678 Midterm proposal

1. Personal Statement:

My career goal is to work as a data analyst in a professional soccer club. When data analysis first appeared on the soccer field, it was met with a lot of rejection. There was a traditional belief that there was a lot of improvisation on the soccer field. That data should not rigidify an athlete's performance but encourage more of an individual's talent. However, nowadays, data analysis is becoming an indispensable part of many top soccer clubs. By referring to the findings of the data analysts, the manager can decide the starting lineup for each game, develop tactics, reduce or increase the number of passes and increase or reduce the number of fouls on the field to improve the team's winning percentage. Even in the daily management of the team, data analysts can use their findings to advise the head coach whether to reduce or increase the amount of daily training to reduce the team's injury situation. Most of Europe's top leagues are becoming more predictable with the increased dominance of one or two teams (Juventus in Serie A, PSG Paris Saint-Germain in Ligue 1, Barcelona and Real Madrid in La Liga, Bayern and Dortmund in the Bundesliga). It must be admitted that data analysis is gradually changing and promoting the development of soccer. My project analyzes the on-field performance of each team in the Premier League from 2006-2018, such as shots, fouls, off_side, etc., to predict a team's end-of-season point total. This project will allow me to have initial contact with the Football Data Analyst position.

2. Question:

How does the Premium League team's on-field performance in different seasons affect the final points of the season? (goals, yellow cards, red cards ,free kick)

3. Data Resource:

https://www.kaggle.com/zaeemnalla/premier-league?select=stats.csv

4. Proposed Time Line to Work:

11.5 – 11.9. EDA
11.10-11.17 Data Processing
11.18-11.24 Modeling and Validation
11.25-11.30 Finish final report