



# **FOOTBALL DATA VISUALIZATION**

**EXPLORING INSIGHTS FROM FOOTBALL DATASETS**

**Course: Data Visualization using Power BI**

**Lecturer: Chan Sophal**

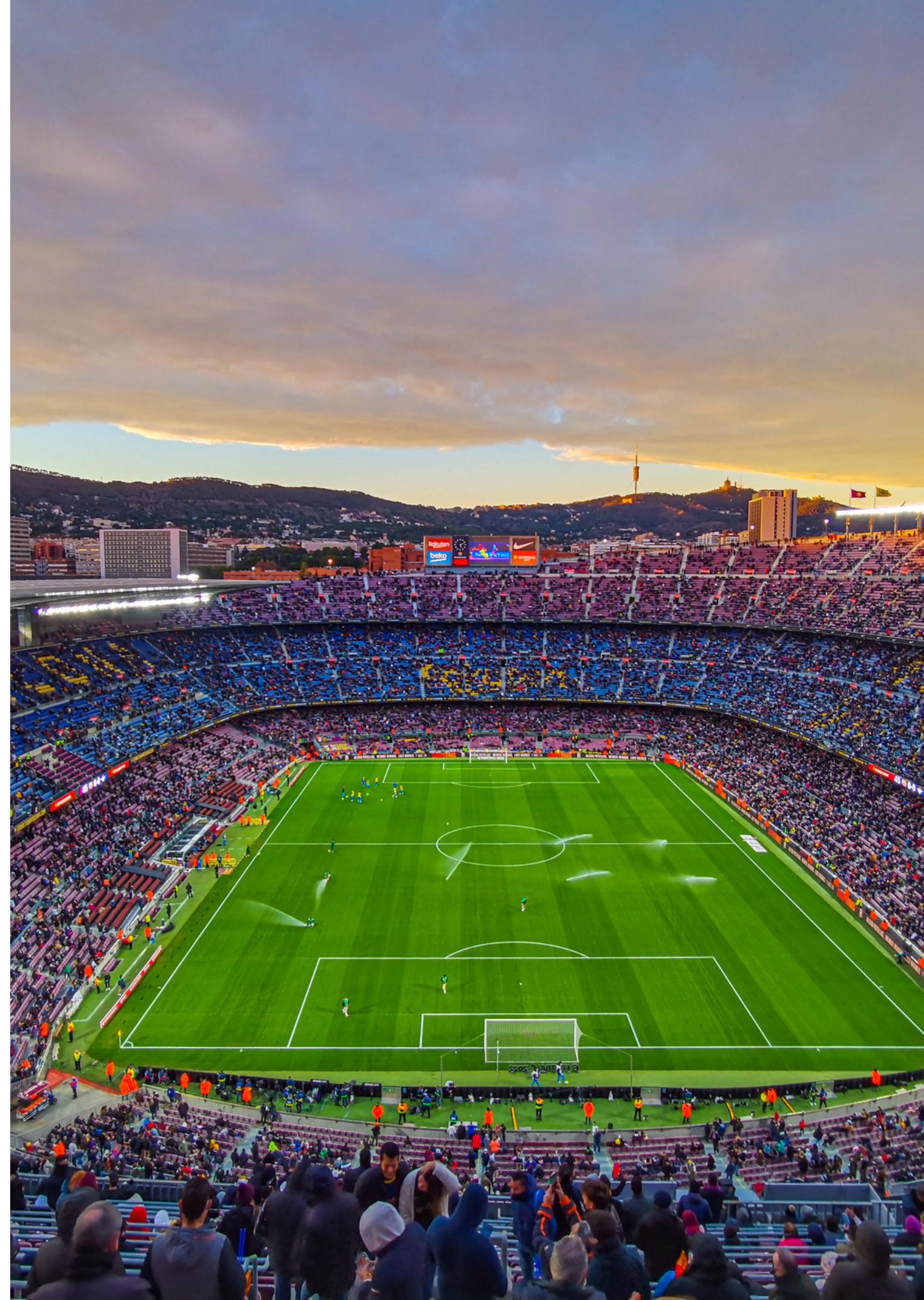
**Teacher Assistant: Thear Sophal**

**Date: 21st November 2023**

# Table of Content

---

- 1 INTRODUCTION
- 2 OBJECTIVES
- 3 DATA SOURCE
- 4 DATA CLEANING
- 5 DATA VISUALIZATION
- 6 SUMMARY INSIGHT
- 7 CONCLUSION



# Our Team

---



SO  
KIMLANG



SONGEAM  
KANHA



TEANG  
LAKSEY



SUN  
SOKSEKA



EAV  
SARIN



# Introduction

- The football dataset at hand is a comprehensive collection of match, player, and team-related data.
- It includes information on matches, countries, leagues, teams, players, and their respective attributes.
- The dataset spans multiple seasons, providing a rich source for in-depth analysis.

# Objectives

---



**Exploration of  
Football Insights**

**Enhance Betting  
Strategies**



**Inform Decision-  
Making**

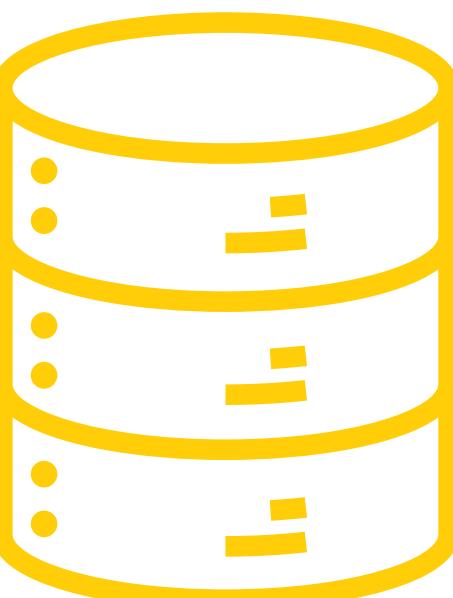
# Data Source



## EUROPEAN SOCCER DATABASE

- +25,000 matches
- +10,000 players
- Seasons 2008 to 2016
- Players and Teams' attributes\*
- Team line up with squad formation (X, Y coordinates)
- Betting odds from up to 10 providers
- Detailed match events (goal types, corner, cross, fouls, cards etc...) for +10,000 matches

