMAT([Date],"MMMM"))
Week of Month (Column)	Week of Month = VAR num_ = WEEKNUM('Date'[Date],2) - WEEKNUM(EOMONTH('Date'[Date], -1) +1,2) + 1 RETURN CONCATENATE("Week ", num_)
WMY Conc (Column)	WMY Conc = 'Date'[Week of Month] & ", " & 'Date'[Month Name] & " " & 'Date'[Year]

Appendix Table 2 - Athlete Measures

Measure	Code
Age years	Age years = CONCATENATE(MIN(Athlete[Age])," years")
Athlete count	Athlete count = VAR count_ = COUNTROWS(Athlete) Return IF(ISBLANK(count_),"0",count_)
Height m	Height m = CONCATENATE(MIN(Athlete[Height])," m")
Sport	Sport = MINX(Sports,Sports[Sport])
Weight Kg	Weight Kg = CONCATENATE(MIN(Athlete[Weight])," kg")

Appendix Table 3 - Correlation Measures

From Table	To Table(s)
S. Correlation all	S. Correlation all = VAR table1_ = ADDCOLUMNS('Merged Training_Game', "Rank Training",RANK.EQ('Merged Training_Game'[Normalised MinMax],'Merged Training_Game'[Normalised MinMax]), "Rank Event",RANK.EQ('Merged Training_Game'[Overall Rating],'Merged Training_Game'[Overall Rating])) VAR table2_ = ADDCOLUMNS(table1_, "Dif square", ABS([Rank Event]-[Rank Training])^2) VAR n_count_ = COUNTROWS(table2_) VAR d_square = SUMX(table2_,[Dif square]) RETURN 1-DIVIDE(6*d_square,n_count_*(n_count_^2-1))
S. Correlation abs	S. Correlation abs = ABS([S. Correlation all])

Appendix Table 4 - Distribution Measures

From Table	To Table(s)
Age performance	Age performance = AVERAGE('Game Events'[Overall Rating])
Avg Switch 2	Avg Switch 2 = SWITCH(SELECTEDVALUE(SwitchDimensions2[Parameter Fields]), "'Athlete'[Height]", AVERAGE(Athlete[Height]), "'Athlete'[Weight]", AVERAGE(Athlete[Weight]), "'Athlete'[BMI]", AVERAGE(Athlete[BMI]), "'Athlete'[Age]", AVERAGE(Athlete[Age]))
Best performance	Best performance = MAX('Game Events'[Overall Rating])
Max Effort	Max Effort = MAX('Training Data'[Original Value])
Max Switch 2	Max Switch 2 = SWITCH(SELECTEDVALUE(SwitchDimensions2[Parameter Fields]), "'Athlete'[Height]", max(Athlete[Height]), "'Athlete'[Weight]", max(Athlete[Weight]), "'Athlete'[BMI]", max(Athlete[BMI]), "'Athlete'[Age]", max(Athlete[Age]))

Min Switch 2	Min Switch 2 = SWITCH(SELECTEDVALUE(SwitchDimensions2[Parameter Fields]), "'Athlete'[Height]", Min(Athlete[Height]), "'Athlete'[Weight]", Min(Athlete[Weight]), "'Athlete'[BMI]", Min(Athlete[BMI]), "'Athlete'[Age]", min(Athlete[Age]))
Selected Dimension A	Selected Dimension A = VAR vSelected = SWITCH(SELECTEDVALUE(SwitchDimensions2[Parameter Fields]), "'Athlete'[Height]", "Height", "'Athlete'[Weight]", "Weight", "'Athlete'[BMI]", "BMI", "'Athlete'[Age]", "Age") RETURN CONCATENATE("Selected Dimension: ",vSelected)
Selected Index	Selected Index = VAR vIndex = SELECTEDVALUE('Index'[Index]) VAR vUnit = MIN('Index'[Unit]) RETURN CONCATENATE(CONCATENATE("Selected Index: ",vIndex), CONCATENATE(" Unit of Measure: ", vUnit))
Worst performance	Worst performance = MIN('Game Events'[Overall Rating])

Appendix Table 5 – Switch Dimensions

SwitchDimensions1	SwitchDimensions1 = {
	("Athlete", NAMEOF(Athlete[Name]),0),
	("Athlete Age", NAMEOF(Athlete[Age]), 1),
	("Athlete Gender", NAMEOF(Athlete[Gender]), 2),
	("Week of Year", NAMEOF('Date'[Week of Year]), 3),
	("Sport", NAMEOF('Sports'[Sport]), 4),
	("Exercise Type", NAMEOF('Exercise Type'[Exercise Type]), 5)
	}

SwitchDimensions2	SwitchDimensions2 = {
	("Athlete Age", NAMEOF(Athlete[Age]), 0),
	("Athlete Height", NAMEOF(Athlete[Height]), 1),
	("Athlete Weight", NAMEOF(Athlete[Weight]), 2),
	("Athlete BMI", NAMEOF(Athlete[BMI]), 3)
	}

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