



FACULTY OF SCIENCES OF THE  
UNIVERSITY OF PORTO

DEPARTMENT OF COMPUTER SCIENCE

Advanced Topics in Databases

Practical Assignment

Andreia Genro      Aníbal Silva      Rui Fernandes

June, 2022

## **Abstract**

This report describes the practical assignment of the Advanced Topics in Databases course.

This practical assignment consists in creating a data warehouse and conducting data analysis on it, as well as creating graphical reports using the Python library `matplotlib`.

In this report, we briefly describe our approach to the problem and discuss the decisions we made.

## Contents

## List of Figures

# 1 Introduction

The data warehouse contains data from national swimming competitions at the master level (*i.e.* class of competitive swimming for swimmers 25 years and older), namely Troféu Pescada 2021 and the Summer 2021 Championship.

## Structure of the Report

The remainder of the report is structured as follows:

- In ??, ??, we provide some insight into the data
- Finally, ??, ??, concludes the report and suggests remarks for future work.

## 2 Data Model

In this section, we describe the data contained in the data warehouse.

## 3 Data Analysis & Visualization

### 3.1 Number of Athletes by Age

To determine to determine the average age of the athletes, we can run the following SQL query:

From this, we can see that the average age of the athletes is 46 years, 6 months and 31 days.

We can also determine who's the youngest athlete by running the following SQL query:

- **Name:** Ana Mónica Eloi
- **Gender:** F
- **Birthdate:** 29/12/1996
- **Age:** 25 years

On the other hand, we can learn information about the oldest athlete by running the following SQL query:

- **Name:** Virgílio Zacarias Costa
- **Gender:** M
- **Birthdate:** 21/07/1931
- **Age:** 90 years

Finally, to determine the number of athletes by age, we can run the following SQL query using the PostgreSQL's built-in **age** function:

We can then plot the result, as illustrated in ??.

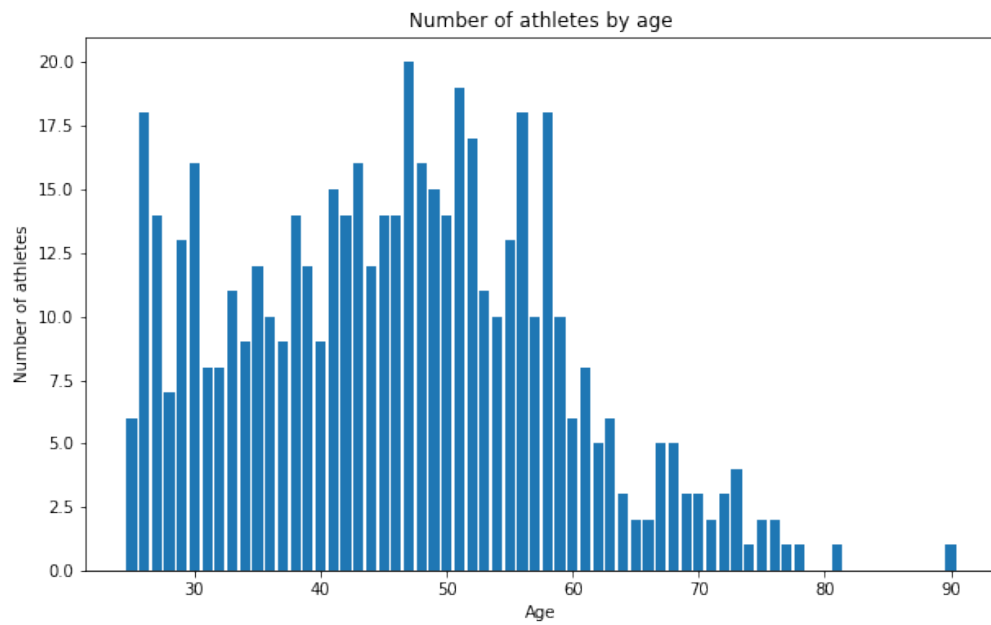


Figure 1: Number of athletes by age

### 3.2 Number of Athletes by Nation

To determine the number of athletes by nation, we can run the following SQL query:  
 We can then plot the result, as illustrated in ??.



