EU Soccer Betting

DMBI Assignment #2

Agenda

01 Introduction & Business Goal

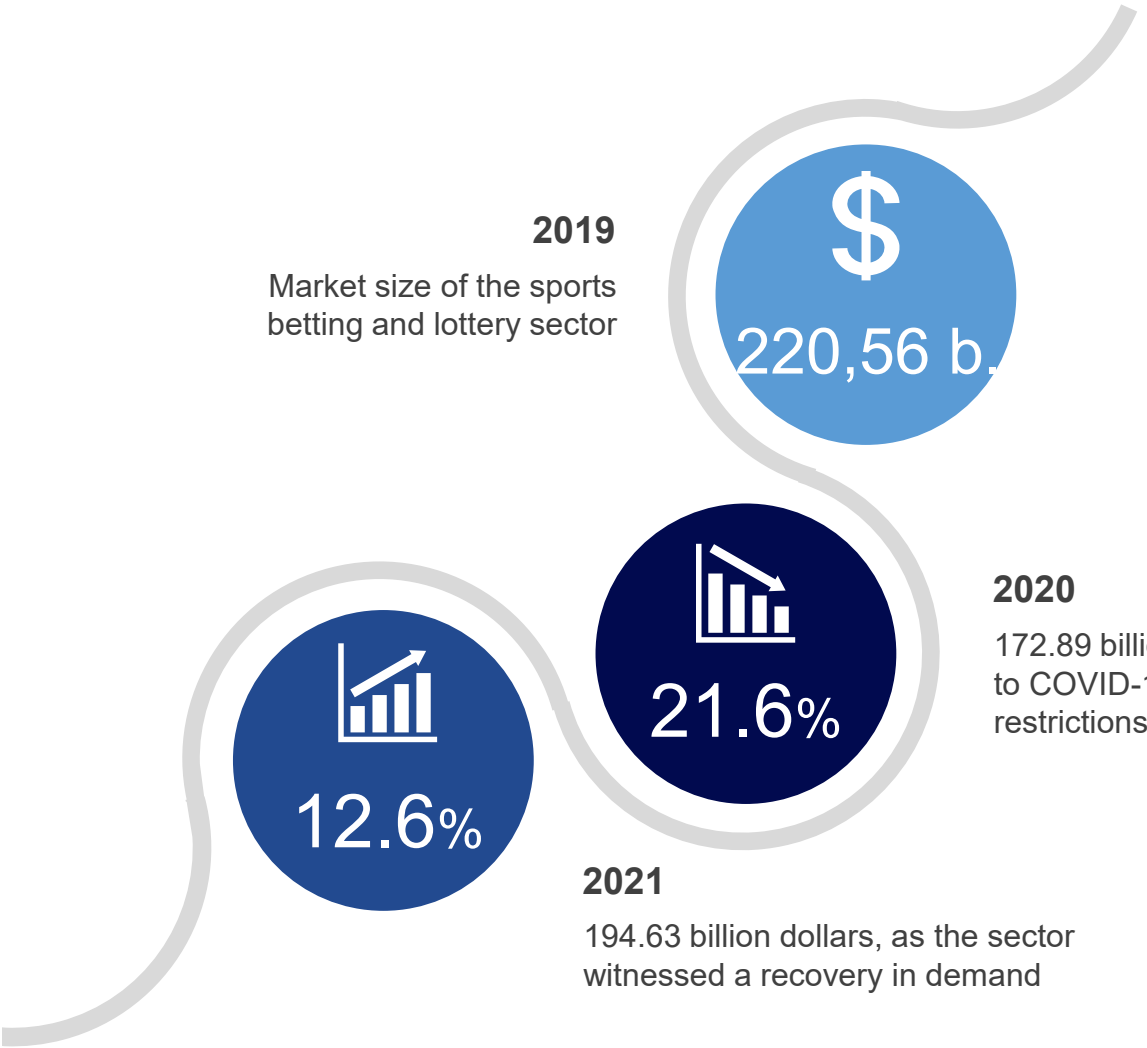
02 Dataset & ETL Procedure

03 Analysis Services (SSAS)

04 Visualizations

Introduction & Business Goal

Introduction



2019

Market size of the sports betting and lottery sector

2020

172.89 billion dollars, due to COVID-19 pandemic restrictions

2021

194.63 billion dollars, as the sector witnessed a recovery in demand

“People should not gamble/bet”

- Not an ethical/philosophical statement

Predicting sports results and making a bet on the outcome is known as sports betting. The frequency and types of sports betting varies according to culture and country of residence, with most bets being placed on amateur and professional levels of football/soccer, American football, basketball, baseball, hockey, track cycling, car racing, mixed martial arts, and boxing.

Owning or investing on a betting/gambling company can be very profitable. If sport betting companies gain large profits, it is due to the existence of betters who bet their money on their odds and of course they lose their money. And sometimes an enormous amount of it.

Source: https://www.statista.com/statistics/1186240/sports-betting-and-lottery-industry-market-size-global/

Business Goal

Not Gambling

Persuade people not to gamble/bet

People should earn their money from their regular jobs. Gambling or betting can lead to addiction.

One solution to this problem and the most efficient one is to persuade betters not to bet again in their life.

Due to the aspect of “easy money” and the fun of being lucky (rarely), it is a rather difficult case for someone to turn the most efficient solution to feasible.

Gambling

Take advantage of big data

Big data that are flooded everywhere in our “web universe”. Someone could collect them and exploit them, making use of

- databases to store them and analyze them,
- and statistics to extract some derivatives making use of descriptive, predictive, or even prescriptive statistics.

In that case, he/she would not avoid losing some of his money but could minimize that amount as close to zero as possible.

Dataset & ETL Procedure

Timeline of ETL Procedure

1

Download

Find the dataset and download it

2

Drivers

Download the required drivers (SQLite db file)

3

Extract

Create the Schema and intermediate (staging) tables and load the initial data in the staging tables

4

Transform

Clean the data, convert data types, deduplicating, etc.

5

Load

Data is loaded “securely” into dimension and fact tables making business-ready data available

Download Dataset

kaggle

Search

Create

Home

Competitions

Datasets

Code

Discussions

Learn

More

Your Work

RECENTLY VIEWED

European Soccer Da...

Match Outcome Pre...

The Most Predictabl...

