



# BUDT 703\_506\_12 Team Project Moneyball Metrics

## Team Members

Gunjan Suriya  
Harish Bhupathiraju  
Sai Devara  
Solayappan Ganesaan

12/7/2023

Last edited: 12/7/2023



FEARLESSLY  
FORWARD

# Background

The users would be the following

*Analyst and Administrators :*

Manages overall system configuration, user accounts, and database maintenance.

*Coach and Players :*

Accesses schedules and historical results.

Data Source:

<https://umterps.com/sports/baseball/schedule>



# Introduction

## Mission Statement

*To study the Terrapin Men's Baseball Team's past games and assist team leadership and analysts in gaining insights into the team's competitiveness over the years by analyzing matches played against opponents across different venues and leagues.*

## Mission Objective

- Find the ratio of game wins and losses based on home, away and neutral sites and identify the presence of a possible home-ground advantage.
- Calculate the team's average margin of wins (or losses) to understand the intricate improvements in offense or defense across the years.
- Find out the team's performance in BigTen conference over the years.
- Evaluate the team's performance by opponents to identify historically tough opponents.
- Identify the states where our team traditionally performed well versus not to prepare better for tough travels.



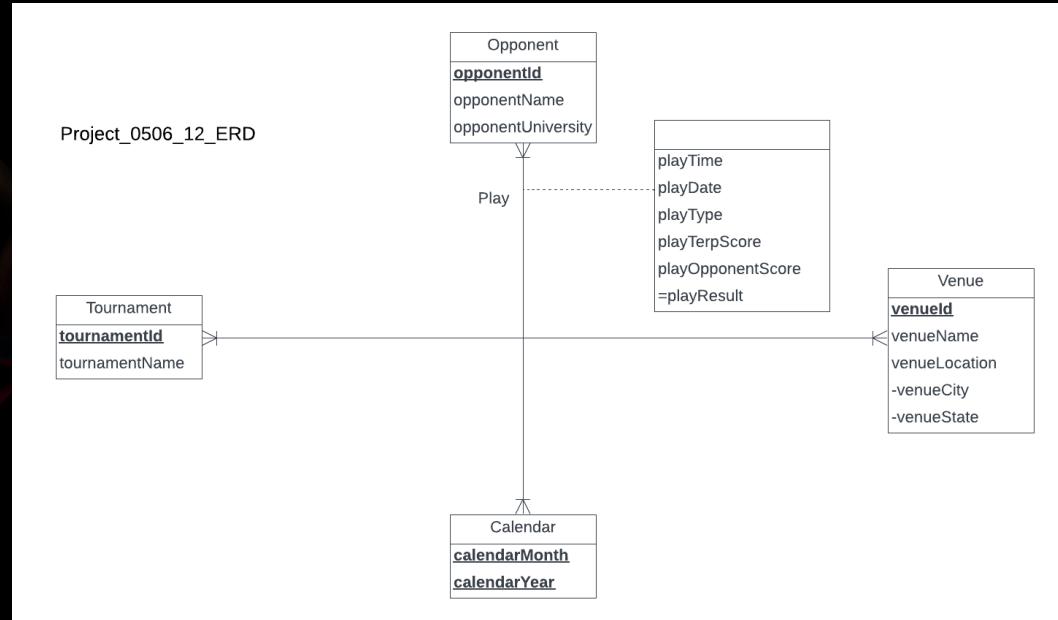
# Conceptual Database Design

## ER Diagram

The Entities are as follows

Opponent  
Tournament  
Calendar  
Venue

All of them are linked together by a 4<sup>th</sup> degree one to many relation with the following attributes: Play type, the scores, the Date and the time



# Logical Database Design

Opponent(opponentId, opponentName, opponentUniversity)

Tournament(tournamentId, tournamentName)

Calendar(calendarMonth, calendarYear)

Venue(venueId, venueName, venueCity, venueState)

Play(playTime, playDate, calendarMonth, calenderYear, opponentId,  
tournamentId, venueId, playType, playTerpScore, playOpponentScore)



# CREATE Table

*Opponent* -

```
CREATE TABLE [Moneyball.Opponent] (
    opponentId CHAR(4) NOT NULL,
    opponentName VARCHAR(50),
    opponentUniversity VARCHAR(99),
    CONSTRAINT pk_Opponent_opponentId PRIMARY KEY (opponentId)
);
```

*Tournament* -

```
CREATE TABLE [Moneyball.Tournament] (
    tournamentId CHAR(3) NOT NULL,
    tournamentName VARCHAR(50),
    CONSTRAINT pk_Tournament_tournamentId PRIMARY KEY (tournamentId)
);
```



# CREATE Table

*Calendar* -

```
CREATE TABLE [Moneyball.Calendar](
    calendarMonth CHAR(3) NOT NULL,
    calendarYear DATE NOT NULL,
    CONSTRAINT pk_Calendar_calendarYear_calendarMonth PRIMARY KEY
        (calendarYear, calendarMonth)
);
```

*Venue* -

```
CREATE TABLE [Moneyball.Venue](
    venueId CHAR(5) NOT NULL,
    venueName VARCHAR(50),
    venueCity VARCHAR(50),
    venueState CHAR(2),
    CONSTRAINT pk_Venue_venueId PRIMARY KEY (venueId)
);
```



