DataScience

# Football Data Analysis

DataScience

DataScience

# Overview

You have been recently hired as a data specialist at a football analyses company. You have been recently given access to their datasets on international team matches starting at 1902 until 2024. Your task will be to work on ingesting, cleaning, transforming and standardizing the dataset to be displayed in a dashboard for upper management, so they could get a full view of each team stats.

Your work will be split into two main parts:

## 1. Data Integration

After reviewing the dataset, you come to the realization that the data needs to be cleaned and standardized to be able to get the information you need from it. You set out to do the following steps:

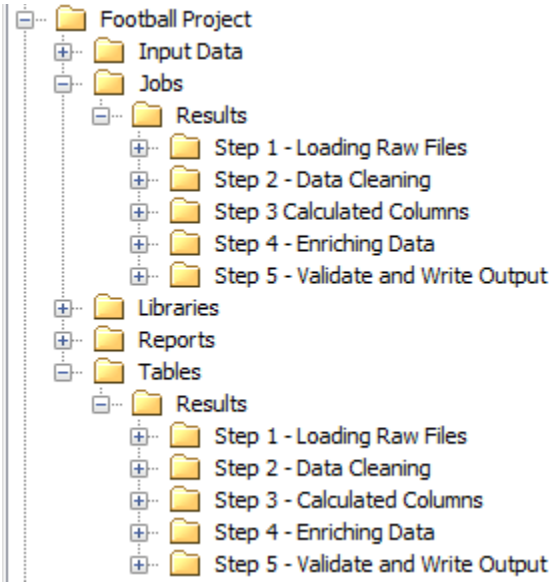
### Steps

1. Import the input files
    - Check the length of each column to make sure data is being read fully from the file into the table.
    - Check the column names to be compatible with the table format
    - Review file object metadata
2. Load raw files
    - Check the mappings of columns
    - Review output table metadata
3. Data Cleaning
    - Write an expression to convert the date column to the desired SAS format Date9.
    - Write an expression to convert all tournament names into English characters instead of Latin.
    - (Bonus) Write an expression to convert all city names into English characters only.
    - (Bonus) Write an expression to convert all country names into English characters only.
4. Adding extra columns with calculated information
    - Write an expression to create extra calculated column Winner. This column will calculate for each match which team won. If both teams scored the same number of goals, keep the value ‘Tie’

5. Enriching data
    - Using the latest results table, join with the given dimensions to replace lookup values with actual names
6. Validate and Write output
    - Using the data validation transformation, make sure that the latest dataset contains no missing values for all fields. If any row contains a missing value, then move it to an error table.
    - Create a new SAS BASE library to point to a data directory.
    - Create a job to write the output of the results to a SAS Table
    - Write the output of the results table to a CSV file

## Structure

Using a consistent structure makes it easier to identify what goes where. Especially when working with other people, where things easily get confused and lost in the noise. This is why we advise you to use the following folder structure where you have a folder for each object type, as well as a folder for each step of the above.



DataScience™

## 2. Visualization

Moving on to the next step. You need to design a dashboard to show different charts for the data extracted from the cleaned dataset. The charts will contain insights to assist in making informed data-driven decisions

### Data Source

First step into building a new report is to select the data source that you will use. You will need to import the final data set after all the ETL steps have been executed successfully. Once you have the data source imported, you will be ready to start building your reports.

### Dashboards

You are asked to create three pages of reports in your dashboard. While you have the freedom to pick the chart type for each of the points. **Avoid** using the same chart type twice. Also, make sure the charts are presented correctly and organized throughout the dashboard. The pages are as follows:

*International Games – General Information:*

This page should contain 2 charts.

1. Show the top 10 countries by number of wins in each tournament.
    - Exclude ties from showing up in the chart
2. Number of games in each tournament.
    - Exclude friendlies from showing in the chart

*International Games – City:*

1. Show top 30 cities by the number of games played in the city for each tournament.
    a. Add a filter for the chart, to filter data based on the selected tournament.
    b. In case any city names contain Latin characters. Standardize the city names to show only English characters in the dashboard (Only 1 example is enough)

*International Games – Country:*

1. Add a required drop-down list, which contains country names. The filter will be used on all objects in the page.
2. Show the number of goals scored Vs the number of goals received for each selected country from the drop-down list.
3. Show the number of goals scored Vs goals received in each year for the selected country from the drop-down list.

*International Games – Map:*

1. Show the number of matches played in each country on a map. The size of the bubble should map to the number of matches, while the bubble colour should be for the tournament
    a. Add a filter for the tournament

*Mystery*

On the last page, surprise your management with a creative mystery chart! Use this opportunity to explore the data deeper and find interesting correlations or information you want to highlight in one or more charts.

## Colour Theme

Having a consistent colour scheme throughout the report can be very helpful in identifying common values quickly. This is why you are advised to configure the following display rules on the report level. Refer to the table below for the exact colour codes

| Tournament | Theme colour code |
|---|---|
| FIFA World Cup | #005391 |
| African Cup of Nations | #189E4B |
| UEFA Euro | #00088E |
| AFC Asian Cup | #FFC72C |
| Copa America | #33A3FF |
| Friendly | #DD5757 |

N.B: This case study is just the beginning. Feel free to share any creative insights or alternative approaches that you think could provide valuable additional context.