

Hypothesis Testing on the European Soccer Database

Daniel Martinez, William McMillan, Zach Timmons

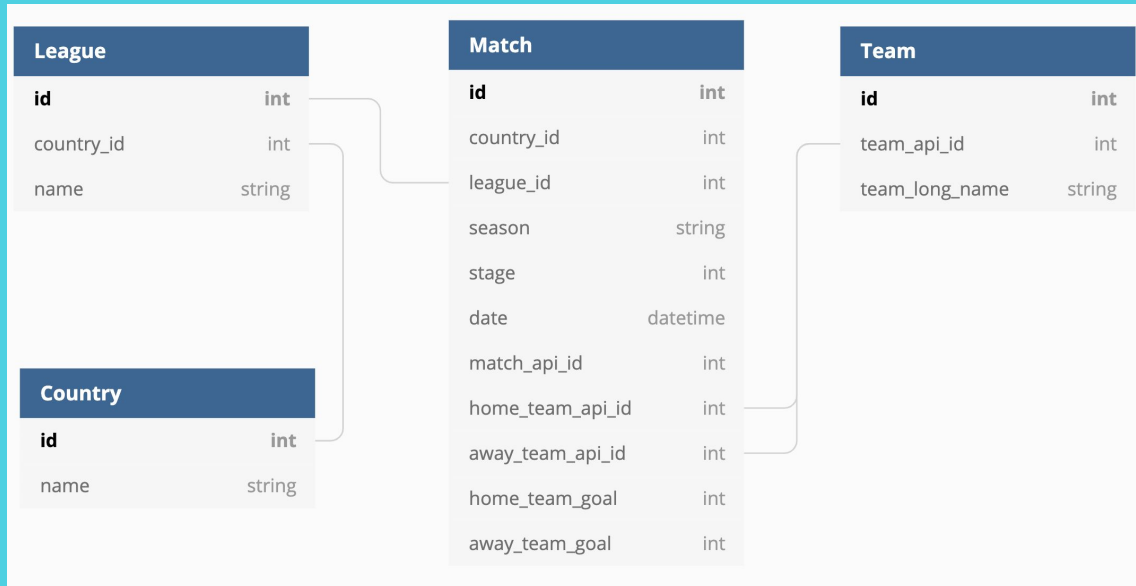
Outline

- **Database**
- **Hypothesis Testing**
- **Home/Away**
- **England Home**
- **Scotland Away**
- **Conclusions**
- **Future Improvements**

Database

European Soccer Database:

<https://www.kaggle.com/hugomathien/soccer>

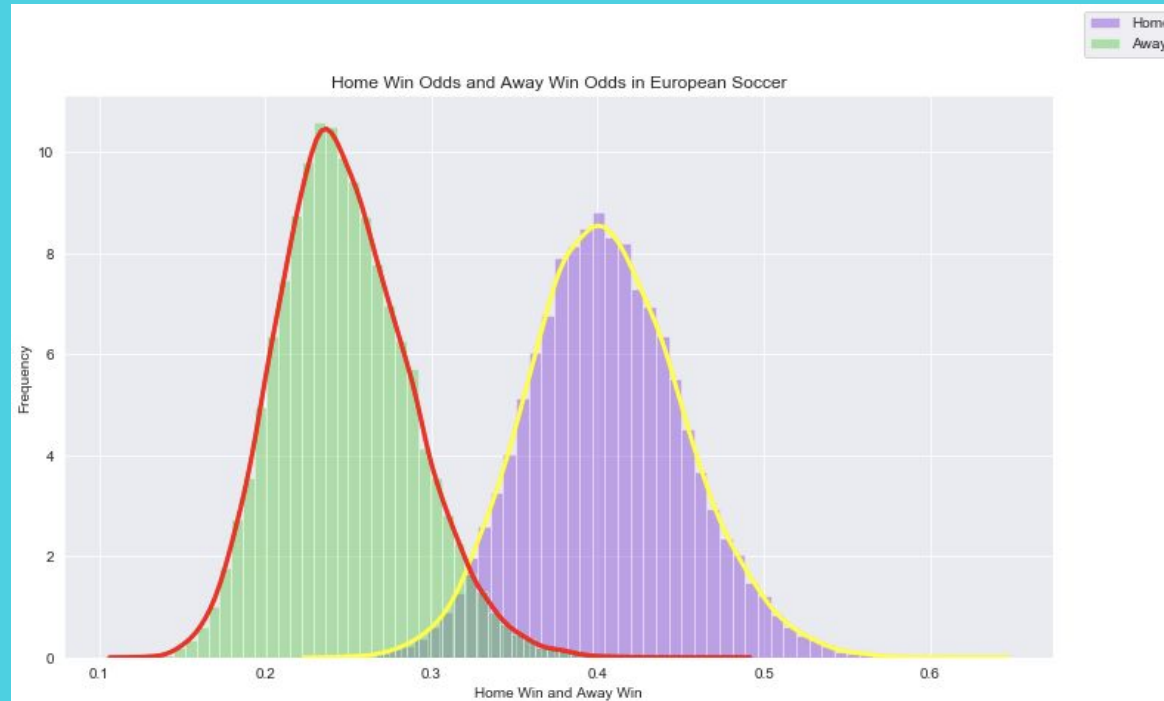


Hypothesis Testing

Three Questions asked during our Hypothesis Testing:

- Is there a statistical difference when a team plays in front of a home crowd?
- Is there a statistical significance in the odds of winning at home in England than in the rest of Europe?
- Is the higher away winning percentage in the Scottish Premier League statistically significant?

Do European Soccer Teams Win More at Home?



