



Prediction of match results in English Premier League based on football statistics

Final Project

L. T. OZGUR YILDIRIM



Ironhack
Data Analytics Bootcamp

November 4th, 2022

Introduction



Premier
League

Introduction

- The main objectives of this study are to;
 - analyze relationships between football statistics
 - predict the results of football matches based on football statistics using classification models.
- The dataset consists of:
 - Premier League standings for the last 22 seasons
 - football statistics of 27 teams in the premier league ranging from 2012-2013 season to 2021-2022 season.



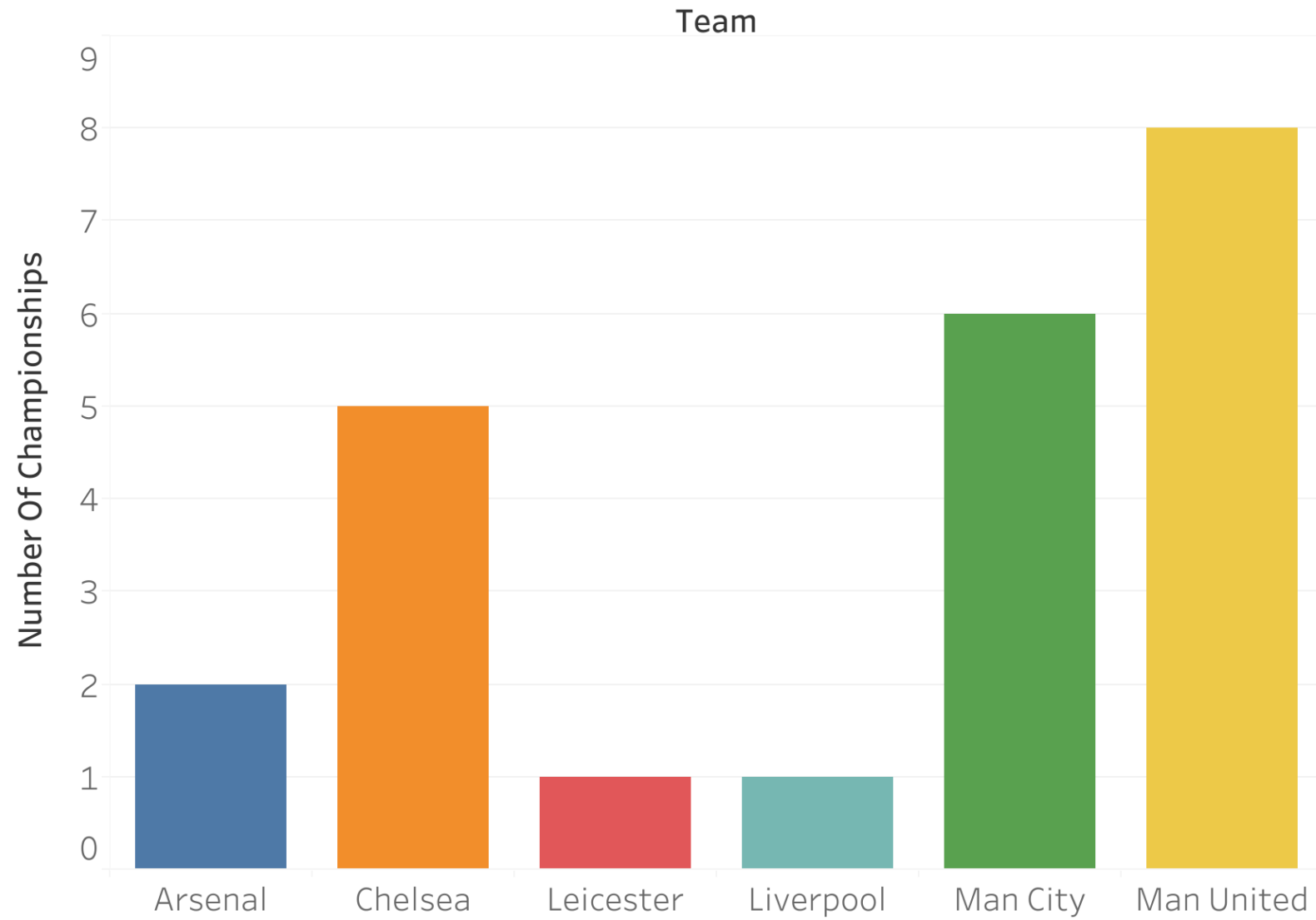
**Premier
League**

Football Statistics



**Premier
League**

Championship numbers

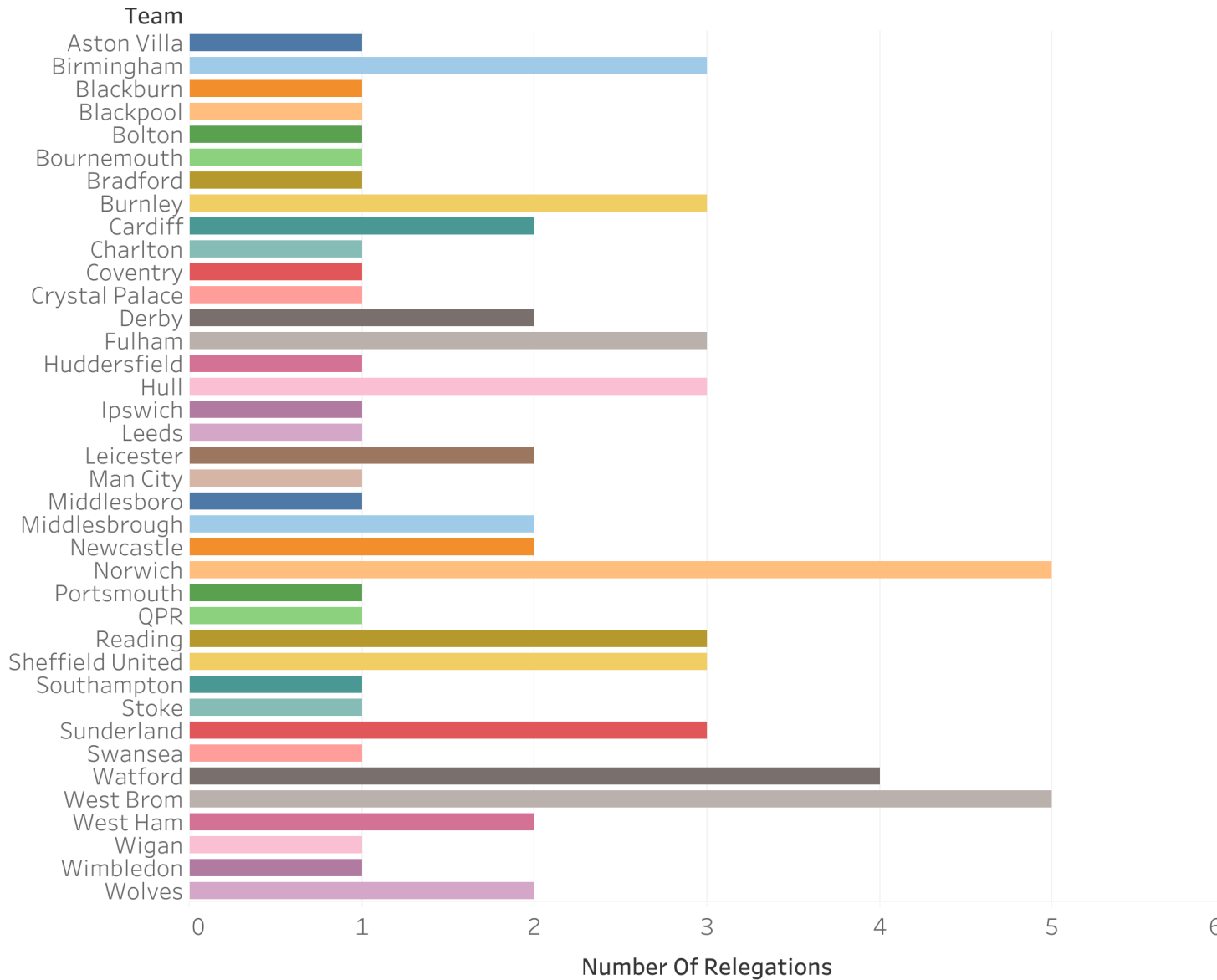


14 titles went to Manchester since 2000.



