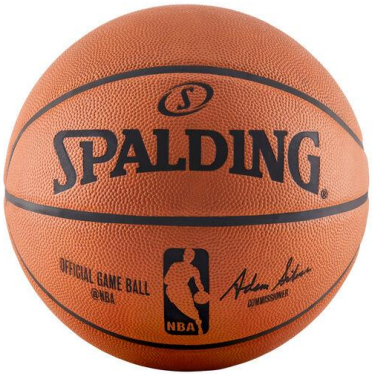
Predicting Winners and Losers for NBA games at Halftime

J. Garrecht Metzger

(Metis Winter Cohort 2021)



Overview

Overview

1. Background

Overview

1. Background
2. Dataset

Overview

1. Background
2. Dataset
3. Model

Overview

1. Background
2. Dataset
3. Model
4. Summary & Future Work

Inspiration for study?

Inspiration for study?

- Common complaint is only the later half of the game matters

Inspiration for study?

- Common complaint is only the later half of the game matters
  - Is this true?

Inspiration for study?

- Common complaint is only the later half of the game matters
  - Is this true?

- Use Case: Betting

Inspiration for study?

- Common complaint is only the later half of the game matters
  - Is this true?

- Use Case: Betting
  - Maybe make some avocado money off your friends?

Structure

Structure

- Split into 4 quarters = 2 halves.

Structure

- Split into 4 quarters = 2 halves.
- Every game *must* end with a winner and a loser.

Structure

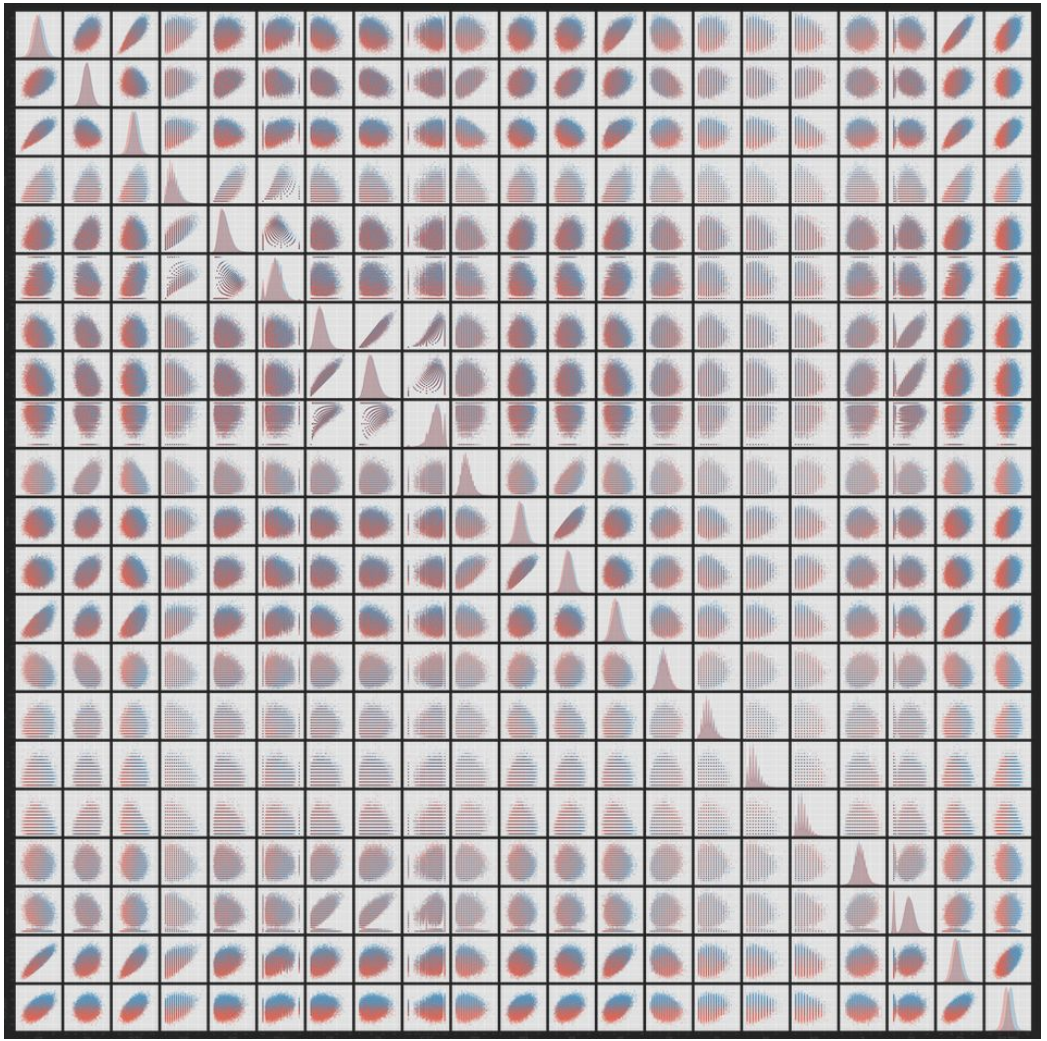
- Split into 4 quarters = 2 halves.
- Every game *must* end with a winner and a loser.
    - Balanced targets

Structure

- Split into 4 quarters = 2 halves.
- Every game *must* end with a winner and a loser.
  - Balanced targets
    - (i.e., count(win) = count(loss)

