

Dataset Football Data from Transfermarkt

Tuesday, March 3, 2026

10:04 AM

Tasks

- ✓ Ingest into Power BI
- ✓ Review dataset
 - Clean
 - De-dup
 - Check nulls and blanks
 - Check data types
 - Check field names
 - Case
 - Clarity
 - Normalize
 - ~~Aggregate~~—Not Necessary for this dataset
 - Set up exception process
 - Add basic calculations
 - ~~RowCount~~—Included in Metadata
- ✓ Power BI model
 - Make table relationships
 - Star or Snowflake Schema
 - Make Calendar table(s)
 - Create dynamic calculations
 - Create dynamic data dictionary & META data
 - Make pages
 - Create page template & theme
 - Include footer, header, logo and Navigation
 - © 2026 JPByDesign
 - Main report
 - Focus on KPI
 - Exception reporting
 - Details (Drill through)
 - Additional pages to answer questions/stories
 - Data dictionary & META page
 - Add Documentation page/link
 - Include
 - License
 - Data owner(s)
 - Data description
 - Source & expected refresh frequency
 - Identify
 - KPIs/Goals/Story
 - Intended target audience
 - Initial results/discovery
 - ~~Add security model~~—Not necessary for this dataset - public

- Do Performance Analysis and Optimization
- ✓ Documentation
 - Purpose - See below
 - Story - See below
 - Metadata - see dynamically created Metadata in report
 - Data Dictionary - see dynamically created Data Dictionary in report
 - Relationship model - See Below
 - Performance Analysis results
 - ~~Instructions~~
 - ~~How to run the file~~
 - ~~How to update excel file and PBI~~
 - ~~How to share file~~
- ✓ Deliver

Issues Questions

- In [Games] table, [Game Aggregate] (formerly [aggregate]) is a time value... but doesn't really make sense. What should this represent?
- In [Players] table, [Players Last Season Start Year] (formerly [season]) could be either the last season player played or Players intended last season
 - I made an assumption that this was last season player played - which means that if, as of March 2026, they had 2025 - they are not retired. All players that had a year before that would be retired.
 - Based the [Is Retired] calculation upon this assumption

Additional Steps

- Added Game Calendar - joined to Game table
- Added Player Age Calculation
- Added [Is Retired] calculation
- Found Club logo from <https://tmsl.akamaized.net/images/wappen/head/281.png> and was able to expand that to most other clubs.
- Used DAX Studio to render Metadata
- Made Data Dictionary update dynamically
- Added calculation weights table
- Created Club Opponents from Clubs table – reversing the main set of values

Dataset Description

David Cariboo · Updated 12 hours ago

Football Data from Transfermarkt

Football (Soccer) data scraped from Transfermarkt website



Football Data from Transfermarkt

Code

Download



About Dataset

TL;DR

Clean, structured and **automatically updated** football data from [Transfermarkt](#), including

- 60,000+ games from many seasons on all major competitions
- 400+ clubs from those competitions
- 30,000+ players from those clubs
- 400,000+ player market valuations historical records
- 1,200,000+ player appearance records from all games

and more!

What does it contain?

The dataset is composed of multiple CSV files with information on competitions, games, clubs, players and appearances that is automatically updated **once a week**. Each file contains the attributes of the entity and the IDs that can be used to join them together.

For example, the appearances file contains **one row per player appearance**, i.e. one row per player per game played. For each appearance you will find attributes such as goals, assists or yellow_cards and IDs referencing other entities within the dataset, such as player_id and game_id.

Idea: Chekout the mate [data.world project Football Data from Transfermarkt](#) for examples of how to use the resources in this dataset to create different types of analysis.

How did we build it?

The source code that maintains this dataset, as well as the data pipeline, [is available in Github](#). On a high level, the project uses [transfermarkt-scraper](#) to pull the data from [Transfermarkt website](#) and a set of Python scripts and SQL to curate it and publish it here.

Idea: Watch the [transfermarkt-datasets Github repository](#) for updates on improvements and bugfixes on this dataset

What is the status?

This dataset is a live project subject to regular updates and new enhancements. The best way to find out about the different initiatives within this project and their status is to check the `Issues` section on [transfermarkt-datasets](#).

Contact

The source of truth for the code and data for this project is Github and that is the best place to start a conversation, please checkout the community guidelines [here](#).

