

# TOP PLAYER RANKINGS IN THE PREMIER LEAGUE

ANGEL GUILLERMO LOPEZ DELGADO  
ALBERTO CARLOS NAVARRETE GARCIA

PLAY



The background of the slide features a vibrant green grass texture at the top and bottom, with a central area of white paper that has a torn, deckled edge effect.

# **INTRODUCTION**

Objective: To predict whether a player is a "Top Player" based on their performance (goals, assists, and other metrics).

Dataset: Premier League player statistics 2024-2025.





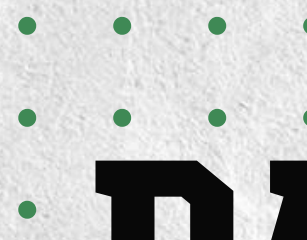
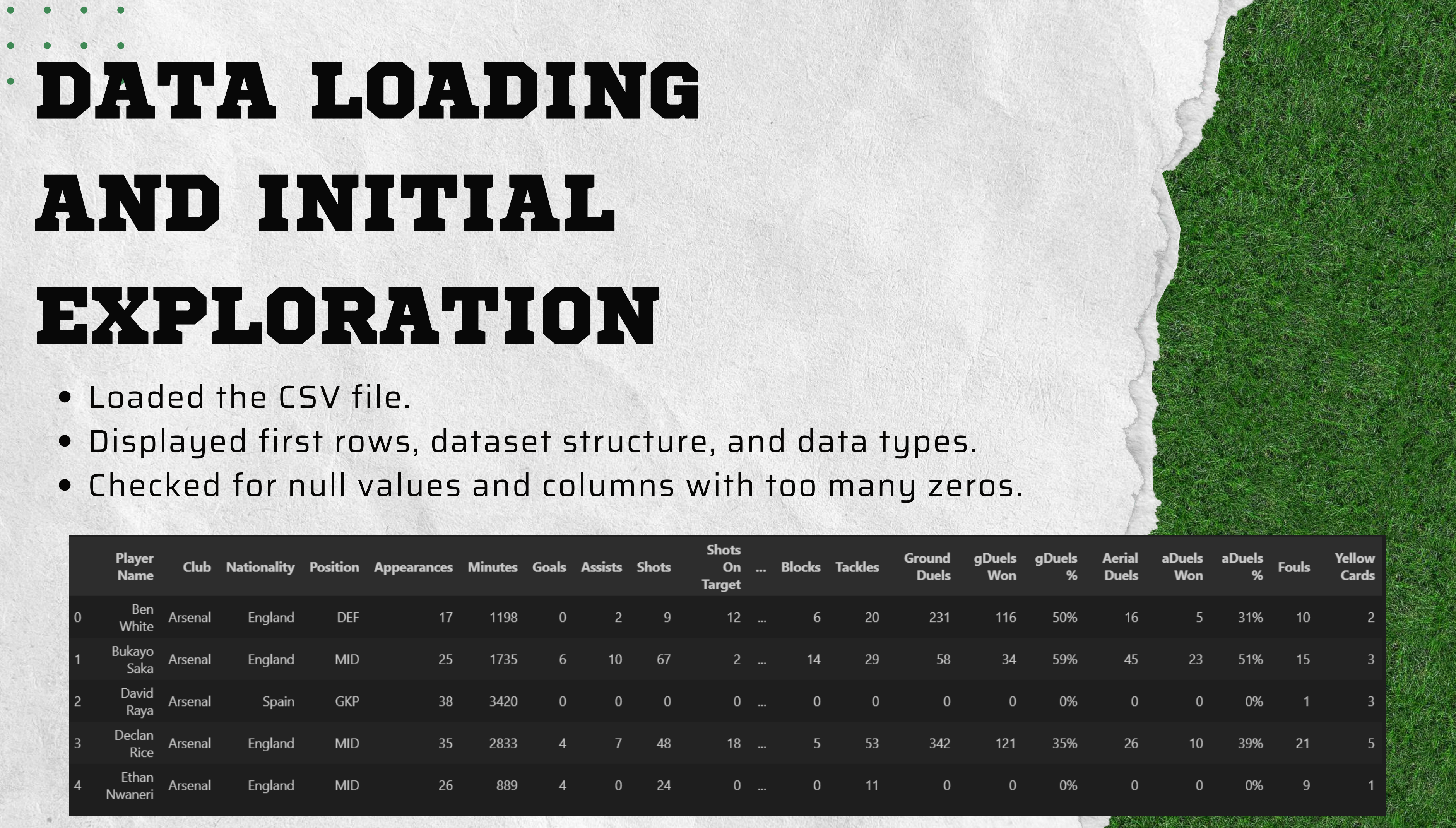
# LIBRARY IMPORT

- Pandas, Numpy: Data manipulation.
- Matplotlib, Seaborn: Data visualization.
- Scikit-learn: Data preprocessing, model building, and evaluation metrics.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import (
    confusion_matrix, ConfusionMatrixDisplay,
    accuracy_score, precision_score, recall_score, f1_score
)
```

✓ 33.2s





# DATA LOADING AND INITIAL EXPLORATION

- Loaded the CSV file.
