

# **Exploratory Data Analysis**

**EPL Dataset from 2018-2019**



# Today's Agenda

## Presentation Outline

Background

Data Understanding

Objectives

Exploration



# Background

**Exploratory Data Analysis from dataset  
English Premier League 2018-2019**

We want to explore dataset to get insight and answer the objectives which we discussed and agreed before.



# Data Understanding

Index: 20 entries, Manchester City to Huddersfield  
Data columns (total 43 columns)

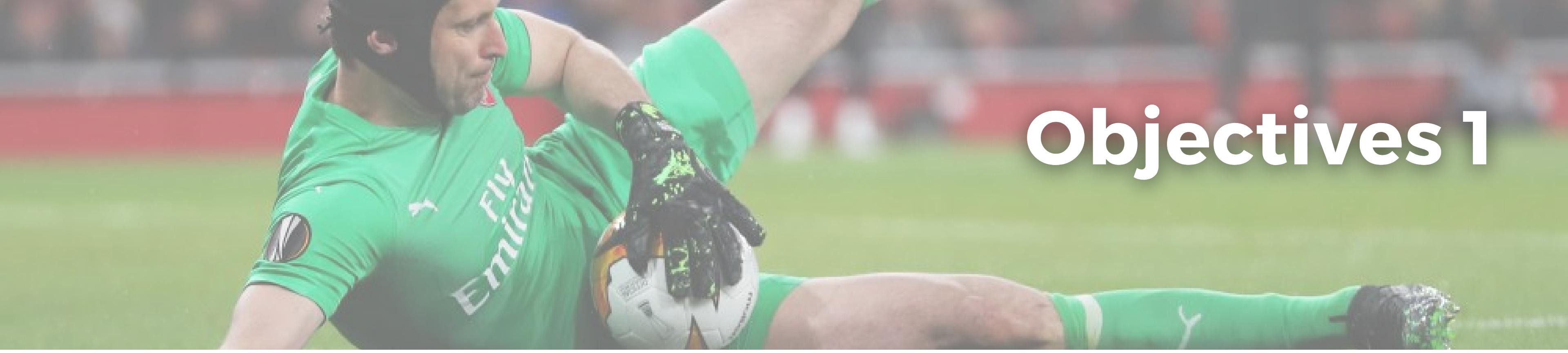
dtypes: float64(3), int64(35), object(5)

This dataset contains 20 teams with 43 columns of which 35 are integer type, 3 column is float type (38 numeric) and 5 column is object type.

# Objectives

The question that we need to solve to get insight

- 1. IS THIS DATA CLEAN?**
- 2. HOW IS THE POINT DISTRIBUTION OF THE EPL TEAM? AND WHICH TEAM IS AN ANOMALY? BEST ATTACK TEAM**
- 3. BEST ATTACK TEAM**
- 4. BEST DEFENCE TEAM**
- 5. TEAM WITH GOOD FINANCIAL ASPECT**
- 6. EXPLORE THE DATASET**



# Objectives 1

## Change type of data

From the dataset info, there are 4 datas that should not be of type object. We have to change it to integer same as another same variable.

