Andrew Tremante
STAT 231 - Data Science
Final Project Proposal

Mapping Value: Nationality and Its Economic Implications in European Football

Introduction

In August 2021, the transfer of English footballer Jack Grealish to Manchester City for $138 million reignited passionate conversations about the valuation of English players in the Premier League. Prior to this transfer Grealish was valued at only $82 million, raising questions about why Manchester City needed to overpay by $56 million to acquire him. This deal reinforced the notion of an English Tax, which suggests that players' nationalities, particularly English, unjustly inflate their market value. This project seeks to broaden this inquiry to Europe's most prominent football leagues, utilizing data from FBref and FootStats to explore the impact of nationality on players' market values and team affiliations. Through its analysis, the project aims to offer a more nuanced understanding of the international football market, enhancing discussions on player valuation and recruitment strategies.

Data

This project will source data through HTML scraping from FBref and FootStats. FBref will provide team and player statistics and FootStats will be used for information on player market values. By combining these sources, the final dataset will include performance metrics, demographic details, and financial valuations.

Analysis and Product

The main focus of the project will be on spatial data analysis, aiming to map and interpret the geographical distribution of players' nationalities and their market values. By implementing interactive Leaflet visualizations in a Shiny app, the project will allow users to explore different relationships between player's nationalities, leagues and individual statistics. The web application and analysis will seek to identify the correlation between a player's nationality and their market value and whether certain leagues favor players from specific regions. The project will also employ statistical methods such as 2-Way ANOVA and multiple linear regression to investigate the effects of age and nationality on a player's market value and to explore patterns in the international mobility of players.

The Shiny app will feature three tabs with unique visualizations and analysis. The first will be an interactive map highlighting the distribution of player nationalities and market values. The second will be a heat map for visualizing the concentration of players from different countries

within specific leagues or teams. The final visualization will be dynamic map that traces the global movement of players based on their country of origin. These visualizations will each offer their own unique insight into the relationships between players and their nationalities, but they will all contribute to answers the overarching questions the project seeks to resolve.

Conclusion

By combining advanced spatial analysis and statistical modeling, this project hopes to uncover the ways in which nationality influences the careers and market perceptions of professional football players in Europe. It hopes to shed light on common debates in modern football, such as the English Tax, and also provide insight into the recruitment methodologies and approaches of the top clubs and leagues around the world.

Plan

The following is an outline for the steps needed to complete the project:

1. Scrape and wrangle data from FBref and FootyStats
2. Research Leaflet and possibilities for Shiny App
3. Implement visualizations (pages in app)
4. Conduct statistical analysis (2-Way ANOVA, Multiple Linear Regression)
5. Combine app visualizations with statistical analysis results
6. Finalize app and finish report
7. Present in-class and complete peer feedback
8. Complete project with reflection