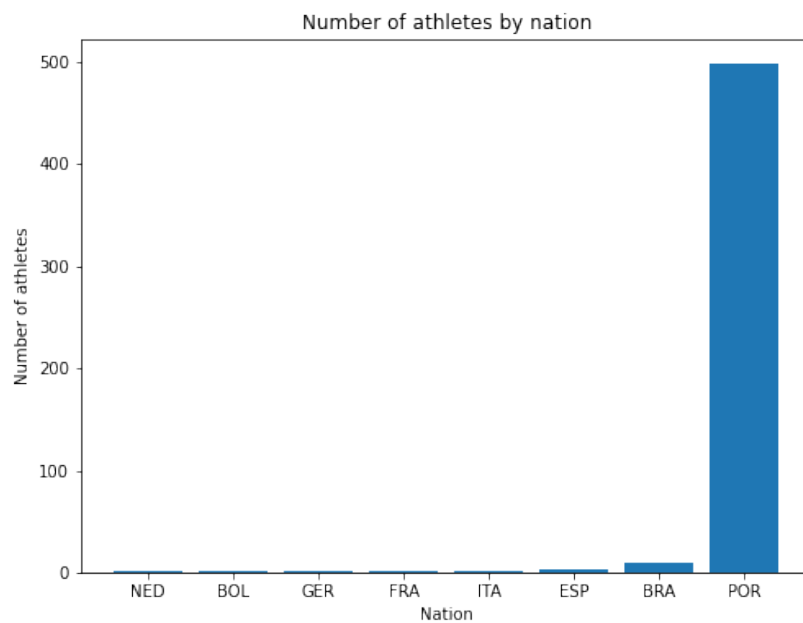


Figure 2: Number of athletes by nation

To have another perspective, we can also plot in a pie chart, as illustrated in ??.

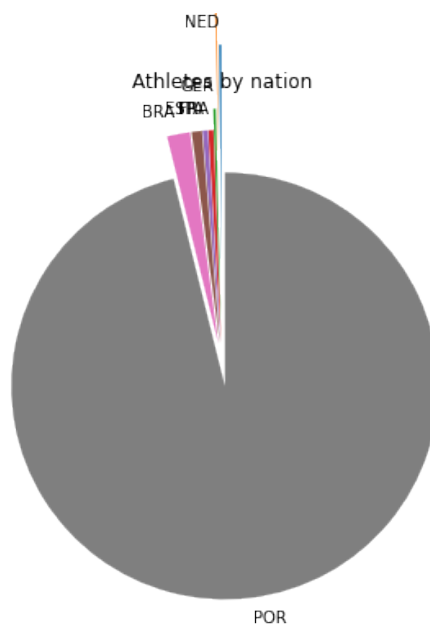


Figure 3: Number of athletes by nation

### 3.3 Number of Athletes by Gender

To determine the number of athletes by gender, we can run the following SQL query:  
We can then plot the result, as illustrated in ??.

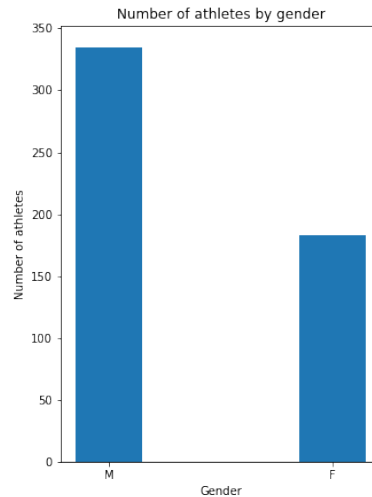


Figure 4: Number of athletes by gender

We can also plot this in a pie chart, as illustrated in ??.

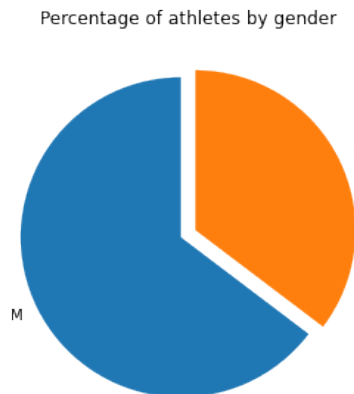


Figure 5: Percentage of athletes by gender

### 3.4 Number of Events by Gender

To determine the number of events by gender, we can run the following SQL query:

We can then plot the result, as illustrated in ??.