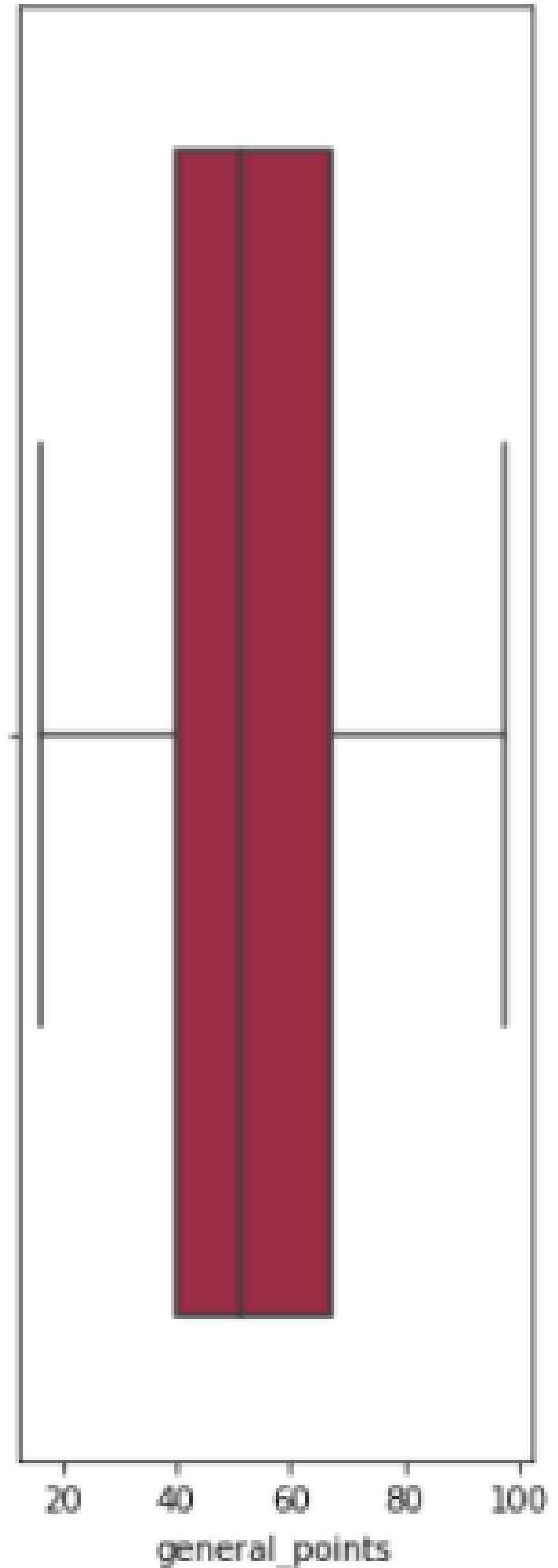
## Change Column Name

We have to renaming column without space, to better processing for modelling.

