

Euros Final 2024: Spain vs. England - Exploratory Data Analysis

This presentation showcases an exploratory data analysis (EDA) of the 2024 Euros final between Spain and England. The analysis investigates key match events, team performance, and individual player contributions to reveal insights into the game's dynamics.

● by **Ausaaf Shaikh**

Introduction to the Data

The Euros 2024 final match between Spain and England offered a thrilling display of high-level football. This EDA analyzes a dataset containing 1,383 events from the match, using Python libraries such as pandas and matplotlib to uncover insightful patterns and trends.

Event Data

The dataset includes detailed information on player actions, including passes, shots, tackles, and other key events. This data is essential for understanding the tactical strategies employed by both teams.

Spatial Data

X and Y coordinates within the dataset represent the location of each event on the pitch. This spatial information allows for a deeper understanding of where events occurred and how teams utilized different areas of the field.

Player Contributions

The dataset captures the individual actions of each player throughout the match, revealing their impact on the game and highlighting their specific roles and responsibilities.

Objective

This EDA aims to uncover valuable insights from the match data. The analysis focuses on understanding the distribution and frequency of different event types, analyzing the spatial distribution of events on the pitch, comparing the performance of both teams, and identifying key players and their contributions to the game.

1 Event Distribution

Understanding the frequency and types of events occurring throughout the match provides insights into the game's overall flow and intensity.

2 Spatial Analysis

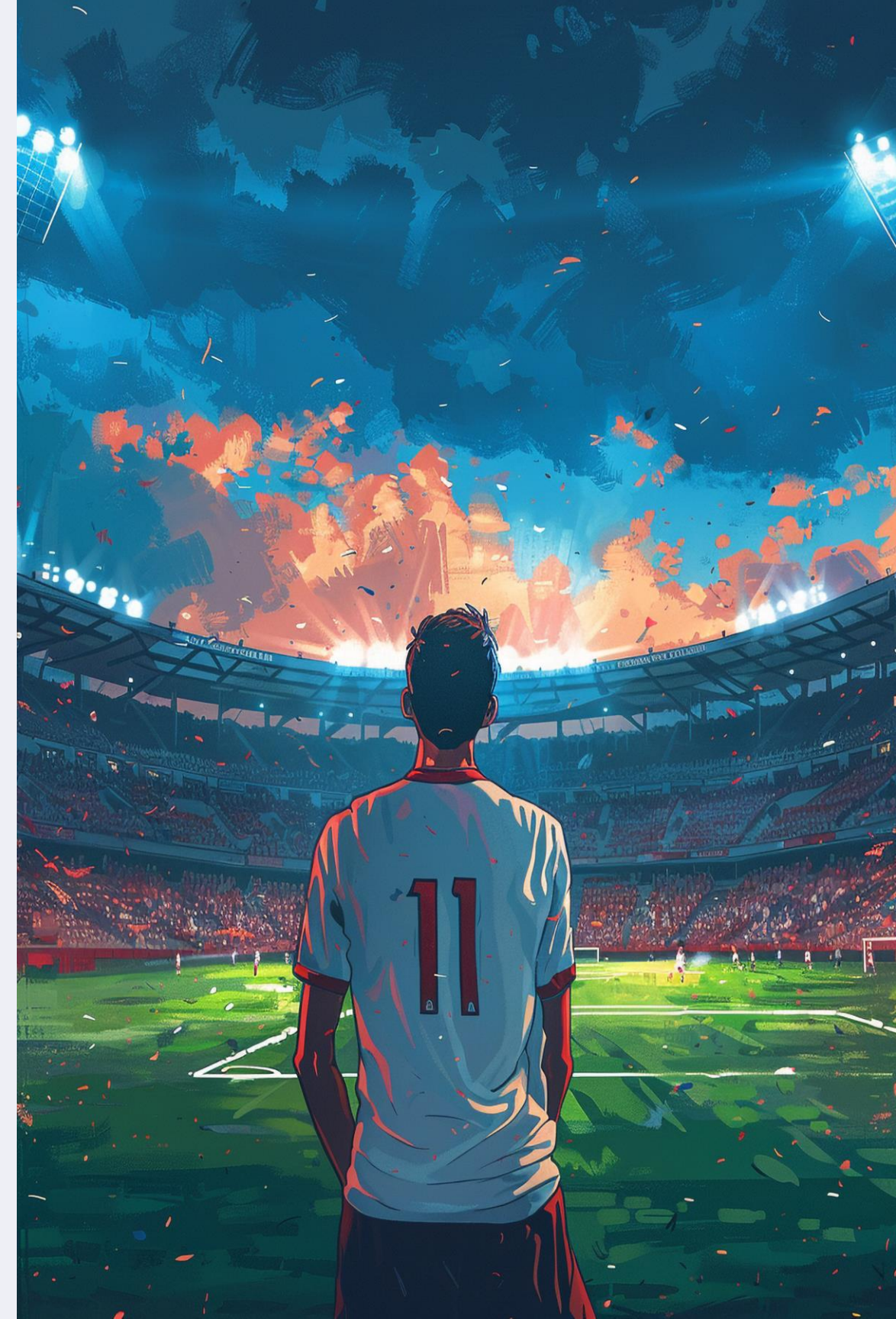
Analyzing the location of events on the pitch reveals strategic patterns and team tendencies in utilizing different areas of the field.

