Rice Data Analytics Bootcamp – Spring 2023

Project 1

Title: Soccer Players Values and Their Statistics

Team Members: Alexandr Gorbulin, Rocky Owens, and Francis Shobo

Project Description/Outline: Merged data from transfermarkt.de and fbref.com regarding soccer players from top 5 European leagues.

Dataset: Soccer layers values and their statistics (Kaggle.com)

Research questions to answer:

1. Which nationality has the most players on the list?
2. Which squad has the most players on the list?
3. Which position plays the most minutes?
4. Which age has the highest value?
5. Which position has the most red cards?
6. Which league has the most valuable players?
7. What is the distribution of player values among the squad?
8. What is the average age and birth year of the players in the squad?
9. How does player position and foot preference relate to their value and age?
10. How do the number of games played, minutes played, and goals scored vary among players of different nationalities?
11. What is the distribution of goals per 90 minutes among players in different positions?
12. What is the correlation between xG and actual goals scored for the players in the squad?
13. How does the league in which the player plays affect their value, goals scored, and minutes played?
14. How does height relate to position, value and other performance metrics?
15. How does the number of tackles and tackles won relate to the player's position?
16. How does the goalkeeper's performance, such as saves percentage, relate to the number of goals against and clean sheets?

Breakdown of Tasks:

1. Data cleaning and formatting
2. Create Jupyter notebook
3. Perform & illustrate analysis
4. Create visualizations
5. Draw conclusions
6. Create report of findings and presentations