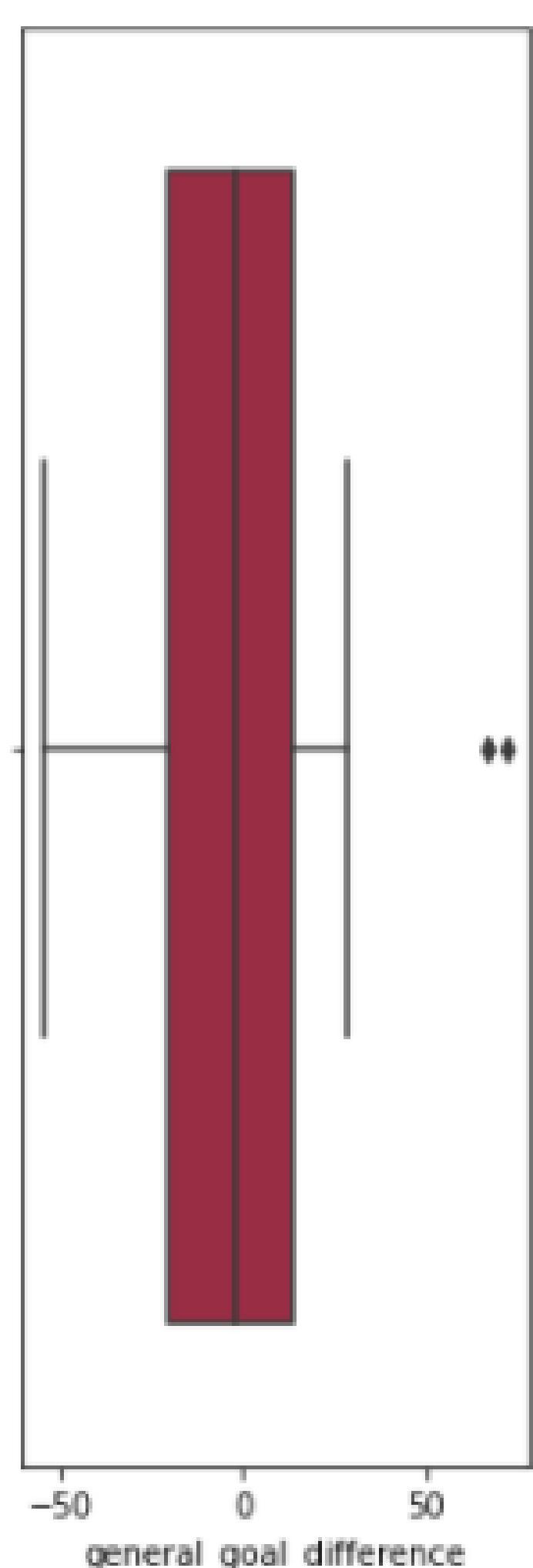
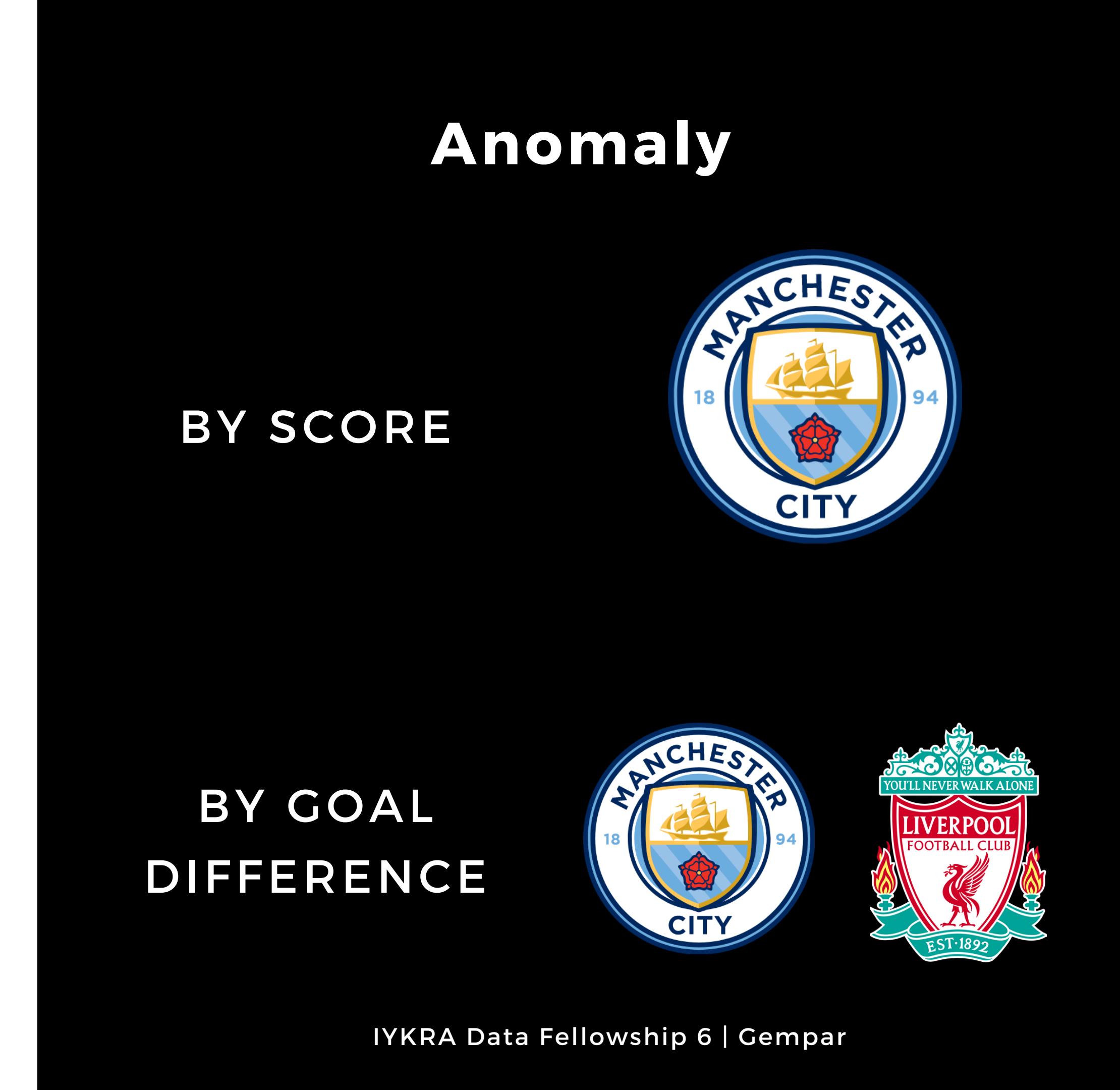