# CREATE Table

*Play* -

```
CREATE TABLE [Moneyball.Play](
    playTime CHAR(8) NOT NULL,
    playDate INT NOT NULL,
    calendarMonth CHAR(3) NOT NULL,
    calendarYear DATE NOT NULL,
    opponentId CHAR(4),
    tournamentId CHAR(3),
    venueId CHAR(5),
    playType VARCHAR(10),
    playTerpScore INT,
    playOpponentScore INT,
    CONSTRAINT pk_Play_playTime_playDate_calendarMonth_calendarYear PRIMARY KEY
        (playTime,playDate,calendarYear,calendarMonth),
    CONSTRAINT fk_Play_opponentId FOREIGN KEY (opponentId)
        REFERENCES [Moneyball.Opponent](opponentId)
        ON DELETE NO ACTION ON UPDATE CASCADE,
    CONSTRAINT fk_Play_tournamentId FOREIGN KEY (tournamentId)
        REFERENCES [Moneyball.Tournament](tournamentId)
        ON DELETE NO ACTION ON UPDATE CASCADE,
    CONSTRAINT fk_Play_calendarYear_calendarMonth FOREIGN KEY (calendarYear, calendarMonth)
        REFERENCES [Moneyball.calendar](calendarYear, calendarMonth)
        ON DELETE NO ACTION ON UPDATE CASCADE,
    CONSTRAINT fk_Play_venueId FOREIGN KEY (venueId)
        REFERENCES [Moneyball.Venue](venueId)
        ON DELETE NO ACTION ON UPDATE CASCADE);
```





# Transaction 1

What is the ratio of game wins based on home, away, and neutral sites?

# SQL



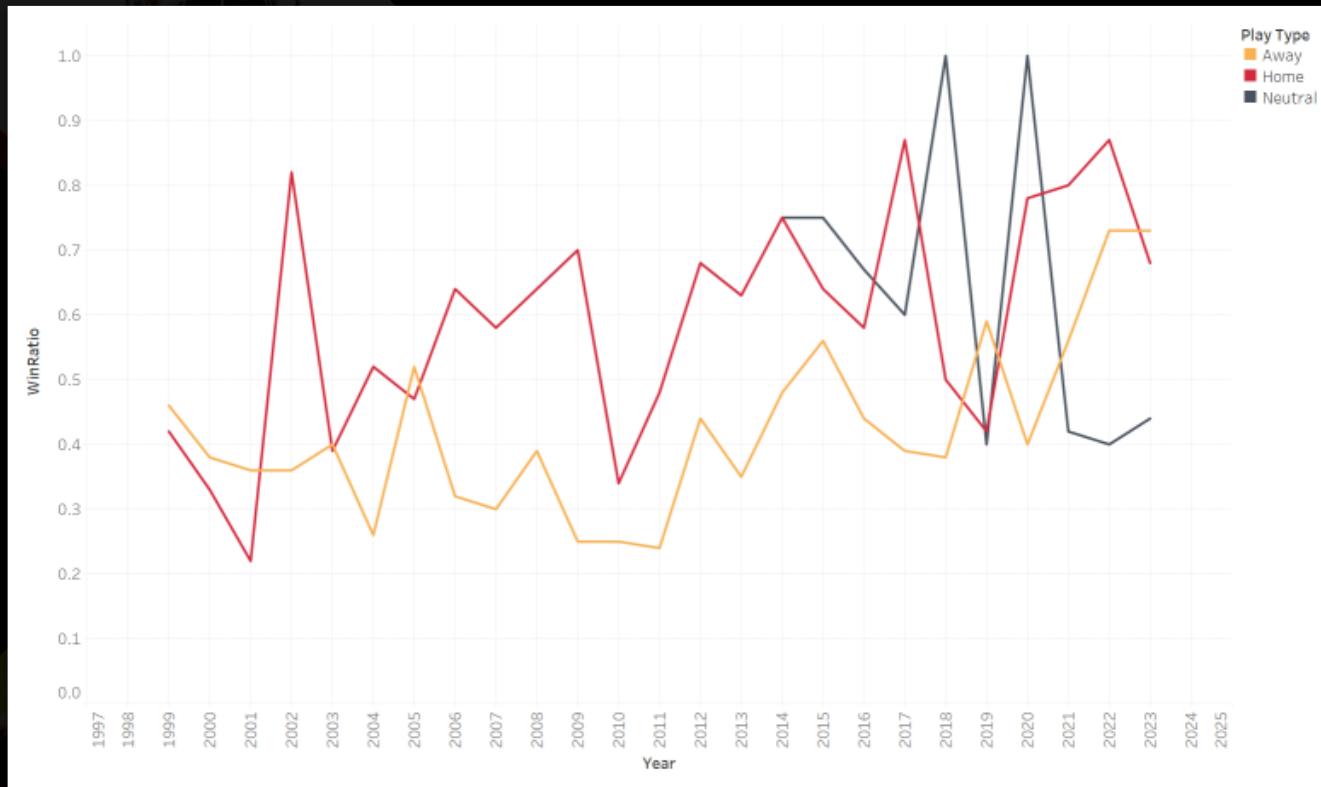
```
SELECT YEAR(p.calendarYear) as 'Year', p.playType AS 'Play
Type',
       ROUND(CAST(SUM(CASE WHEN playResult = 'W'
THEN 1 ELSE 0 END) AS FLOAT) / COUNT(*), 2) AS
WinRatio
FROM [Moneyball.Play] p
GROUP BY p.playType, YEAR(p.calendarYear)
ORDER BY YEAR(p.calendarYear);
```