**Premier
League**

Relegation numbers

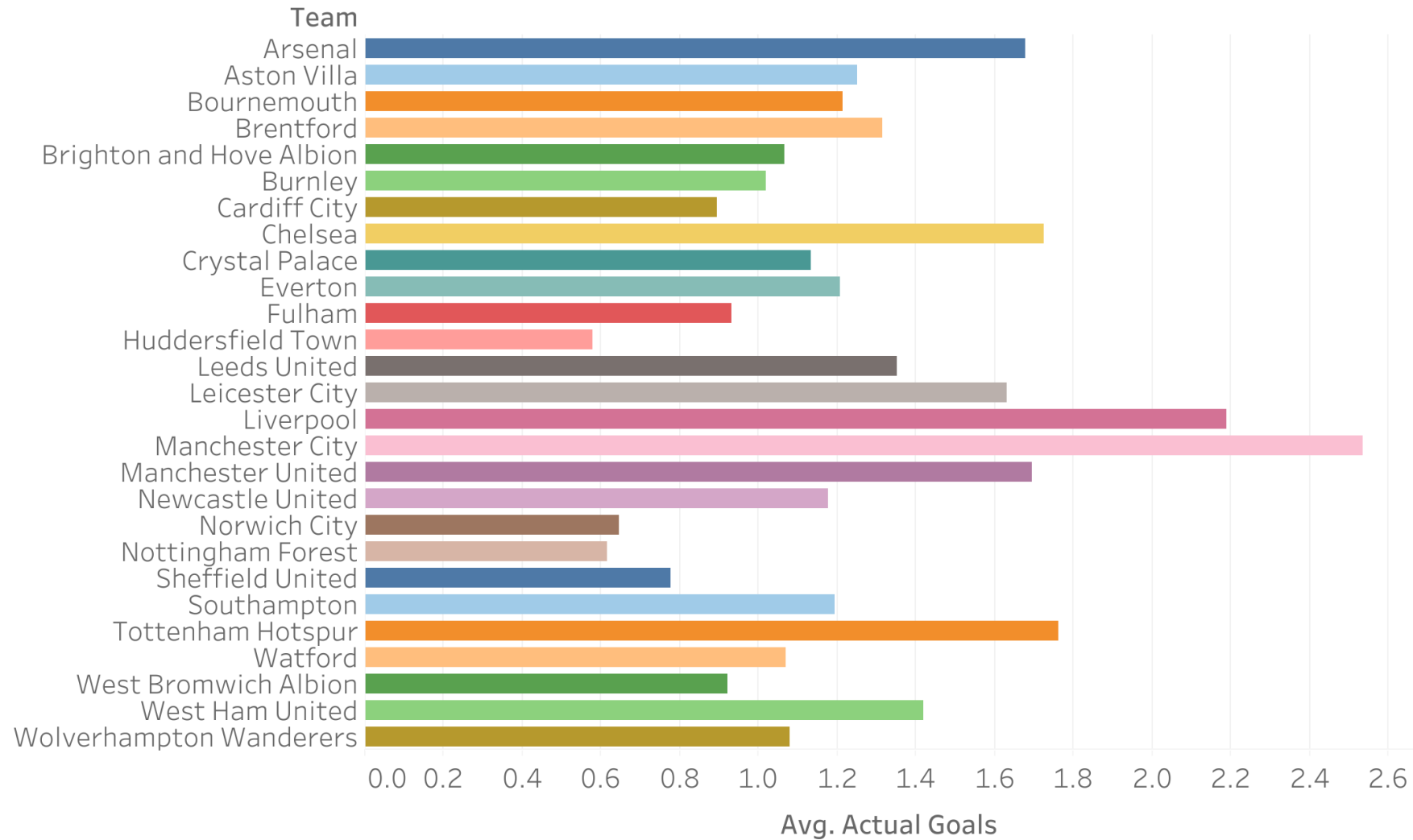


Even Manchester City relegated for 1 time.