If you have a github user I kindly ask that you post your questions there and not in the **Discussion** tab in this website.

View less

Usability

9.41

License

[CC0: Public Domain](#)

Expected update frequency

Weekly

Tags

[Computer Science](#)[Programming](#)[Sports](#)[Games](#)

From <https://www.kaggle.com/davidcariboo/player-scores?select=game_lineups.csv>

File link:

[Dataset Football Data from Transfermarkt.zip](#)

Name	Type	Compressed size	Password ...	Size
 appearances.csv	Microsoft Excel Comma S...	38,206 KB	No	
 club_games.csv	Microsoft Excel Comma S...	1,679 KB	No	
 clubs.csv	Microsoft Excel Comma S...	25 KB	No	
 competitions.csv	Microsoft Excel Comma S...	2 KB	No	
 game_events.csv	Microsoft Excel Comma S...	37,332 KB	No	
 game_lineups.csv	Microsoft Excel Comma S...	98,202 KB	No	
 games.csv	Microsoft Excel Comma S...	4,209 KB	No	
 player_valuations.csv	Microsoft Excel Comma S...	4,917 KB	No	
 players.csv	Microsoft Excel Comma S...	2,793 KB	No	
 transfers.csv	Microsoft Excel Comma S...	2,956 KB	No	

Purpose

Story

From a team perspective,

~~What can we expect against our next opponent?~~

- ~~What is the most likely lineup?~~
- ~~Who are the most dangerous players?~~
- ~~Has anyone transferred recently that would know our schemes?~~
- ~~How have we fared historically?~~
- ~~How many games have been played in last weeks?~~
 - ~~Possibly average games for all teams in that competition for a +/- value~~
- ~~What is the win/loss streak in last weeks?~~

~~From a position perspective what can we expect?~~

- ~~Who plays in the opposite position?~~
- ~~Who has had success historically/recently? Failure?~~
- ~~Did they play in last game(s)?~~
- ~~How many primary and secondary subs do they have for that position?~~

Revised Story

From a team perspective,

- ~~From a historical perspective, how does a primary team compare to selected other teams~~
 - ~~Season points/rank~~
 - ~~Previous period (ex: month) number of Goals & Assists~~
 - ~~Penalty counts (as a negative)~~
 - ~~Historical points against selected other teams~~
 - ~~Other (TBD)~~

-
- ~~Distributed — Divide by # of games played~~
 - ~~Weighted (TBD)~~
 - ~~From a position perspective, how does a primary team compare to selected other teams at each position~~
 - ~~Compare Primary player and Substitutes for each position~~
 - ~~Previous period number of Goals & Assists~~
 - ~~Previous period number of Penalty counts (as a negative)~~
 - ~~Other (TBD)~~
-
- ~~Distributed — Divide by # of games played in previous period~~
 - ~~Weighted (TBD)~~

Revised Story2

Managers and Coaches sometimes need a place to start when considering an upcoming opponent.

All things being equal, where does the team have an advantage over a selected opponent:

- From a Historical perspective
 - How many wins (vs ties and losses) has the team experienced against our selected opponent?
 - How does that help our spirit?
- From a Season perspective
 - How do we compare Points (win = 3, tie = 1, loss = 0) per game to our selected opponent?
 - How should this indicate how we should play?
 - How do we compare Card Penalties (Red = 3, Yellow = 1) per game to our selected opponent?
 - What should we focus on to be ahead of the other team?
- From a recent play perspective
 - How do we compare in goals for the last 30 days?
 - If we are scoring well, don't mix up the scoring line

This is a rich dataset, and I have barely scraped the surface of what could be done.

Some future opportunities might include:

From a team perspective:

What can the team expect against its' selected opponent?

- What is the most likely lineup?
- Who are the most dangerous players?
- Has anyone transferred recently that would know our schemes?
- How have we fared historically?
- How many games have been played in last weeks?
- What is the win/loss streak in last weeks?

From a position perspective:

what can we expect from a historical perspective?

- Who plays in the opposite position?
- Who has had success historically/recently? Failure?

- Did they play in last game(s)?
- How many primary and secondary subs does the team have for that position?