	Year	playType	WinRatio		Year	playType	WinRatio
1	1999	Away	0.46	31	2014	Away	0.48
2	1999	Home	0.42	32	2014	Home	0.75
3	2000	Away	0.38	33	2014	Neutral	0.75
4	2000	Home	0.33	34	2015	Away	0.56
5	2001	Away	0.36	35	2015	Home	0.64
6	2001	Home	0.22	36	2015	Neutral	0.75
7	2002	Away	0.36	37	2016	Away	0.44
8	2002	Home	0.82	38	2016	Home	0.58
9	2003	Away	0.4	39	2016	Neutral	0.67
10	2003	Home	0.39	40	2017	Away	0.39
11	2004	Away	0.26	41	2017	Home	0.87
12	2004	Home	0.52	42	2017	Neutral	0.6
13	2005	Away	0.52	43	2018	Away	0.38
14	2005	Home	0.47	44	2018	Home	0.5
15	2006	Away	0.32	45	2018	Neutral	1
16	2006	Home	0.64	46	2019	Away	0.59
17	2007	Away	0.3	47	2019	Home	0.42
18	2007	Home	0.58	48	2019	Neutral	0.4
19	2008	Away	0.39	49	2020	Away	0.4
20	2008	Home	0.64	50	2020	Home	0.78
21	2009	Away	0.25	51	2020	Neutral	1
22	2009	Home	0.7	52	2021	Away	0.56
23	2010	Away	0.25	53	2021	Home	0.8
24	2010	Home	0.34	54	2021	Neutral	0.42
25	2011	Away	0.24	55	2022	Away	0.73
26	2011	Home	0.48	56	2022	Home	0.87
27	2012	Away	0.44	57	2022	Neutral	0.4
28	2012	Home	0.68	58	2023	Away	0.73
29	2013	Away	0.35	59	2023	Home	0.68
30	2013	Home	0.63	60	2023	Neutral	0.44

SQL Output



# What is the ratio of game wins based on home, away, and neutral sites?



# **Transaction 2**

What is the team's average margin  
of wins and losses in the previous  
years?

# SQL

```
SELECT YEAR(calendarYear) as 'Year',
```

```
    ROUND(AVG(CASE WHEN playresult = 'W' THEN  
        CAST(playTerpScore – playOpponentScore AS FLOAT) ELSE  
        NULL END),1) AS AvgWinMargin,
```

```
    ROUND(AVG(CASE WHEN playresult = 'L' THEN  
        CAST(playTerpScore - playOpponentScore AS FLOAT) ELSE  
        NULL END),1) AS AvgLossMargin
```

```
FROM [Moneyball.Play]  
GROUP BY YEAR(calendarYear)  
ORDER BY YEAR(calendarYear);
```

	Year	AvgWinMargin	AvgLossMargin
1	1999	6.4	-5.5
2	2000	4.9	-4
3	2001	5.8	-4.7
4	2002	6.6	-5.9
5	2003	4	-6
6	2004	5.1	-5.1
7	2005	4.8	-4.2
8	2006	2.9	-5
9	2007	5.9	-5.4
10	2008	5	-5.2
11	2009	4.9	-5.5
12	2010	3.8	-6.7
13	2011	4.3	-5.1
14	2012	4.4	-3.3
15	2013	3.8	-3.5
16	2014	4.7	-4.9
17	2015	4.4	-2.8
18	2016	3.7	-3.1
19	2017	4.7	-4.1
20	2018	4.9	-4.7
21	2019	3.7	-5.1
22	2020	6.9	-5.8
23	2021	4.5	-4.1
24	2022	6.1	-4.9
25	2023	5.7	-4.4

SQL Output



# Why not Run differential?

Season	Game	TerpScore	OppScore
1	1	2	1
1	2	2	5
2	1	1	2
2	2	1	2

Run differential:

Season1: -2

Season2: -2

AvgLossMargin: AvgWinMargin:

Season1: -3

Season1: 1

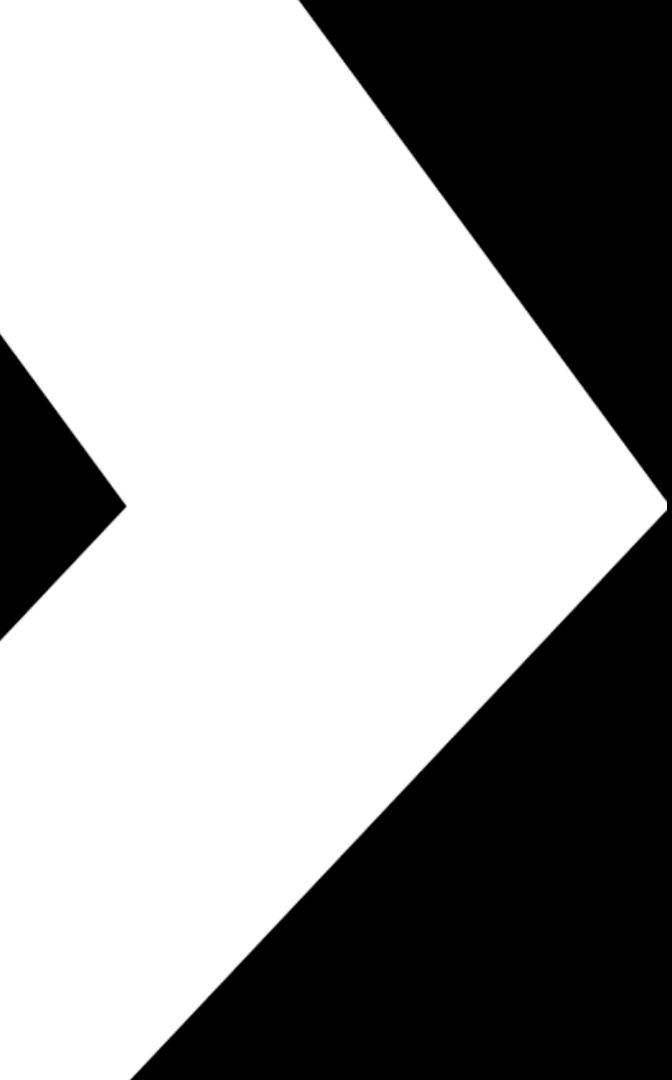
Season2: -1

Season2: 0



# What is the team's average margin of wins and losses in the previous years?





# Transaction 3

What is the ratio of game wins in Big Ten conference games?

# SQL

```
SELECT YEAR(p.calendarYear) as 'Year',
```

```
    ROUND(CAST(SUM(CASE WHEN playResult = 'W' THEN 1  
ELSE 0 END) AS FLOAT) / COUNT(*), 2) AS WinRatio
```

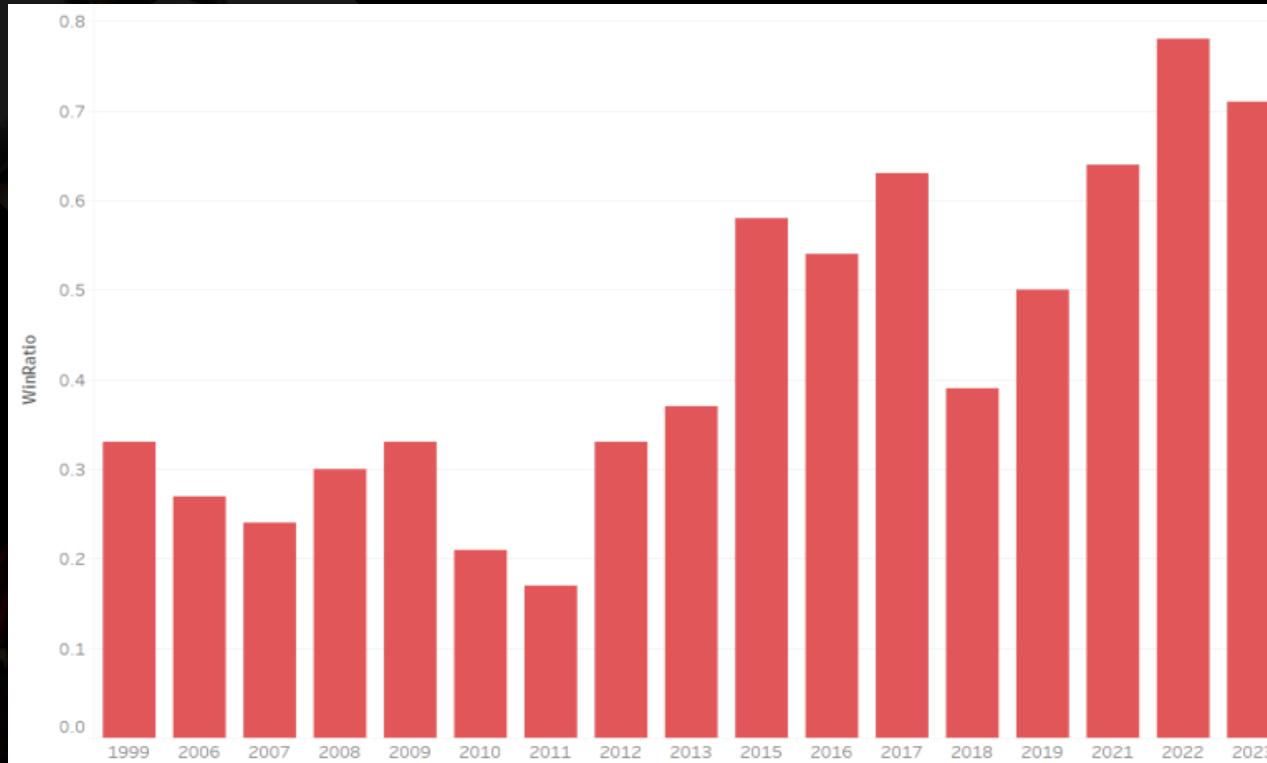
```
FROM [Moneyball.Play] p, [Moneyball.Tournament] t  
WHERE p.tournamentId = t.tournamentId AND t.tournamentName = 'Big Ten'  
GROUP BY t.tournamentName, YEAR(p.calendarYear)  
ORDER BY YEAR(p.calendarYear);
```