3 Team Comparison

Comparing the event frequency and spatial distribution between Spain and England highlights the strategic differences in their game plan and execution.

4 Player Contributions

Analyzing individual player events reveals their impact on the game and identifies key performers who significantly influenced the outcome.



Dataset Overview

Dataset Name

EurosFinal2024-Spain vs England.csv

Total Entries

1324 events

Total Features

44 columns

Key Columns

minute, second, teamId, type, outcomeType, playerId, x, y, teamName, position, xT

	Unnamed: 0.1	Unnamed: 0	id	eventId	minute	second	teamId	x	y	expand
0	0	0	2.701068e+09	2	0	0.0	338	0.000	0.000	
1	1	1	2.701068e+09	2	0	0.0	345	0.000	0.000	
2	2	2	2.701068e+09	3	0	0.0	345	52.500	34.000	
3	3	3	2.701068e+09	4	0	4.0	345	29.505	38.556	
4	4	4	2.701068e+09	4	0	33.0	338	5.250	37.332	

	start_zone_value_xT	end_zone_value_xT	xT
count	727.000000	727.000000	727.000000
mean	0.006182	0.018661	0.001185
std	0.020744	0.017575	0.014425
min	-0.257454	0.006383	-0.149403
25%	-0.010595	0.011105	-0.001587
50%	0.011263	0.014736	0.000000
75%	0.018703	0.021221	0.001698
max	0.108051	0.257454	0.227921



Data Summary

The Euros final data spans from minute 1 to minute 94, encompassing a range of events including passes, shots, tackles, fouls, and more. The data provides a detailed look at the action on the pitch, revealing crucial tactical decisions and individual player performance.

1

Event Types

The dataset includes a variety of event types, providing a comprehensive view of the actions taken by both teams. These include passes, shots, tackles, fouls, offsides, and other relevant events.

2

Spatial Data

X and Y coordinates within the dataset provide a precise location for each event on the pitch, allowing for detailed analysis of spatial distribution and tactical patterns.

3

Team Performance

The data captures the actions of both Spain and England, allowing for a comparison of their performance based on event frequency and location.

Approach and Tools

Python was the primary language used for this EDA, leveraging its powerful data manipulation capabilities. Libraries such as pandas, matplotlib, and seaborn played crucial roles in data cleaning, analysis, and visualization.

Data Cleaning

Cleaning and preparing the data ensures accuracy and consistency in the analysis. This includes handling missing values, correcting data entries, and standardizing formats.

Data Analysis

Pandas was used for data **manipulation**, allowing for efficient analysis of the dataset. This involved exploring data patterns, calculating statistics, and identifying key trends.

Data Visualization

Matplotlib and seaborn facilitated the creation of various visualizations to communicate the insights derived from the data. These visualizations included histograms, scatter plots, heatmaps, and more.



Analytical Approach

The analysis employed descriptive statistics to summarize the dataset, followed by data visualization techniques to reveal underlying patterns and insights.