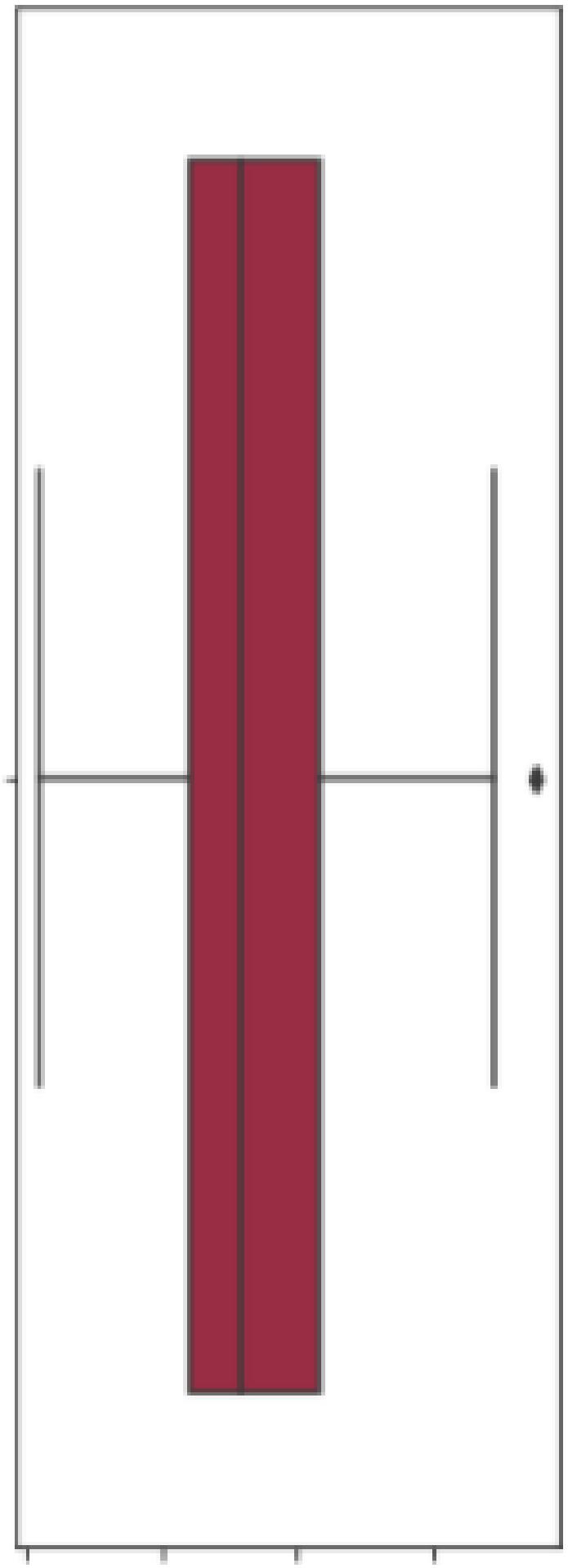
Ex : "finance \_team\_market"  
to "finance\_team\_market"