View Active Events

HUGO MATHIEN · UPDATED 6 YEARS AGO

4057

New Notebook

Download (34 MB)

European Soccer Database

25k+ matches, players & teams attributes for European Professional Football

Data | Code (1547) | Discussion (115)

About Dataset

The ultimate Soccer database for data analysis and machine learning

What you get:

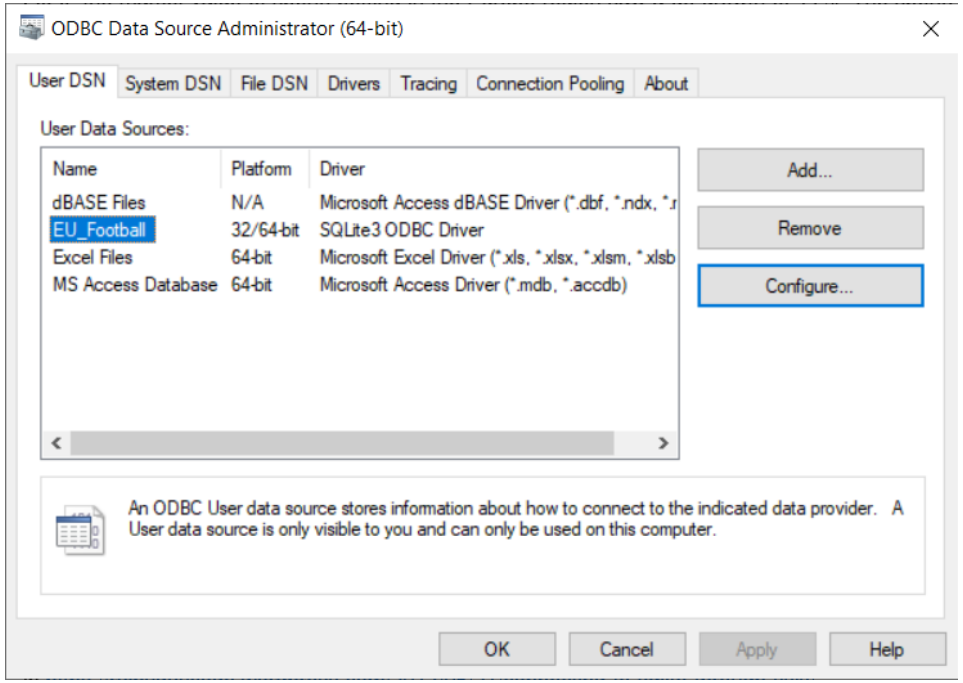
- +25,000 matches
- +10,000 players
- 11 European Countries with their lead championship
- Seasons 2008 to 2016
- Players and Teams' attributes* sourced from EA Sports' FIFA video game series, including the weekly updates

Usability
7.06

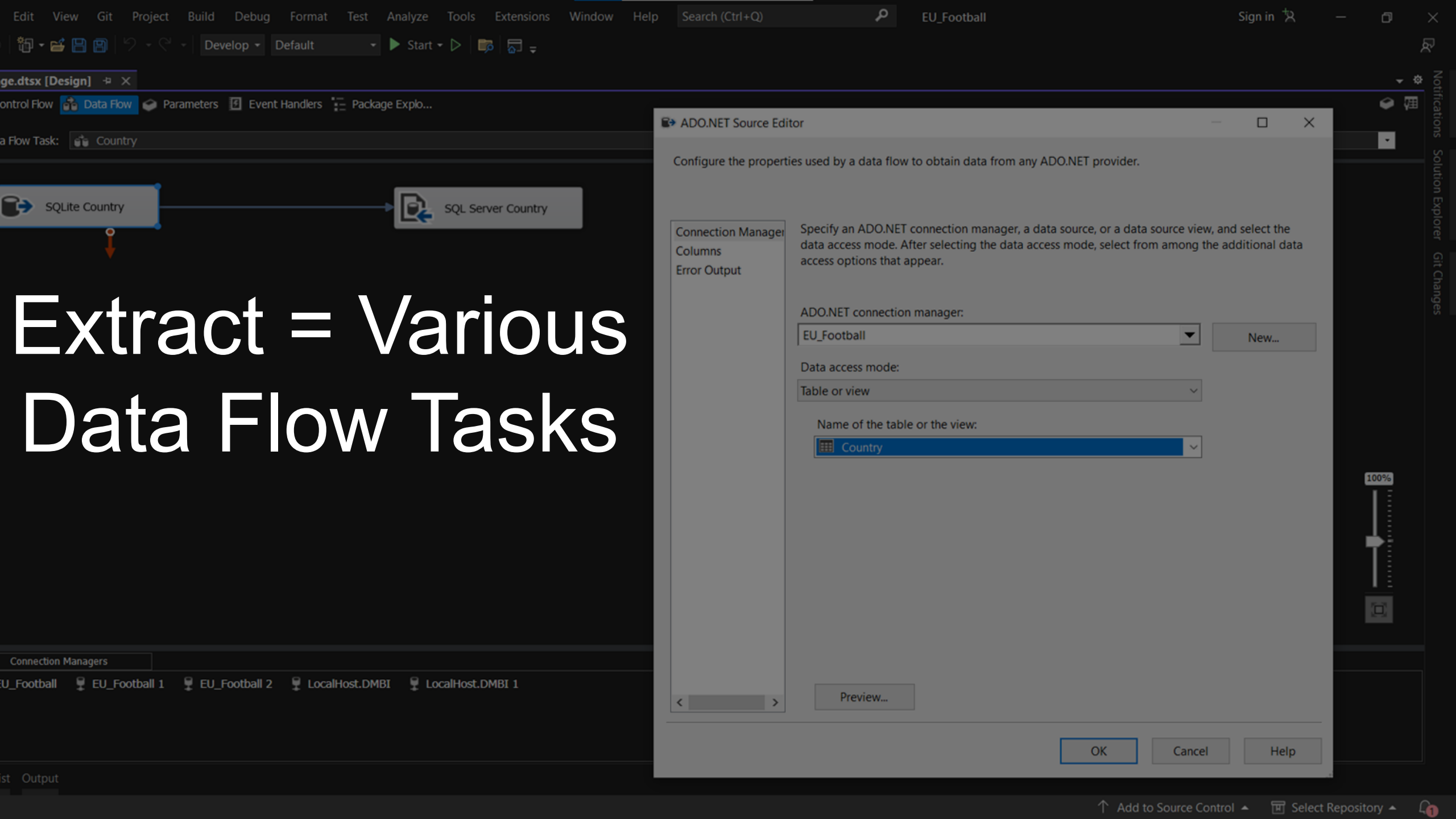
License
Database: Open Database, Cont...