- Displayed first rows, dataset structure, and data types.
- Checked for null values and columns with too many zeros.

|   | Player Name   | Club    | Nationality | Position | Appearances | Minutes | Goals | Assists | Shots | Shots On Target | ... | Blocks | Tackles | Ground Duels | gDuels Won | gDuels % | Aerial Duels | aDuels Won | aDuels % | Fouls | Yellow Cards |
|---|---------------|---------|-------------|----------|-------------|---------|-------|---------|-------|-----------------|-----|--------|---------|--------------|------------|----------|--------------|------------|----------|-------|--------------|
| 0 | Ben White     | Arsenal | England     | DEF      | 17          | 1198    | 0     | 2       | 9     | 12              | ... | 6      | 20      | 231          | 116        | 50%      | 16           | 5          | 31%      | 10    | 2            |
| 1 | Bukayo Saka   | Arsenal | England     | MID      | 25          | 1735    | 6     | 10      | 67    | 2               | ... | 14     | 29      | 58           | 34         | 59%      | 45           | 23         | 51%      | 15    | 3            |
| 2 | David Raya    | Arsenal | Spain       | GKP      | 38          | 3420    | 0     | 0       | 0     | 0               | ... | 0      | 0       | 0            | 0          | 0%       | 0            | 0          | 0%       | 1     | 3            |
| 3 | Declan Rice   | Arsenal | England     | MID      | 35          | 2833    | 4     | 7       | 48    | 18              | ... | 5      | 53      | 342          | 121        | 35%      | 26           | 10         | 39%      | 21    | 5            |
| 4 | Ethan Nwaneri | Arsenal | England     | MID      | 26          | 889     | 4     | 0       | 24    | 0               | ... | 0      | 11      | 0            | 0          | 0%       | 0            | 0          | 0%       | 9     | 1            |



Loads the original dataset  
(epl\_player\_stats\_24\_25.csv) and displays  
the first rows for a quick overview of the  
contents.

```
df = pd.read_csv('epl_player_stats_24_25.csv')
df.head()
```

|   | Player Name   | Club    | Nationality | Position | Appearances | Minutes | Goals | Assists | Shots | Shots On Target | ... | Fouls | Yellow Cards | Red Cards | Saves | Saves % | Penalties Saved | Clearances Off Line | Punches | High Claims | Goals Prevented |
|---|---------------|---------|-------------|----------|-------------|---------|-------|---------|-------|-----------------|-----|-------|--------------|-----------|-------|---------|-----------------|---------------------|---------|-------------|-----------------|
| 0 | Ben White     | Arsenal | England     | DEF      | 17          | 1198    | 0     | 2       | 9     | 12              | ... | 10    | 2            | 0         | 0     | 0%      | 0               | 0                   | 0       | 0           | 0.0             |
| 1 | Bukayo Saka   | Arsenal | England     | MID      | 25          | 1735    | 6     | 10      | 67    | 2               | ... | 15    | 3            | 0         | 0     | 0%      | 0               | 0                   | 0       | 0           | 0.0             |
| 2 | David Raya    | Arsenal | Spain       | GKP      | 38          | 3420    | 0     | 0       | 0     | 0               | ... | 1     | 3            | 0         | 86    | 72%     | 0               | 0                   | 8       | 53          | 2.1             |
| 3 | Declan Rice   | Arsenal | England     | MID      | 35          | 2833    | 4     | 7       | 48    | 18              | ... | 21    | 5            | 1         | 0     | 0%      | 0               | 0                   | 0       | 0           | 0.0             |
| 4 | Ethan Nwaneri | Arsenal | England     | MID      | 26          | 889     | 4     | 0       | 24    | 0               | ... | 9     | 1            | 0         | 0     | 0%      | 0               | 0                   | 0       | 0           | 0.0             |



Explore the dataset structure:  
Displays size, data types, null values, and columns that might be irrelevant (with too many zeros).

```
df.shape # Ver cantidad de filas y columnas
df.dtypes # Tipos de datos por columna
df.describe(include='all').T # Estadísticas generales
df.isnull().sum().sort_values(ascending=False) # Valores nulos
(df == 0).sum().sort_values(ascending=False) # Columnas con muchos ceros
```

✓ 0.2s

|                           |     |
|---------------------------|-----|
| Carries Ended with Goal   | 470 |
| Carries Ended with Assist | 437 |
| Hit Woodwork              | 397 |
| Assists                   | 362 |
| Carries Ended with Shot   | 334 |
| Through Balls             | 323 |
| Successful Crosses        | 321 |
| Carries Ended with Chance | 302 |
| Goals                     | 301 |
| Blocks                    | 299 |
| Crosses                   | 281 |
| Offsides                  | 280 |
| Shots On Target           | 279 |
| Big Chances Missed        | 278 |
| aDuels Won                | 268 |
| Aerial Duels              | 267 |
| Ground Duels              | 267 |
| Possession Won            | 267 |
| Interceptions             | 267 |
| Carries                   | 267 |
| Passes                    | 267 |
| Successful Passes         | 267 |
| gDuels Won                | 267 |
| Successful fThird Passes  | 267 |
| Progressive Carries       | 267 |



# DATA CLEANING

- Dropped irrelevant columns (mainly goalkeeper stats).