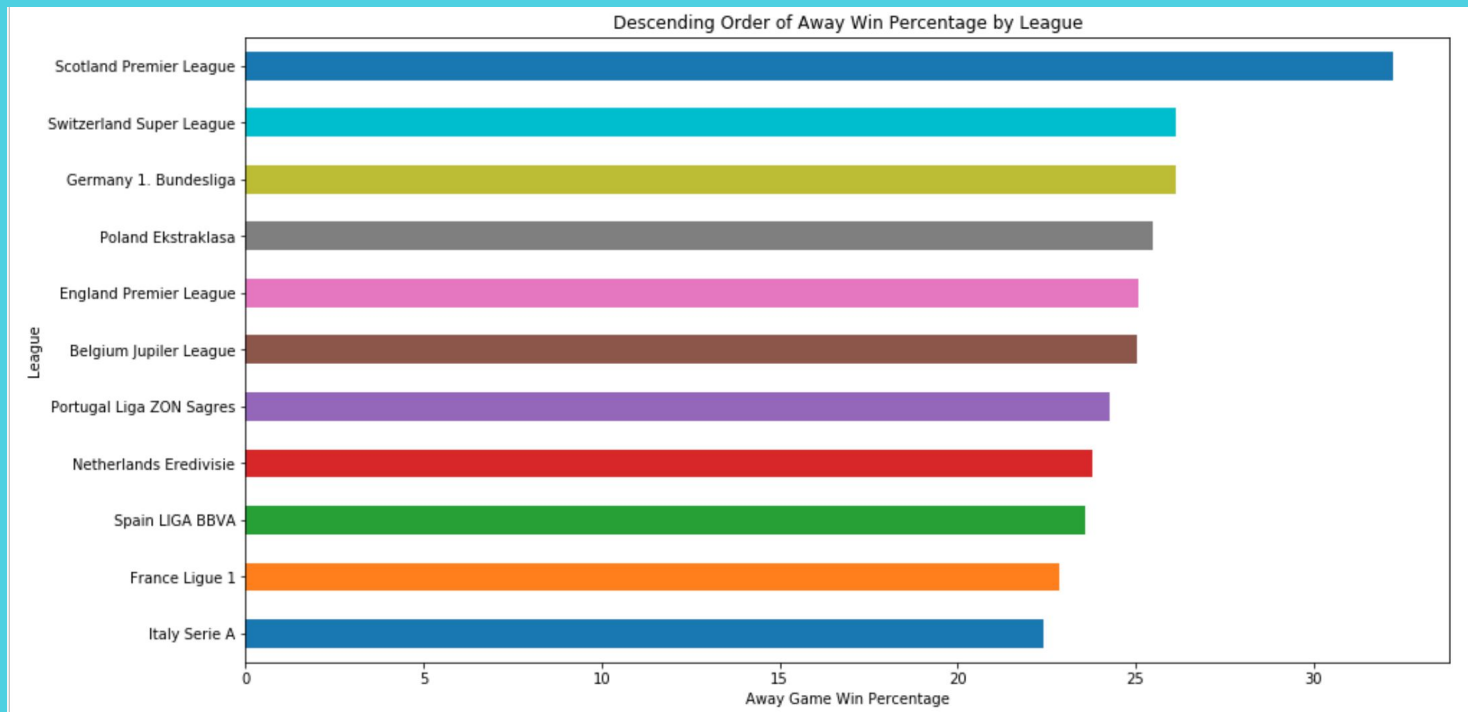
England Home Win %

H_0 : There is no difference in home win percentage between the English Premier League and the rest of Europe

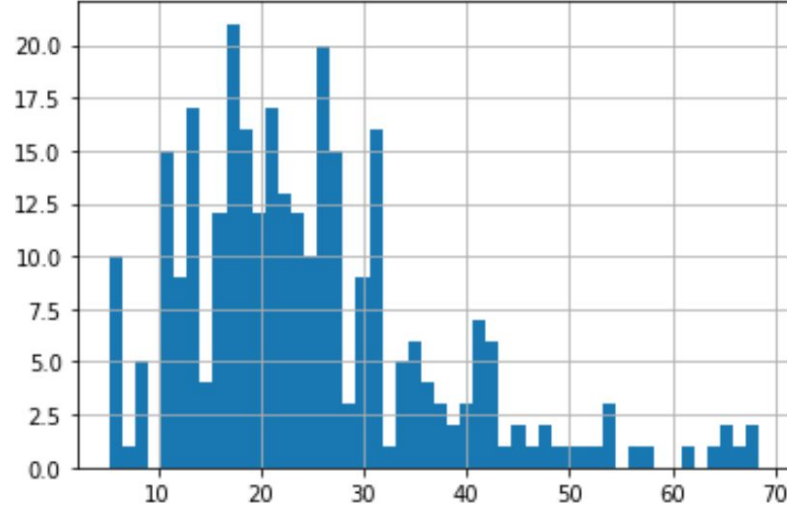
H_a : The English Premier League has a higher home win percentage than the rest of Europe



A Stark Difference in Away Win %



Population Distribution of Away Win Percentage

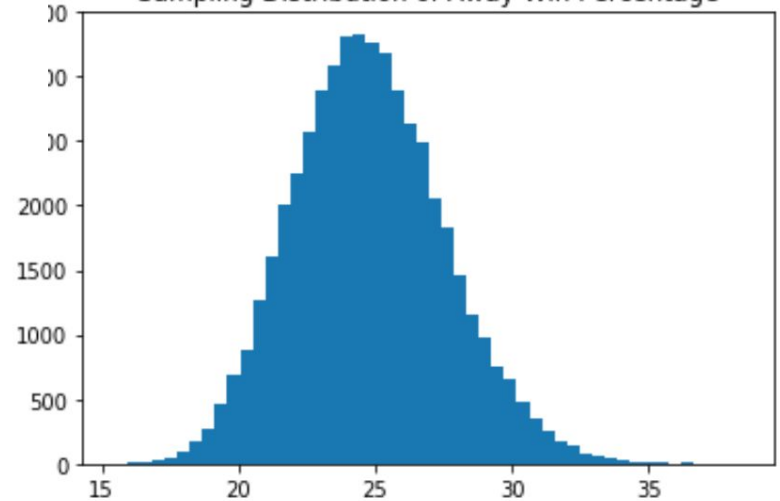


'Away_Win_%' Population Mean = 24.77
Standard Deviation = 12.42

$p = 0.006725$

Statistically significant at the 0.05 level

Sampling Distribution of Away Win Percentage



Conclusions

- There is indeed evidence for the proverbial 'home field advantage'
- There is a surprising amount of parity in European soccer
- Parity is even more pronounced in less renowned leagues

Future Improvements

- **Spend more time with data**
- **Look into betting company data**
- **Create better predictors using Fifa data**
- **Spend more time with data**