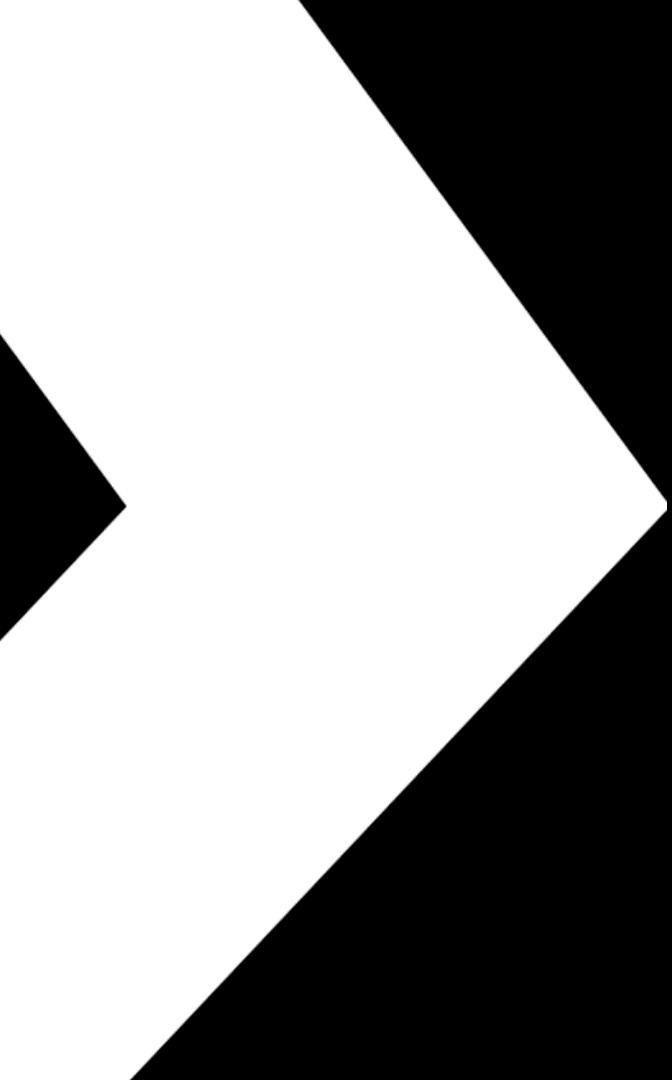
	Year	WinRatio
1	1999	0.33
2	2006	0.27
3	2007	0.24
4	2008	0.3
5	2009	0.33
6	2010	0.21
7	2011	0.17
8	2012	0.33
9	2013	0.37
10	2015	0.58
11	2016	0.54
12	2017	0.63
13	2018	0.39
14	2019	0.5
15	2021	0.64
16	2022	0.78
17	2023	0.71

SQL Output



# What is the ratio of game wins in Big Ten conference games?





# Transaction 4

What is the team's Winning rate  
against opponents where atleast 10  
games being played?