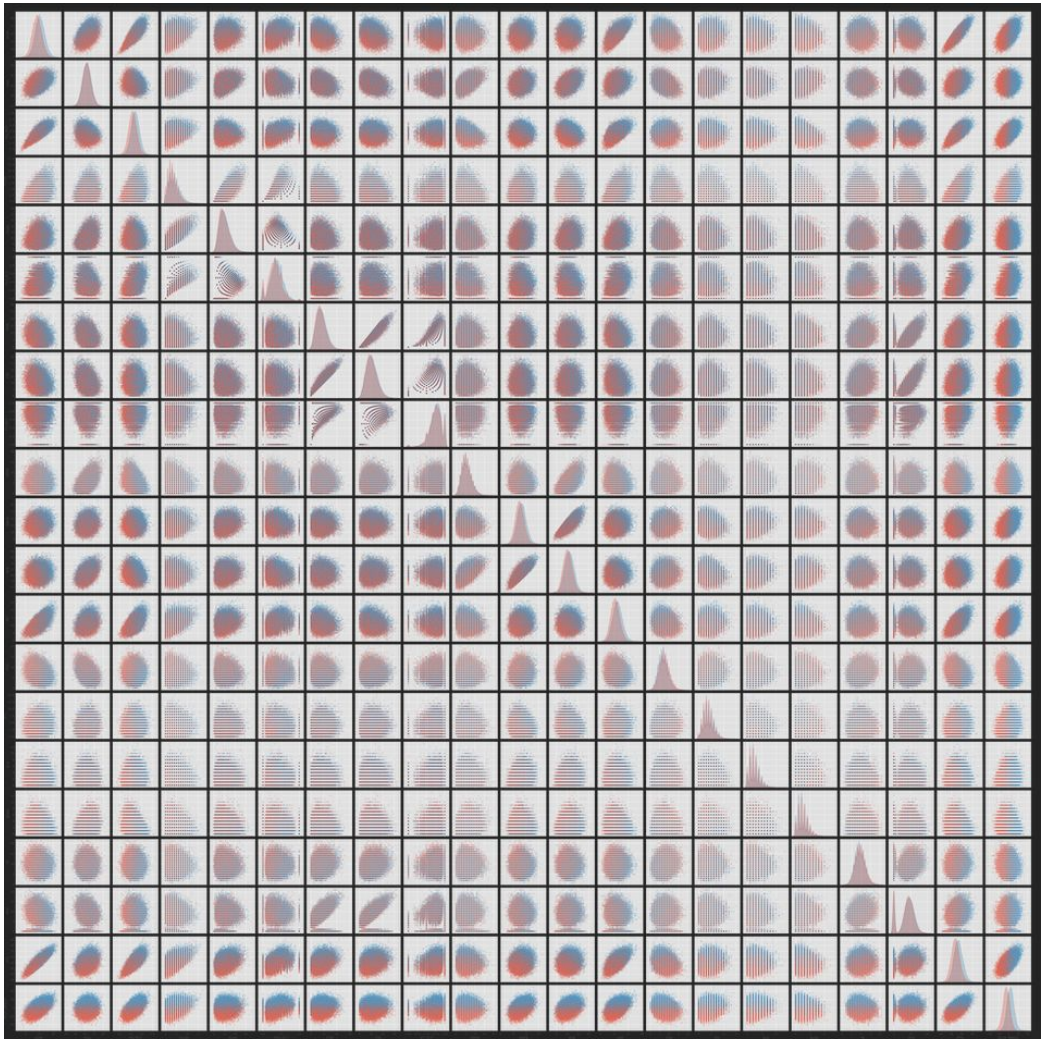
Data

Data



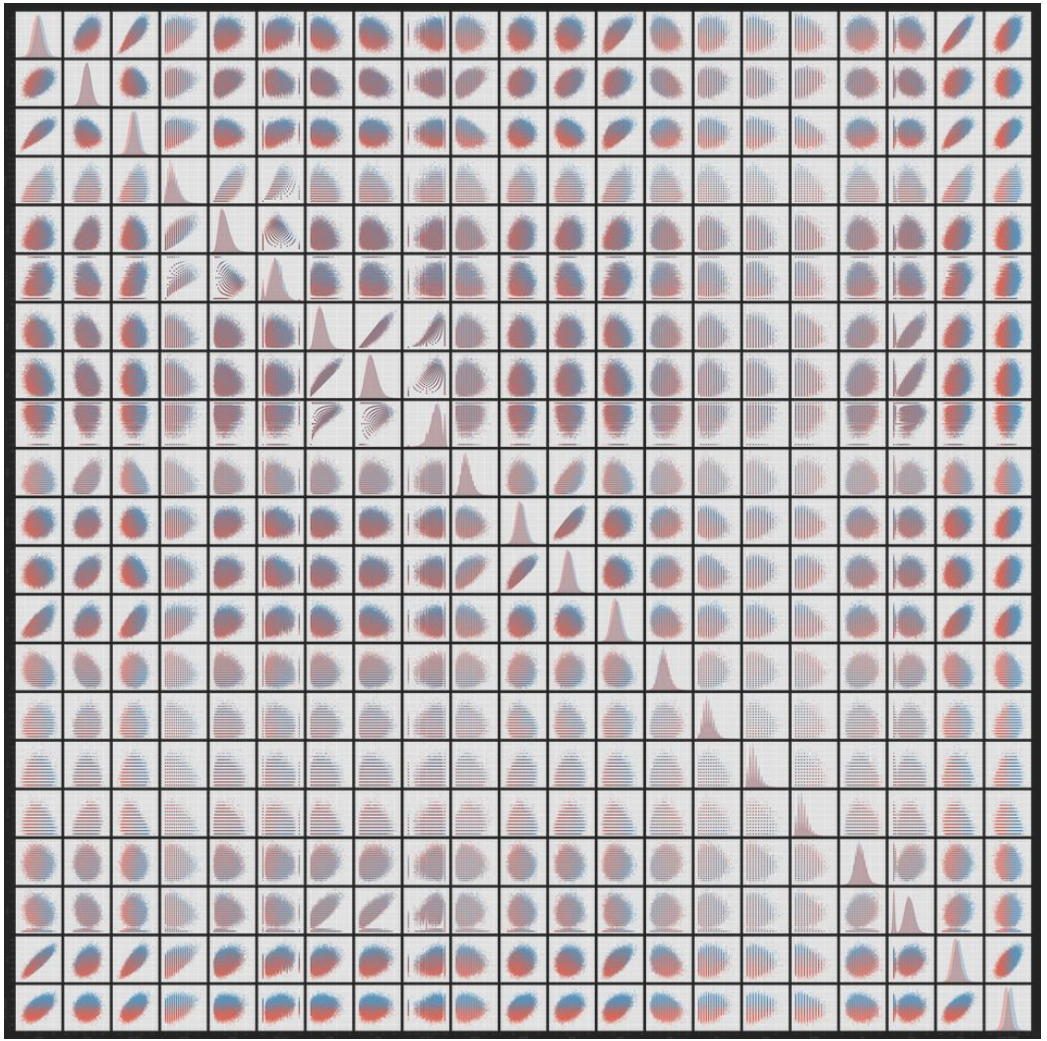
Data

- **20** Years



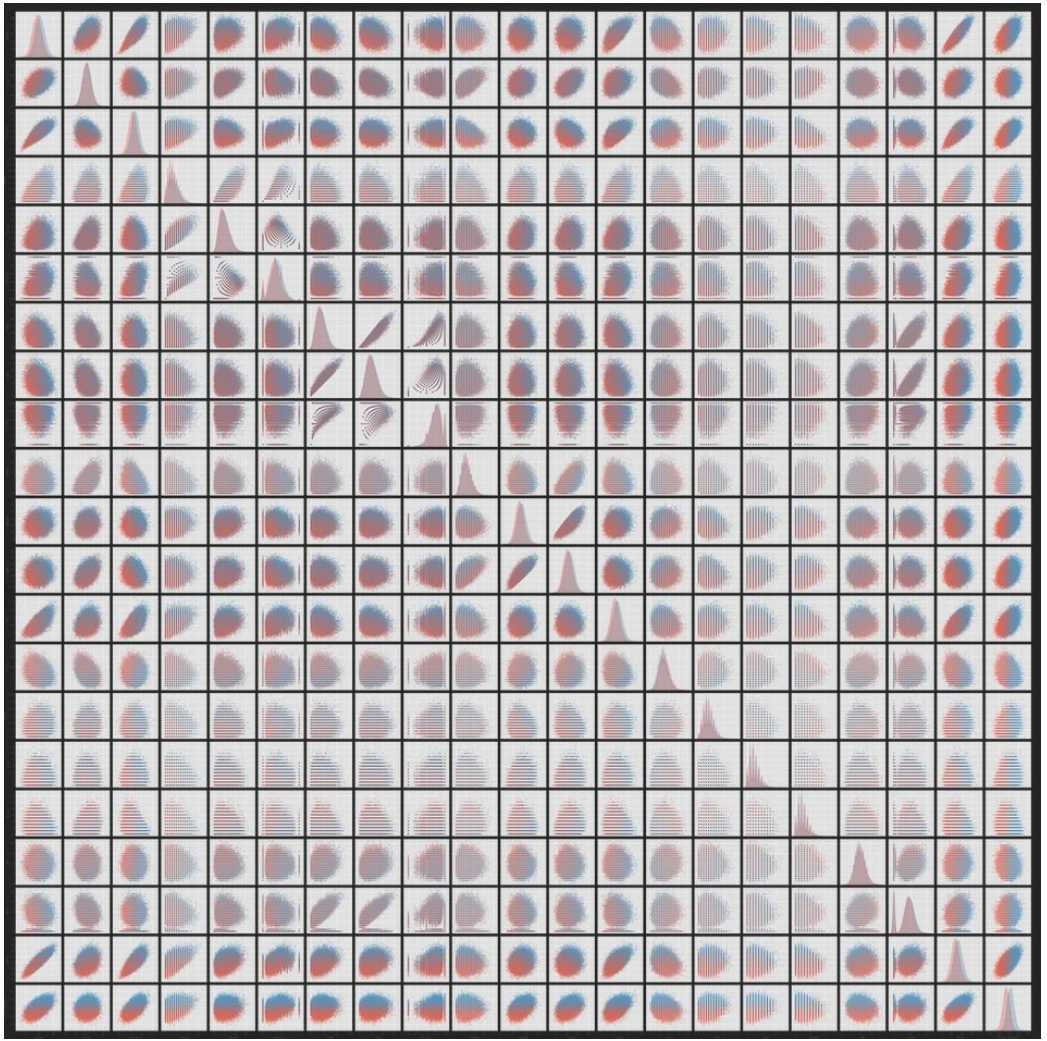
Data

- **20** Years
- **20** Seasons



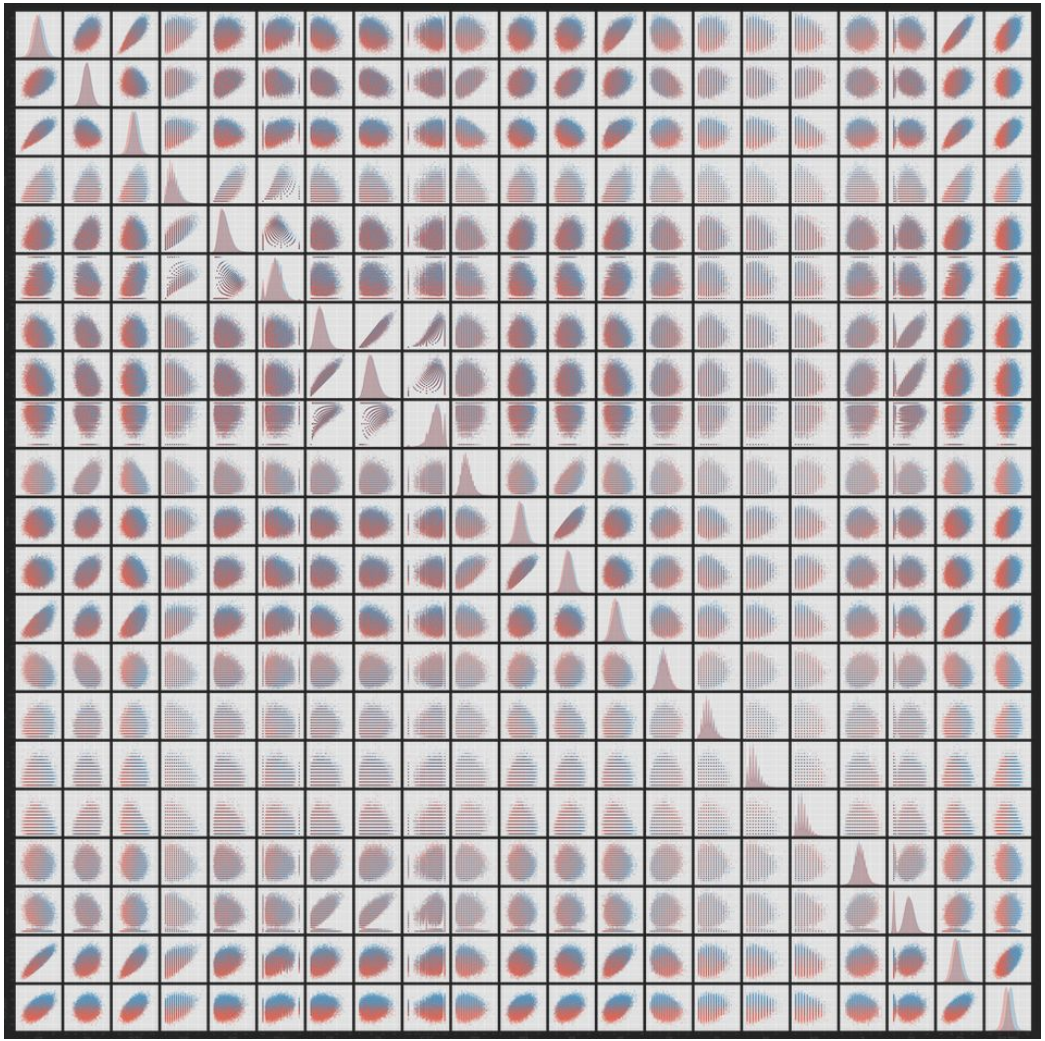
Data

- **20** Years
- **20** Seasons
- **30** Teams



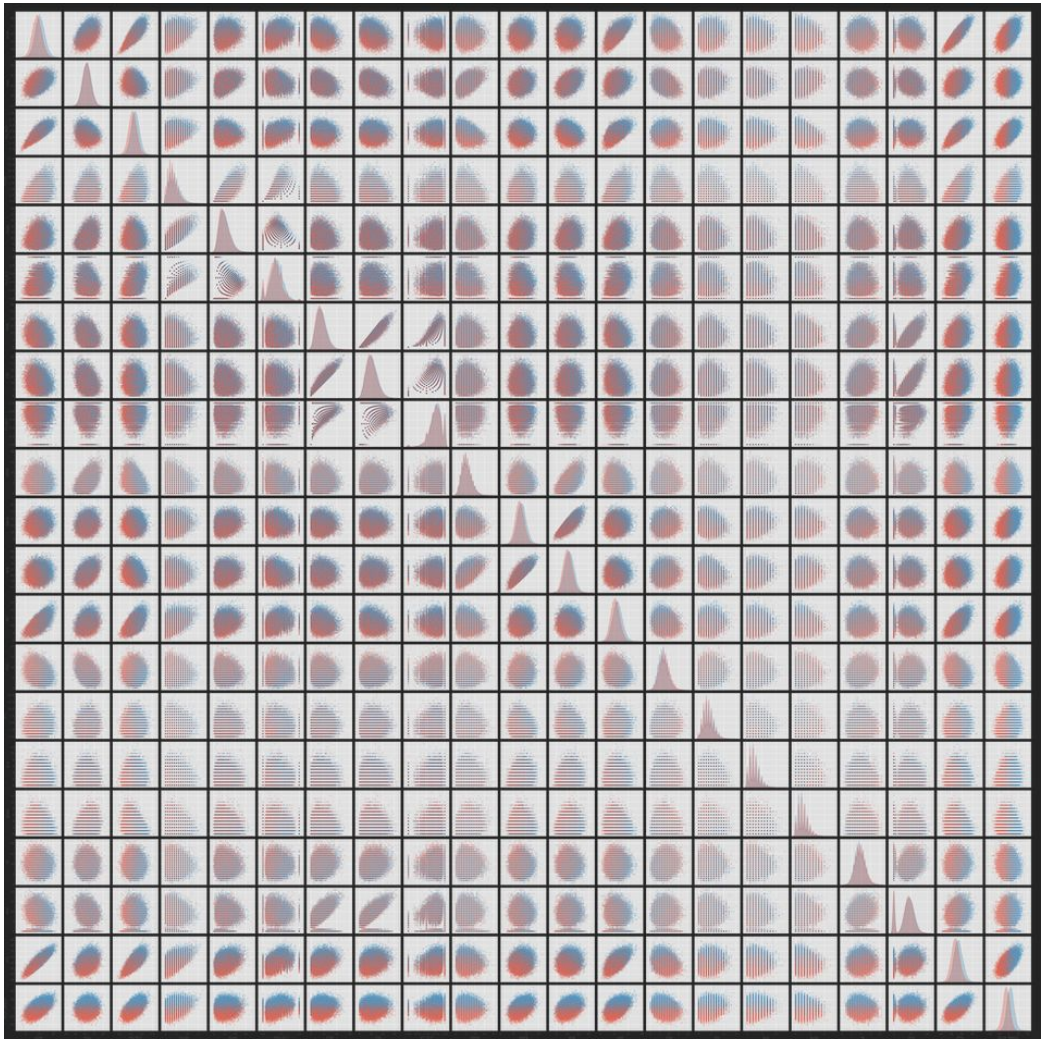