# Dataset summary



**database.sqlite** (313.09 MB)

↓ >

Table	Total Rows	Total Columns
Country	11	2
League	11	3
Match	25979	115
Player	11060	7
Player_Attributes	183978	42
Team	299	5
Team_Attributes	1458	25



# Import Dataset Process

---

Using DB Browser (SQLite)

**DATASET IS FILE SQLITE** → **CONVERT TO FILE SQL**

**IMPORT FILE SQL TO SQL SERVER BY CREATE DATABASE**

**IN POWER BI: GET DATA** → **CHOOSE SQL SERVER**

→ **INPUT NAME OF SERVER AND DATABASE**



# Remove Duplicates

Identify and remove duplicate records to ensure each observation in the dataset is unique.

# Select Relevant Columns

Identify the columns in the dataset that are relevant to your analysis. Exclude unnecessary variables to simplify the dataset.



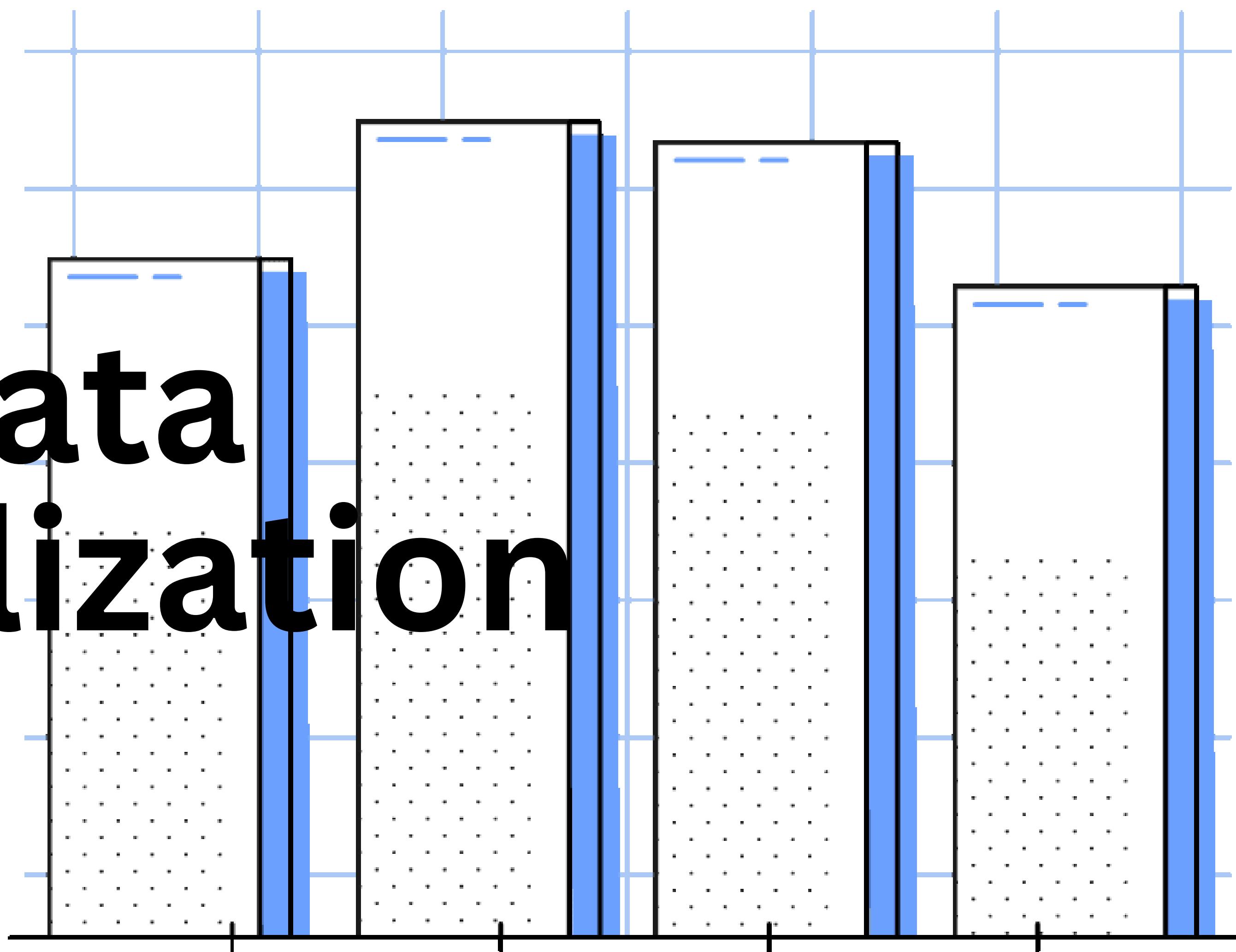
# Data Modeling

- **Structuring the Dataset:** structuring the football dataset for effective analysis, define the Dataset's Purpose
- **Defining Relationships:** Relationships were established between tables such as matches, players, teams, and attributes, enhancing data connectivity.