Premier
League

Avg. Actual Goals vs. Team

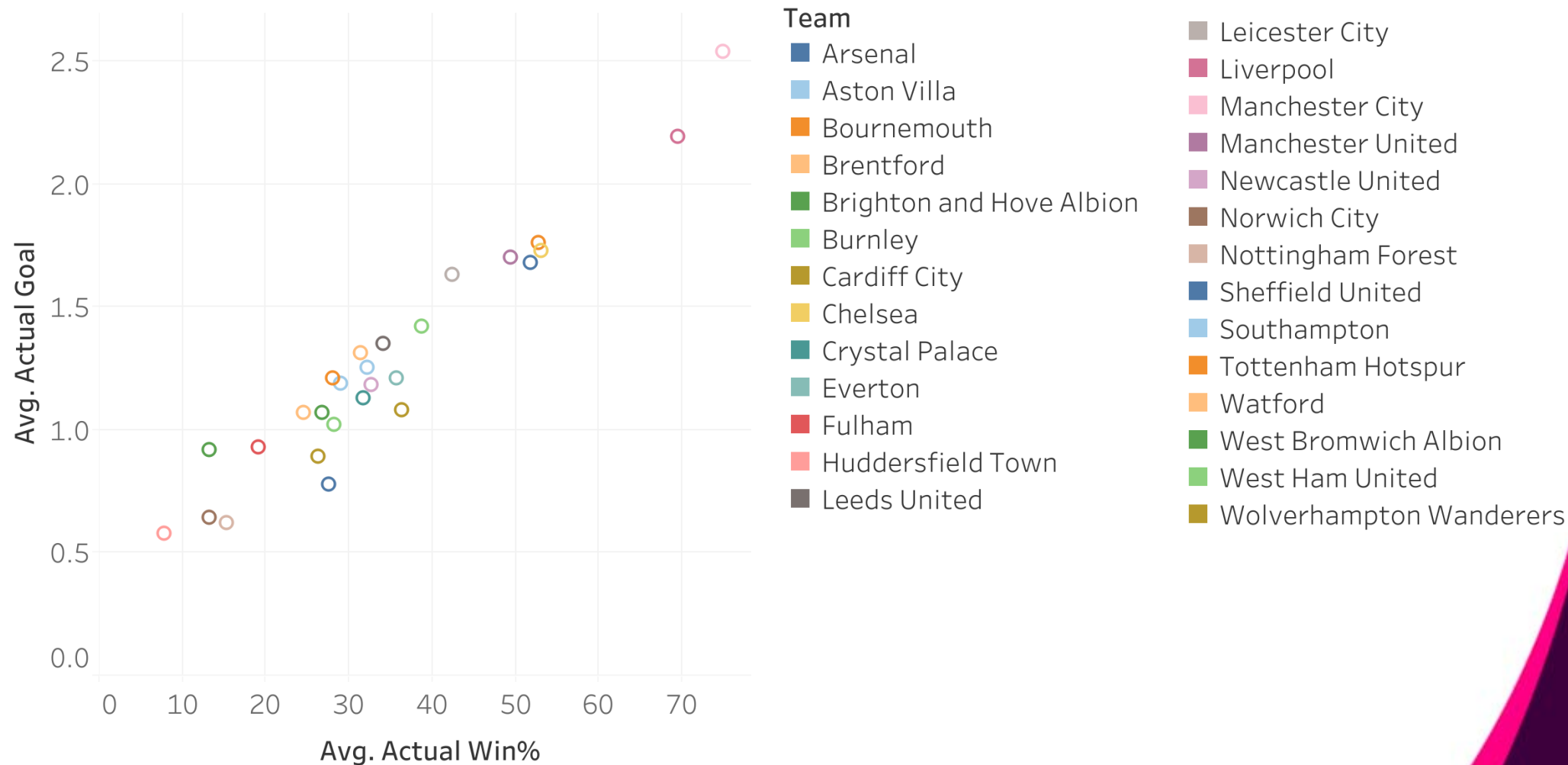


Top 6 teams have avg. goals more than 1.5.



Premier
League

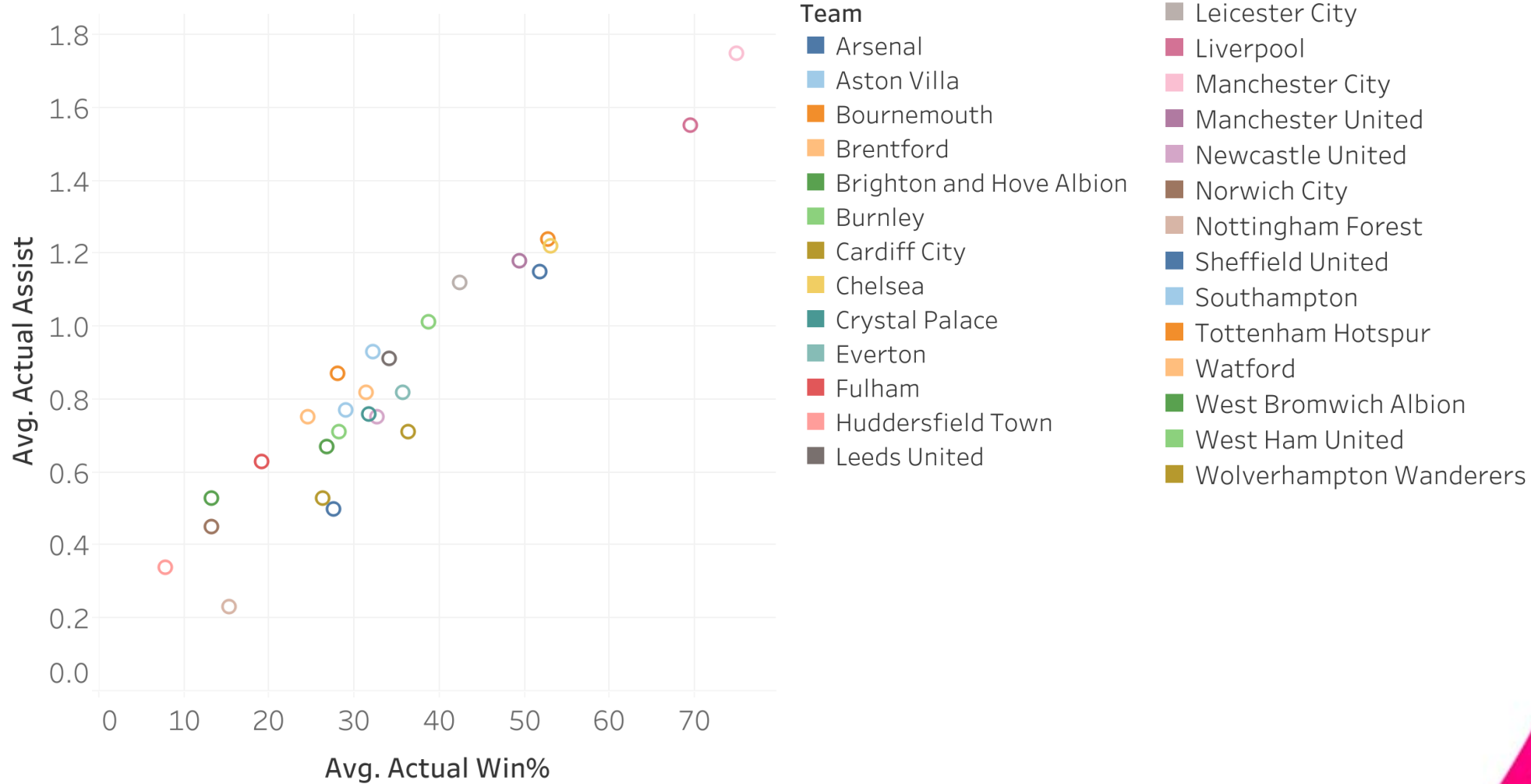
Avg. Actual Goal vs. Avg. Actual Win%



Linear relationship between goals scored by teams and their win%.