Data

- **20** Years
- **20** Seasons
- **30** Teams
- **23,144** Games



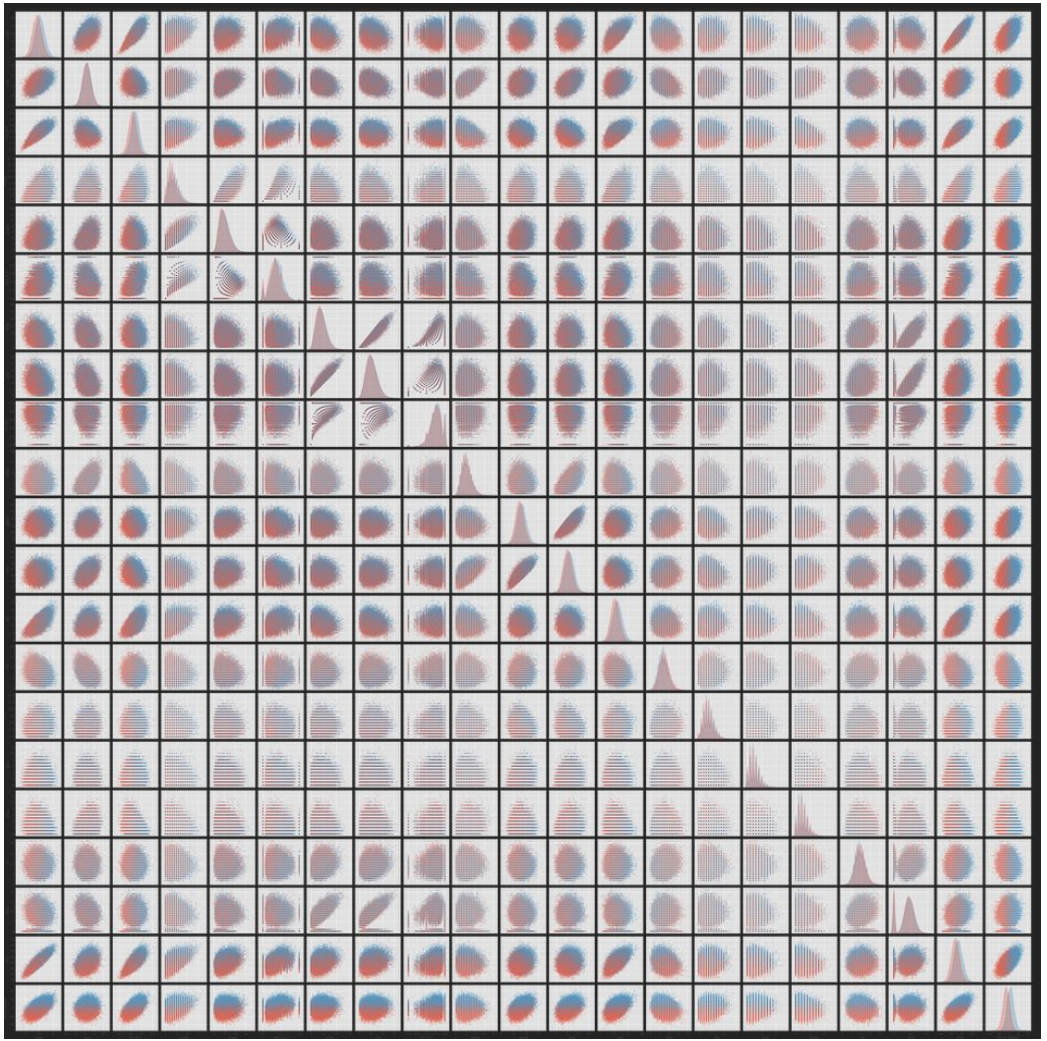
Data

- **20** Years
- **20** Seasons
- **30** Teams
- **23,144** Games
- **46,288** Rows



Data

- **20** Years
- **20** Seasons
- **30** Teams
- **23,144** Games
- **46,288** Rows
- **21** Features



Model Selection