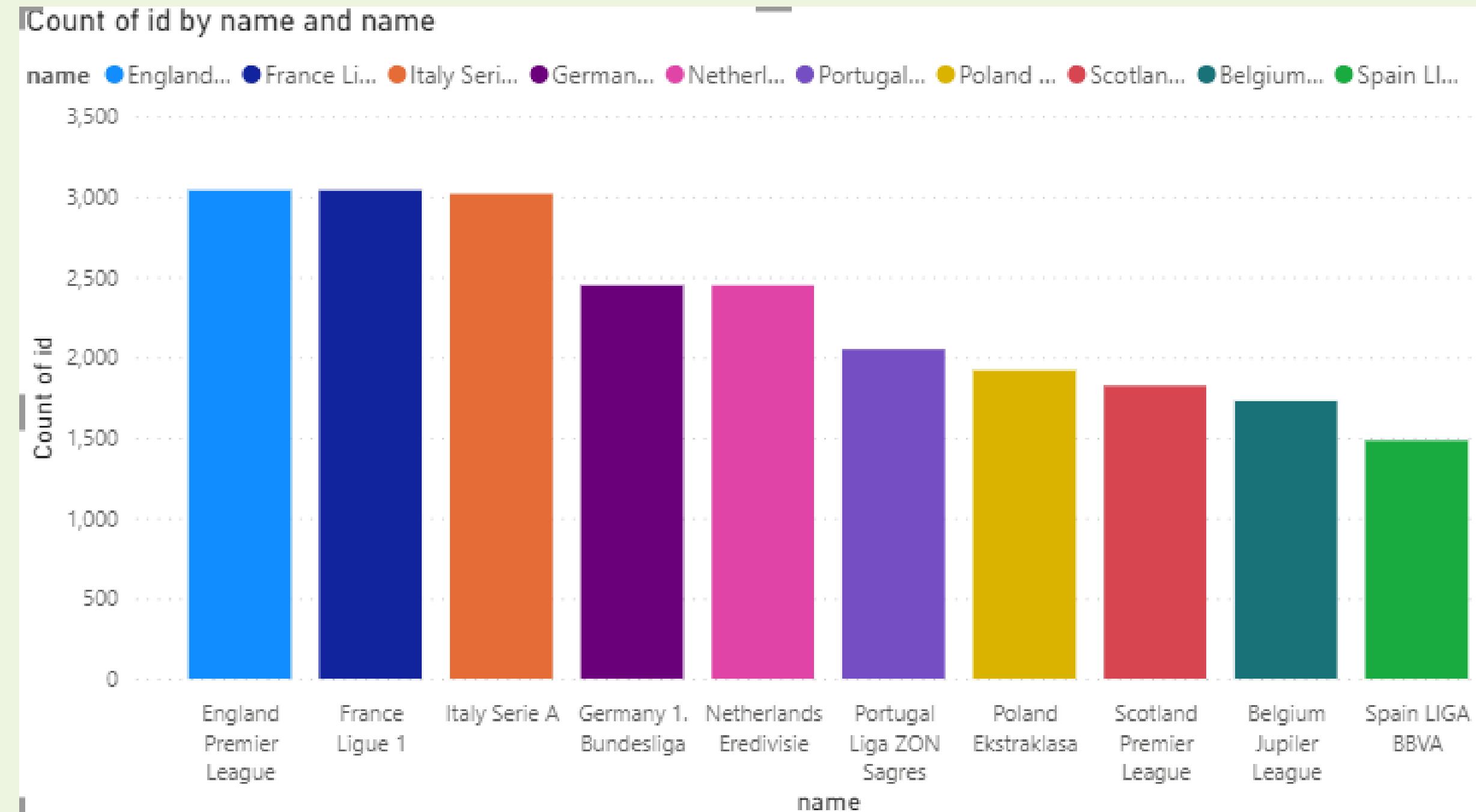
# Missing Value

Address missing values by dropping rows with missing value

# Data visualization



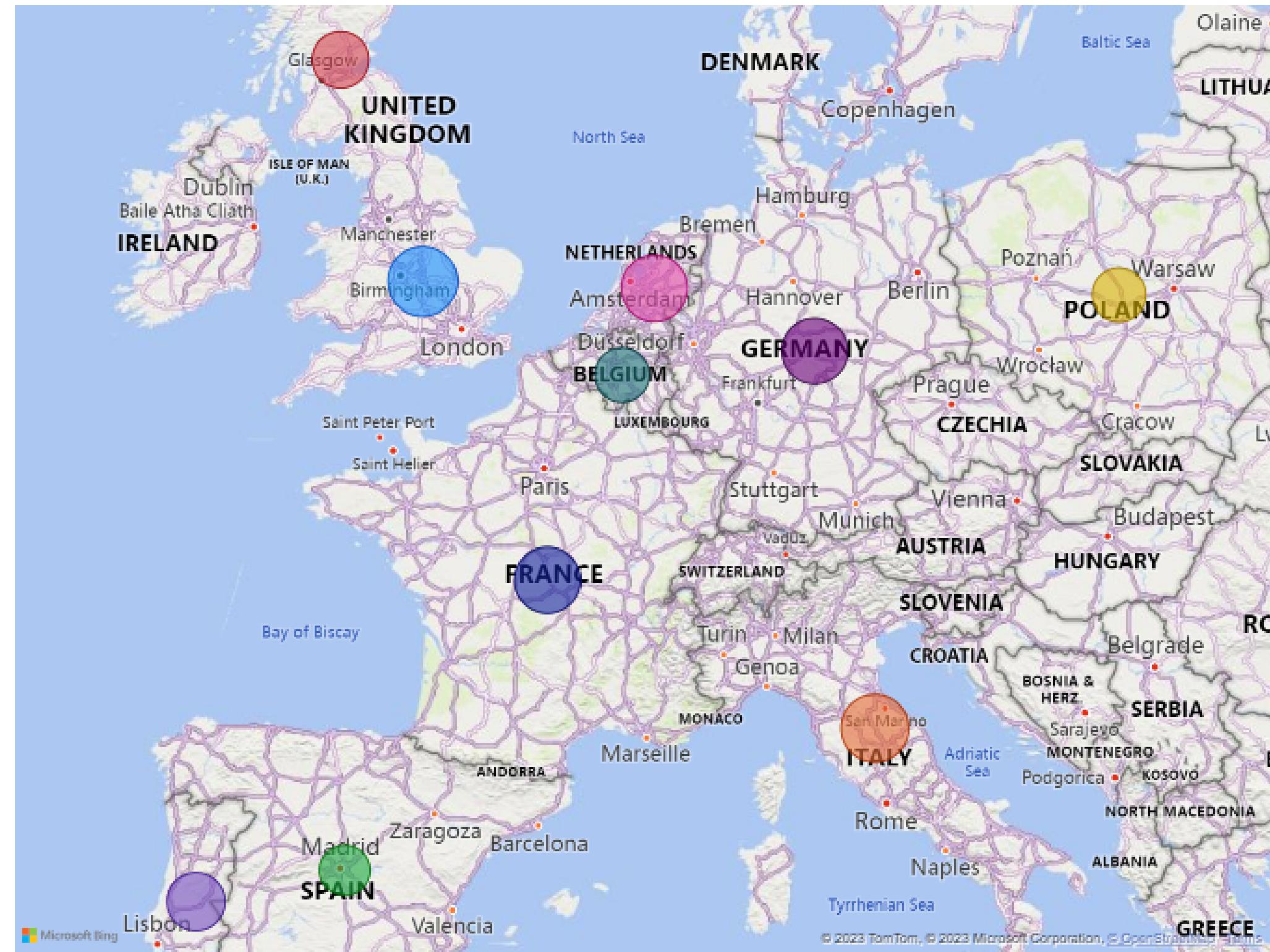
# Amount of match in each league



# Total goals in each league

## Amount of Total Goals in each League and Country

name ● Belgium J... ● England ... ● France Li... ● Germany ... ● Italy Seri... ● Netherl... ● Poland ... ● Portugal ... ● Scotlan... ● Spain Li...



## League Name

name  