# SQL

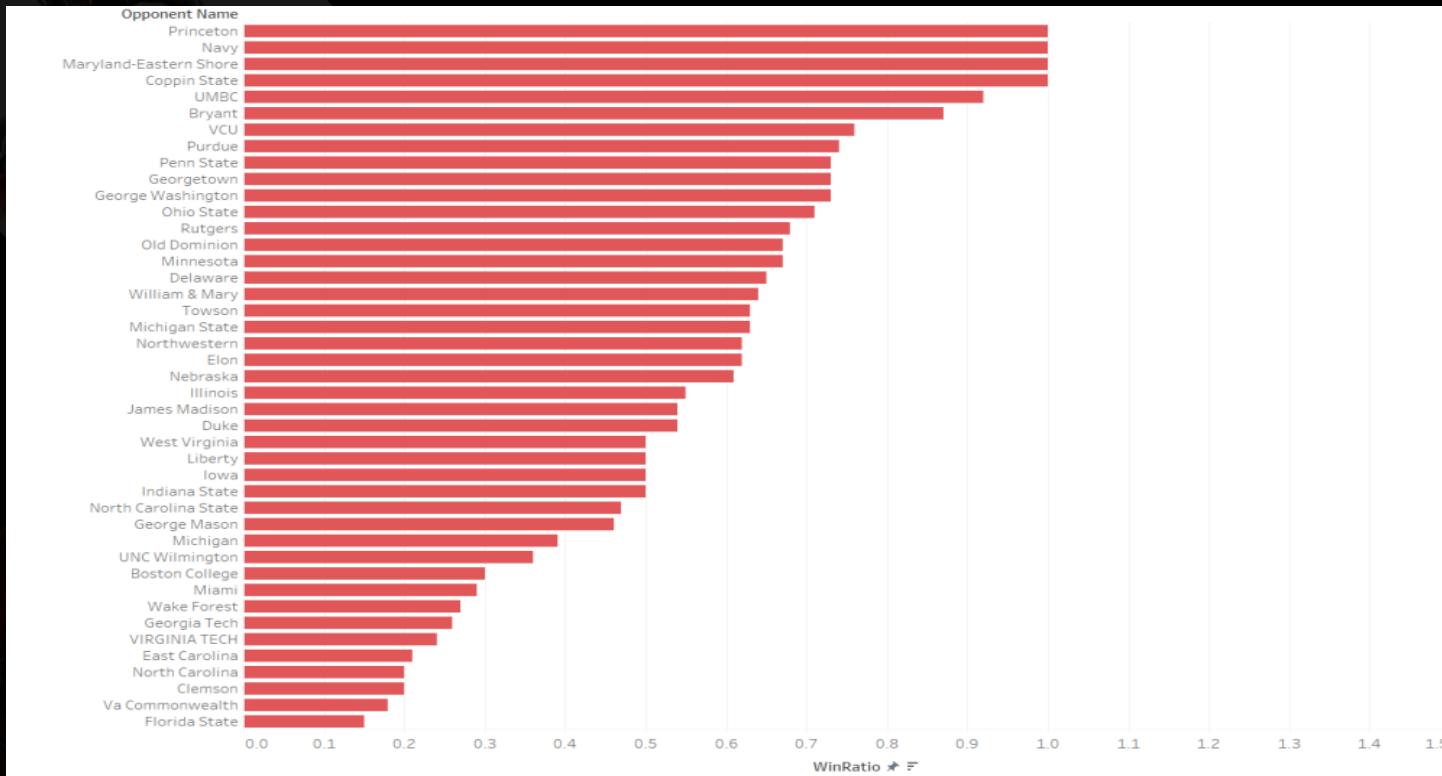
```
SELECT o.opponentName AS 'Opponent Name',
       ROUND(CAST(SUM(CASE WHEN p.playResult = 'W' THEN
                           ELSE 0 END) AS FLOAT) / COUNT(*), 2) AS WinRatio
    FROM [Moneyball.Play] p
  JOIN [Moneyball.Opponent] o ON p.opponentId = o.opponentId
 GROUP BY p.opponentId, o.opponentName
 HAVING COUNT(*)>=10
 ORDER BY WinRatio DESC;
```

	Opponent Name	WinRatio	Opponent Name	WinRatio	
1	Coppin State	1	22	Nebraska	0.61
2	Maryland-Eastern Shore	1	23	Illinois	0.55
3	Navy	1	24	James Madison	0.54
4	Princeton	1	25	Duke	0.54
5	UMBC	0.92	26	West Virginia	0.5
6	Bryant	0.87	27	Liberty	0.5
7	VCU	0.76	28	Iowa	0.5
8	Purdue	0.74	29	Indiana State	0.5
9	Penn State	0.73	30	North Carolina State	0.47
10	Georgetown	0.73	31	George Mason	0.46
11	George Washington	0.73	32	Michigan	0.39
12	Ohio State	0.71	33	UNC Wilmington	0.36
13	Rutgers	0.68	34	Boston College	0.3
14	Old Dominion	0.67	35	Miami	0.29
15	Minnesota	0.67	36	Wake Forest	0.27
16	Delaware	0.65	37	Georgia Tech	0.26
17	William & Mary	0.64	38	VIRGINIA TECH	0.24
18	Towson	0.63	39	East Carolina	0.21
19	Michigan State	0.63	40	North Carolina	0.2
20	Northwestern	0.62	41	Clemson	0.2
21	Elon	0.62	42	Va Commonwealth	0.18
22	Florida State	0.15	43		

SQL Output



# What is the team's Winning rate against opponents where atleast 10 games being played?





# Transaction 5

What is the win ratio over the years  
across different States?