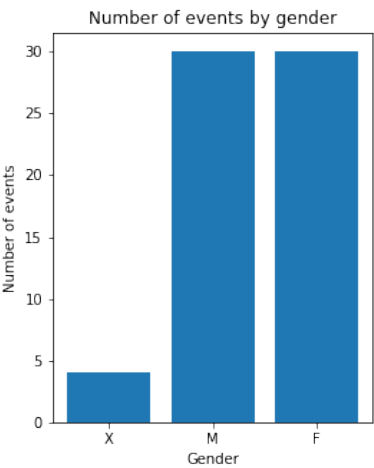


Figure 6: Number of events by gender

Here, the value **X** refers to events that allow athletes from both genders to participate. We can also plot this in a pie chart, as illustrated in ??.

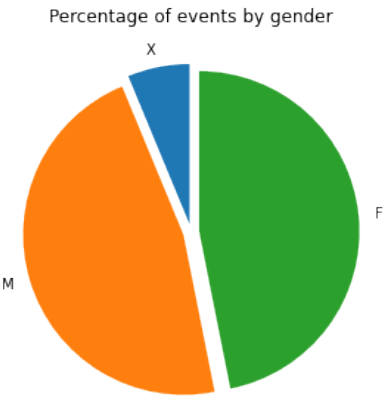


Figure 7: Percentage of events by gender

**3.5 Number of Clubs by Nation**

We can determine the number of clubs by each nation by running the following SQL query:  
**COMPLETAR COM TEXTO.**

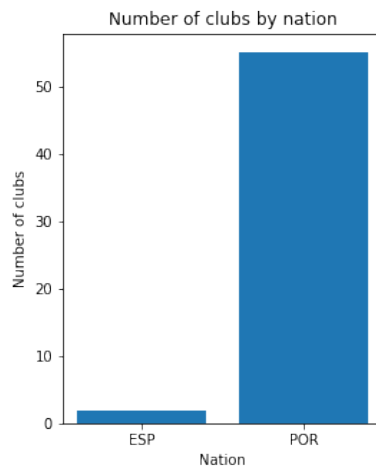


Figure 8: Number of clubs by nation

COMPLETAR COM TEXTO.



Figure 9: Percentage of clubs by nation

### 3.6 Number of Clubs by Region

COMPLETAR COM TEXTO.

COMPLETAR COM TEXTO.

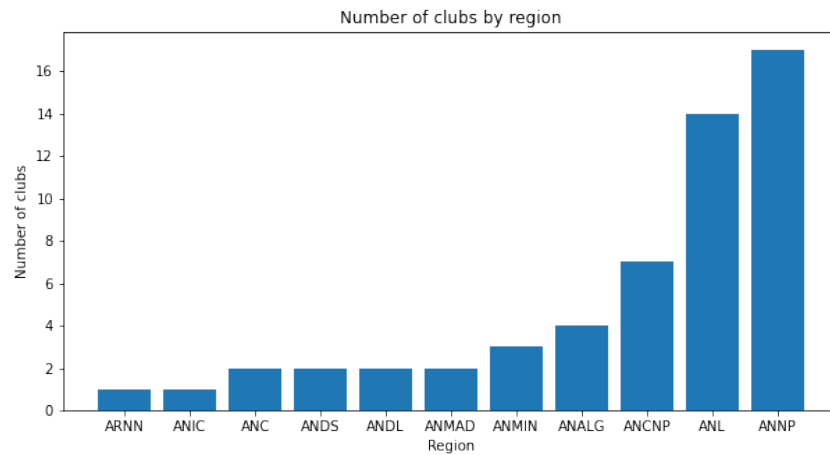


Figure 10: Number of clubs by region

### 3.7 Swim Styles

COMPLETAR COM TEXTO (style com mais distance) É FRESSTYLE.

COMPLETAR COM TEXTO (style com menos distance) FLY.

### 3.8 Results

#### 3.8.1 Average Swim Time

COMPLETAR COM TEXTO (resultado = 00:02:23.769068).

#### 3.8.2 Average Number of Points

COMPLETAR COM TEXTO (resultado = 340).

## 4 Conclusions & Future Work