Model Selection

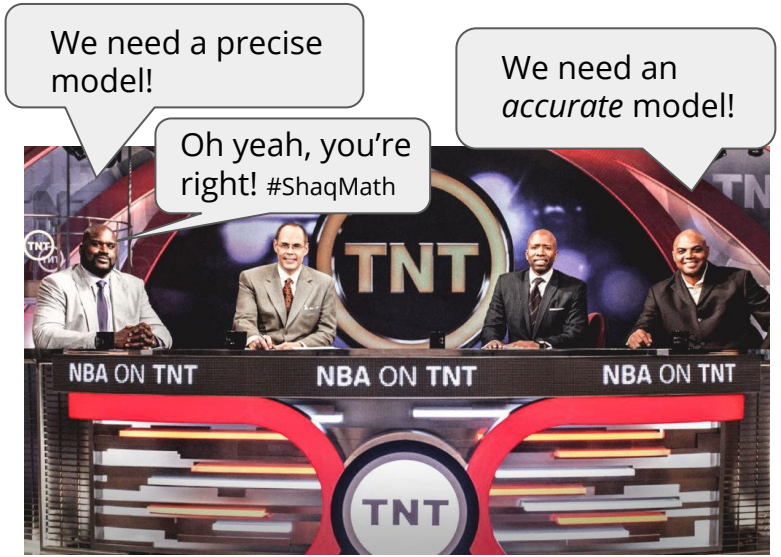
- What do we need?

Model Selection

- What do we need?
  - *Accuracy*

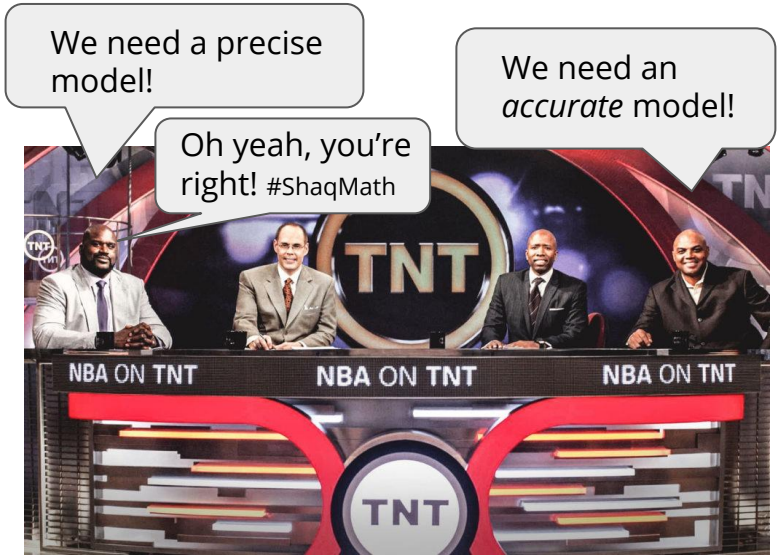
Model Selection

- What do we need?
  - *Accuracy*
  - Interpretable



Model Selection

- What do we need?
  - *Accuracy*
  - Interpretable



Model Selection

- What do we need?
