

# Analyzing Euro 2024



# Overview

**The Euro 2024 Data Analytics project** aims to analyze and visualize data related to the UEFA European Football Championship, scheduled to take place in Germany from June 14 to July 14, 2024. The objective is to provide insights in-depth analysis: focus on game events and goals.

# Problem Statement

The primary goal is to analyze a soccer game events dataset to gain insights into goal-scoring patterns, particularly focusing on late-game substitutions and their impact on goal-scoring.

# Goals

**1.**

**Understand Goal-Scoring  
Patterns:**

Analyze the distribution of goals across different game periods.

**2.**

**Identify "Joker" Goals:**

Determine if there's a correlation between late-game substitutions and subsequent goals.

**3.**

**Explore Game-Level Statistics:**

Analyze the distribution of goals per game and identify games with the highest number of goals.

# Datasets

The dataset is a CSV file named `game_events.csv`. It's assumed to contain information about various game events, including goals, substitutions, and other relevant details. The specific source of the data is not explicitly mentioned.

The dataset was obtained from <https://docs.sportmonks.com/football>

# Outputs

Exploratory data analysis --> Examine the data to understand its structure and identify any potential issues.

Data modeling --> Transform the data into a format that facilitates easier analysis.

In-depth analysis: Focus on game events and goals.

# FINDINGS

## Goal Distribution:

- Most goals are scored in the latter half of the game, especially in the final 15 minutes and extra time.
- A significant number of goals are scored in the 90th minute, which might be due to late-game pressure or added time.

	event_type	events
0	Yellowcard	32
1	Substitution	29
2	Goal	12
3	Redcard	2
4	VAR	1

The events that happen during the 90th minute.

	minute	events
	0	67
	1	83
	2	81
	3	52
	4	11
	...	...
	105	105
	106	20
	107	36
	108	12
	109	98
110 rows × 2 columns		

"Joker" Goals:

- The analysis can identify potential "joker" goals by analyzing substitutions and subsequent goals within a certain time frame.
- However, a more detailed analysis is needed to confirm the effectiveness of late-game substitutions in directly leading to goals.

	game_period_bucket	goal_type	cnt
0	5	normal goal	4
1	10	normal goal	4
2	15	normal goal	6
3	20	normal goal	8
4	25	normal goal	5
5	30	normal goal	7
6	35	normal goal	4
7	40	normal goal	4
8	45	normal goal	1
9	45 +	normal goal	3
10	50	normal goal	2
11	55	normal goal	8
12	60	normal goal	7
13	65	normal goal	2
14	70	normal goal	9
15	70	joker goal	1
16	75	normal goal	4
17	80	joker goal	2
18	80	normal goal	7
19	85	joker goal	1
20	85	normal goal	3
21	90	normal goal	2
22	90 +	normal goal	8
23	90 +	joker goal	4
24	OT	normal goal	2



Game-Level Statistics:

- The dataset can be analyzed to identify games with the highest number of goals.
- The distribution of goals per game can provide insights into the overall scoring rate and variability.

Which Game Had the Most Goals?

	game	goals
0	Germany vs Scotland (2024-06-14)	6
1	Netherlands vs Austria (2024-06-25)	5
2	Spain vs Georgia (2024-06-30)	5
3	Croatia vs Albania (2024-06-19)	4
4	Poland vs Austria (2024-06-21)	4
5	Turkey vs Georgia (2024-06-18)	4
6	Hungary vs Switzerland (2024-06-15)	4
7	Portugal vs Czech Republic (2024-06-18)	3
8	Spain vs Germany (2024-07-05)	3
9	Turkey vs Portugal (2024-06-22)	3

How many goals are there usually per game?

	goals	cnt
0	6	1
1	5	2
2	4	4
3	1	7
4	3	13
5	2	15

# Insights

## 1.Goal-Scoring Patterns:

- Peak Scoring Periods: Identify specific time periods (e.g., 75-90 minutes, extra time) where goals are more likely to occur.
- Impact of Extra Time: Analyze how extra time influences goal-scoring rates.

## 2.Joker" Goals:

- Effectiveness of Late Substitutions: Quantify the impact of late-game substitutions on goal-scoring.

## 3.Game-Level Statistics:

- Goal Distribution: Analyze the distribution of goals per game to identify high-scoring and low-scoring matches.