Avg. Actual Assist vs. Avg. Actual Win%

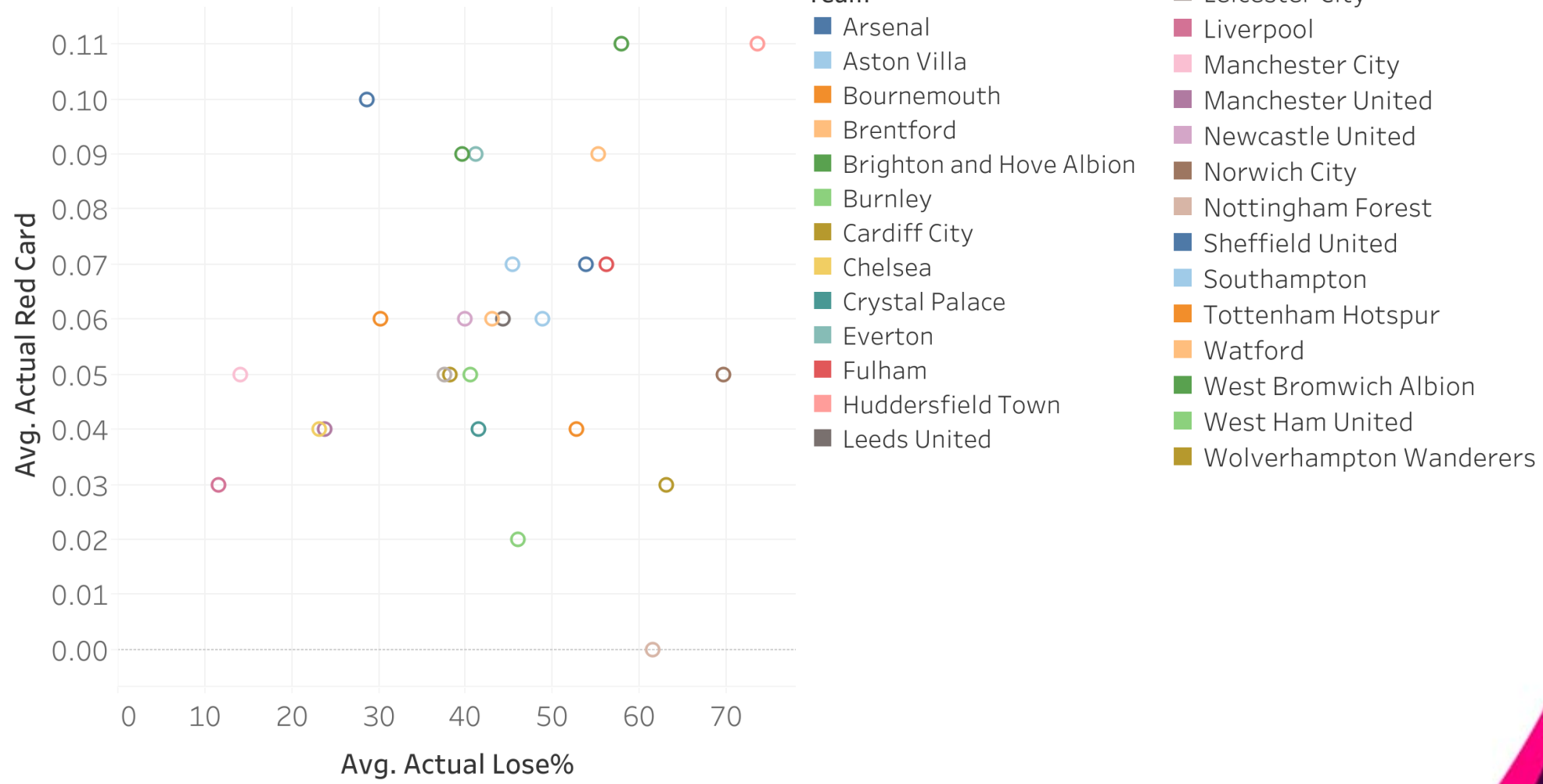


Linear relationship between assists made by teams and their win%.



