SHE PLAYS

Apoorva Angre,
Sandya Madhavan
Sanjana Rajagopala
Shefali Vajramatti

PROBLEMS AND OBJECTIVES

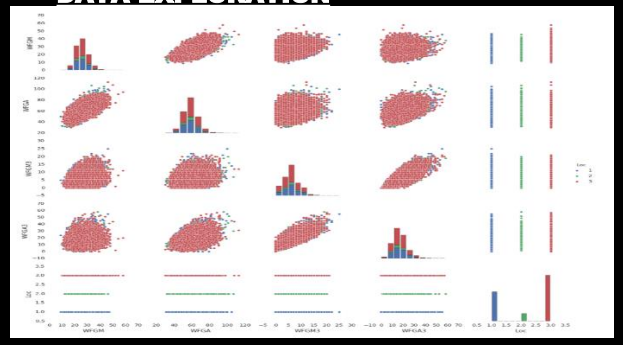
Sports Analytics is defining new opportunities to maximize the potential of players for a better outcome. Today, sports agencies depend on analytical platforms to devise game strategies and to engage investors more. This is majorly missing in the Women's sports section.

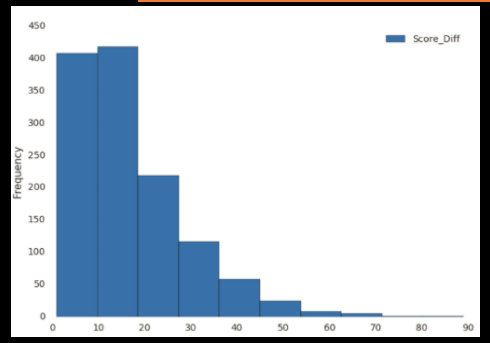
Our project aims at developing a predictive model based on the historical data of Women's basketball tournaments from 1998-2017. This model estimates the win probabilities of teams with lower ID in a particular match game. Our project can help in understanding which women teams might require more exposure for professional training/coaching.

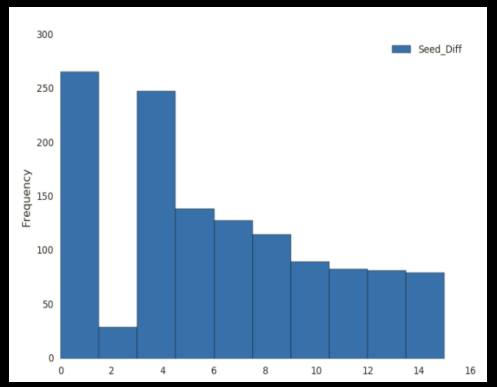
DATA DESCRIPTION

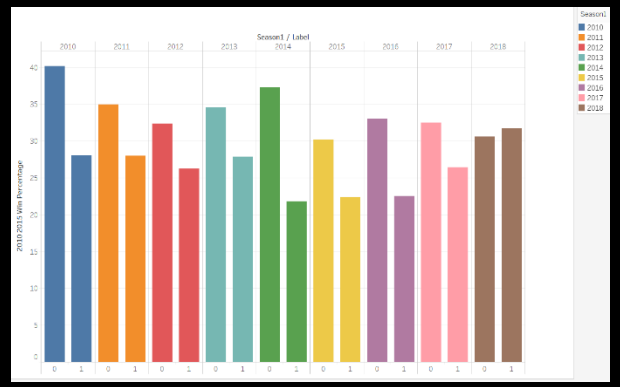
The dataset was acquired from the Google Cloud & NCAA ML Competition 2018- Women's on Kaggle. It includes three datasets categorized based on years- 1998-2017, 2010-2017 and 2018. We integrated mapping between team ids, names, seeds with the above datasets to form the consolidated data.

DATA EXPLORATION









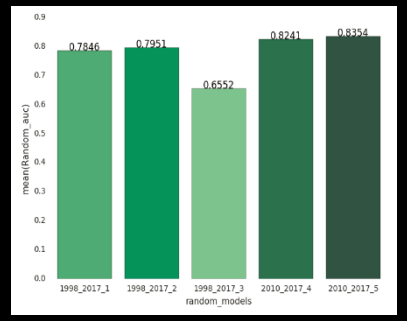
Score & Seed Diff, Location, Overtime, Win % Range, Goals, 3-Pointers, Free throws, Rebounds, Fouls, Steals, Blocks, Assists, Turnovers