Belgium Jupiler League  
England Premier League  
France Ligue 1  
Germany 1. Bundesliga  
Italy Serie A  
Netherlands Eredivisie  
Poland Ekstraklasa  
Portugal Liga ZON Sagres  
Scotland Premier League  
Spain LIGA BBVA

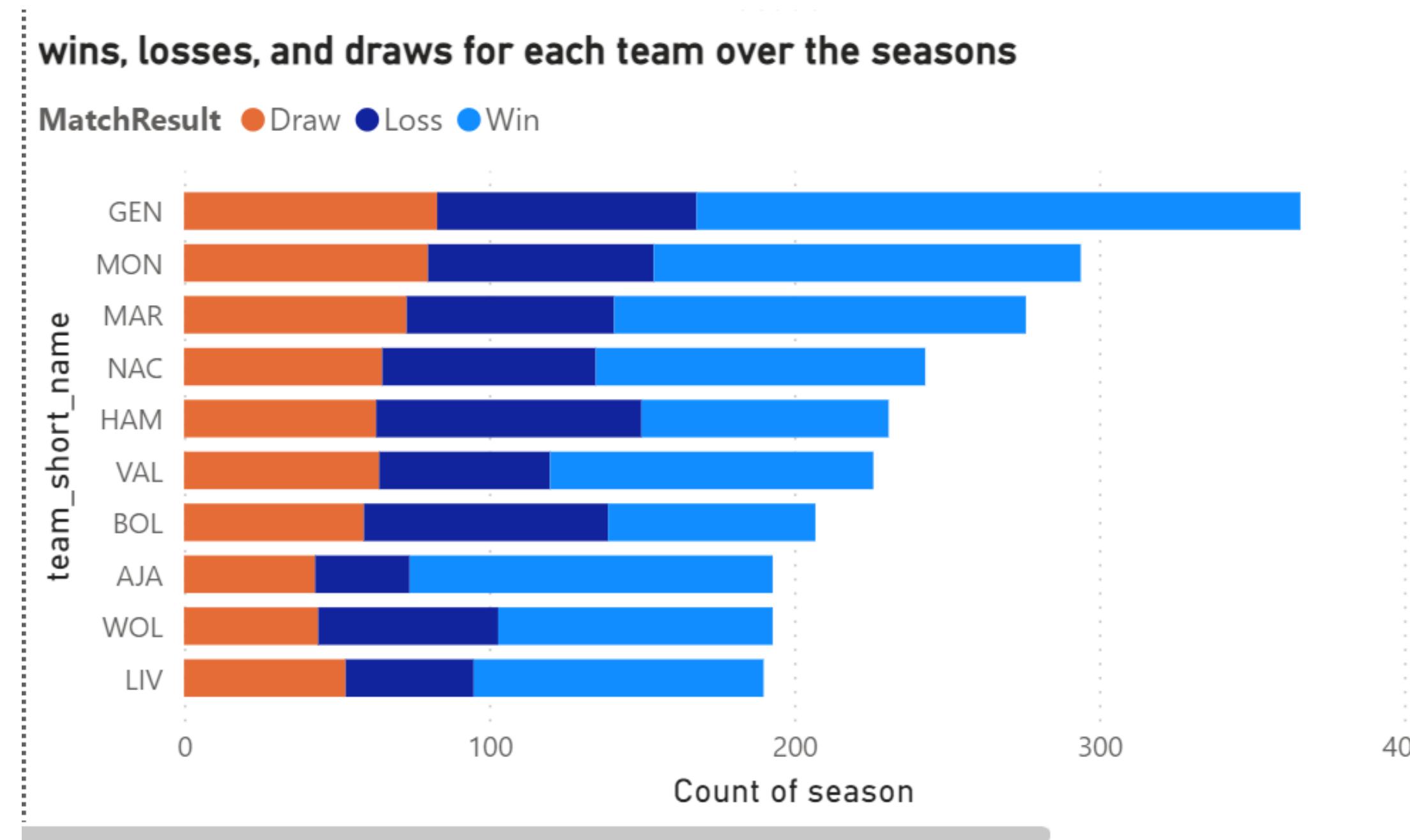
## Team Name

team\_long\_name  
1. FC Kaiserslautern  
1. FC Köln  
1. FC Nürnberg  
1. FSV Mainz 05  
Aberdeen  
AC Ajaccio  
AC Arles-Avignon  
AC Bellinzona  
Académica de Coimbra  
ADO Den Haag  
AJ Auxerre