Premier
League

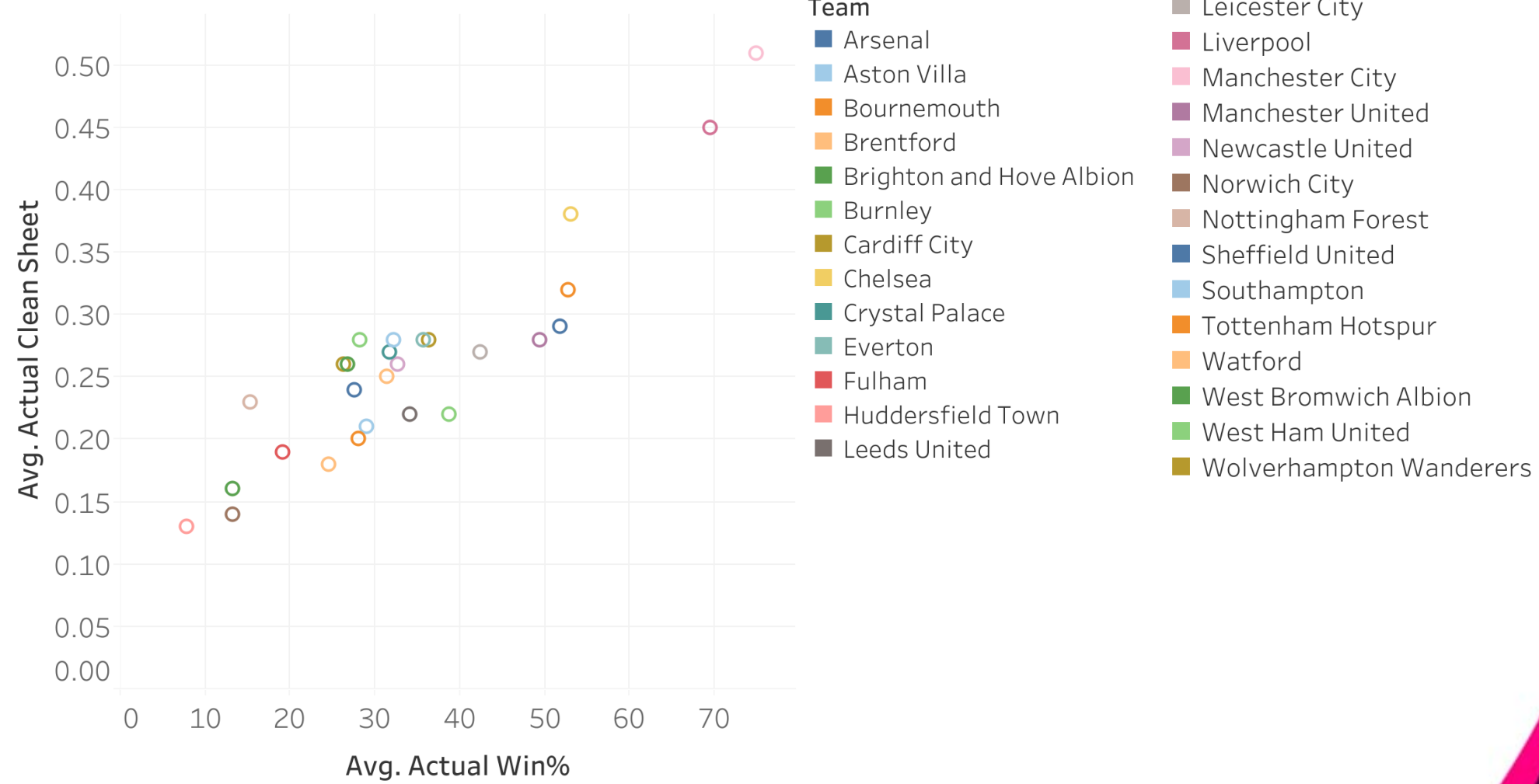
Avg. Actual Red Card vs. Avg. Actual Lose%



Top 6 teams have the least red cards and lose%.



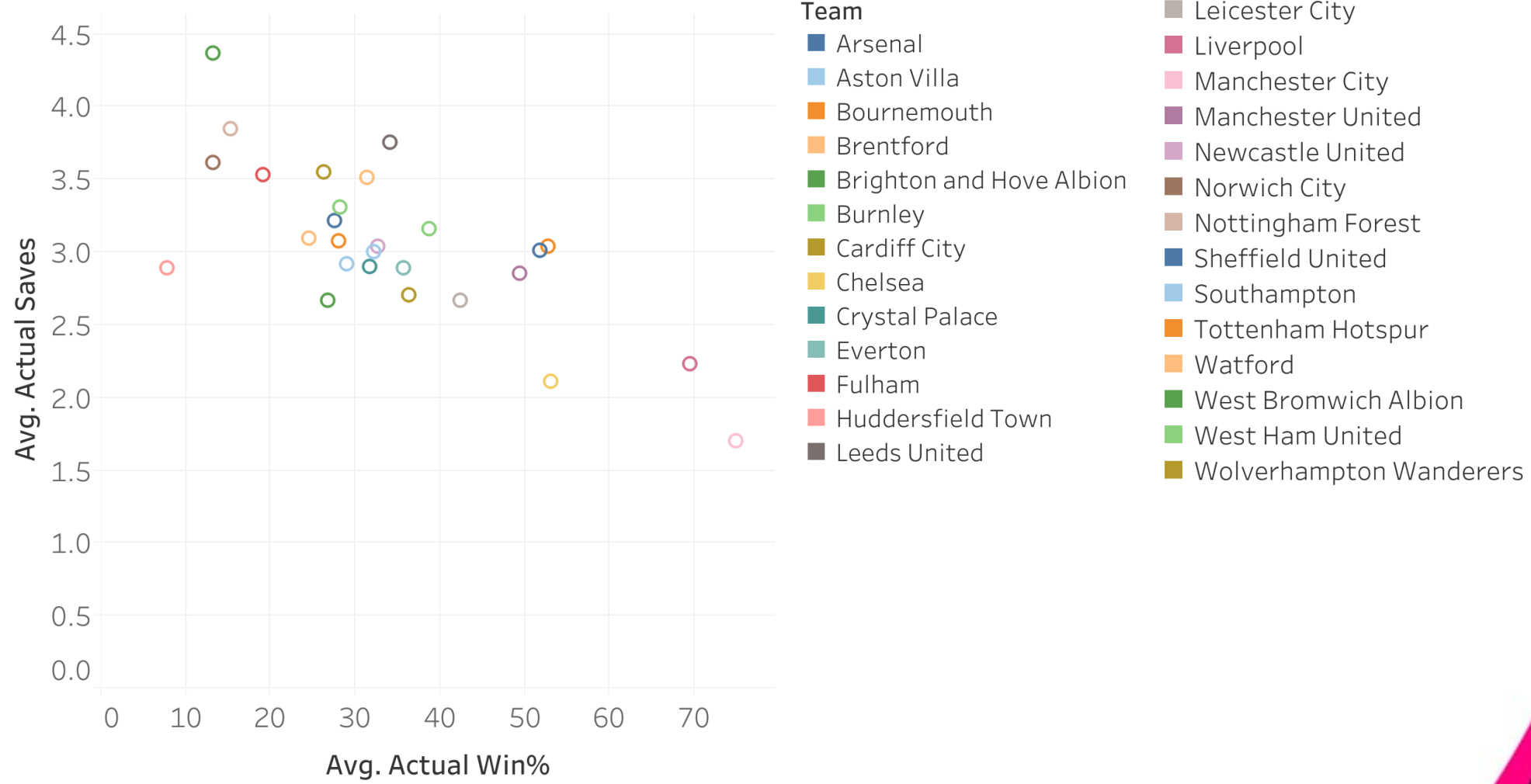
Avg. Actual Clean Sheet vs. Avg Actual Win%



