Premier League 2018-19 Performance Analysis

Group Members

- Jasmine Zhang (yz4720)
- Yuxin Zhang (yz4719)
- Tianyou He (th3072)
- Xiaoyu Wu(xw2962)
- Xiaoyu Huang (xh2636)

Motivation

As the most popular football league worldwide, the Premier League consists 20 competitive football clubs in England and attracts the most famous football players. Statistical analysis is widely used in order to achieve better game performance for the club. We set our interest on the PL season 2018/19 because it is the season pre-pandemic and due to data availability. We tend to investigate the relationship between transfer market status, player characteristics, match style and the final standings of the season.

Intended Final Products

Our intended final products include different types of plots

Anticipated Data Resource

https://datahub.io/sports-data/english-premier-league#readme (38 match detail) https://www.kaggle.com/datasets/thesiff/premierleague1819/ (financial and match data) Reference: https://www.kaggle.com/code/thesiff/premier-league-2018-2019-review-by-eda/script https://fbref.com/en/comps/9/2018-2019/stats/2018-2019-Premier-League-Stats#all_stats_standard (player stats)

```
tranfer_df = read_csv("./data/transfer.csv")
player_df = read_excel("./data/player_game.xlsx", sheet = 1)
match_df = read.csv("./data/match_data.csv")
```

Planned Analyses

- Line graph: Transfer market spending over the past decade for the PL clubs
- Scatterplot comparison: Match style (possession, short pass, long pass) of the PL clubs
- Bar plot / boxplot: player (nationality, age, minutes played) data analysis

Planned timeline