DSC540 Final Course Term Project:

**Analyzing Factors Influencing Team Performance in Soccer Leagues**

Pragathi Porawakara Arachchige

Bellevue University

DSC540-T301 Data Preparation (2251-1)

Dr. Catherine Williams

11/15/2024

#### Introduction

A team's performance in professional soccer is influenced by many different factors. These consist of financial investments, team tactics, player statistics, and more. Teams can increase performance by making better judgments by having a better understanding of these aspects. The goal of this project is to check whether soccer player overall ratings will have a significant affect with player ages. I successfully integrated and examined three cleaned datasets Premier League Matches, UEFA Champions League, and FIFA Players during this project. Python was used to visualize the combined data after the datasets were loaded into a SQLite database and joined using SQL.

**Database Management Skills**

I learned how to structure and manage a relational database, loading each dataset into individual tables. Writing SQL queries to join the datasets demonstrated the power of relational databases for combining diverse data sources into a single view for analysis. I could identify relationships between player nationalities, team performance, and historical championship winners by writing SQL joins and queries.

**Visualization Skills**

Using Python libraries like Matplotlib and Seaborn, I created five visualizations that provided insights into player demographics, match outcomes, attendance trends, and more. I produced combination visualizations that included several datasets, such as examining high-scoring Premier League games and connecting the nationalities of FIFA players to UEFA champions. I learned how to extract ideas from combined data and successfully communicate them thanks to the graphics.

**Ethical Considerations**

The project also focused on ethical data handling. Transformations such as replacing and removing missing values or deleting duplicates were all documented to maintain clarity and avoid bias. I derived data in a ethical way from credible sources such as FIFA datasets, Wikipedia (UEFA) and APIs (Premier League games), so that the data was reliable. To limit the risk, I wrote down all assumptions and protected the integrity of the original data.

**Obstacles and Solutions**

One roadblock was the integration of data between datasets based on different types and formats. For example, mapping FIFA data players by nationalities to UEFA winners needed data correlation. I was able to overcome this by reading the data and creating SQL joins to match them up.

**Conclusion**

Overall, this project helped me to step up my technical understanding of database administration, SQL, data wrangling, and visualization while emphasizing the importance of sound data governance. These skills will be used to create future data science and analytics initiatives.