- Created the target variable is\_top\_player (players with ≥10 goals + assists).
- Filtered the dataset to work only with top players.

|    | Player Name        | Club    | Nationality | Position | Appearances | Minutes | Goals | Assists | Shots | Shots On Target | ... | Tackles | Ground Duels | gDuels Won | gDuels % | Aerial Duels | aDuels Won | aDuels % | Fouls | Yellow Cards | is_top_player |
|----|--------------------|---------|-------------|----------|-------------|---------|-------|---------|-------|-----------------|-----|---------|--------------|------------|----------|--------------|------------|----------|-------|--------------|---------------|
| 1  | Bukayo Saka        | Arsenal | England     | MID      | 25          | 1735    | 6     | 10      | 67    | 2               | ... | 29      | 58           | 34         | 59%      | 45           | 23         | 51%      | 15    | 3            | 1             |
| 3  | Declan Rice        | Arsenal | England     | MID      | 35          | 2833    | 4     | 7       | 48    | 18              | ... | 53      | 342          | 121        | 35%      | 26           | 10         | 39%      | 21    | 5            | 1             |
| 7  | Gabriel Martinelli | Arsenal | Brazil      | MID      | 33          | 2300    | 8     | 4       | 55    | 12              | ... | 23      | 237          | 111        | 47%      | 72           | 25         | 35%      | 16    | 1            | 1             |
| 11 | Kai Havertz        | Arsenal | Germany     | FWD      | 23          | 1874    | 9     | 3       | 53    | 2               | ... | 16      | 127          | 59         | 47%      | 30           | 8          | 27%      | 38    | 5            | 1             |
