

Final Project Proposal: UEFA EURO Analysis

Group Members: Matthew Galvez

2024-11-18

What dataset?

We will analyze the dataset titled “**Football - Soccer - UEFA EURO, 1960 - 2024**”, which is publicly available on Kaggle. The dataset can be accessed [here](#). This dataset includes match results, team statistics, player performance, and country rankings, providing a comprehensive resource for analyzing historical and recent trends in the UEFA EURO championships.

What questions are we trying to answer?

1. **Team Performance:** Which teams have consistently performed well over the years, and what factors contribute to their success?
2. **Player Impact:** How do individual player performances (goals, assists, defensive actions) correlate with overall team success?
3. **Country Comparisons:** Which countries have shown significant improvement or decline in performance over time?
4. **Predictions for 2024:** Can we predict the likely outcomes of UEFA EURO 2024, including group stage winners and the eventual champion?
5. **Trend Analysis:** How has the game evolved over time in terms of goals scored, defensive strategies, and player statistics?

What is your hypothesis about the result?

- Teams with a history of strong defensive performance are more likely to progress to the later stages of the tournament.
- Star players significantly influence their teams' chances of success, but teamwork metrics (e.g., assists and passes) play an equally crucial role.
- Countries with stable coaching staff and infrastructure investments exhibit consistent improvement over time.
- For UEFA EURO 2024, historically dominant teams like Germany, Spain, and France are predicted to perform well based on historical trends and recent form.

How are you going to analyze the data?

1. **Data Cleaning and Exploration:**
 - Verify the completeness of match results, player statistics, and country performance data.

- Handle missing values and standardize fields (e.g., team names and match dates).
2. **Descriptive Statistics:**
 - Analyze team win rates, goal differences, and average player ratings.
 - Compare historical data to identify trends in team and player performance.
 3. **Visualizations:**
 - Create time-series plots to illustrate performance changes over decades.
 4. **Predictive Modeling:**
 - Train linear regression models using historical data (team strengths, player stats, and past results) to predict future match outcomes.
 5. **Hypothesis Testing:**
 - Apply statistical tests to validate the correlation between team success and individual contributions.
 6. **Outcome Predictions for 2024:**
 - Simulate group stage outcomes using predictive models.
 - Estimate the probabilities of team progression to the knockout rounds and ultimate victory.