Image Soccer ball

- Source: <https://www.publicdomainpictures.net/en/view-image.php?image=70135&picture=soccer-ball-2>
- License: [CC0 Public Domain](#)
- Notes:

Piotr Siedlecki has released this “Soccer Ball 2” image under Public Domain license. It means that you can use and modify it for your personal and commercial projects. If you intend to use an image you find here for commercial use, please be aware that some photos do require a model or a property release. Pictures featuring products should be used with care.



Custom Theme

Color Palette:

- Derived from <https://color.adobe.com/create/image> based off of JPByDesign logo (fun)
 - Theme Colors:

#A36BF2 [Example](#)

#7B44F2 [Example](#)

#38F254 [Example](#)

#0D0D0D [Example with dark highlight](#)

#0DF205 [Example](#)

#0597F2 [Example](#)

#020659 [Example](#)

#05F2DB Example

- Sentiment Colors:
 - Negative: #D7041A Example
 - Positive: #1AAB40 Example
 - Neutral: #FFA600 Example

Default text:

General:

Font family: Calibri
Standard size: 11pt
Default Font color: #3B3B3A

Title:

Font family: Calibri
Standard size: 24pt
Default Font color: #000000

Cards and KPIs:

Font family: Calibri
Standard size: 20pt
Default Font color: #3B3B3A

Tab Headers:

Font family: Segoe UI Semibold
Standard size: 15pt
Default Font color: #3B3B3A

Page:

Wallpaper:

Color: #A36BF2
Transparency: 0%

Page Background:

Color: #FFFFFF (white)
Transparency: 0%

<<JPByDesign Theme.json>>

Calculations

See all calculations in Data Dictionary within Report

Based on comparison between teams

- Current Season = prior September 1 to refreshed date
- Period to Date = specified period (example: last 3 months) to refreshed date
 - will be determined using parameters/custom table
- Team points - Sum wins(3), ties (1), losses (0) for current season - divide by # of games played in current season
 - ~~Could also use as rank across all teams~~
- Goals & Assists within last period - 1 point for each ~~— divide by # of games~~ in period to date

- Penalty counts within ~~last period~~ for season- sum Red cards (2), Yellow cards (1), other (1) - divide by # of games in period
- Historical Points against specified team(s) - Sum wins(3), ties (1), losses (0) ~~—divide by # games played against~~
- ~~Other (TBD) (ideas: Home/Away records — within period to date, # supporters, etc...)~~

- Weighting of results

Team Points	30%
Goals & Assists	40%
Penalty counts	10%
Historical Points	15%
Other	5%
Total	100%

- Note: this is adjustable

Use Calculation Groups to break up

Published Site

For Smaller Screen (600x373.5)

HTML to paste on website:

```
<iframe title="Football from Transfermarkt report" width="600" height="373.5"
src="https://app.powerbi.com/view?r=eyJrIjojN2E4NGYyNTktNDZhZi00ZTY2LWEwYTktNjMyNjg2MDczZmI2liwidCI6IjBmN2FkMDA3LTQyYjgtNGU2MS1hNzJiLWYyMDY0ZmNjZThhZSJ9&pageName=fb3c721ae450400d8d67" frameborder="0" allowFullScreen="true"></iframe>
```

Link to send via email

[Power BI](#)



Microsoft Power BI

For bigger screens (800x636 px)

Website embed link:

```
<iframe title="Football from Transfermarkt report" width="800" height="636"  
src="https://app.powerbi.com/view?r=eyJrljoiN2E4NGYyNTktNDZhZi00ZTY2LWEwYTktNjMyNjg  
2MDczZmI2liwidCI6IjBmN2FkMDA3LTQyYjgtNGU2MS1hNzJiLWYyMDY0ZmNjZThhZSJ9&pageNa  
me=fb3c721ae450400d8d67" frameborder="0" allowFullScreen="true"></iframe>
```

Email Link:

[Power BI](#)



Microsoft Power BI

Link to website posting:

<https://ipbydesign.com/index.php/2026/04/11/demo-power-bi-site-football-soccer-club-comparison/>

Short Link:

<https://wp.me/p7Ze6V-XQ>

Performance Analysis:

- All testing passed
 - Testing on 1st pass: Less than 1 second on all pages
 - Testing on 3rd pass: Less than 2 seconds on all pages
 - Including with modifications to slicers
- See attached csv for details

Relationship Model

See details in Data Dictionary (in report)