| 13 | Leandro Trossard   | Arsenal | Belgium     | MID      | 38          | 2550    | 8     | 7       | 72    | 2               | ... | 31      | 167          | 89         | 53%      | 59           | 35         | 59%      | 27    | 2            | 1             |



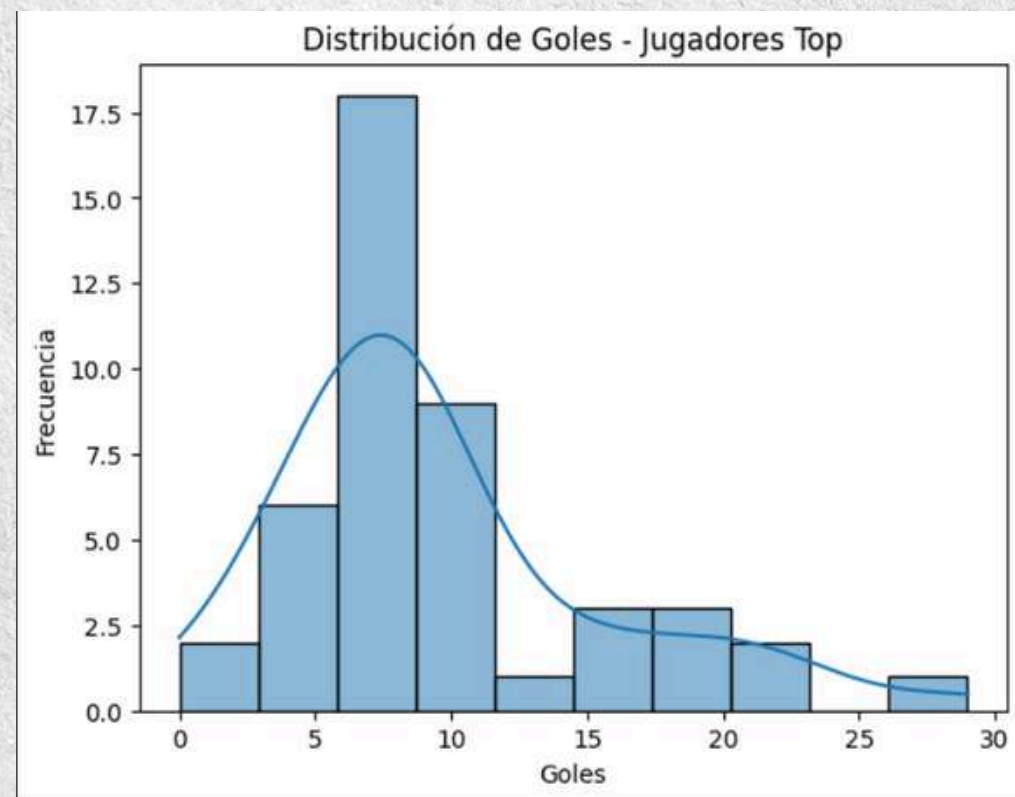


# **EXPLORATORY DATA ANALYSIS - EDA**

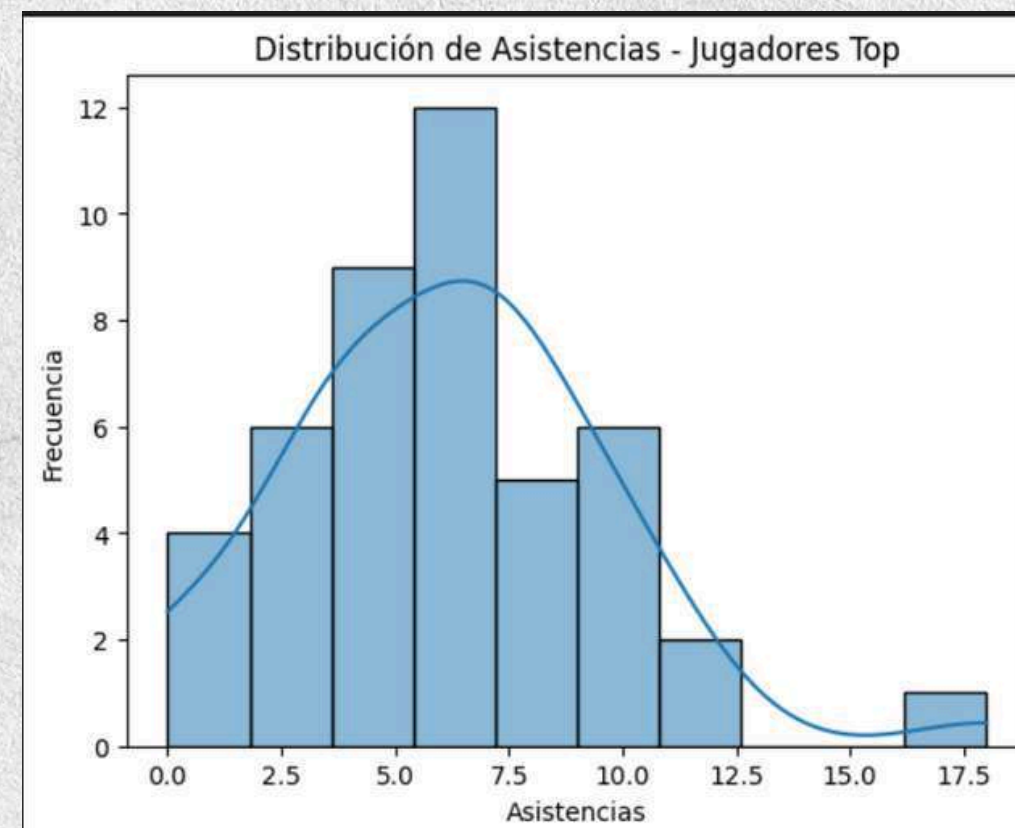
- Individual Distributions:
- Goals Histogram
- Assists Histogram
- Minutes Played Histogram



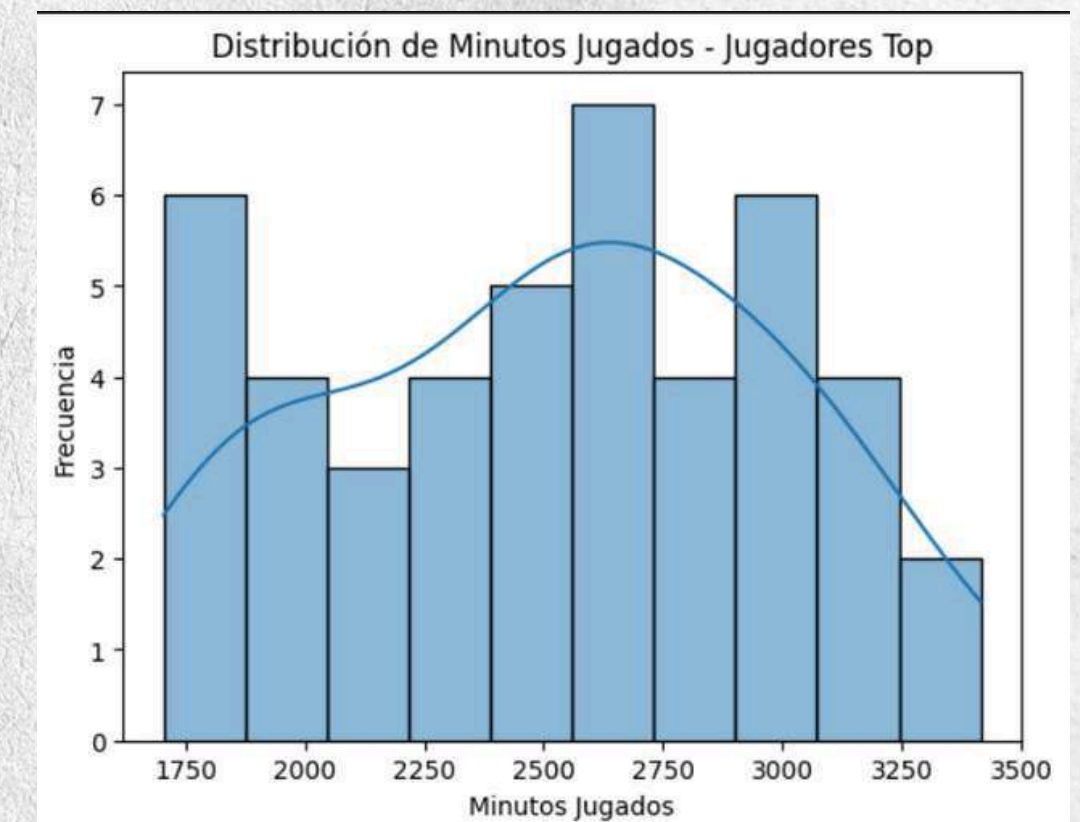
# VISUALIZATIONS



Goals Histogram



Assists Histogram



Minutes Played Histogram





# **EXPLORATORY DATA ANALYSIS - EDA**

Ranking:

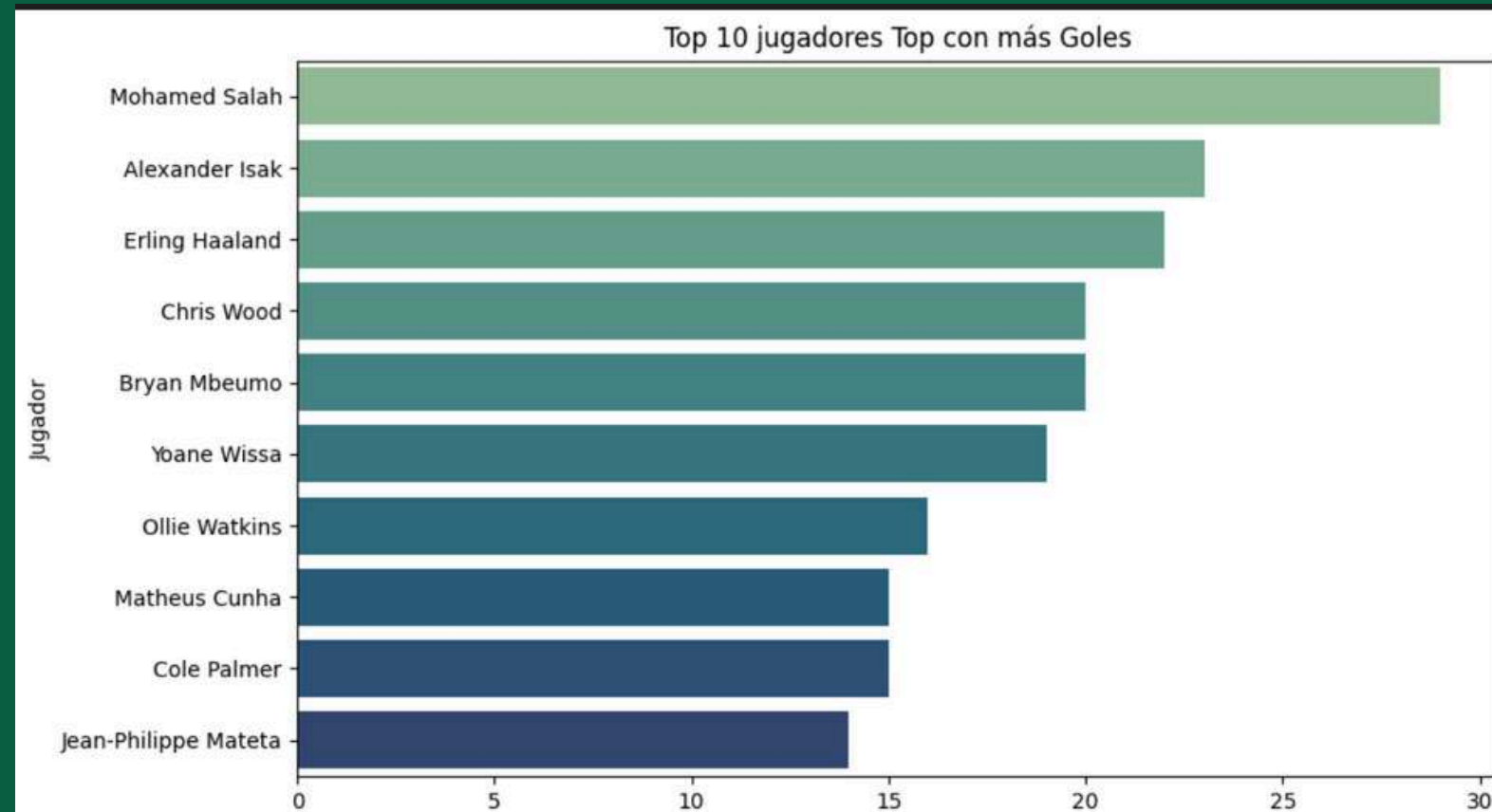
- Top 10 players with most goals