Expected update frequency
Not specified

Source: https://www.kaggle.com/datasets/hugomathien/soccer


ODBC Data Source Administrator (64-bit)
User DSN
System DSN
File DSN
Drivers
Tracing
Connection Pooling
About
User Data Sources:
Name
Platform
Driver
dBASE Files
N/A
Microsoft Access dBASE Driver (*.dbf, *.ndx, *.r
EU_Football
32/64-bit
SQLite3 ODBC Driver
Excel Files
64-bit
Microsoft Excel Driver (*.xls, *.xlsx, *.xlsm, *.xlsb
MS Access Database
64-bit
Microsoft Access Driver (*.mdb, *.accdb)
Add...
Remove
Configure...
An ODBC User data source stores information about how to connect to the indicated data provider. A User data source is only visible to you and can only be used on this computer.
OK
Cancel
Apply
Help

Drivers


Edit View Git Project Build Debug Format Test Analyze Tools Extensions Window Help
Search (Ctrl+Q)
EU_Football
Sign in
Develop
Default
Start
ge.dtsx [Design]
ontrol Flow
Data Flow
Parameters
Event Handlers
Package Explo...
a Flow Task:
Country
SQLite Country
SQL Server Country
Extract = Various Data Flow Tasks
ADO.NET Source Editor
Configure the properties used by a data flow to obtain data from any ADO.NET provider.
Connection Manager
Columns
Error Output
Specify an ADO.NET connection manager, a data source, or a data source view, and select the data access mode. After selecting the data access mode, select from among the additional data access options that appear.
ADO.NET connection manager:
EU_Football
New...
Data access mode:
Table or view
Name of the table or the view:
Country
Preview...
OK
Cancel
Help
Connection Managers
U_Football
EU_Football 1
EU_Football 2
LocalHost.DMBI
LocalHost.DMBI 1
Output
Add to Source Control
Select Repository
Notifications
Solution Explorer
Git Changes
100%

Transform

Convert Datatypes

Converted all NVARCHAR(MAX) columns to VARCHAR because SQL Server treats the first case as TEXT values and constraints UNIQUE KEYS could not be added

Identity Primary Keys

Added IDENTITY PRIMARY KEYS to Dimension Tables

Columns (no ID) to Dimensions

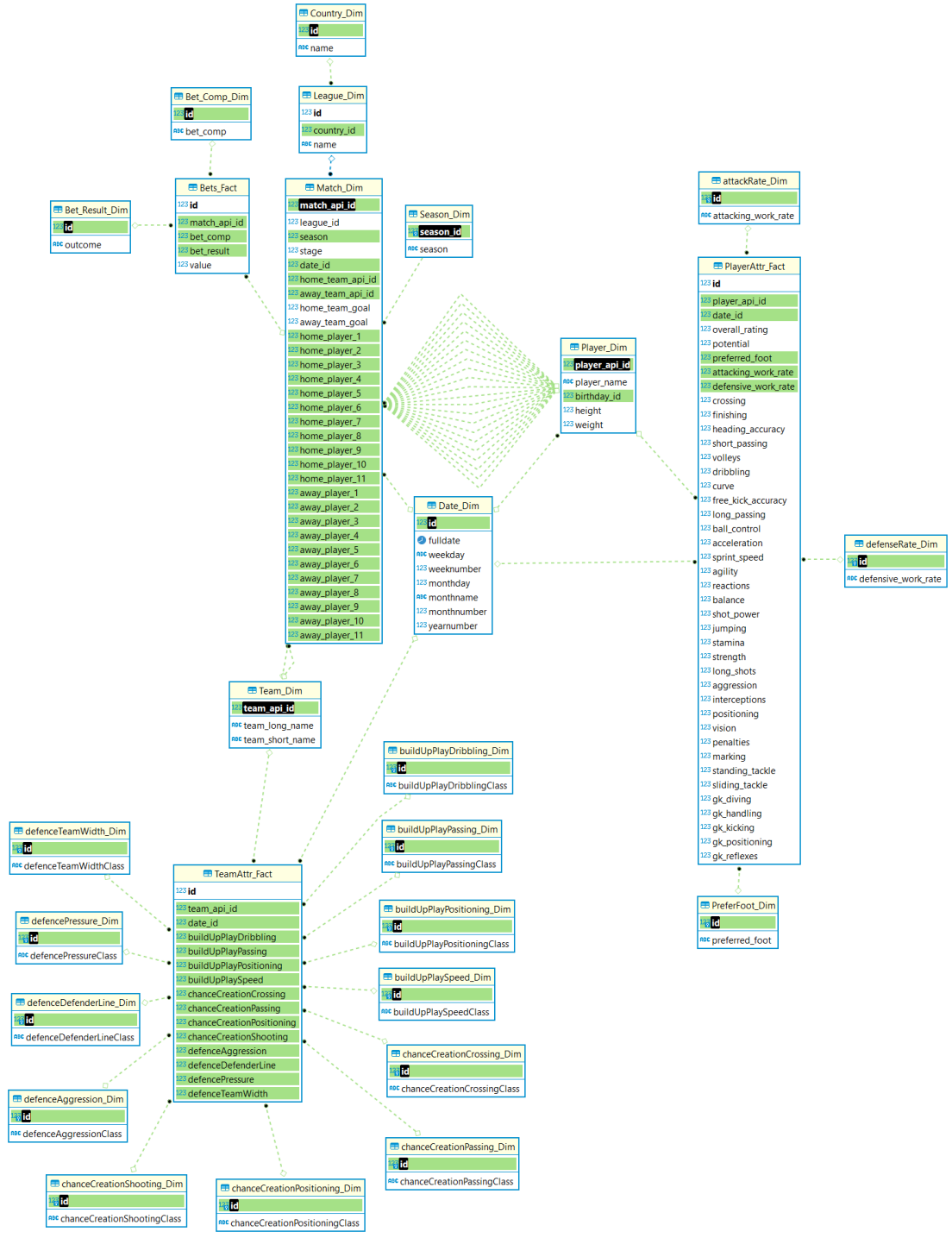
There were some columns that should be Dimensions and a separate table with an id did not exist. So, we took advantage of the Analytic Function “ROW_NUMBER() OVER(ORDER BY column)”.

