Moneyball

Aaryan Bisen
Bhoomi Chavan
Neil Gupta
Srishti Gupta

Presentation : 12/04/2023
Fill update : 12/07/2023

BACKGROUND

Users

1. Coaches and Strategists:

- Analyze match results and performance data to develop effective game strategies.

2. Team Analysts:

- Explore tournament data for insights into the team's overall performance trends.
- Identify strengths and weaknesses based on historical match results.

Data Source:

We collected the women's softball team data using a web scraper on the 'umdterps' website. (https://umterps.com/sports/softball/schedule)

INTRODUCTION

Mission statement:

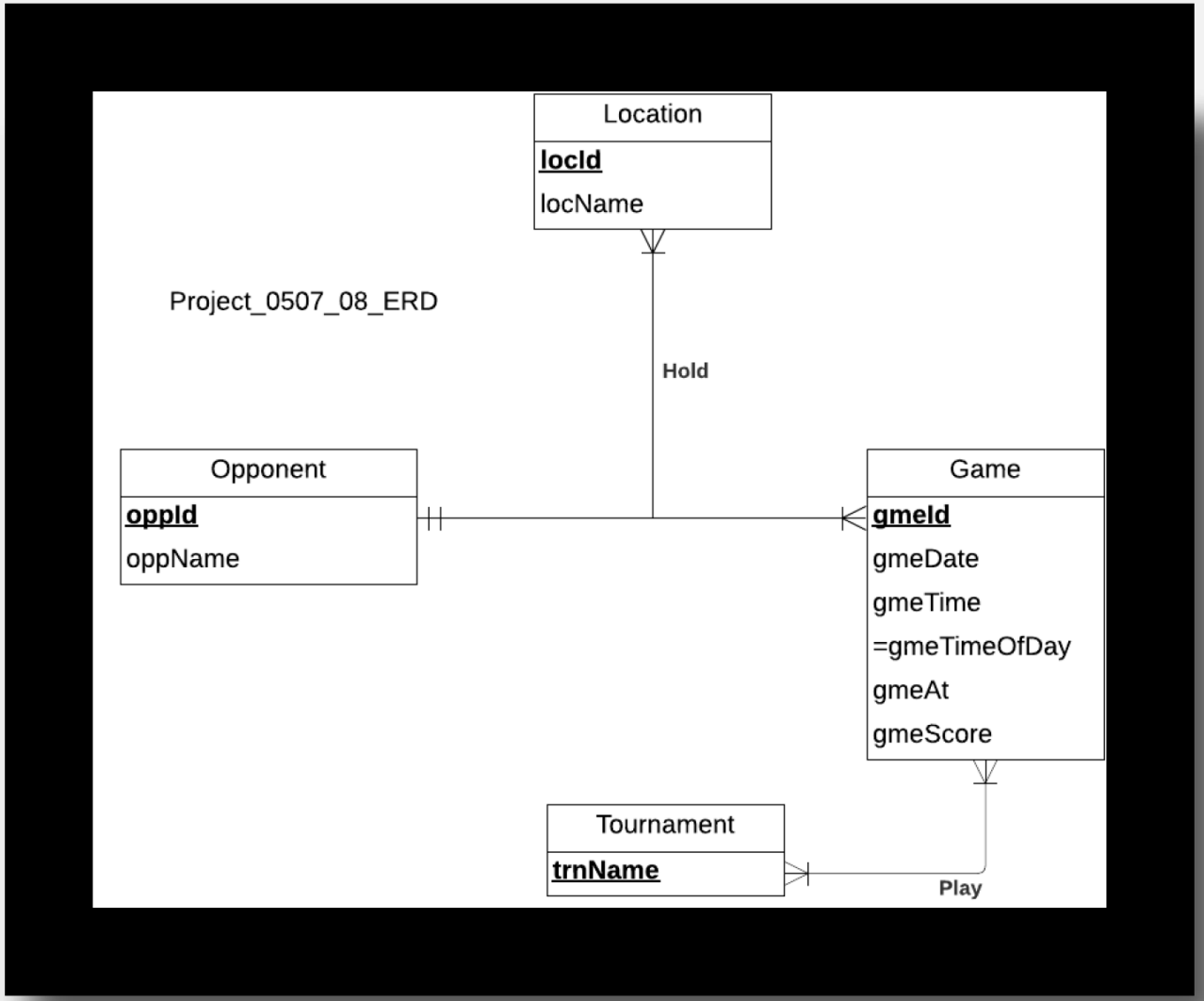
Our mission is to provide the UMD Athletics Department with comprehensive insights into the softball team's performance over the past two decades. By analyzing various metrics, we aim to offer a deep understanding of the team's win rate, helping the department make data-driven decisions to further elevate UMD's success in softball.

