



Introduction

In this project, we aim to identify a set of non-revealing statistics that are predictive of winning in basketball. We selected defensive/offensive rebounding, 3-point attempts, and forced turnovers based on our understanding of the game.

Existing predictive models and analysts are far too inaccurate with live game-by-game predictions and end of season championship odds that we decided a more transparent, simpler model was needed in order to summarize the evolution of the game.

Objectives

Find the correct weights for each desired stat for our team score algorithm.

Create an automated program that will predict the set of stats and declare the winning team with a greater accuracy than existing models.

Deliver data via website.

Materials and Methods

Basketball Reference is a data source that we used in order to gather full season stats from the last decade in order to decide what stats to focus on and what stats to ignore.

NBA-API allows developers to access data and statistics related to the NBA. This is used to gather in season data from past games.

Random Forest Classifier is a machine learning algorithm that can make accurate and robust predictions based on previous data that we use in order to predict stats for upcoming games.

Team Score is the calculated value that we give each teams based on weights given to each stat. This algorithm is updated in order to achieve that highest predictability percentage possible.

For every scheduled game, the program predicts the stats of both teams using the random forest classifier algorithm with data accessed from the NBA-API.

The team score and confidence interval is calculated and posted on to the website.

Results

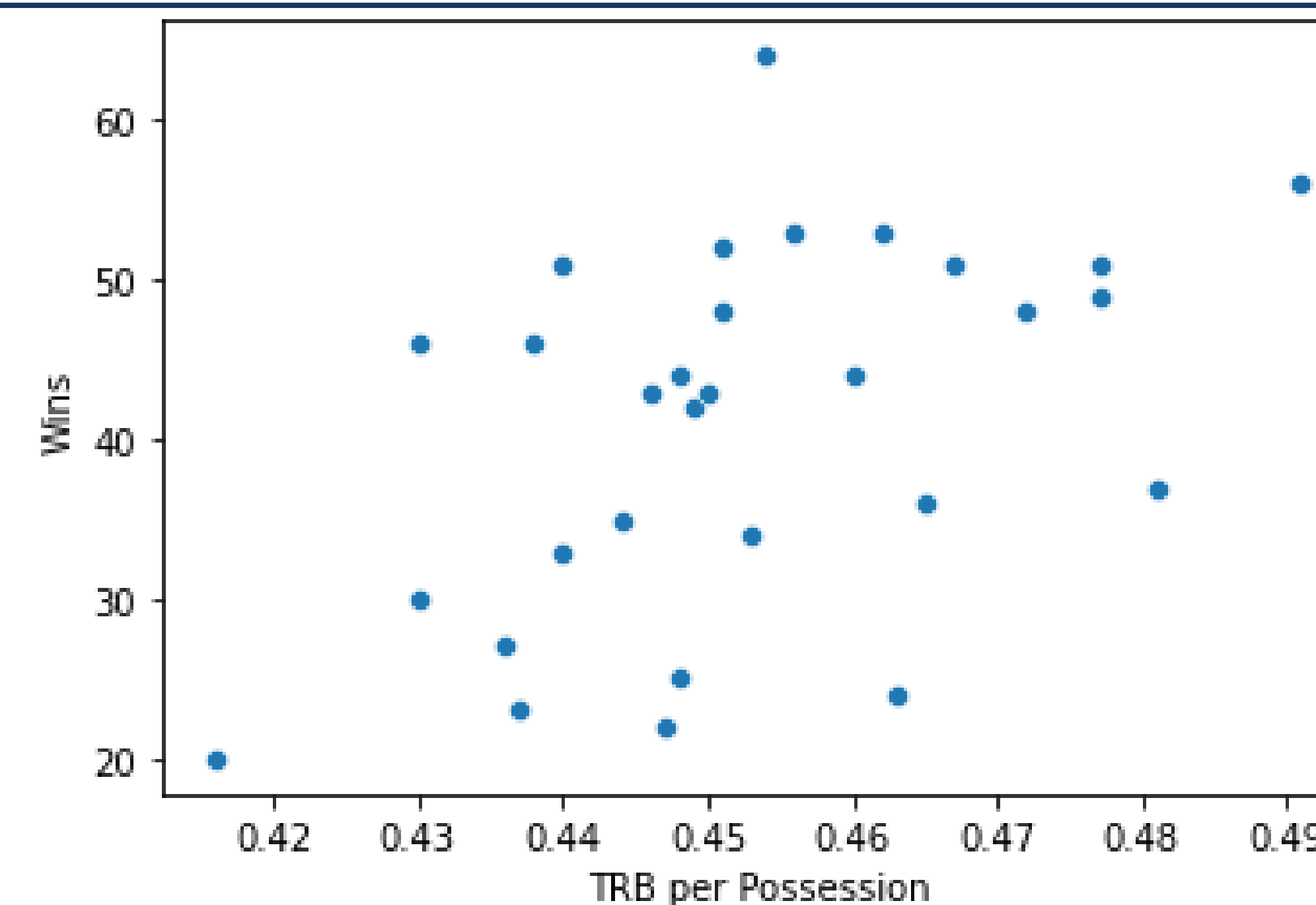


Figure 1 — Total Rebounds / Possession vs Winning (2010 – 2020)

A medium positive correlation of .47 between the number of rebounds to the number of wins in a season

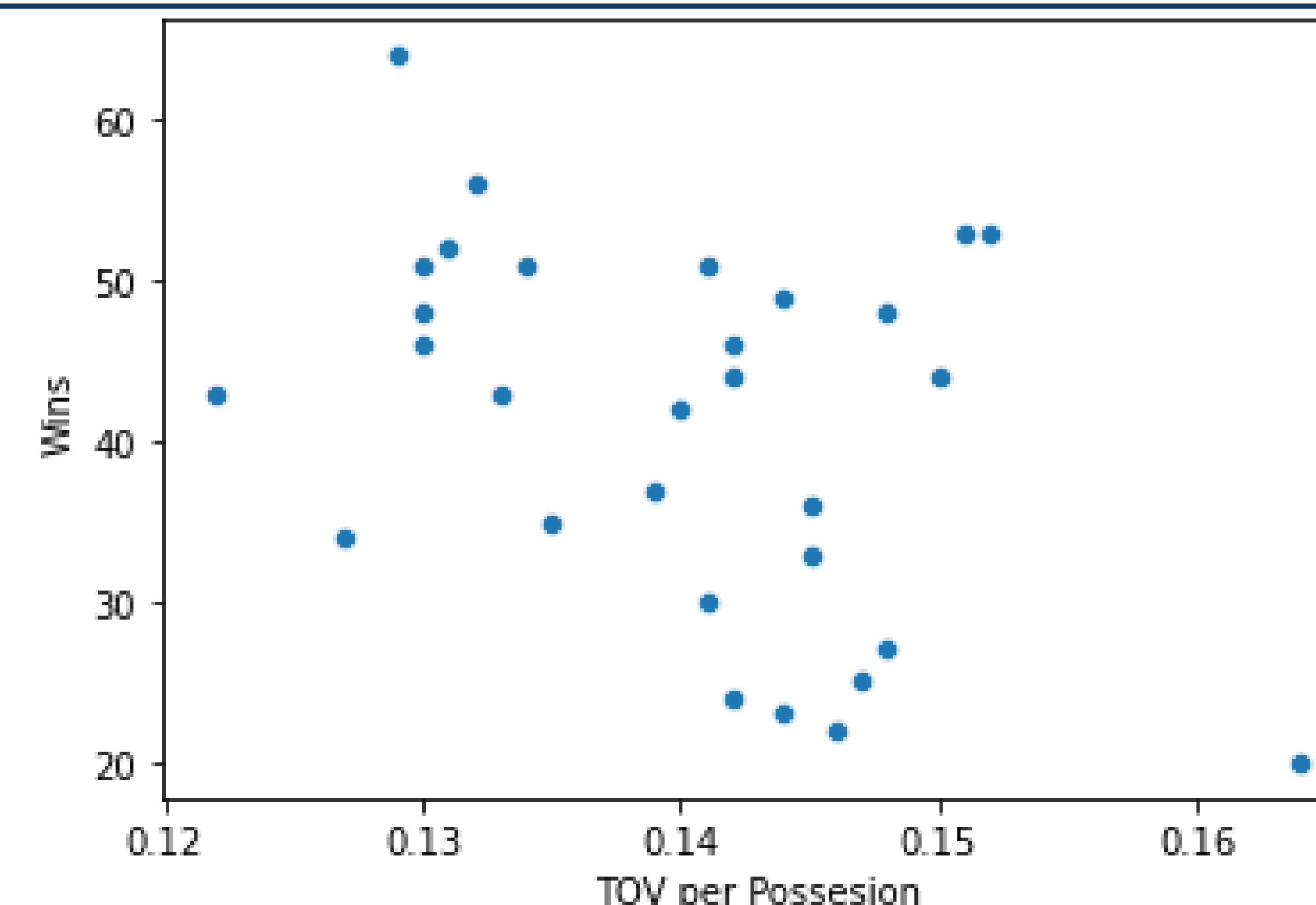


Figure 2 — Total Turnovers / Possession vs Winning (2010 – 2020)