# SQL

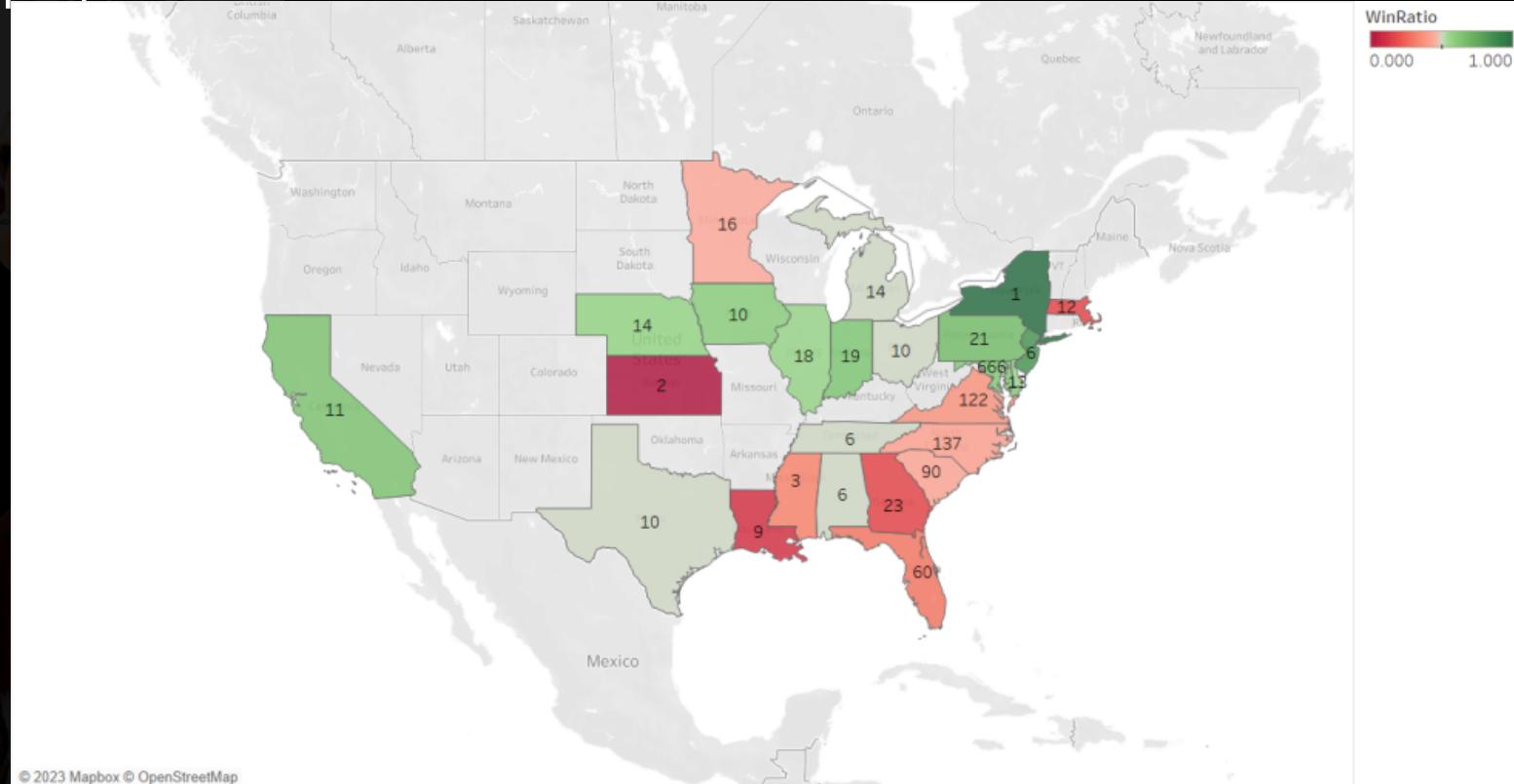
```
SELECT v.venueState AS 'State',  
  
       ROUND(CAST(SUM(CASE WHEN p.playResult = 'W' THEN 1  
                           ELSE 0 END) AS FLOAT) / COUNT(*), 2) AS WinRatio  
  
FROM [Moneyball.Play] p, [Moneyball.Venue] v  
WHERE p.venueld = v.venueld  
GROUP BY v.venueState  
Order BY WinRatio DESC;
```

	State	WinRatio
1	NY	1
2	DC	0.86
3	NJ	0.83
4	PA	0.67
5	CA	0.64
6	IN	0.63
7	MD	0.61
8	IA	0.6
9	NE	0.57
10	IL	0.56
11	DE	0.54
12	AL	0.5
13	MI	0.5
14	OH	0.5
15	TN	0.5
16	TX	0.5
17	MN	0.44
18	SC	0.43
19	NC	0.42
20	VA	0.38
21	MS	0.33
22	NB	0.33
23	FL	0.3
24	GA	0.17
25	MA	0.17
26	LA	0.11
27	KS	0

SQL Output



What is the team's Winning rate against opponents where atleast 10 games being played?