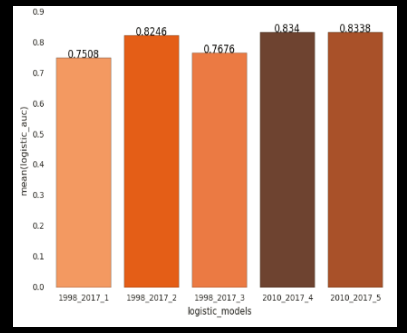
MODEL DESCRIPTION

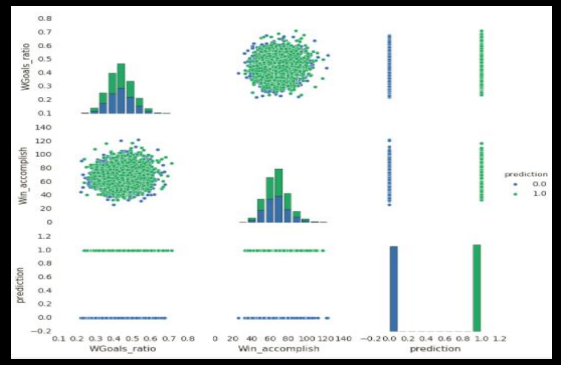
Training:60% Validation:30% Testing:10%

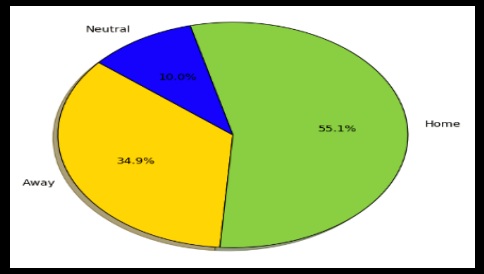
Model Name	Logistic Regression	Random Forest
1998-2017 Model 1	0.7508	0.7846
1998-2017 Model 2	0.8246	0.7951
1998-2017 Model 3	0.7676	0.6552
2010-2017 Model 1	0.8282	0.8287
2010-2017 Model 2	0.834	0.8241
2010-2017 Model 3 (with cross validator)	0.8338	0.8354

MODELS COMPARISON

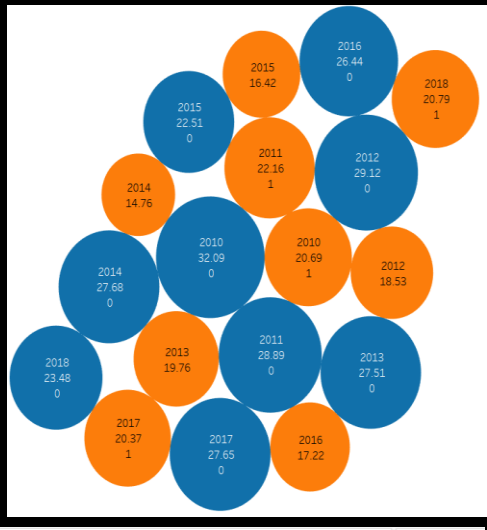


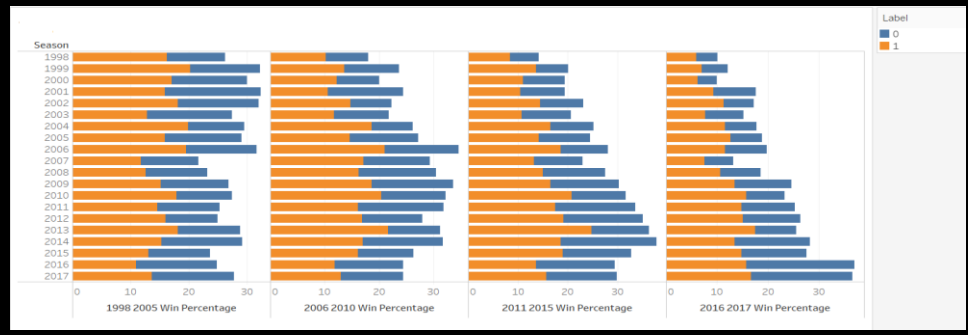






RESULTS & INFERENCE





CONCLUSION

- The percentage of the teams that won on home ground was observed to be significantly more than the ones played outside.
- The match characteristics related to goals, fouls and 3-pointers were found to have a positive effect on the winning probability of the teams in a match game.
- The past performance of the individual teams across different year ranges- 1998-2005, 2006-2010, 2011-2015, 2016-2017 was observed to have a strong impact on the prediction; predominantly the performance in 2011-2015

Random Forest 2010-2017 Model 3 Test AUC: 0.8016

([-0.2886, -0.1441, -0.15, -0.0028, 0.0356, 0.8083, -0.7344, 0.2873, 0.6106, 0.4616, -0.7572, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, -0.756, 0.4703, -0.4535, 0.0, -0.3633, -0.1328, 0.3662, -0.2257, -0.3653, 0.0, 0.1809, 0.1754, 0.4585])