A medium negative correlation of .42 between the number of turnovers to the number of wins in a season

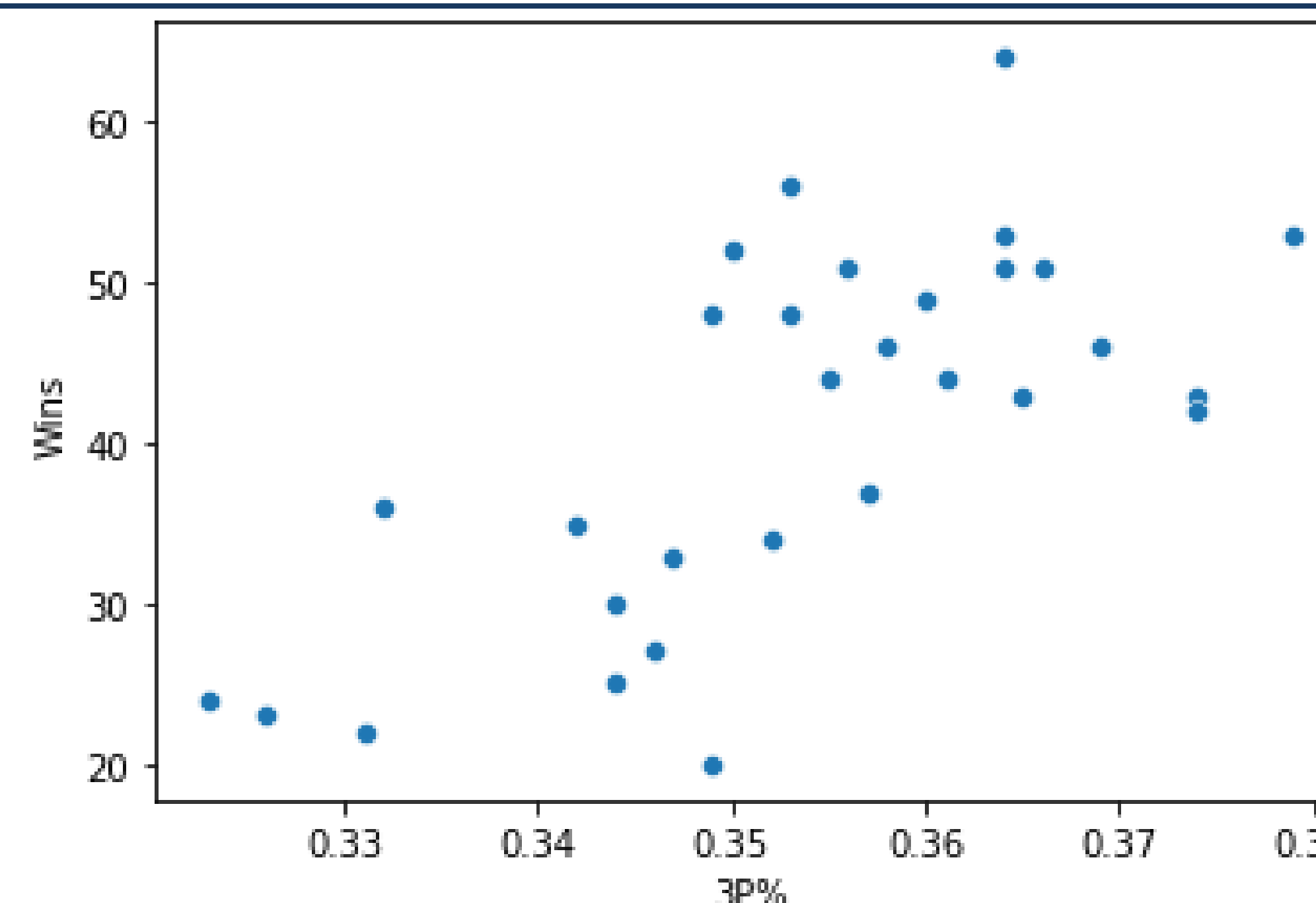


Figure 3 — 3 Point % vs Winning (2010 – 2020)