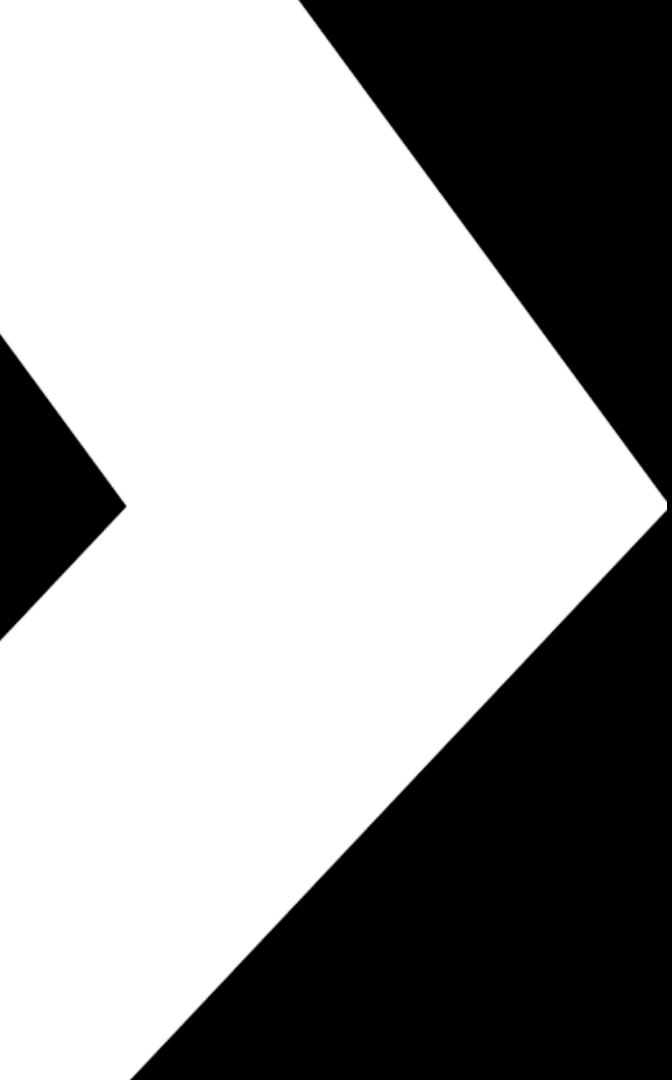


© 2023 Mapbox © OpenStreetMap



FEARLESSLY  
FORWARD

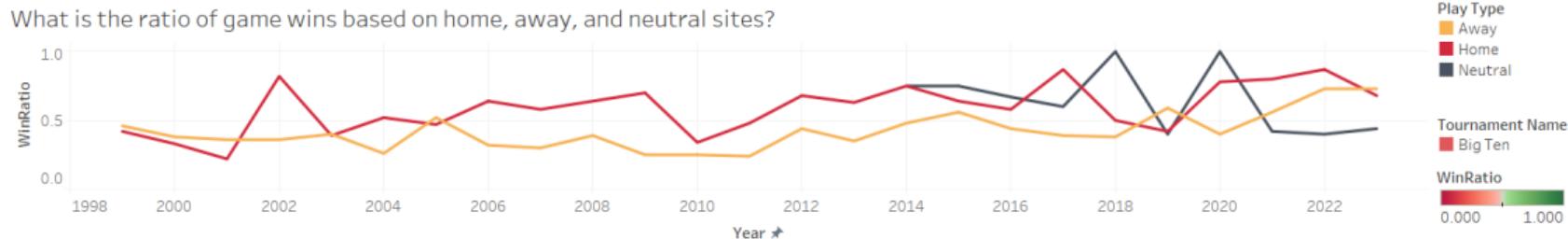




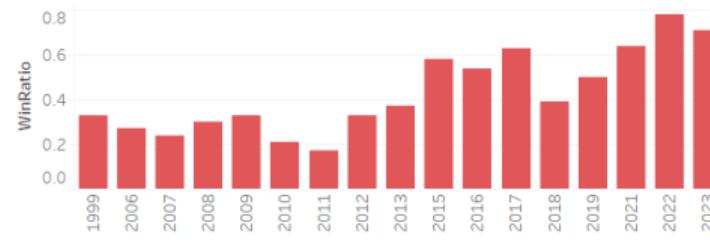
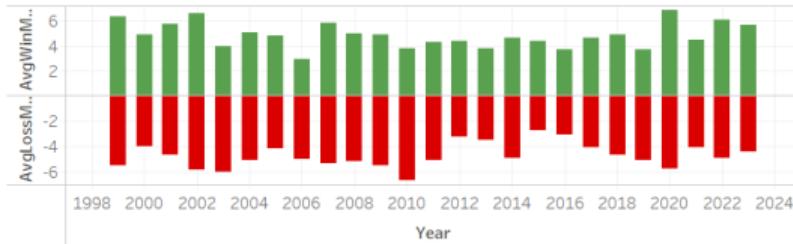
# Tableau Dashboard

# Game Result analysis

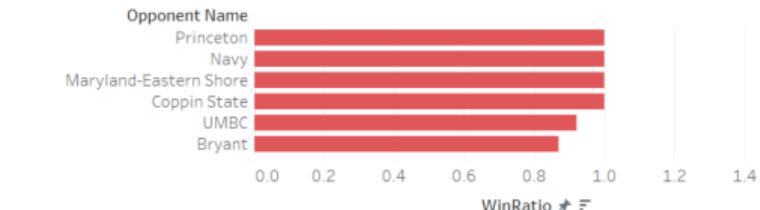
What is the ratio of game wins based on home, away, and neutral sites?



What is the team's average margin of wins and losses across years? What is the ratio of game wins in Big Ten conference games?



What is the team's Winning rate against opponents where atleast 10 games being played?



What is the win ratio over the years across different States?



# Facts and Figures



**1315 Games**  
with results in **25 Years**

**Home Or Away?**  
game is not  
always determined  
by the stadium  
where the match is  
played

**175 Venues**  
reduced to  
**99 Unique Venues** post data clean up

**213 Opponents**  
reduced to  
**145 Unique Opponents** post data clean up

**Identification of double headers**  
leading to time also being used as a primary key

Worried about not having Venue and Tournament details for a game, We got you covered





# Thank You