  - *Accuracy*
  - Interpretable
- **Logistic Model**

Model Selection

- What do we need?
  - *Accuracy*
  - Interpretable
- **Logistic Model**
  - 21→ 2 Features

Model Selection

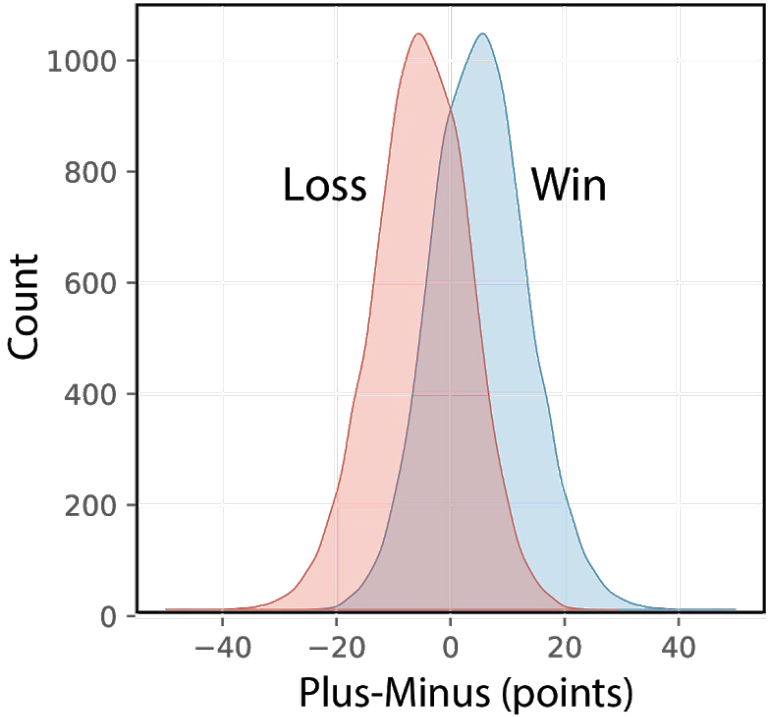
- What do we need?
  - *Accuracy*
  - Interpretable
- **Logistic Model**
  - 21→ 2 Features
    - “PLUS-MINUS”