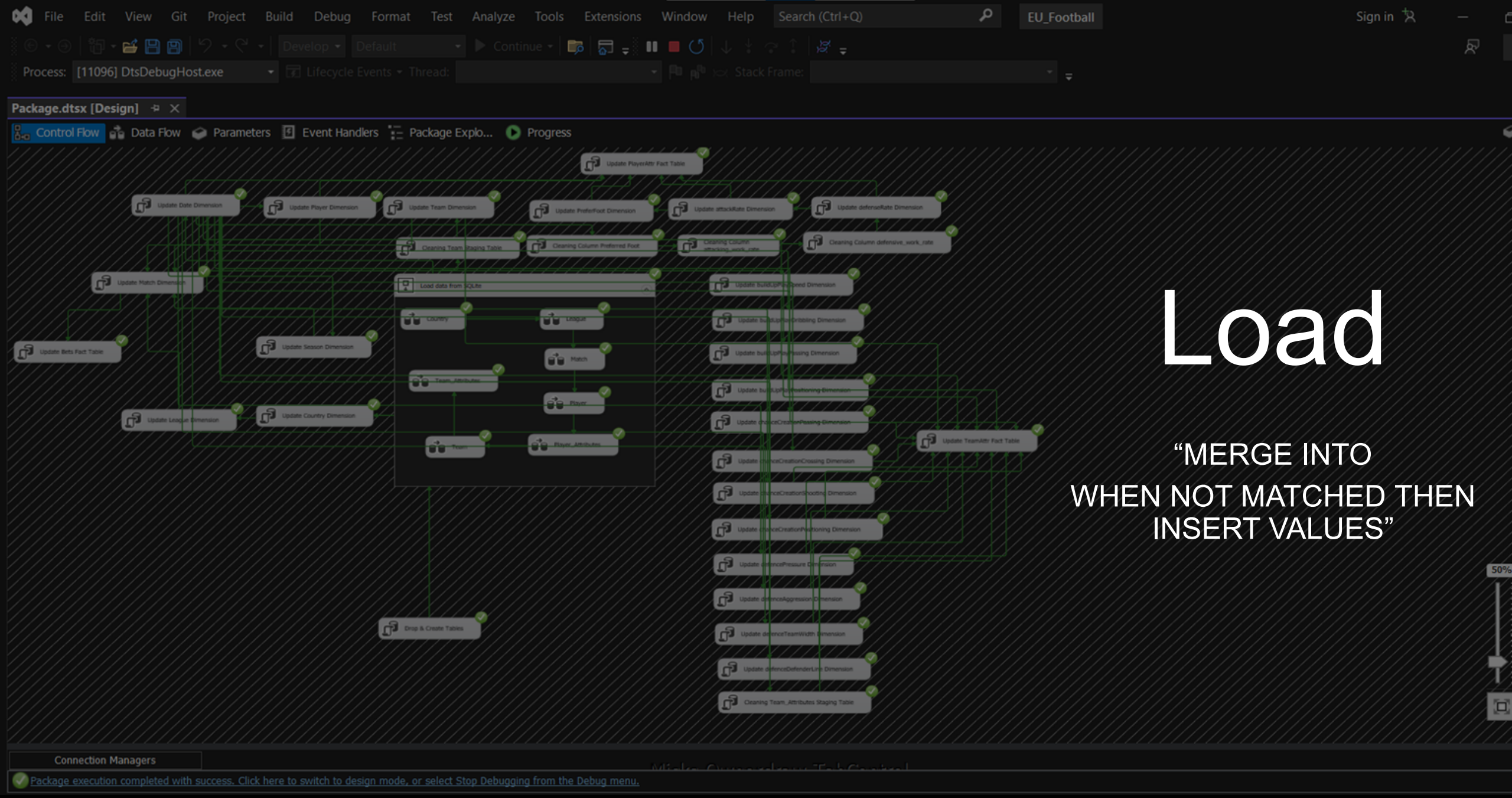
Cleaning

- Wrong character values → N/A
- Duplicate name values with different IDs → Updated to correct name values and deleted where truly was duplicate

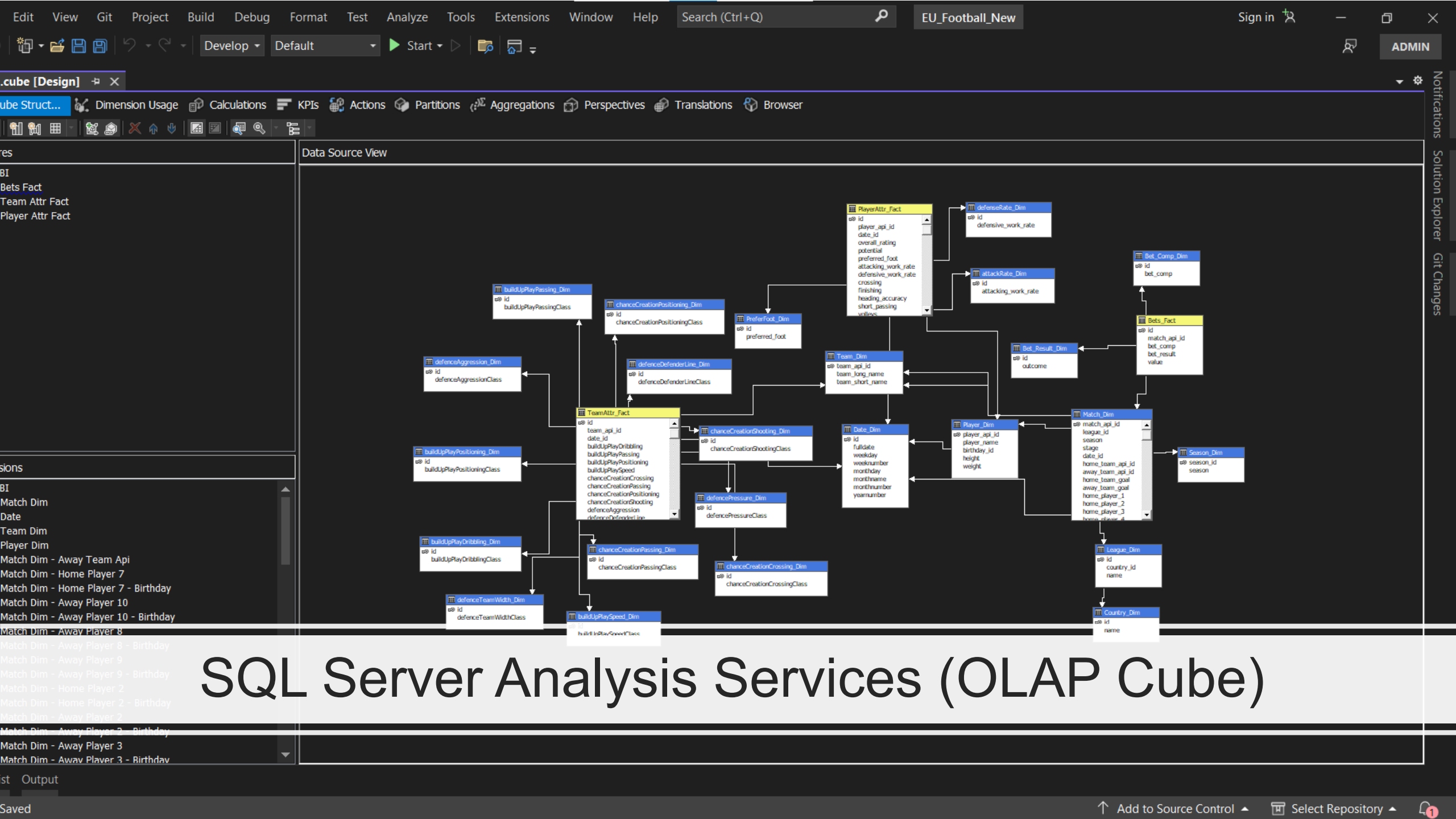
Used UNPIVOT Relational Operator to transform 30 columns of a table to column values discarding the rows that contained null values for all columns