Top 6 teams have the most red cards and win%.



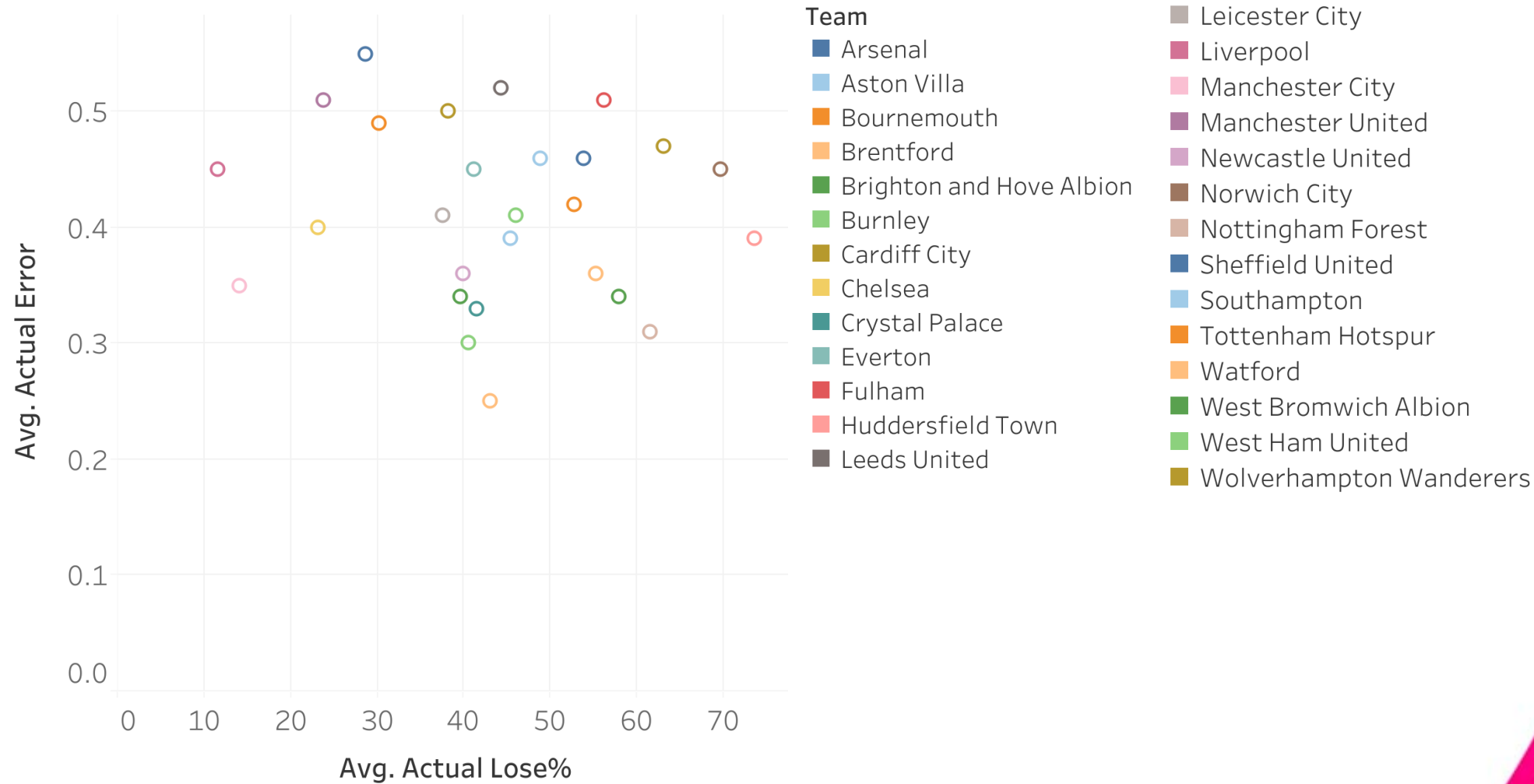
Avg. Actual Saves vs. Avg Actual Win%



Relatively weak teams have the most saves but less win%.



Avg. Actual Error vs. Avg. Actual Lose%



Arsenal has the most number of errors but less lose%.

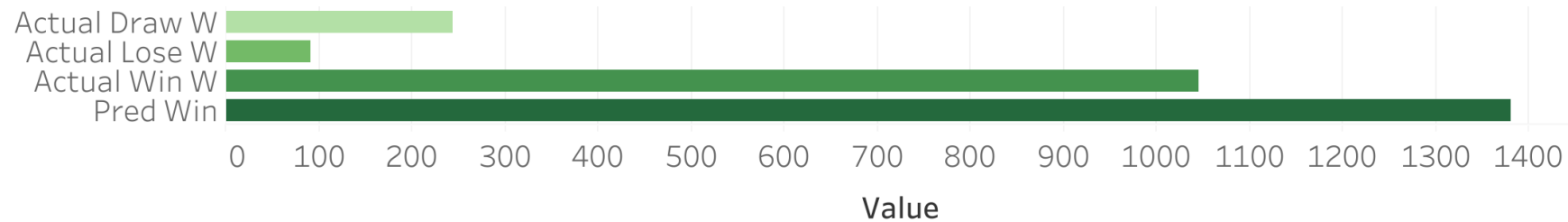
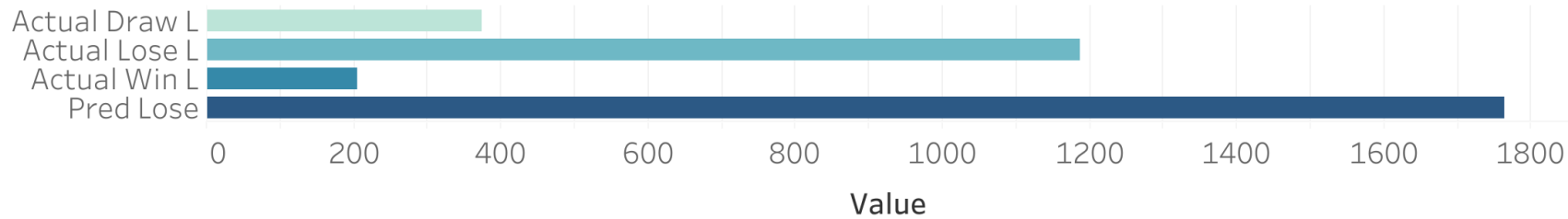
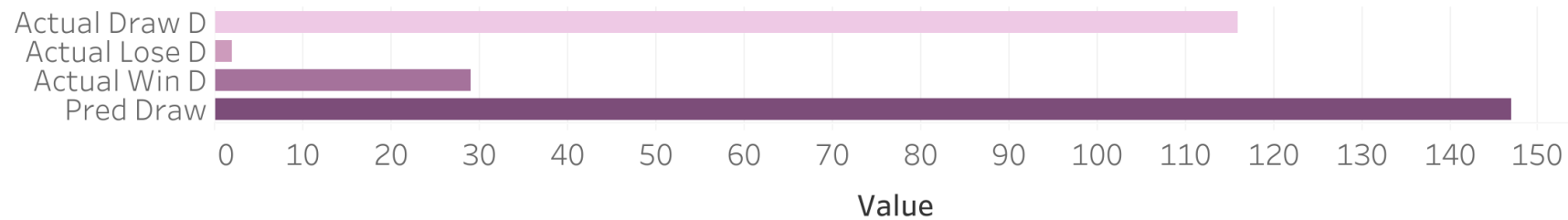


Prediction results of matches



Premier
League

- Results of the classification model:
 - Accuracy score: 0.71



Premier
League

Conclusion



Premier
League

Conclusion

- In this study, the relationships between football statistics are analyzed.
- The match results are predicted by the classification model with 0.71 accuracy score.
- Future studies:
 - Prediction of the number of goals to be scored in each match using regression models.
 - Prediction of the match results which are going to be played in the future.



**Premier
League**

References

- The dataset of football statistics is scraped from <https://fbref.com/en/> website.
- Premier League standings for the last 22 seasons (www.kaggle.com)
- Web scraping videos of Dataquest Youtube Channel.



Premier
League

Thank you & Questions



**Premier
League**