# Match Result

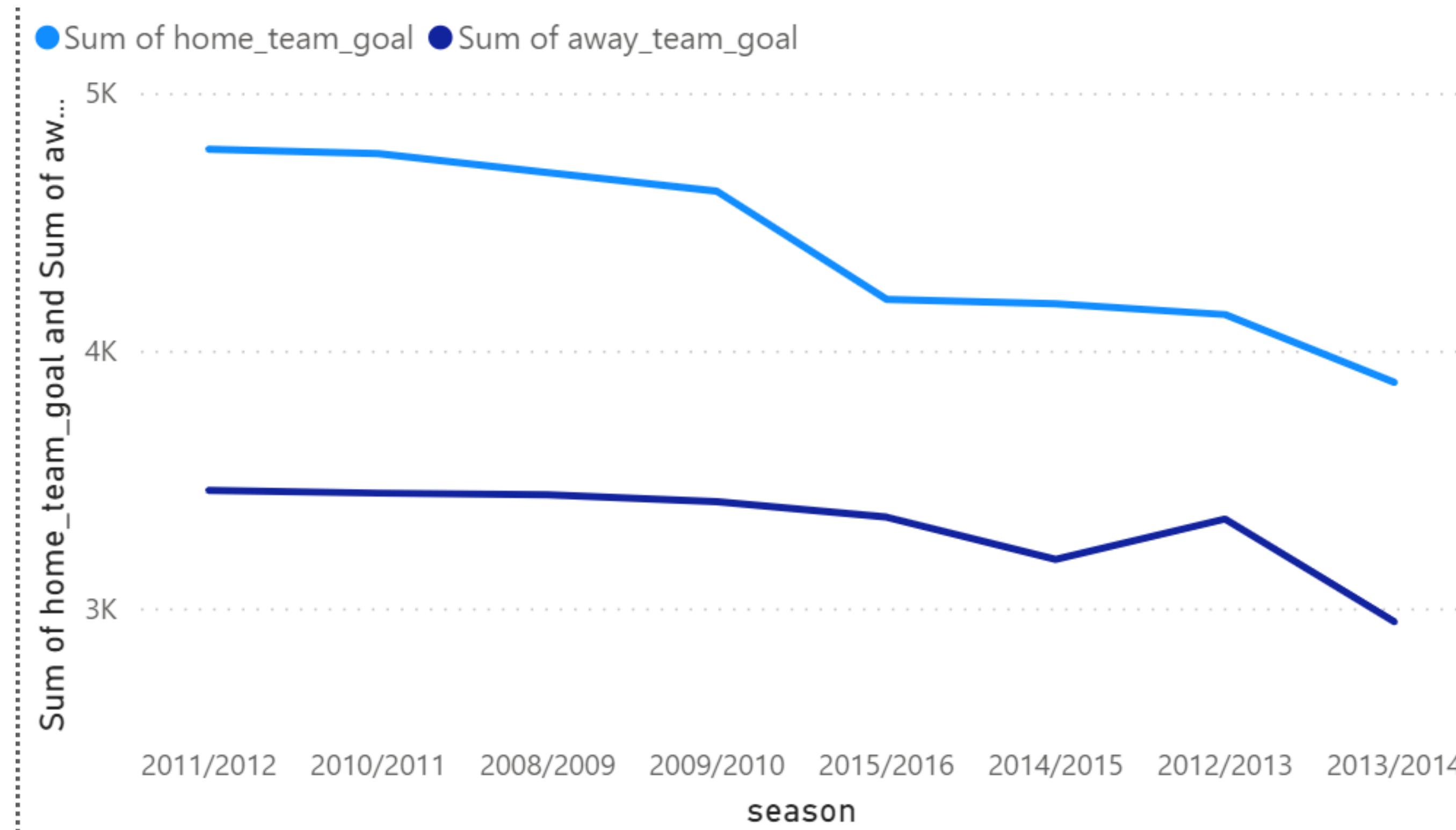
MatchResult =

```
IF('match'[home_team_goal] > 'match'[away_team_goal], "Win",
IF('match'[home_team_goal] < 'match'[away_team_goal], "Loss", "Draw"))
```