Schema


Country_Dim
id
name
League_Dim
id
country_id
name
Bet_Comp_Dim
id
bet_comp
Bets_Fact
id
match_api_id
bet_comp
bet_result
value
Bet_Result_Dim
id
outcome
Match_Dim
match_api_id
league_id
season
stage
date_id
home_team_api_id
away_team_api_id
home_team_goal
away_team_goal
home_player_1
home_player_2
home_player_3
home_player_4
home_player_5
home_player_6
home_player_7
home_player_8
home_player_9
home_player_10
home_player_11
away_player_1
away_player_2
away_player_3
away_player_4
away_player_5
away_player_6
away_player_7
away_player_8
away_player_9
away_player_10
away_player_11
Season_Dim
season_id
season
Player_Dim
player_api_id
player_name
birthday_id
height
weight
attackRate_Dim
id
attacking_work_rate
PlayerAttr_Fact
id
player_api_id
date_id
overall_rating
potential
preferred_foot
attacking_work_rate
defensive_work_rate
crossing
finishing
heading_accuracy
short_passing
volleys
dribbling
curve
free_kick_accuracy
long_passing
ball_control
acceleration
sprint_speed
agility
reactions
balance
shot_power
jumping
stamina
strength
long_shots
aggression
interceptions
positioning
vision
penalties
marking
standing_tackle
sliding_tackle
gk_diving
gk_handling
gk_kicking
gk_positioning
gk_reflexes
defenseRate_Dim
id
defensive_work_rate
Date_Dim
id
fulldate
weekday
weeknumber
monthday
monthname
monthnumber
yearnumber
Team_Dim
team_api_id
team_long_name
team_short_name
PreferFoot_Dim
id
preferred_foot
buildUpPlayDribbling_Dim
id
buildUpPlayDribblingClass
buildUpPlayPassing_Dim
id
buildUpPlayPassingClass
buildUpPlayPositioning_Dim
id
buildUpPlayPositioningClass
buildUpPlaySpeed_Dim
id
buildUpPlaySpeedClass
chanceCreationCrossing_Dim
id
chanceCreationCrossingClass
chanceCreationPassing_Dim
id
chanceCreationPassingClass
chanceCreationPositioning_Dim
id
chanceCreationPositioningClass
chanceCreationShooting_Dim
id
chanceCreationShootingClass
defenceAggression_Dim
id
defenceAggressionClass
defenceDefenderLine_Dim
id
defenceDefenderLineClass
defencePressure_Dim
id
defencePressureClass
defenceTeamWidth_Dim
id
defenceTeamWidthClass
TeamAttr_Fact
id
team_api_id
date_id
buildUpPlayDribbling
buildUpPlayPassing
buildUpPlayPositioning
buildUpPlaySpeed
chanceCreationCrossing
chanceCreationPassing
chanceCreationPositioning
chanceCreationShooting
defenceAggression
defenceDefenderLine
defencePressure
defenceTeamWidth

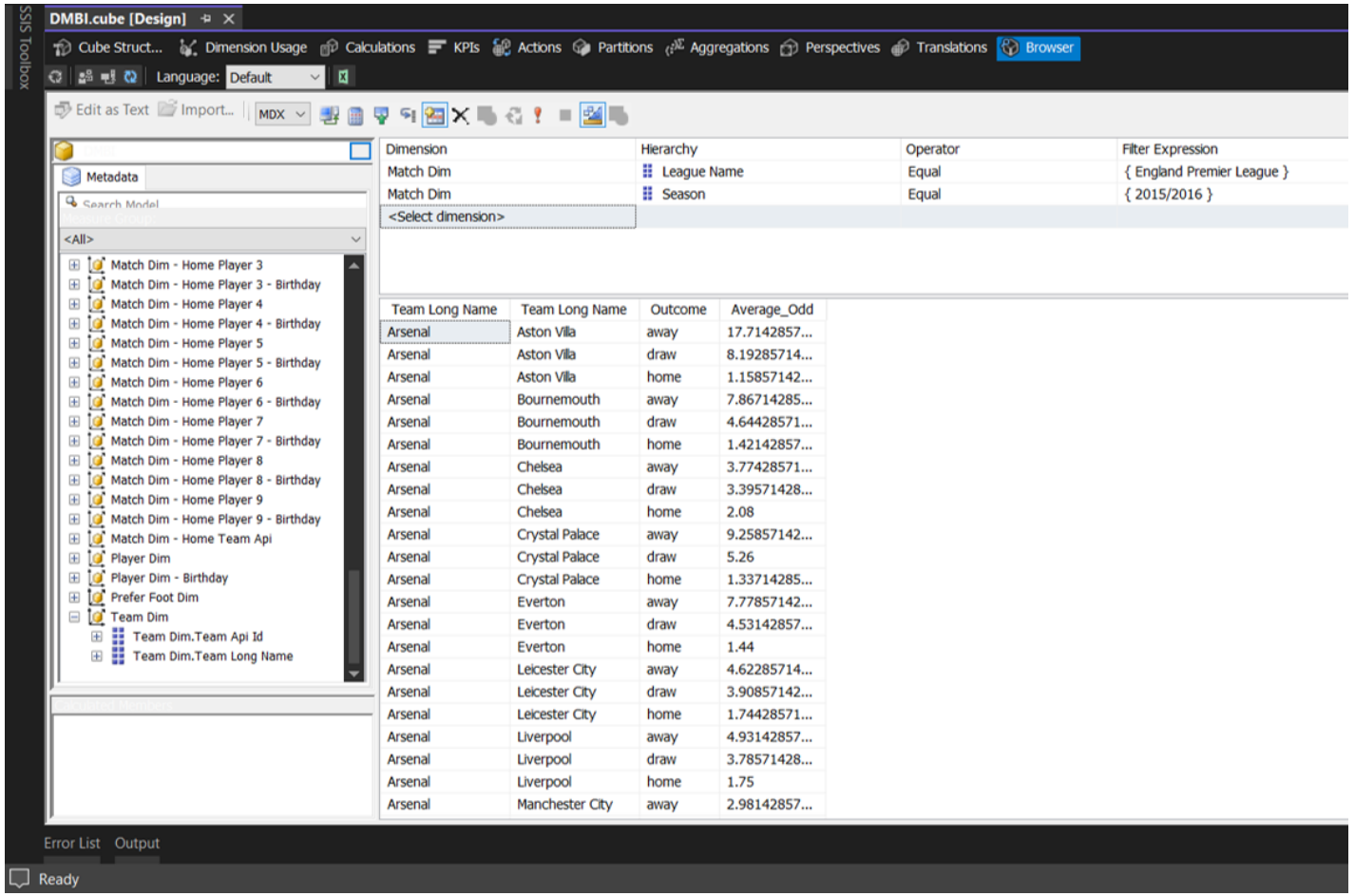

File Edit View Git Project Build Debug Format Test Analyze Tools Extensions Window Help
Search (Ctrl+Q)
EU_Football
Sign in
Process: [11096] DtsDebugHost.exe
Package.dtsx [Design]
Control Flow
Data Flow
Parameters
Event Handlers
Package Explo...
Progress
Load
“MERGE INTO
WHEN NOT MATCHED THEN
INSERT VALUES”
Connection Managers
Package execution completed with success. Click here to switch to design mode, or select Stop Debugging from the Debug menu.