Class Distribution:

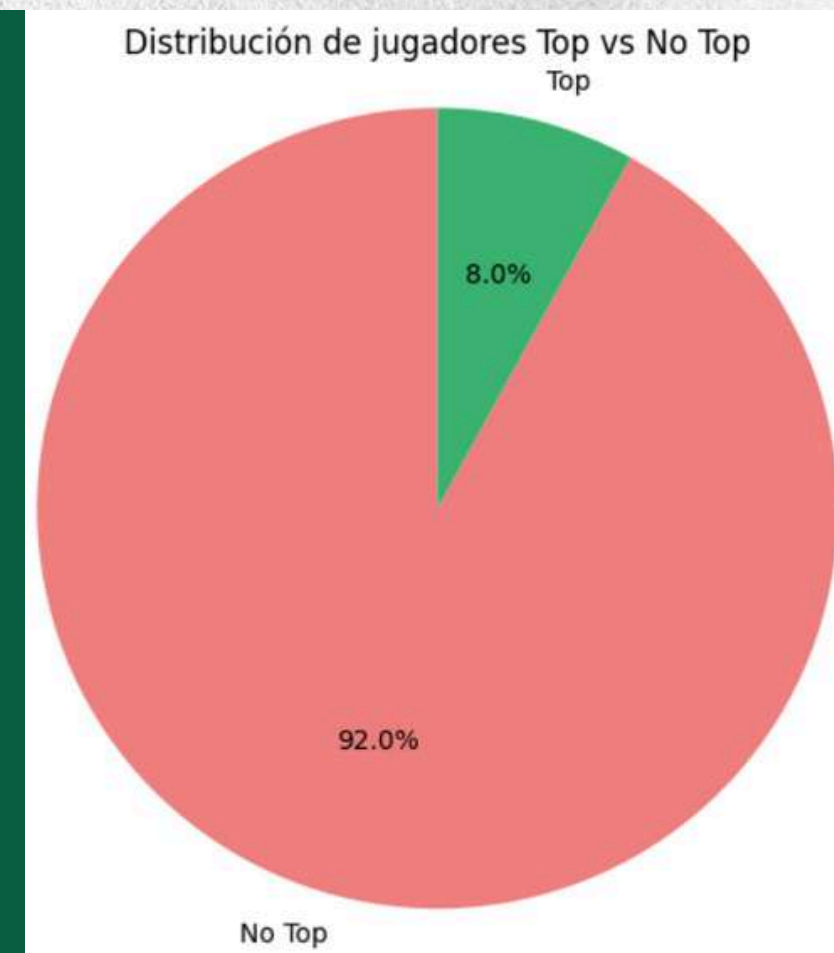
- Pie chart: Top vs Non-Top players



# VISUALIZATIONS



Top 10 players with most goals



Pie chart: Top vs Non-Top players





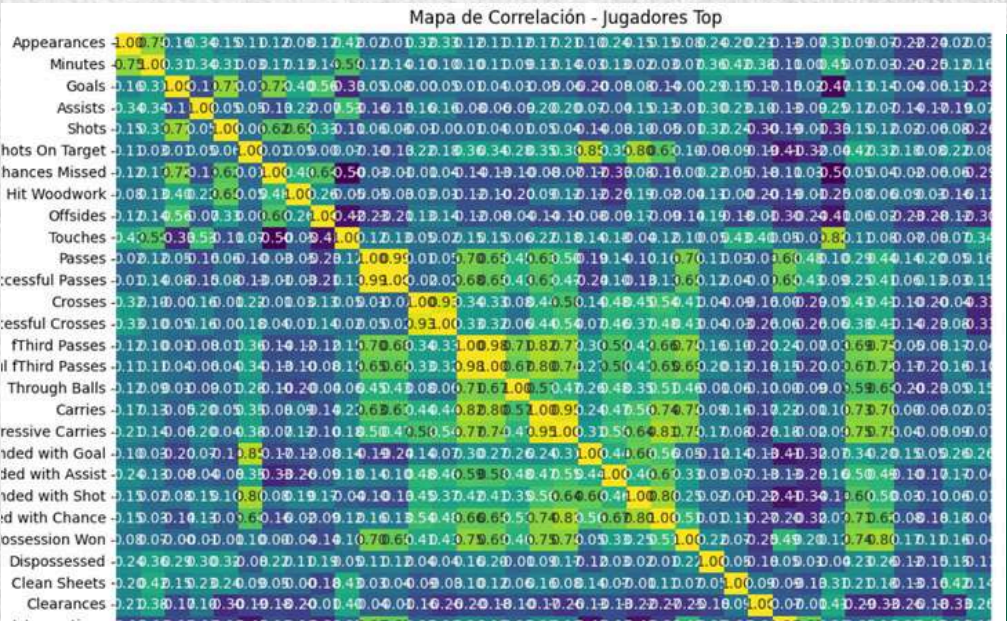
# **EXPLORATORY DATA ANALYSIS - EDA**

Variable Relationships:

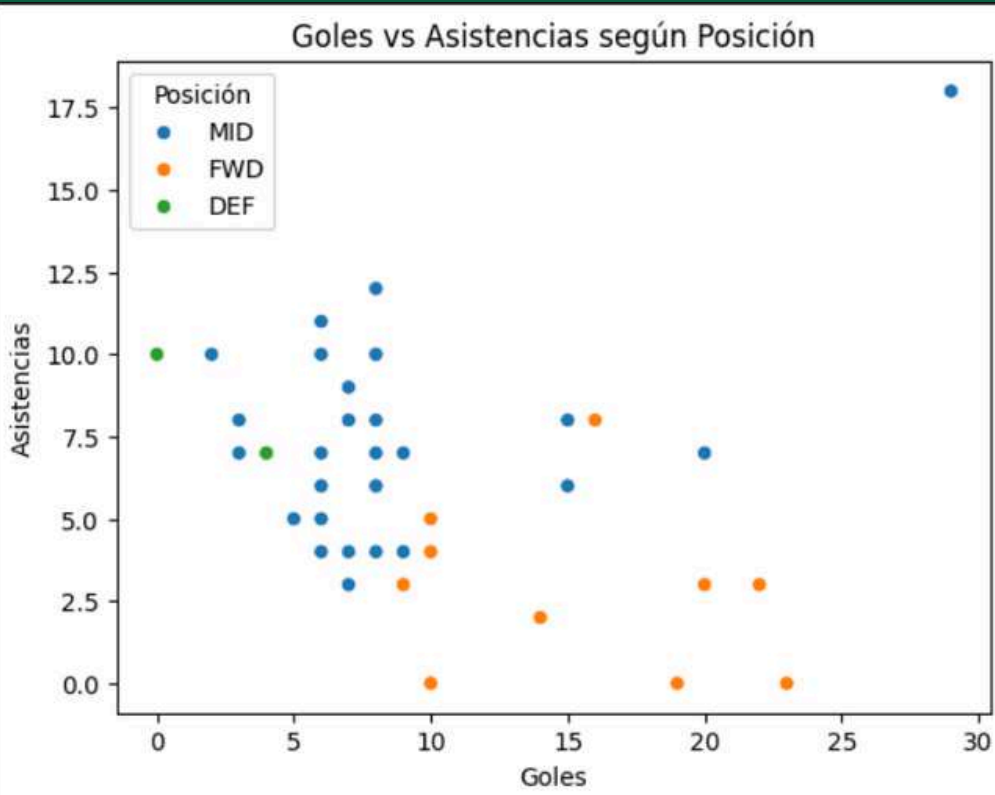
- Correlation Heatmap
- Scatterplot: Goals vs Assists by position
- Average Goals by Position



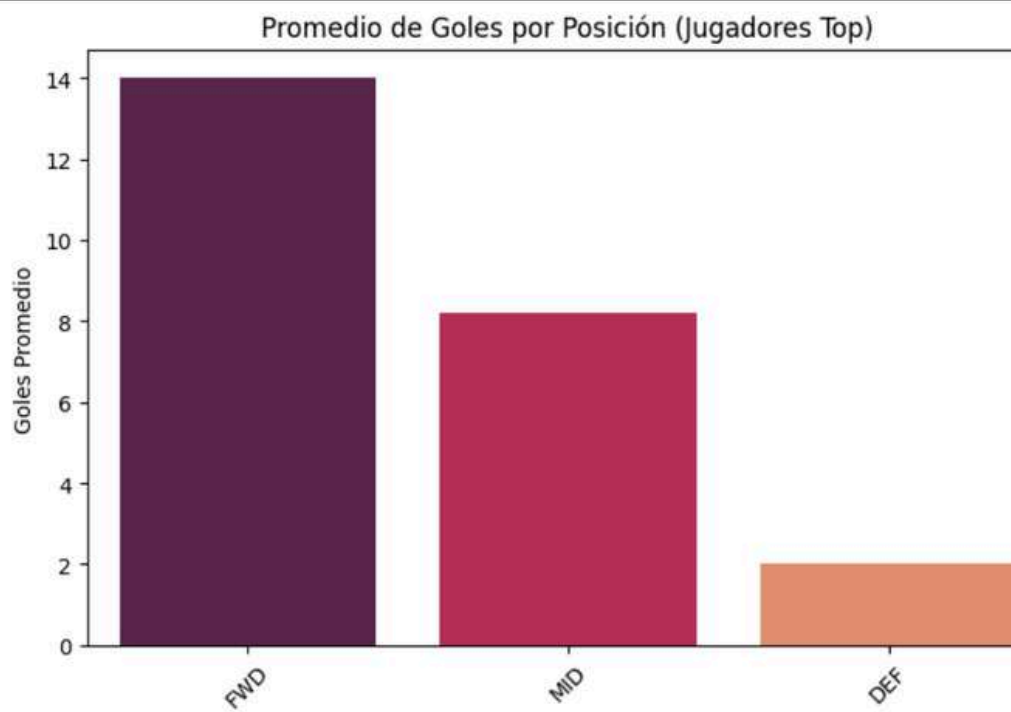
# VISUALIZATIONS



Correlation Heatmap



Scatterplot: Goals vs Assists by position



Average Goals by Position



```
is_top_player  
1    45  
Name: count, dtype: int64
```

```
1    6  
3    4  
7    8  
11   9  
13   8  
14   3  
45   8  
46  16  
51   3  
82  20  
Name: Goals, dtype: int64
```

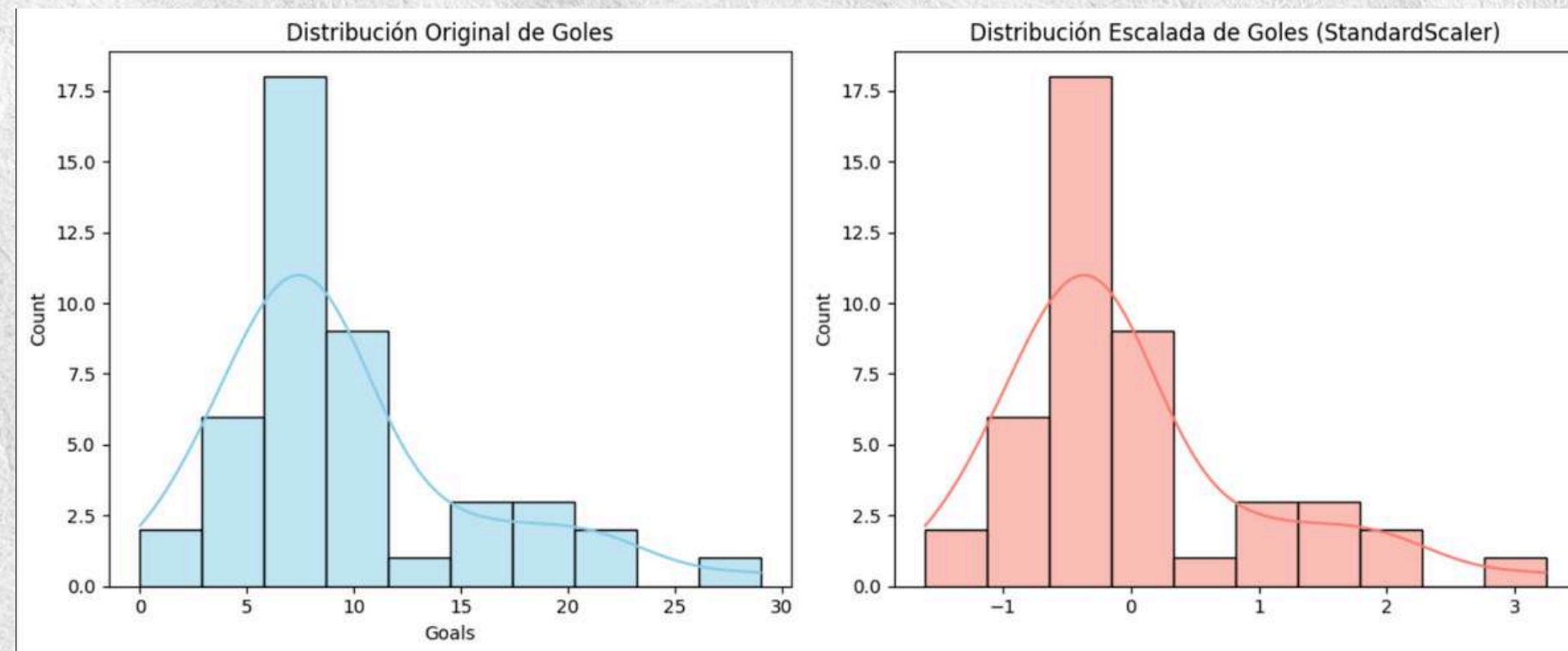