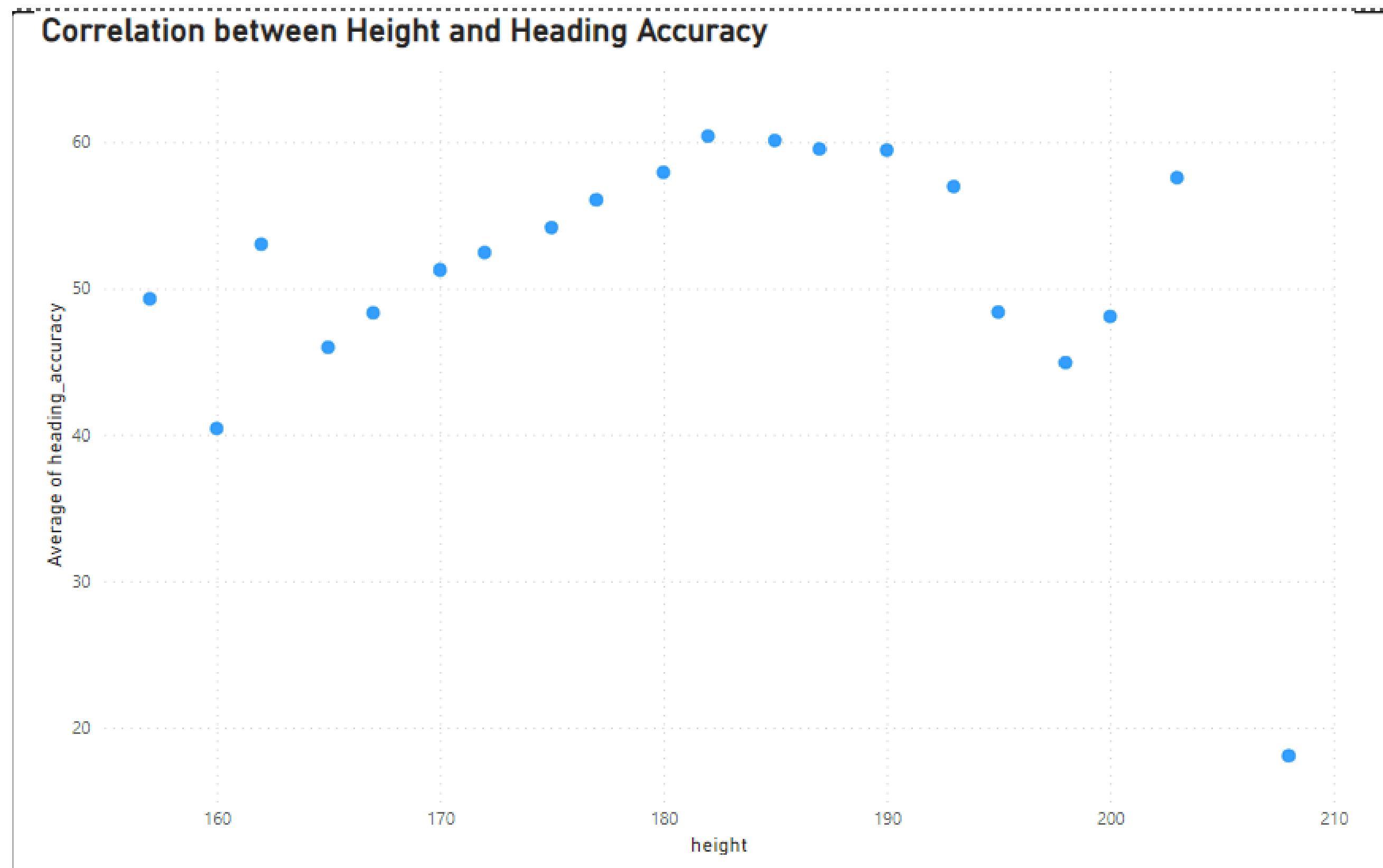


# Home Team goals VS Away Team Goals in each season

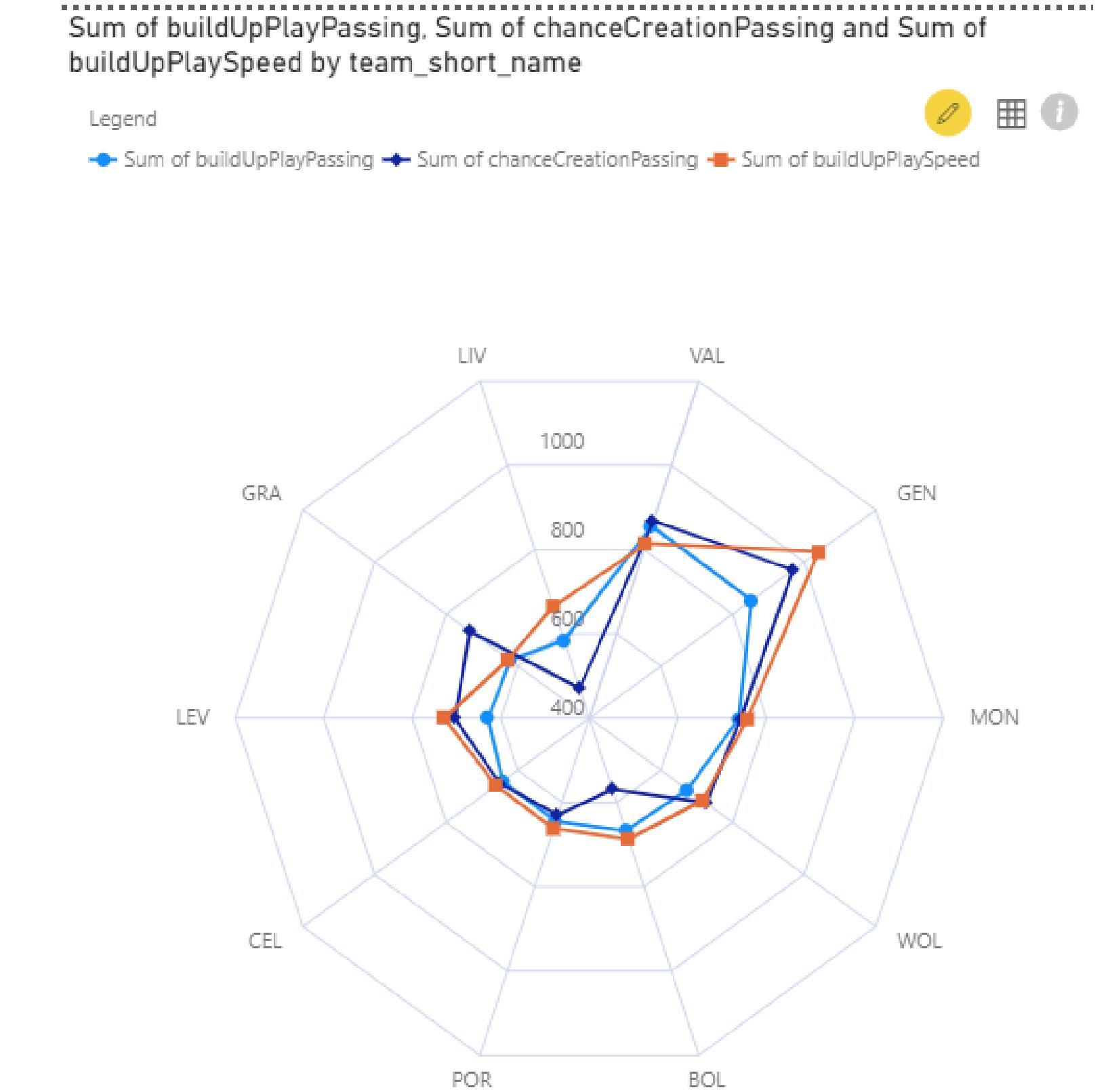
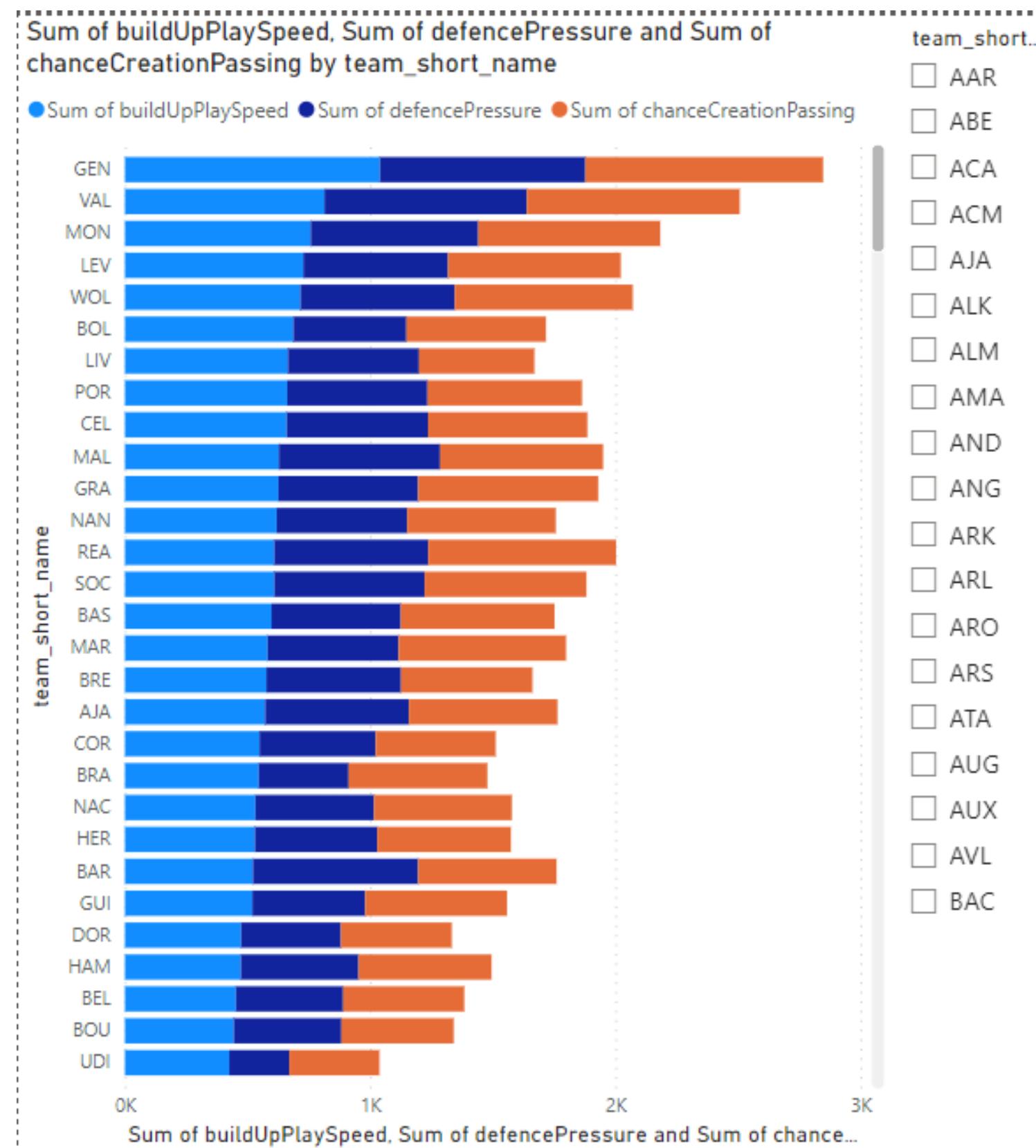
---