1. Descriptive Statistics

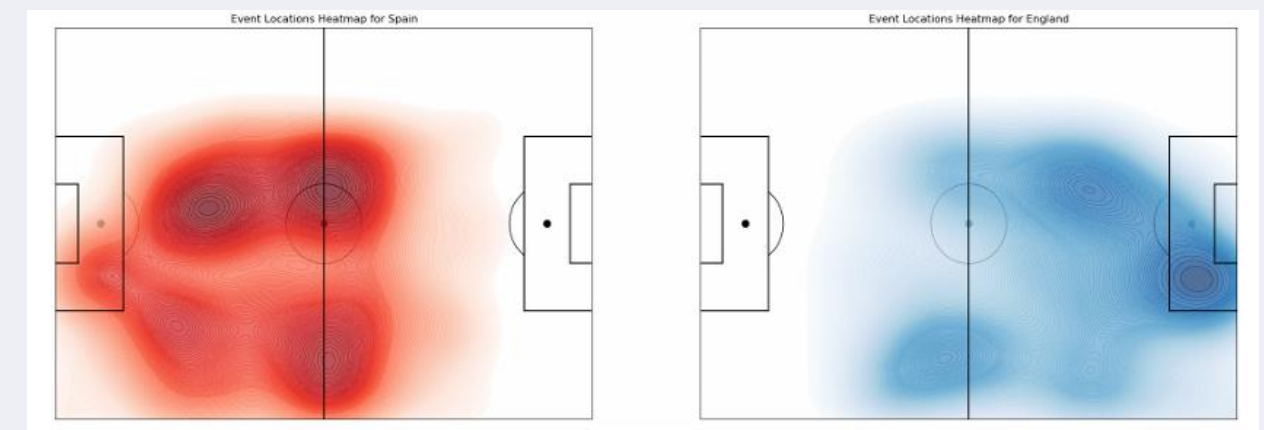
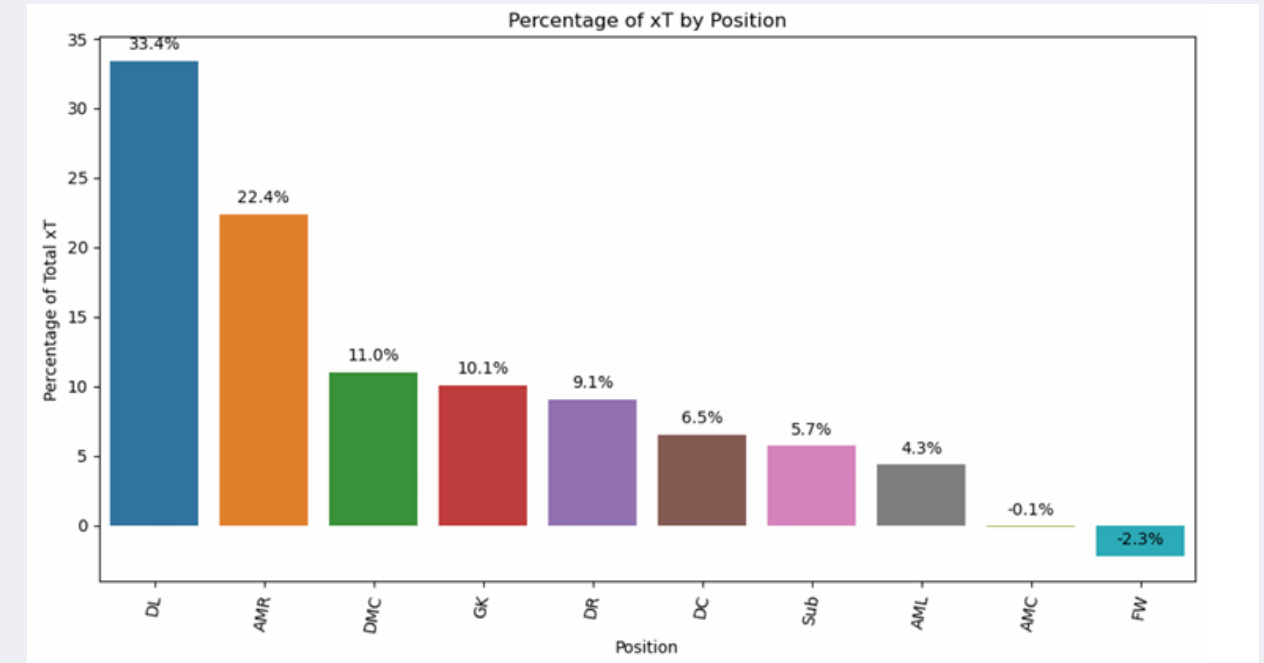
Measures of central tendency (mean, median) and dispersion (variance, standard deviation) were used to summarize the distribution of events and identify key characteristics.

2. Visualizations

Histograms, heatmaps, and bar plots were used to visualize event frequencies, spatial distribution, and team comparisons.