|   | Goals (original) | Goals (escalado) |
|---|------------------|------------------|
| 0 | 6                | -0.601338        |
| 1 | 4                | -0.935414        |
| 2 | 8                | -0.267261        |
| 3 | 9                | -0.100223        |
| 4 | 8                | -0.267261        |
| 5 | 3                | -1.102453        |
| 6 | 8                | -0.267261        |
| 7 | 16               | 1.069045         |
| 8 | 3                | -1.102453        |
| 9 | 20               | 1.737198         |

It was verified that all predictor variables were correctly scaled, a sample of the transformed data was visualized, the distribution of the target variable (is\_top\_player) was reviewed, and the original and scaled values of the Goals variable were compared.

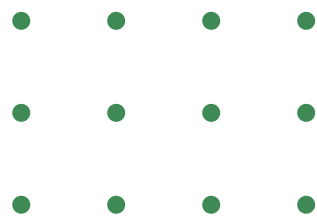


# DISTRIBUTION

We chose Random Forest because it offers an excellent balance between performance and simplicity. It is especially effective on small data sets like this one (45 top players), and handles the mix of coded categorical and numerical variables well. It also has good generalization capabilities and allows for interpreting feature importance.







| Model               | Training Accuracy | Test Accuracy |
|---------------------|-------------------|---------------|
| Decision Tree       | 99%               | 98%           |
| Random Forest       | 100%              | 97%           |
| Logistic Regression | 83%               | 79%           |





# THANKS YOU



Calle Dr. Ignacio Cuesta Barrios # 36 Carretera  
Tepic-Mora Ejido la Cantera, 63506 Tepic, Nay.



311-340-82-29