# Correlation between Height and Heading Accuracy



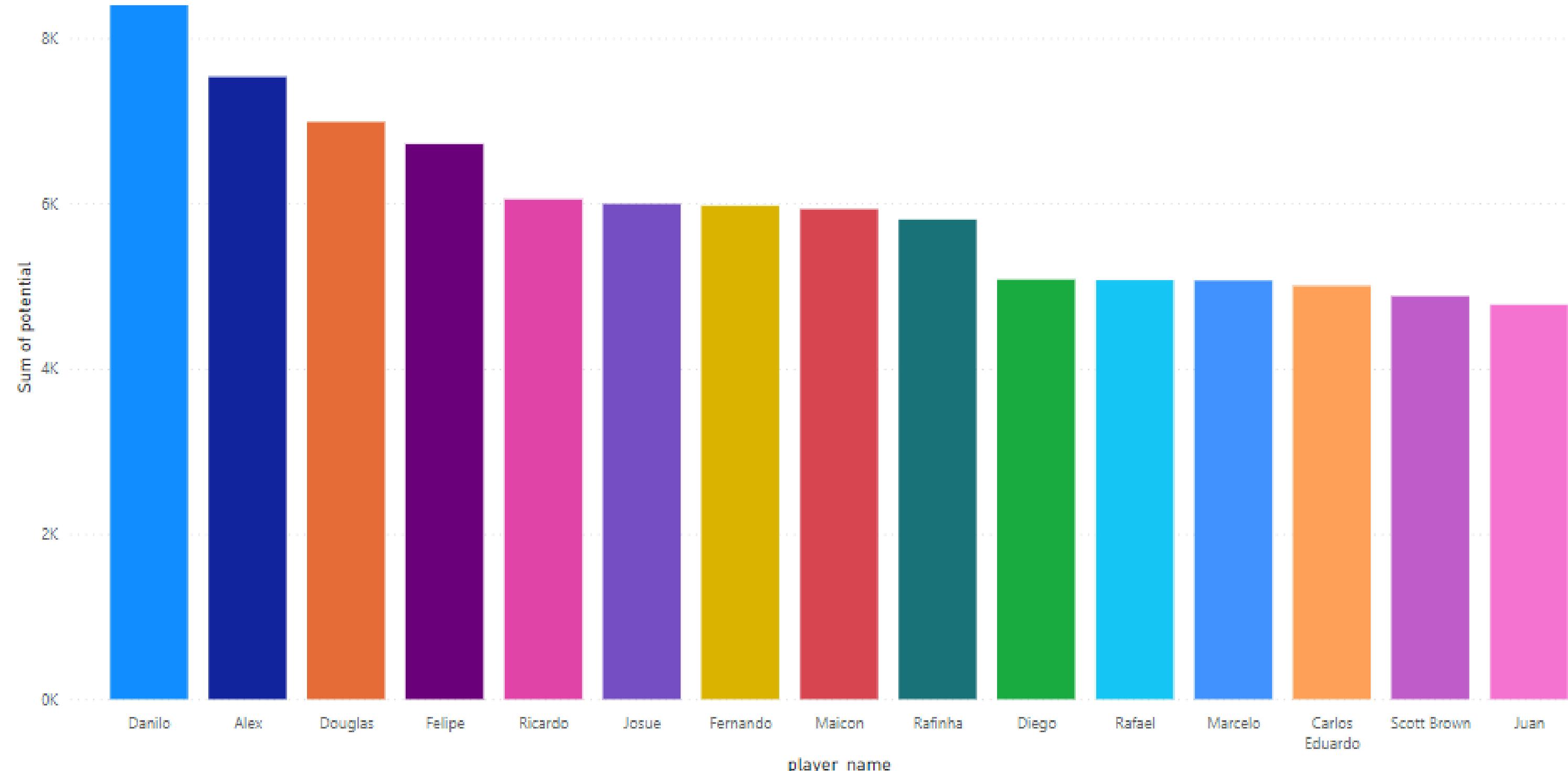
# Each team performances on build up play passing, defense pressure and play speed.



