# Overview

Hi, for this SQL project I am exploring a dataset regarding Cristiano Ronaldo’s soccer career club goals in Europe. I am focusing on Europe because it is widely accepted as the most competitive content for soccer.

The dataset was downloaded from Kaggle and has a usability score of 10.0. I chose this dataset because Cristiano is known worldwide and many people find anything about him interesting, especially his goals. Currently, he is the player with the most official goals in soccer history.

# Excel Data Cleaning and Importation:

The dataset is stored in an Excel worksheet, and I started this data exploration by cleaning and organizing the data. Firstly, moving the most interesting columns to the front of the worksheet. After, I deleted the at\_score column, I couldn’t conclude what it described, nor was it described on the Kaggle page.

## Filtering, Deletions, and Substitutions

**Club column**

Next, I filtered the ‘Club’ column for Al-Nassr and deleted all of the Al-Nassr records directly in the Excel spreadsheet. This is so the dataset only contains European teams as Al-Nassr is from Saudi Arabia.

**Matchday column**

I also the Matchday column and determined it contained data regarding the current matchday of the regular season as well as the current round in a knockout tournament. No cleaning was required here.

**Type column**

The “Type” column contained data regarding the type of goal scored, i.e. header, left-foot, right-foot. If the record had a blank for that field, it means it was scored with some other unorthodox part of the body or unusual fashion. I verified one record for the Manchester United vs Arsenal match on 11/03/2007, here Ronaldo scores with his knee. Here, I changed the value from blank to “other”.

**Playing\_Position column**

The Playing\_Position field also had blanks. I also changed these to “other”. The other values consisted of either LW, RW, or CF (Left Wing, Right Wing, Center Forward).

**Goal\_assist column**

The Goal\_assist field also contained blanks. I changed these values to “None” as that meant Ronaldo either took a penalty, hit a loose ball, or some similar situation.

**Season column**

The Season column was removed as it didn’t describe the season, it seemed to have random data and was not useful for analysis.

**Opponent score column**

This had scores with text in them. For example, 4 on pens. These text in these values was simply removed as none of the values in the team\_score field had text and the data still made sense by removing the text, i.e. if the number without text was larger than the team\_score value, it meant Ronaldo’s team lost.

## Transformations:

The “result” column was formatted as a time datatype, e.g. “03:04”, so the data here had to be converted into text. This was done using the TEXT(expression, format) formula and then using the fill handle to complete the formula for all rows. After, I copy and pasted the new columns VALUES into a new column. Now, the data was being interpreted as Text data and I could delete the Result column interpreted as Time and the calculated TEXT formula column.

NEXT, I used the “Text to Columns” tool from the Data Tools section to split the team’s (ronaldo’s team’s) score from the opponent’s score. I did this to allow me to more easily analyze patterns and trends regarding this data.

Finally, I imported the data to excel and began the SQL exploration. The data was dumped using the “SQL Server 2022 Import and Export Data” wizard and dumped into the “CR7PortfolioProject” database in a table called “CR7Goals”. “CR7” is a nickname for CristianoRonaldo.

# SQL Exploration

Troubleshooting:

After the data was imported, I noticed there were NULL values in the table. This was due to some values datatypes being invalid and being skipped.

For example, the matchday column being read as double data type, but there were text values in the spreadsheet such as “First Round” for knock-out matches.

These data types and other were changed to TEXT in the source excel file, and the importation was re-processed.