3. Comparative Analysis

The data was analyzed to compare the performance of Spain and England based on the frequency and spatial distribution of events.



Key Findings: Event Distribution

The analysis revealed that the majority of events occurred within the first and last 10 minutes of each half. This suggests an increased intensity during these periods, likely due to the start of the game and the final push for victory.

Event Type	Frequency
Passes	High
Shots	Moderate
Tackles	Moderate
Fouls	Moderate





Spatial Analysis

The heatmap of events showed that both teams concentrated their activities primarily in the midfield and attacking thirds. This pattern indicates a focus on ball possession, creating scoring opportunities, and controlling the flow of the game.



Pass Distribution

The heatmap highlighted areas where passes were most frequent, indicating areas of ball possession and tactical play.



Shot Distribution

The analysis revealed the zones where both teams attempted shots, providing insights into their attacking strategies and the effectiveness of their attempts.



Tackle Distribution

Areas with high tackle frequency indicated key areas of defensive pressure and the effectiveness of both teams' tackling strategies.

Team Comparison

Spain exhibited a higher number of events overall, with a notable concentration of events in the attacking third. This indicates a focus on offensive play and possession-based football.

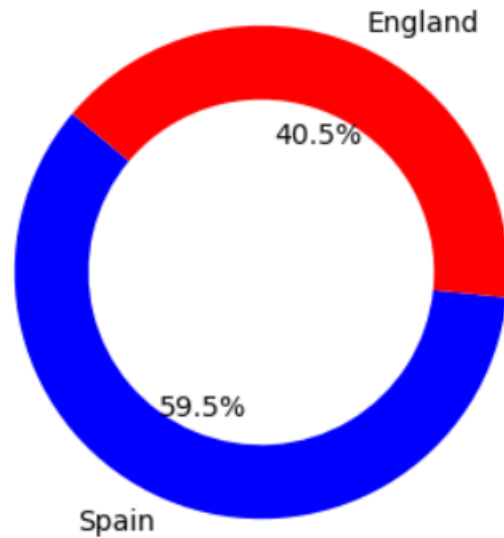
1 Spain

Spain's higher event frequency in the attacking third suggests a strategy focused on possession and attacking opportunities. This could be attributed to a preference for controlling the game through passing and movement.

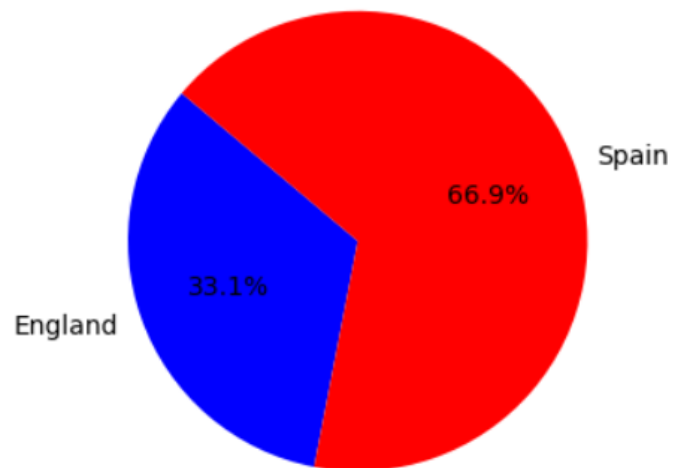
2 England

England's higher number of defensive events reveals a focus on tactical discipline and maintaining a strong defensive line. This approach could be attributed to a more counter-attacking strategy.

Number of Events by Team



Average xT by Team





Appearance

The analysis was presented with a focus on clarity and visual appeal. All visualizations were designed to be easily interpretable and informative, with clear labels, legends, and concise descriptions, allowing for a smooth understanding of the data insights.

This EDA has provided valuable insights into the Euros Final 2024 match between Spain and England. Analyzing the event distribution, spatial patterns, and team performance highlights the strategic differences between the two teams and provides a deeper understanding of the dynamics at play during the game.



Conclusion

This exploratory data analysis of the Euros 2024 final between Spain and England provides valuable **insights** into the dynamics of the match. By analyzing event types, player contributions, and spatial distributions, we can better understand the strategies and performances that defined the game. This analysis can serve as a foundation for further tactical analysis and player performance evaluation



The End

Dive deeper into the data and explore the full code insights from this football EDA project. Visit my GitHub repository for detailed notebooks, visualizations, and further analyses. Check out the full project on GitHub

[GitHub](#)