Model Selection

- What do we need?
  - *Accuracy*
  - Interpretable
- **Logistic Model**
  - 21→ 2 Features
    - “PLUS-MINUS”
    - “SCORE”

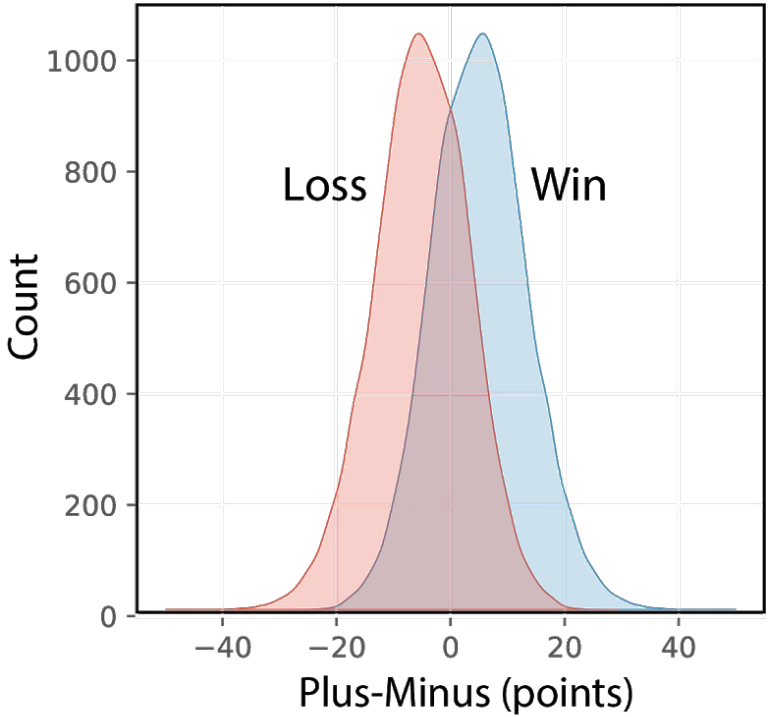
Model Selection

- What do we need?
  - *Accuracy*
  - Interpretable
- **Logistic Model**
  - 21→ 2 Features
    - “PLUS-MINUS”
    - “SCORE”



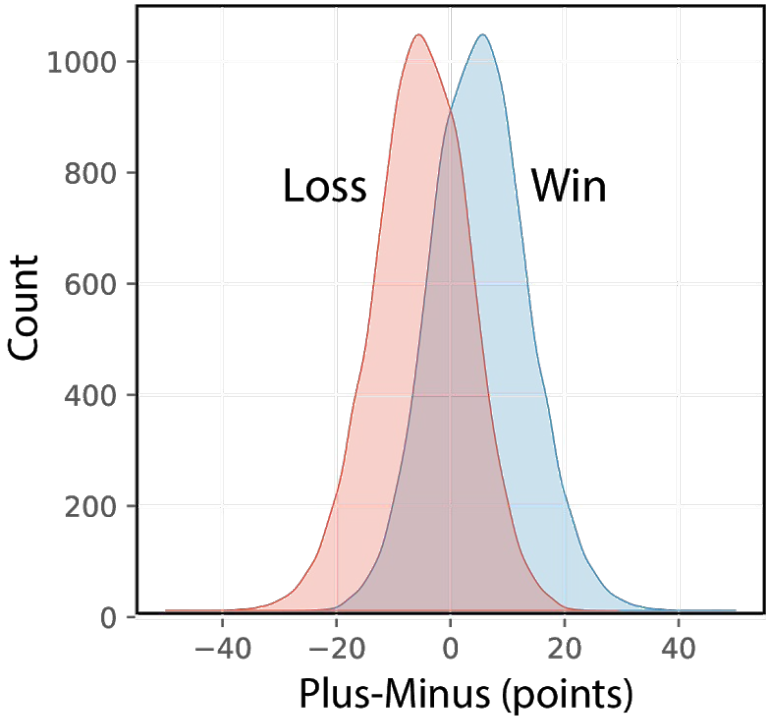
Model Selection

- What do we need?
  - *Accuracy*
  - Interpretable
- **Logistic Model**
  - 21→ 2 Features
    - “PLUS-MINUS”
    - “SCORE”
  - Polynomial transformation



Model Selection

- What do we need?
  - *Accuracy*
  - Interpretable
- **Logistic Model**
  - 21→ 2 Features
    - “PLUS-MINUS”
    - “SCORE”
  - Polynomial transformation
    - Accuracy: 0.72