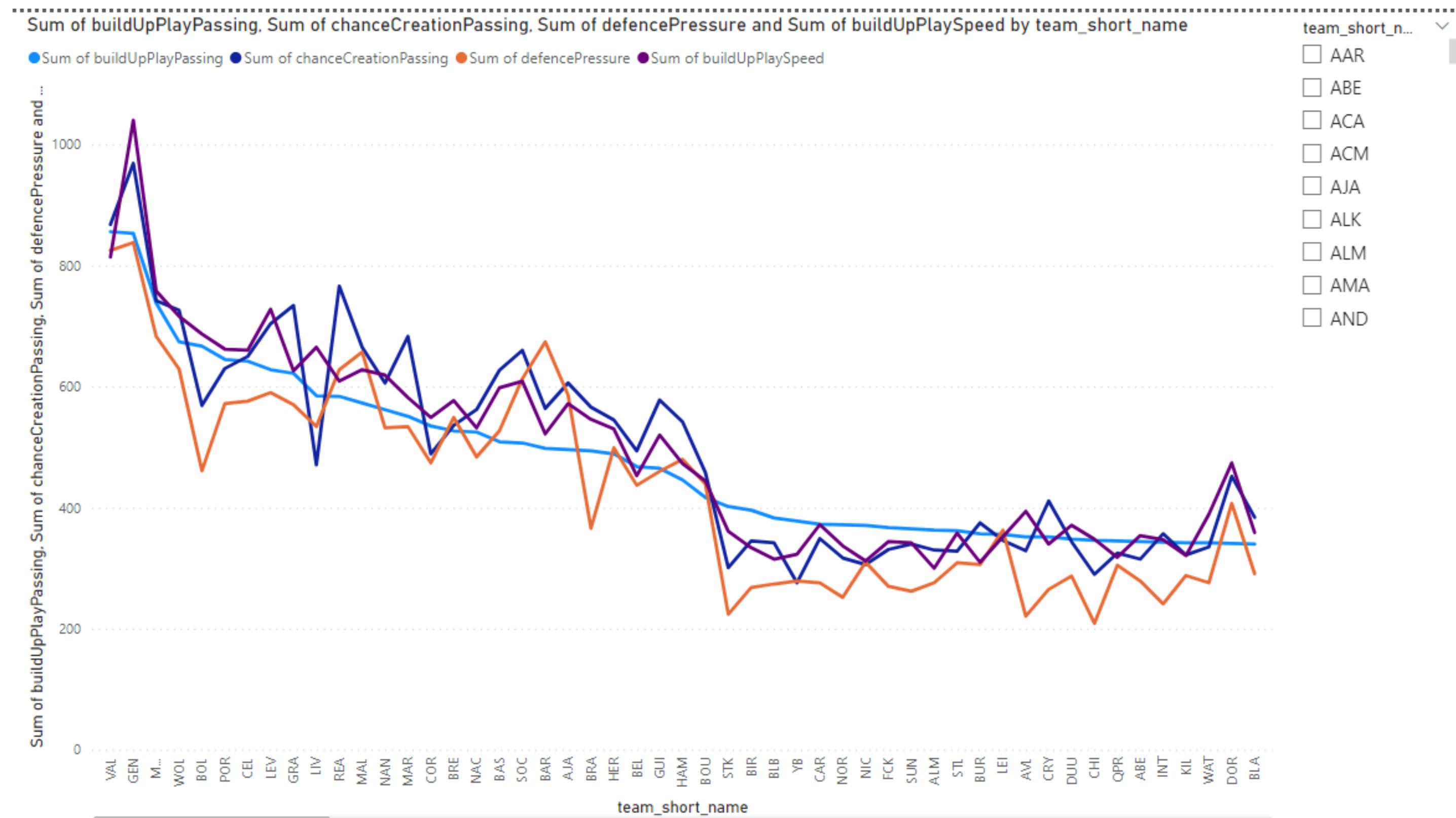
# Potential of each player by league

Sum of potential by player\_name and player\_name

player\_name ● Danilo ● Alex ● Douglas ● Felipe ● Ricardo ● Josue ● Fernando ● Maicon ● Rafinha ● Diego ● Rafael ● Marcelo ● Carlos Eduardo ● Scott Brown ● Juan



# Each team performances on build up play passing, chance creation passing, defense pressure and play speed.



# Conclusions

## Summary Insights

- Predict the result of the match (Bet)
- Top player that has the highest player attributes and characteristics is **Danilo**
- Top team that has the highest team attributes and characteristics is **GEN**
- The correlation that effect the accuracy of the team player to win the match
- Top team and country has the most Winning Result Match is **GEN**

# THANK YOU.

TOTAL

50000

30000

10000

LOADING...