SQL Server Analysis Services

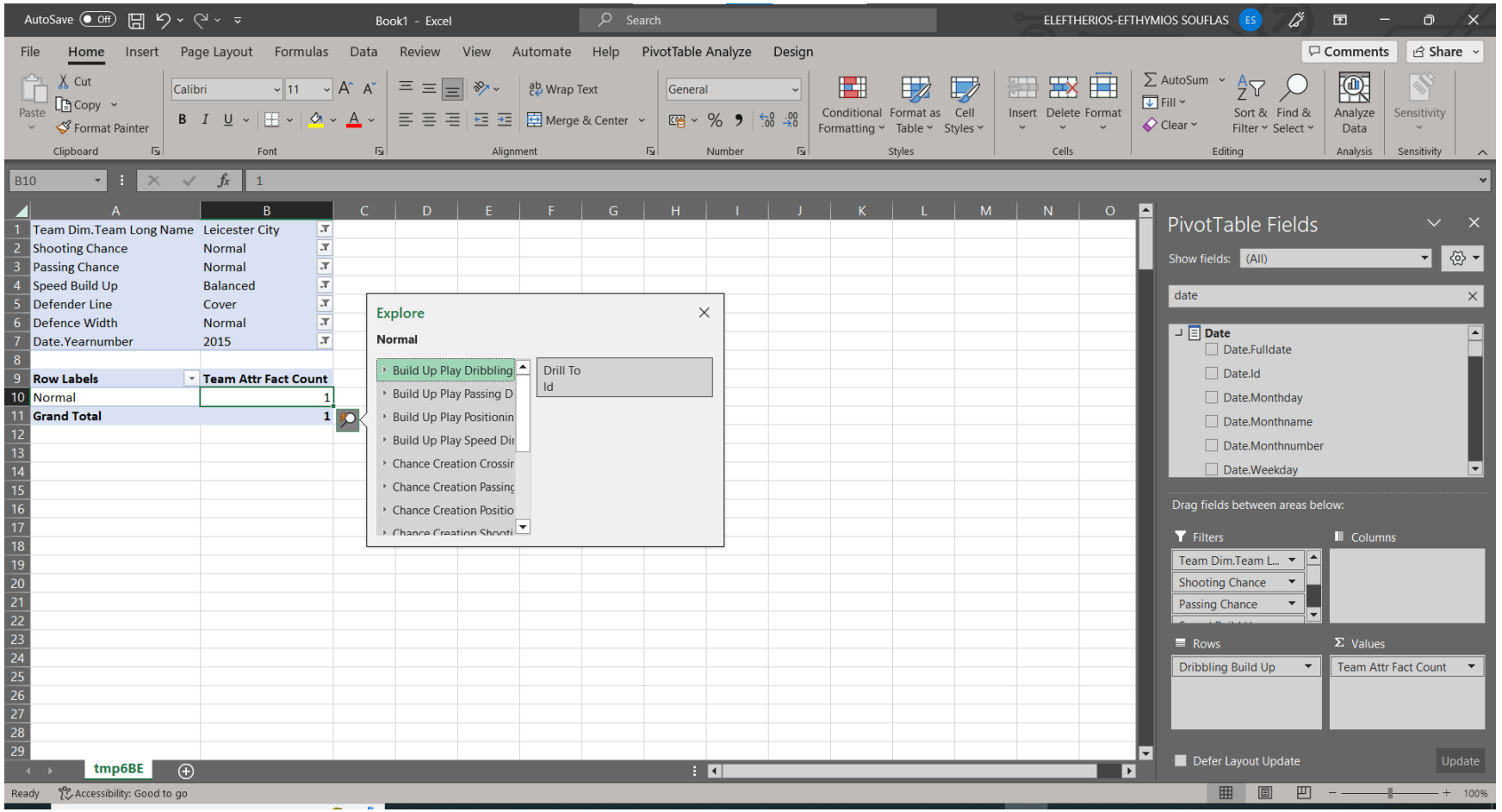


SQL Server Analysis Services (OLAP Cube)

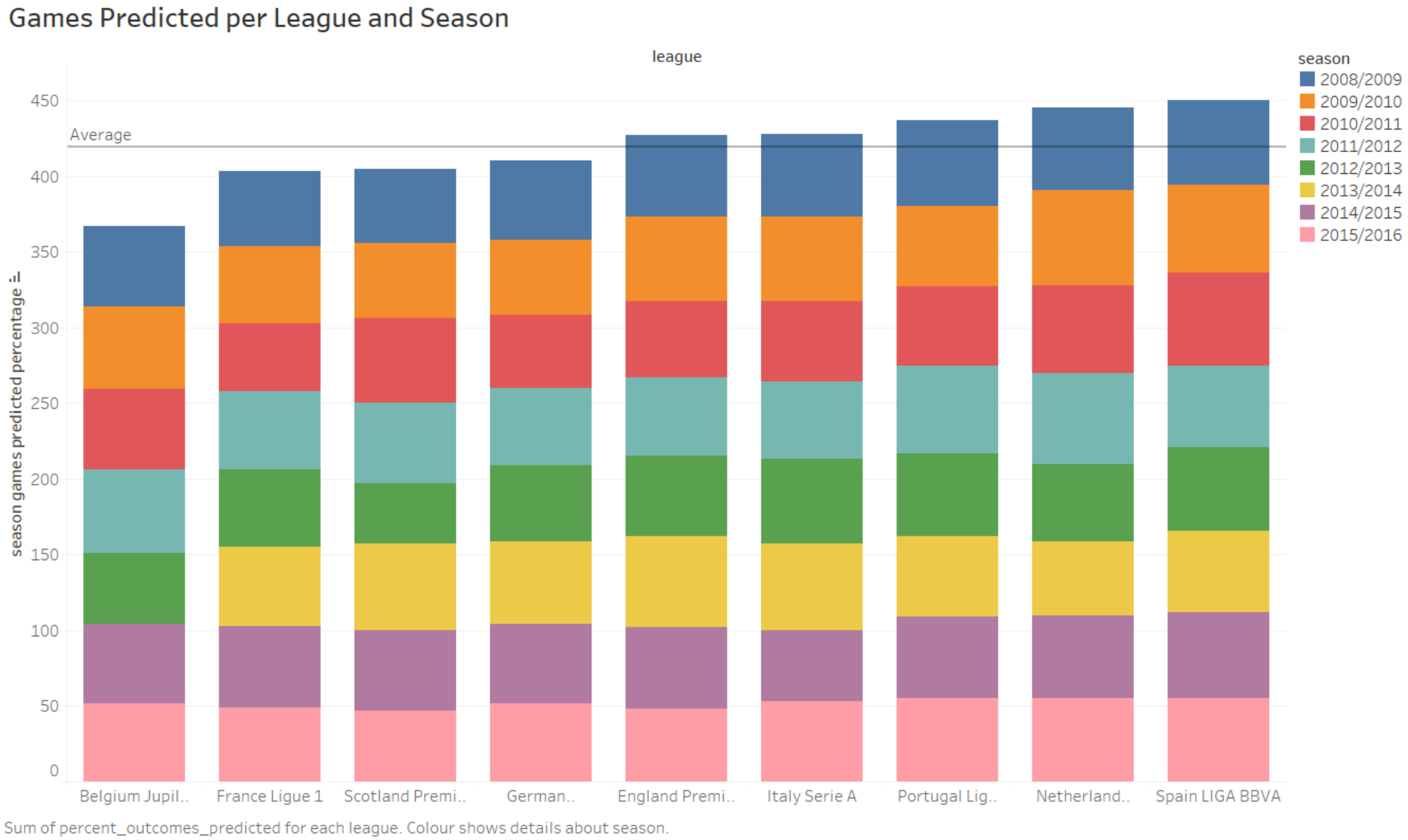
OLAP Report
(Calculated Measure)