# Objectives 2

Point distribution of Team



We see that there are no anomalies in the point distribution data because there are no outliers. but if we dig deeper, there are some anomalies in this data..





# Objectives 3

## Best Attack Team

For the best attack team, I made a 2 new variable.

1. Percentage from shot on target and overall shot
2. Real goal from overall goal minus penalty and freekick



## Best precision shot

MU has 42% precision for shooting to rival.

MU has 225 on target from 526 shoot.



## Best score

ManCity has total 95 goal from 3 penalty, 2 freekick and 90 regular goal.

# Objectives 4



## Best Defend Team

For the best defend team, I made a 2 new variable.

1. Calculate from all defend variabel (saves, block, tackle, clearances)
2. Real Conceded from overall conceded minus penalty



## Best overall defend score

Newcastle has 2668 defend points.

Interestingly, ManCity has the **lowest** defend value with 1496 points



## Fewest Goal Conceded

ManCity has total 19 conceded because their **time is mostly on attack possesion** (their value is 64%).



Watford got revenue \$11.389.553 per match

# Objectives 5

## Best Financial Team

Based on incom  
per live match

We think the best financial is  
the more revenue income, is  
better for the team.

This variabel

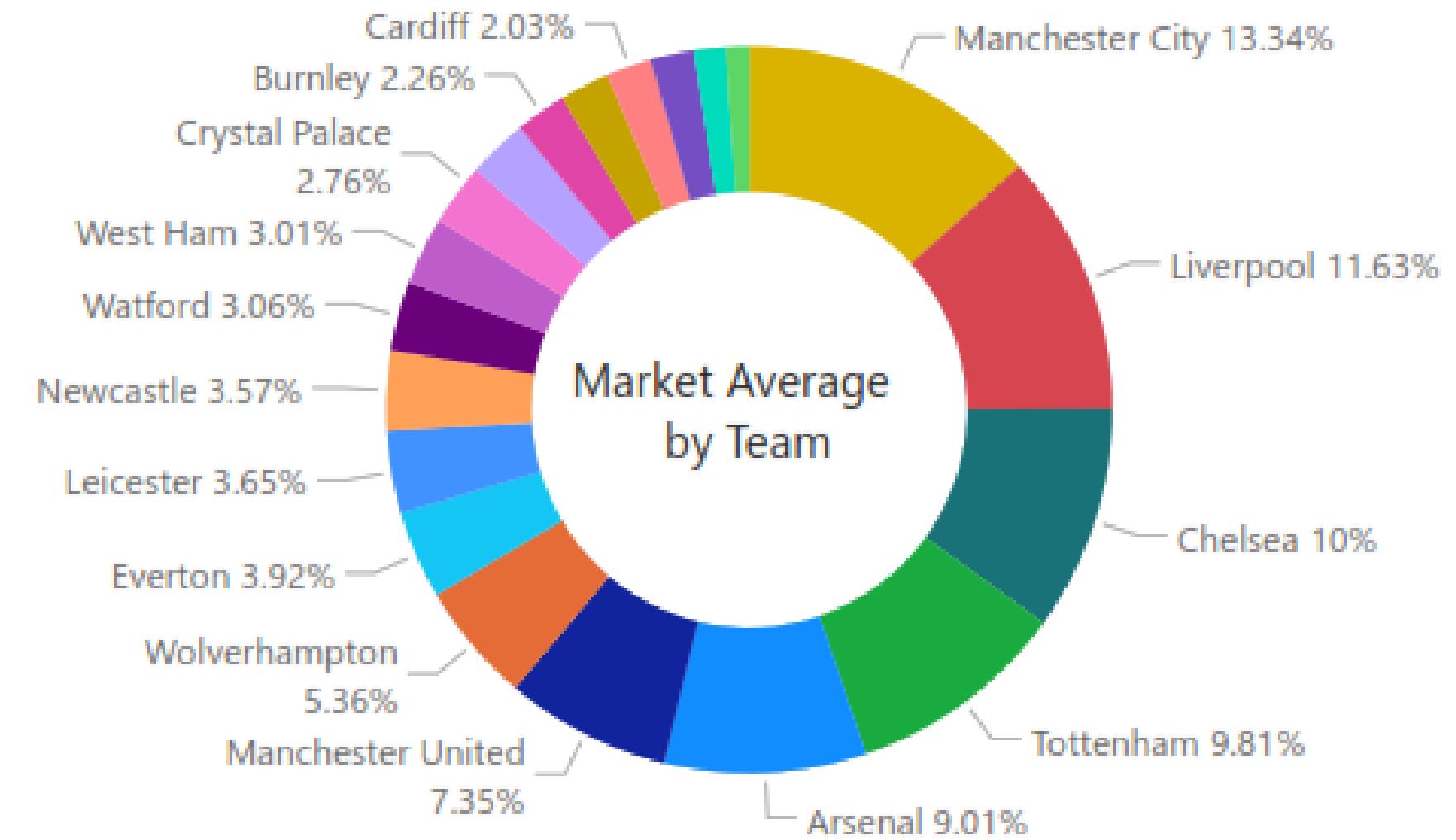
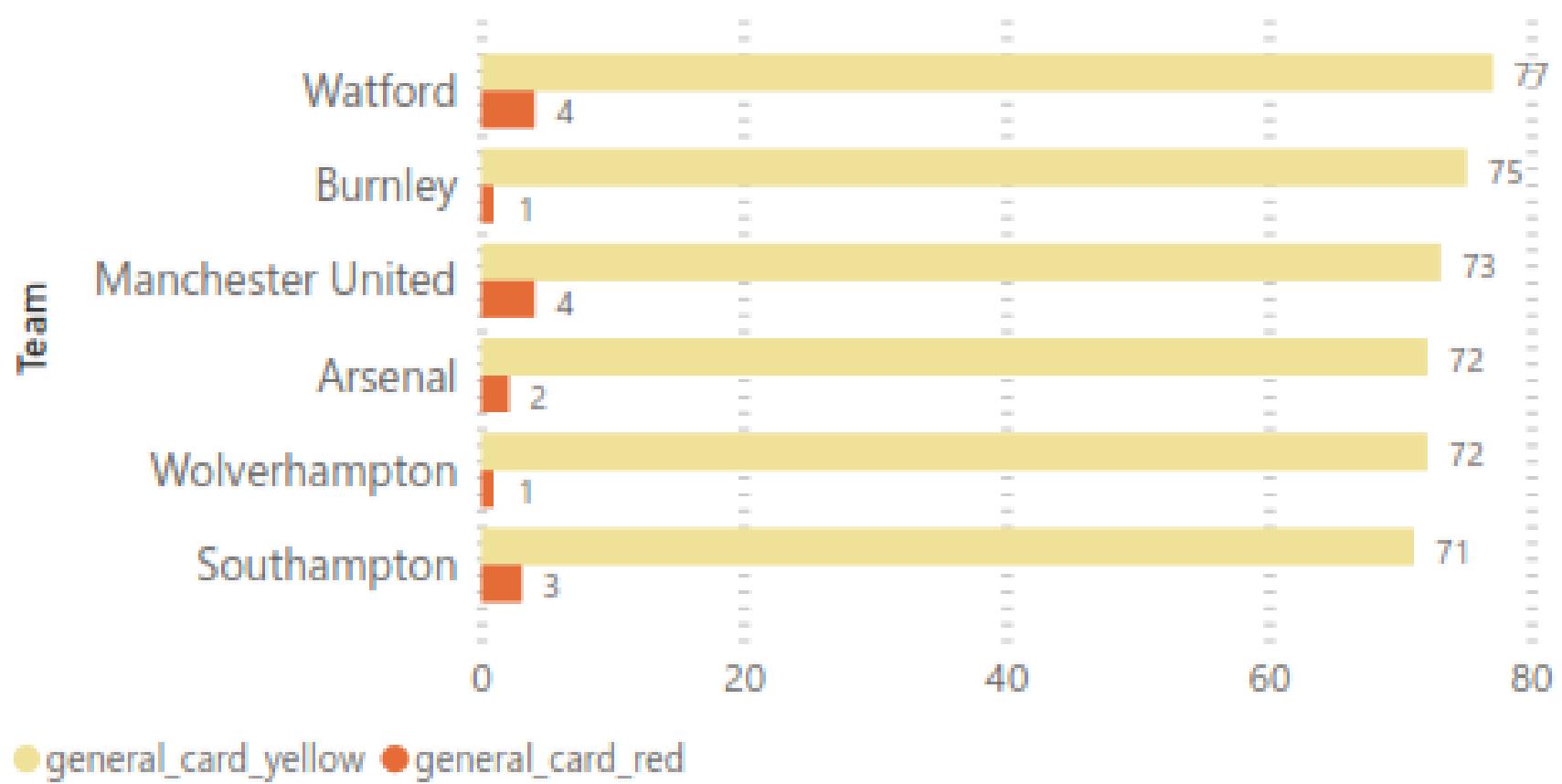
We get this variable from the  
division of total TV revenue by  
the number of matches  
broadcast.