A strong positive correlation of .69 between the 3-point percentage to the number of wins in a season

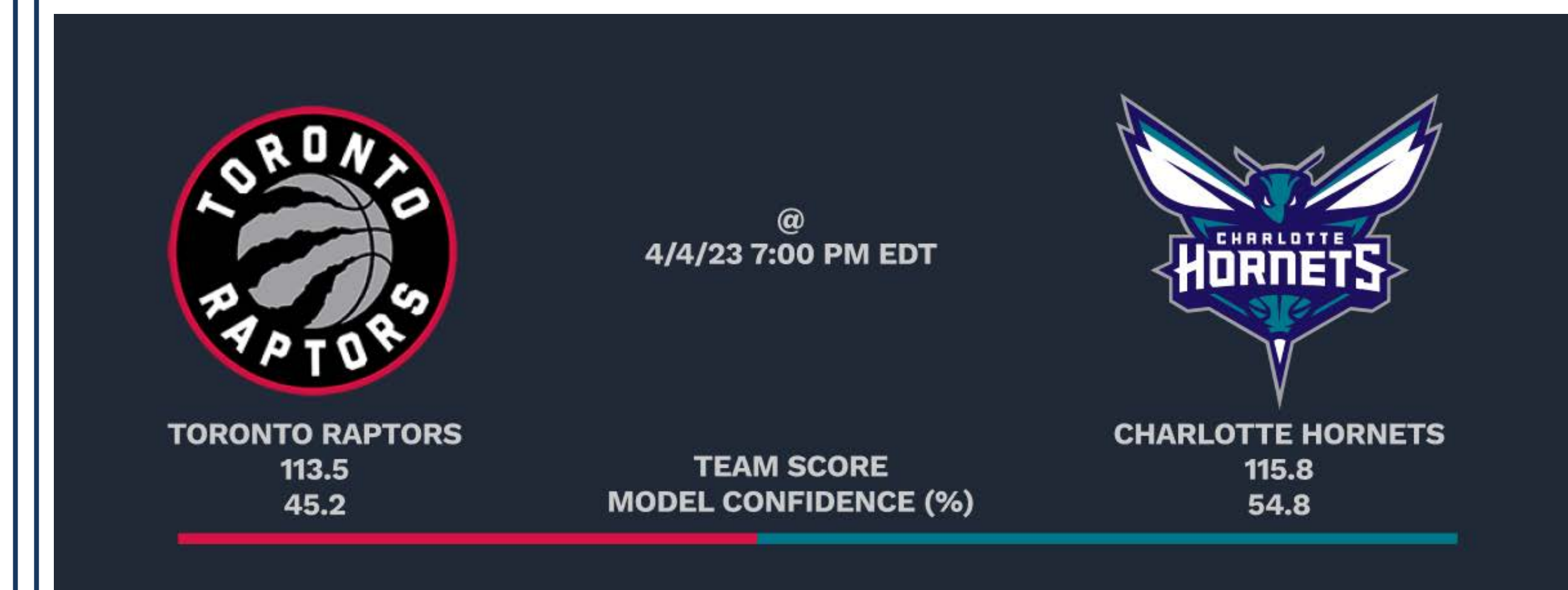


Figure 4 – Game-by-Game Prediction Visual

The scheduled game, team score, and confidence interval is displayed

Conclusion

The algorithm to calculate teams scores have yielded ~60% accuracy with our predictions for games in the 2022-2023 season while existing models have ~65% accuracy. However, our system is automated and with more trial and errors, a greater accuracy rate should be able to be achieved.

Website is up and running with good user interface and a clean design for easier access and navigation to our data.

Future Work

Machine learning can be used in order to find the best weights for our stats. Given multiple trials, we should be able to run a machine learning algorithm that can return the best possible weight values for predictive accuracy.

A new season also means the possibility of changing these stats that we declared as important in the 2022-2023 season. We can look at how the game will evolve once more with rule changes, this season's unexpected variables, and post-season outcome to change what's necessary.

References

- Basketball Statistics & History of every Team & NBA and WNBA players. Basketball. (n.d.). Retrieved 2022, from <https://www.basketball-reference.com/>
- Luszczyszyn, D. (2016, July 13). [web log]. Retrieved from <https://hockey-graphs.com/2016/07/13/measuring-single-game-productivity-an-introduction-to-game-score/#more-9550>.