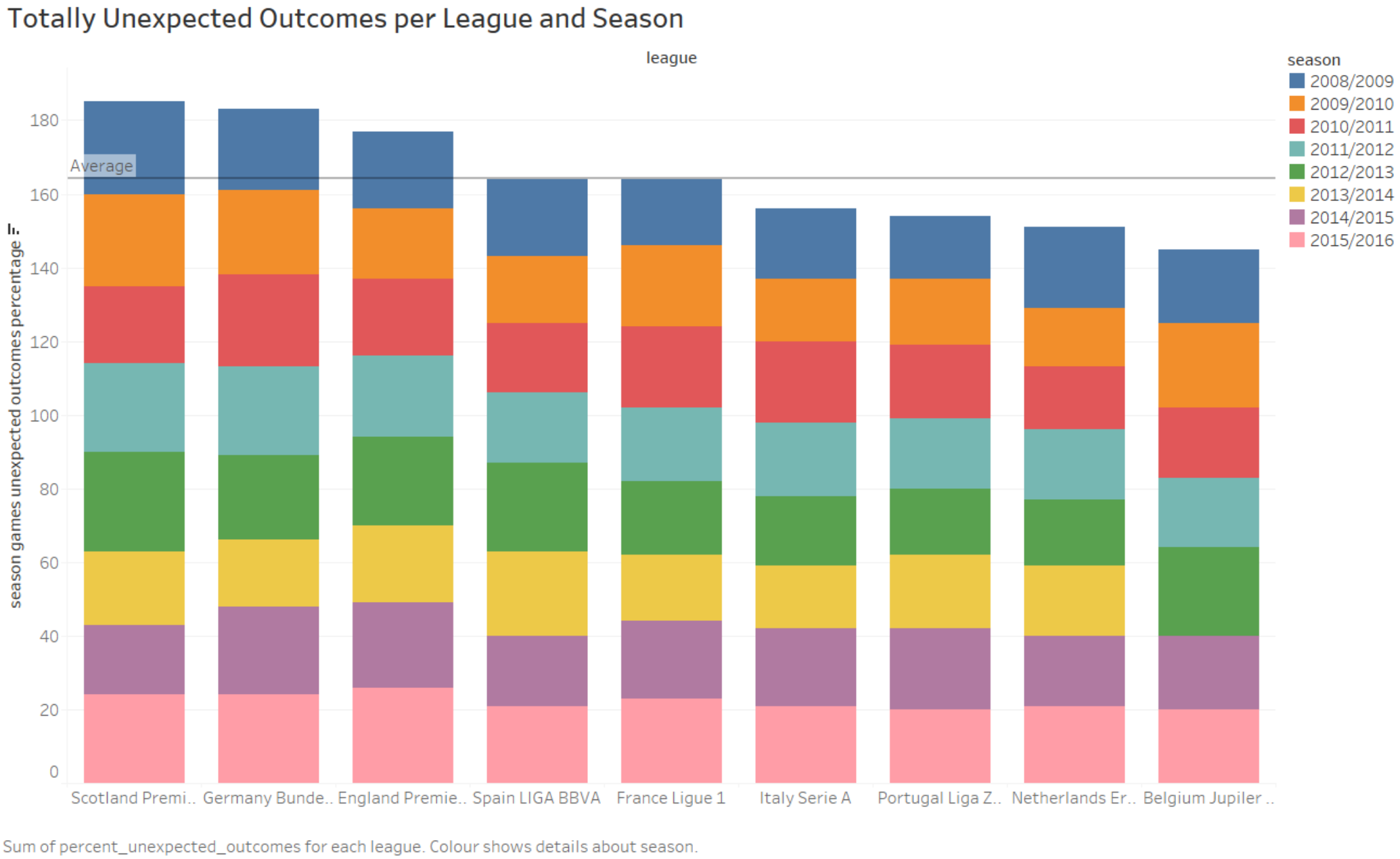
OLAP Report
(Drill-Through Action)