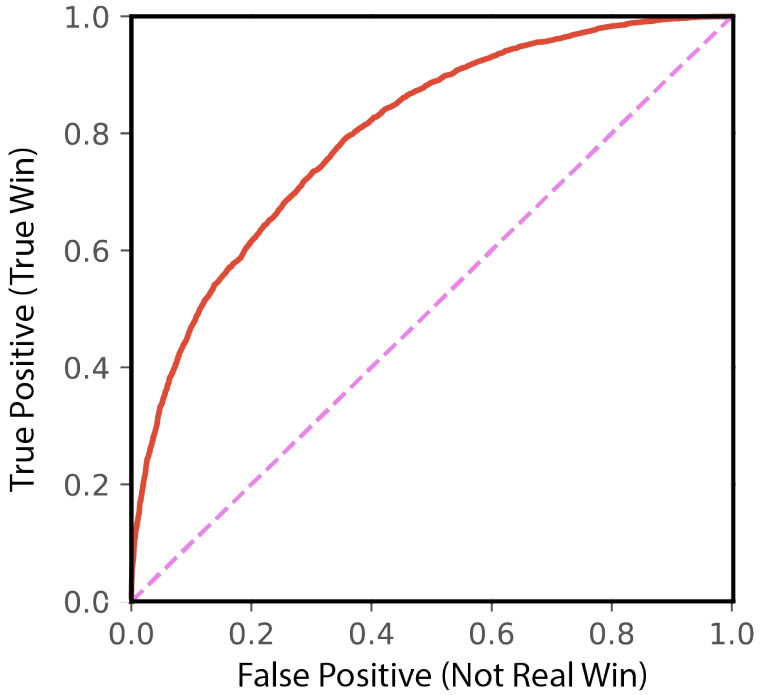
Logistic Model Evaluation

Logistic Model Evaluation

- *Accuracy* Score
  - **0.72**

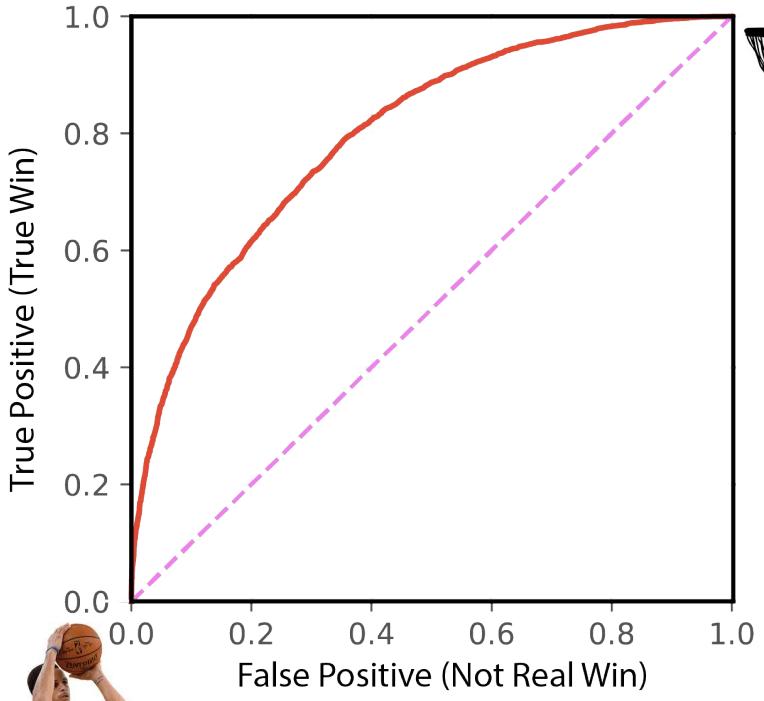
Logistic Model Evaluation

- *Accuracy* Score
  - **0.72**



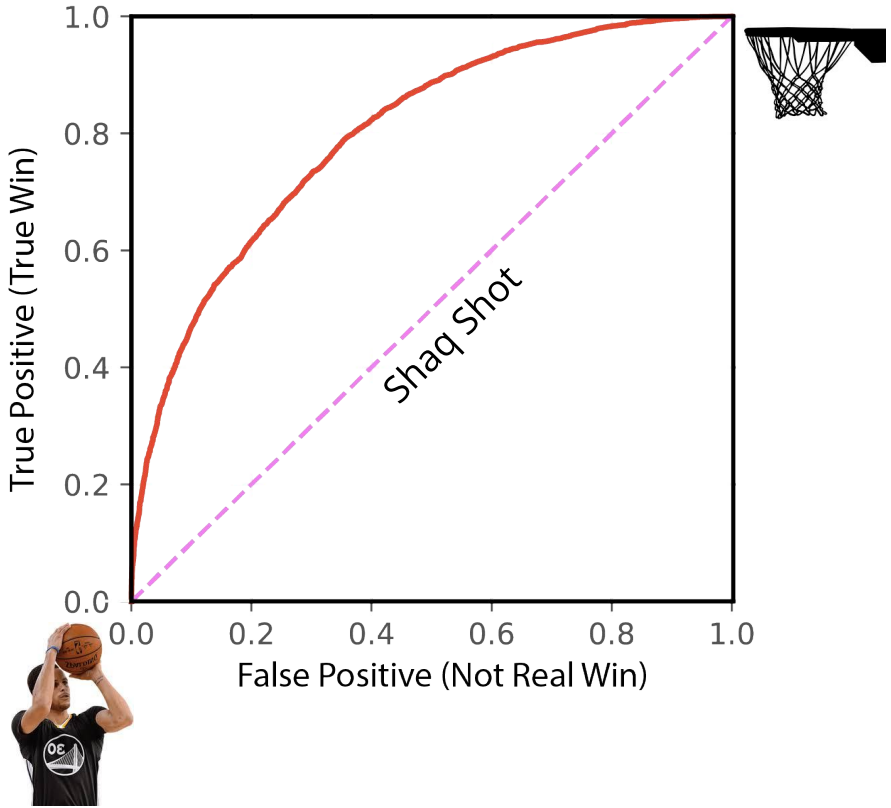
Logistic Model Evaluation

- *Accuracy* Score
  - **0.72**



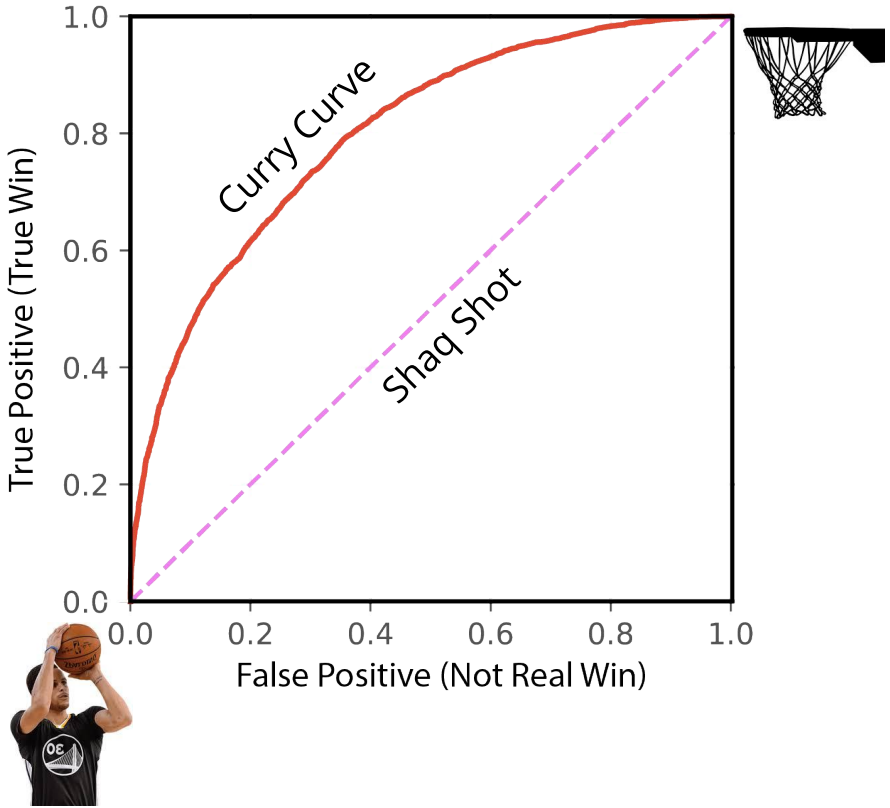
Logistic Model Evaluation

- *Accuracy* Score
  - **0.72**



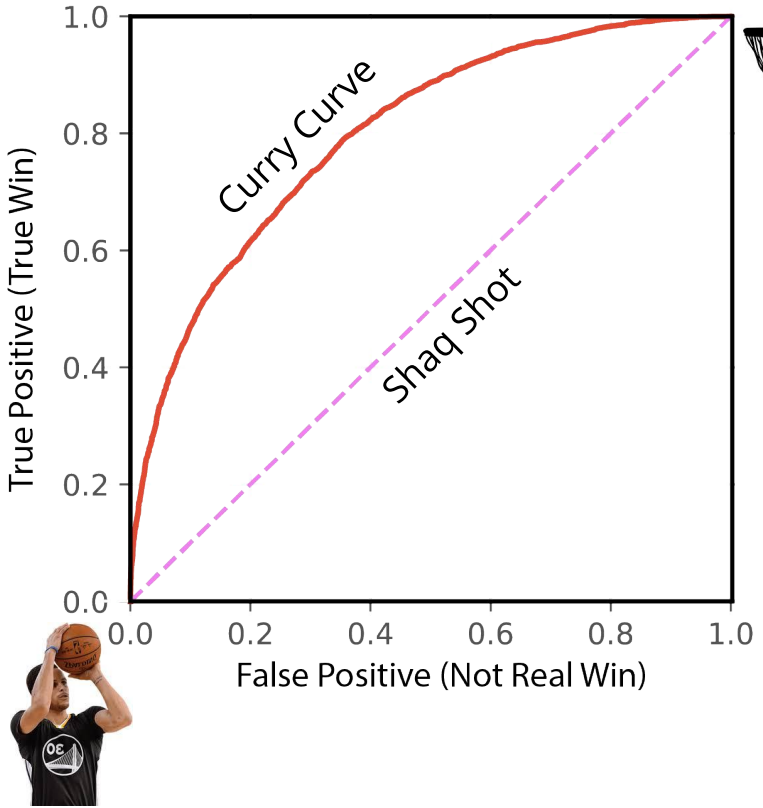
Logistic Model Evaluation

- *Accuracy* Score
  - **0.72**



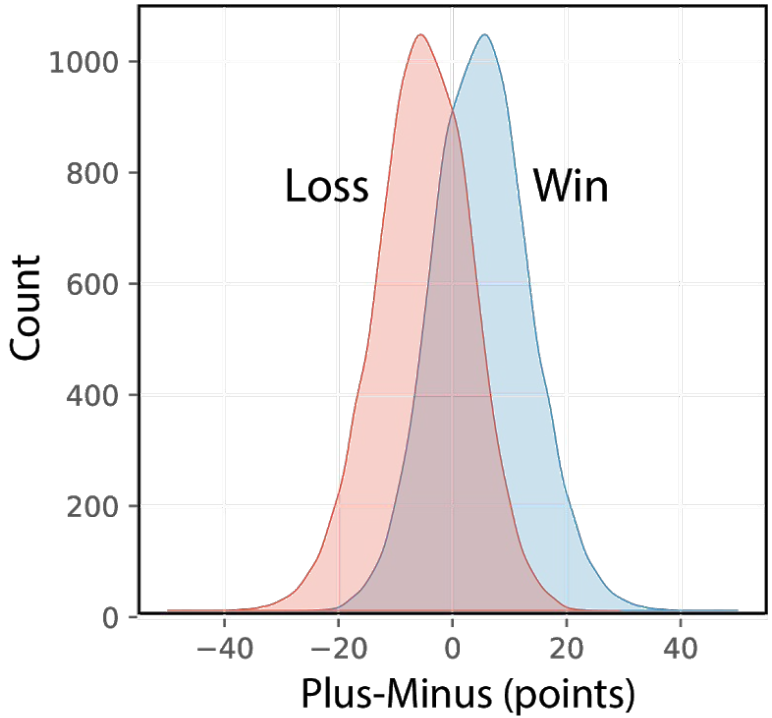
Logistic Model Evaluation

- *Accuracy* Score
  - **0.72**



- *ROAUC* Score
  - **0.8**

Point Difference is the principal feature



What did we learn?

What did we learn?

- The first half matters, a lot!

What did we learn?

- The first half matters, a lot!
- Score difference matters, a lot!

What did we learn?

- The first half matters, a lot!
- Score difference matters, a lot!

What do we want to learn next?

What did we learn?

- The first half matters, a lot!
- Score difference matters, a lot!

What do we want to learn next?

- Feature Engineering

What did we learn?

- The first half matters, a lot!
- Score difference matters, a lot!

What do we want to learn next?

- Feature Engineering
  - Create new time-dependent features

What did we learn?

- The first half matters, a lot!
- Score difference matters, a lot!

What do we want to learn next?

- Feature Engineering
  - Create new time-dependent features
  - More relative features (e.g., PLUS_MINUS, but rebounds).

What did we learn?

- The first half matters, a lot!
- Score difference matters, a lot!

What do we want to learn next?

- Feature Engineering
  - Create new time-dependent features
  - More relative features (e.g., PLUS_MINUS, but rebounds).
- More sophisticated analyses

FIN

Sources

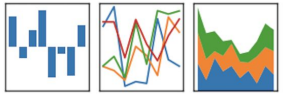
1. Jonathan Ferrey/Getty Images
2. https://pypi.org/project/nba-api/
3. https://github.com/seemethere/nba_py
4. https://github.com/swar/nba_api

Tools Used

NumPy

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



Maybe data goes over too many years or seasons?

“The Game Has Changed”

-*Daft Punk*, Tron: Legacy

Maybe 1 season is all you should consider? Or you could have some factor for what players are still there/composition of the team?

3 pointer has doubled in popularity