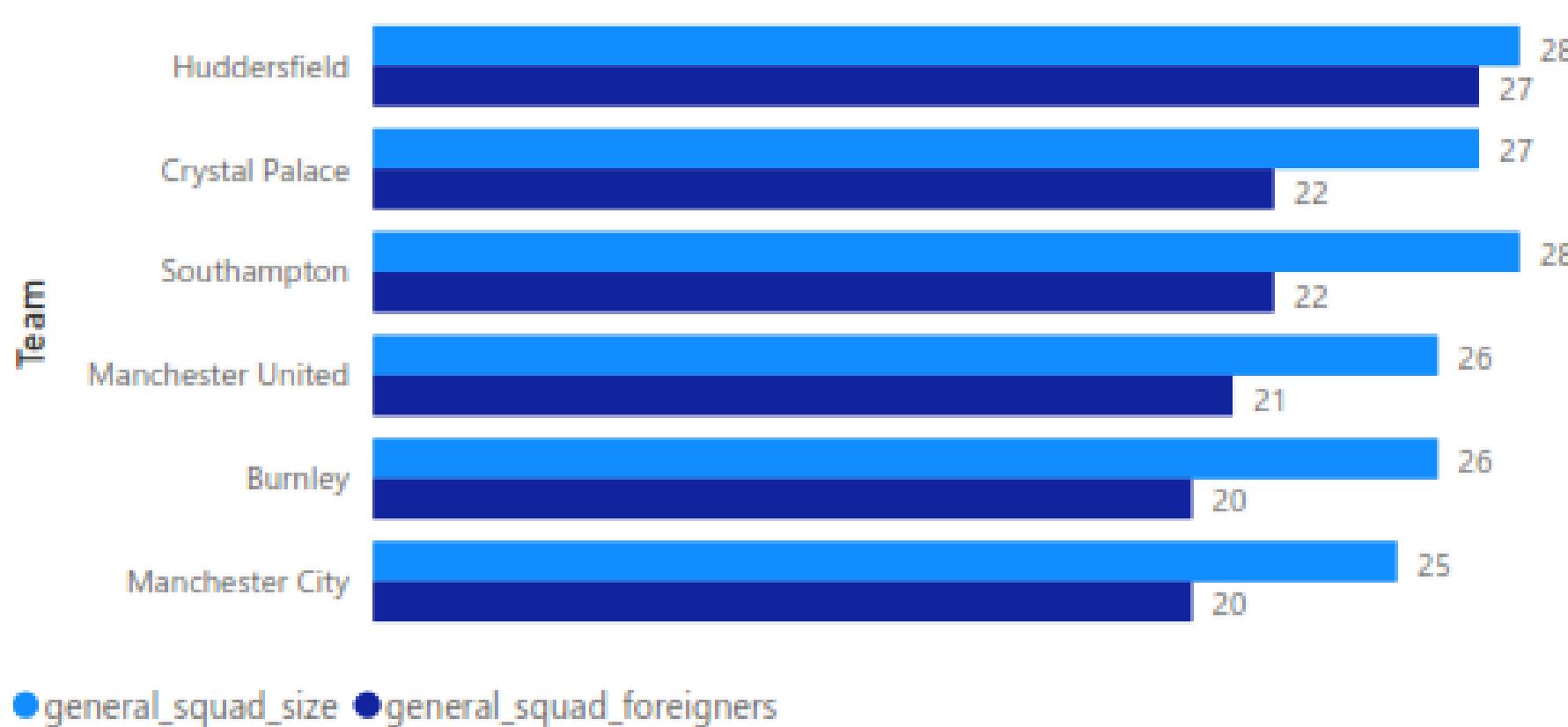
# OBJECTIVES 6

## DATA VISUALIZATION

## The Dirties Play by Team



## Most foreigners in squad



## Correlation between Goal and Shot