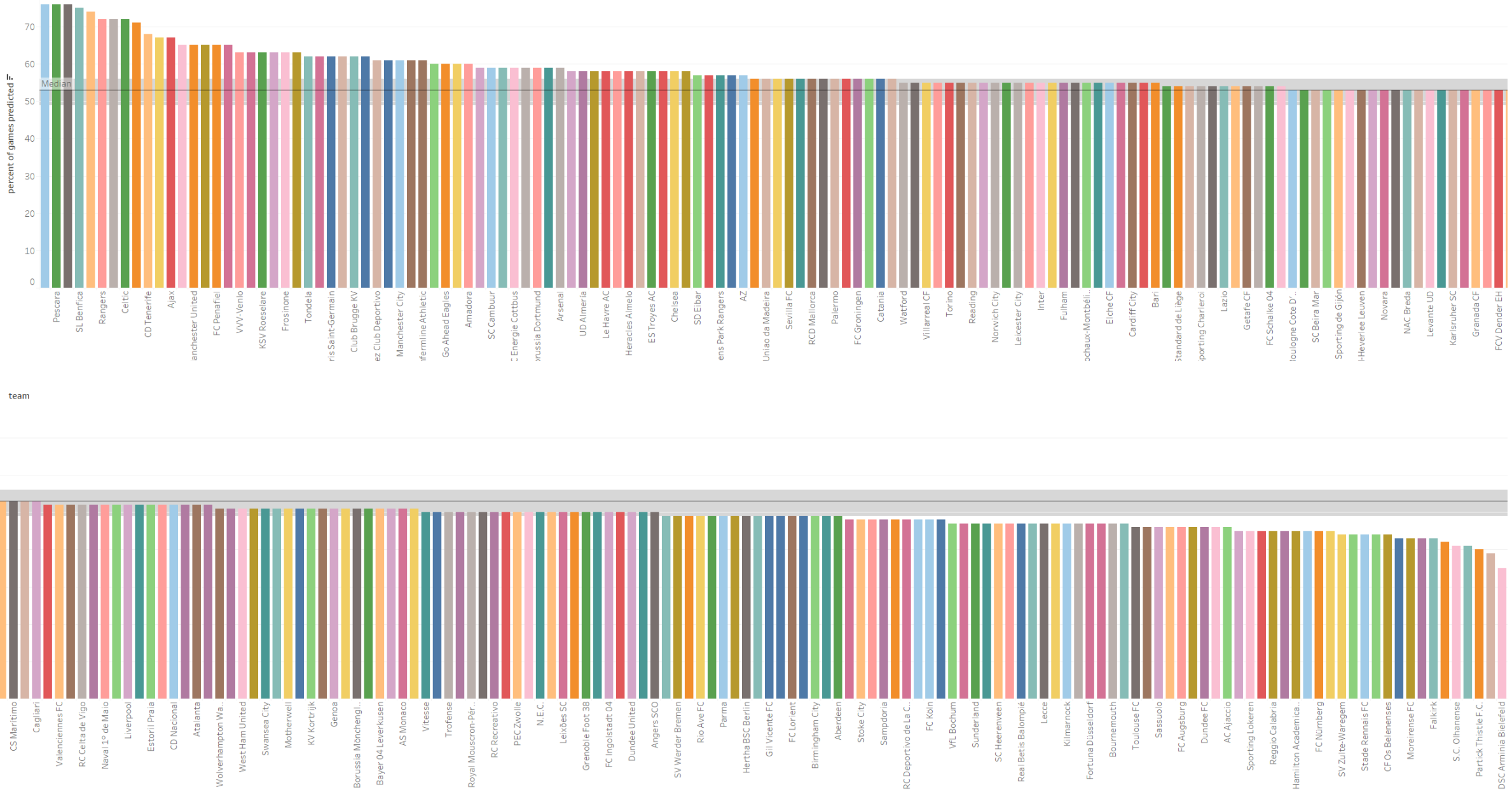
Visualizations



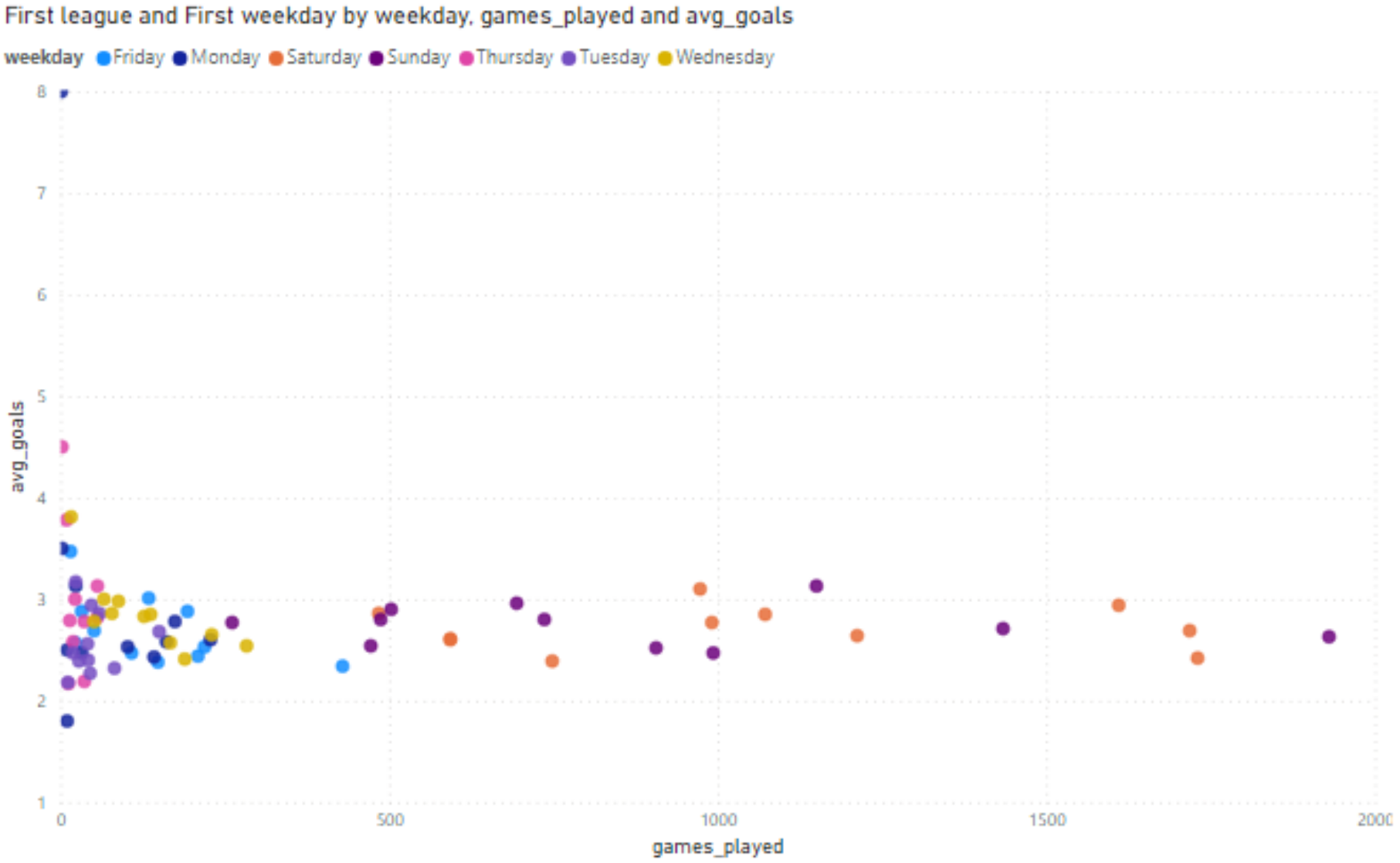
Visualizations (Outcomes Predicted per League and Season)



Visualizations (Unexpected Outcomes per League and Season)



Visualizations (Unexpected Outcomes per League and Season)



Visualizations (Results per Weekday and League)

Filters


Questions ???

Thank You

Gkouma Konstantina (f2822208)
Souflas Eleftherios-Efthymios (f2822217)