# Overview

Hi, for this SQL project I am exploring a dataset regarding Lionel Messi’s soccer career club goals in Europe. I am focusing on Europe because I will use the KPI’s attained from this data to compare with Cristiano Ronaldo’s (CR7) goalscoring ability from a previous project. The dataset was downloaded from Kaggle and has a usability score of 10.0. It was upload by the same user that uploaded the Cristiano Ronaldo Goal’s data set, Azmine Toushik Wasi.

# Excel Data Cleaning and Importation:

## Organization

The dataset is stored in an Excel worksheet, and I started this data exploration by cleaning and organizing the data. Firstly, ordering columns in the same fashion as the CR7 dataset. 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

**at\_score**

I deleted this column, I couldn’t conclude what it described, nor was it explained on the Kaggle page.

**Club column**

This dataset only contained Messi’s European club goals, so no filtering was required.

**Matchday column**

I also the checked 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. There was only one record to verify and I confirmed the goal was scored with a hand.

**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.