MISSION OBJECTIVES:

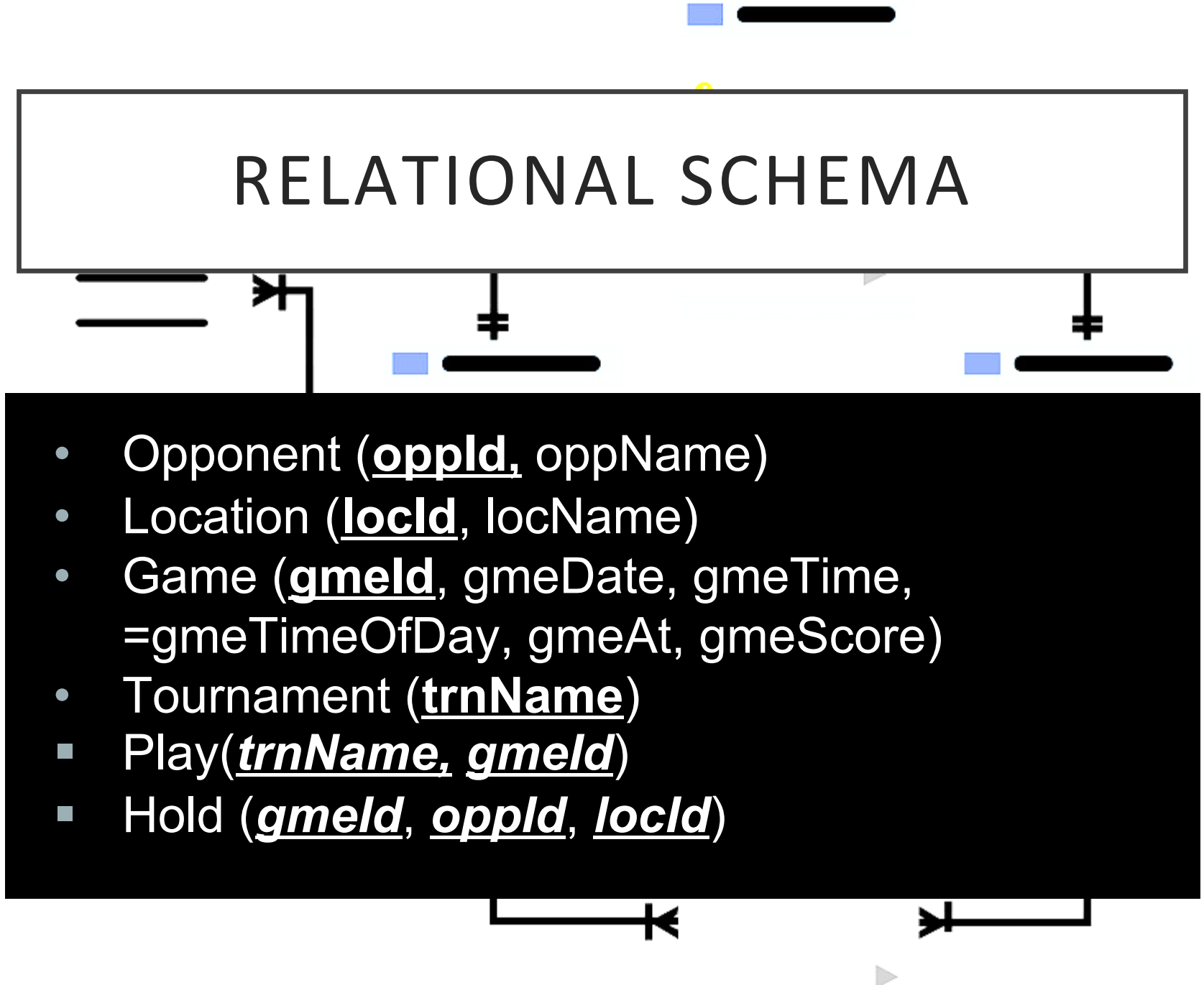
The following mission objectives that our firm deal to elaborate are as follows:

- To determine the top 10 strongest/weakest opponents against UMD.
- To determine the home/away/neutral performance of UMD.
- To find the longest win streak for each year.
- To find the longest losing streak for each year.
- To find the time of the day when the softball team performs the best.
- To find team performance by tournament.

ER DIAGRAM


Location
locId
locName
Project_0507_08_ERD
Hold
Opponent
oppId
oppName
Game
gmeId
gmeDate
gmeTime
=gmeTimeOfDay
gmeAt
gmeScore
Tournament
trnName
Play

RELATIONAL SCHEMA



- Opponent (**oppId,** oppName)
- Location (**locId**, locName)
- Game (**gmeId**, gmeDate, gmeTime, =gmeTimeOfDay, gmeAt, gmeScore)
- Tournament (**trnName**)
- Play(***trnName, gmeId***)
- Hold (***gmeId***, ***oppId***, ***locId***)

PHYSICAL DATABASE DESIGN

-);

```
gmeId VARCHAR (20) NOT NULL,
oppId VARCHAR (20) NOT NULL,
locId VARCHAR (20) NOT NULL,
CONSTRAINT pk_Hold_gmeId_oppId_locId PRIMARY KEY (gmeId, oppId, locId)
CONSTRAINT fk_Hold_gmeId FOREIGN KEY (gmeId)
    REFERENCES [Moneyball.Game] (gmeId)
    ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT fk_Hold_oppId FOREIGN KEY (oppId)
    REFERENCES [Moneyball.Opponent] (oppId)
    ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT fk_Hold_locId FOREIGN KEY (locId)
    REFERENCES [Moneyball.Location] (locId)
    ON DELETE CASCADE ON UPDATE CASCADE
```

```
ATE VIEW TournamentWinRate AS (
 SELECT
     trn.trnName AS 'Tournament',
     COUNT(g.gmeId) AS 'Games Played Against',
     SUM(CASE WHEN CAST(SUBSTRING(g.gmeScore, 1, CHARINDEX('-', g.gmeScore) - 1) AS INT) >
                   CAST(SUBSTRING(g.gmeScore, CHARINDEX('-', g.gmeScore) + 1, LEN(g.gmeScore)) AS INT) THEN 1 ELSE 0 END) AS 'Wins',
     SUM(CASE WHEN CAST(SUBSTRING(g.gmeScore, 1, CHARINDEX('-', g.gmeScore) - 1) AS INT) <
                   CAST(SUBSTRING(g.gmeScore, CHARINDEX('-', g.gmeScore) + 1, LEN(g.gmeScore)) AS INT) THEN 1 ELSE 0 END) AS 'Losses',
     SUM(CASE WHEN CAST(SUBSTRING(g.gmeScore, 1, CHARINDEX('-', g.gmeScore) - 1) AS INT) =
                   CAST(SUBSTRING(g.gmeScore, CHARINDEX('-', g.gmeScore) + 1, LEN(g.gmeScore)) AS INT) THEN 1 ELSE 0 END) AS 'Ties',
     (SUM(CASE WHEN CAST(SUBSTRING(g.gmeScore, 1, CHARINDEX('-', g.gmeScore) - 1) AS INT) >
                   CAST(SUBSTRING(g.gmeScore, CHARINDEX('-', g.gmeScore) + 1, LEN(g.gmeScore)) AS INT) THEN 1 ELSE 0 END) * 1.0 / COUNT(g.gmeI
                   AS 'Tournament Win Rate'
 FROM
     [Moneyball.Tournament] trn
 LEFT JOIN
     [Moneyball.Game] g ON trn.trnName = g.trnName
 GROUP BY
     trn.trnName
```

BUSINESS TRANSACTION

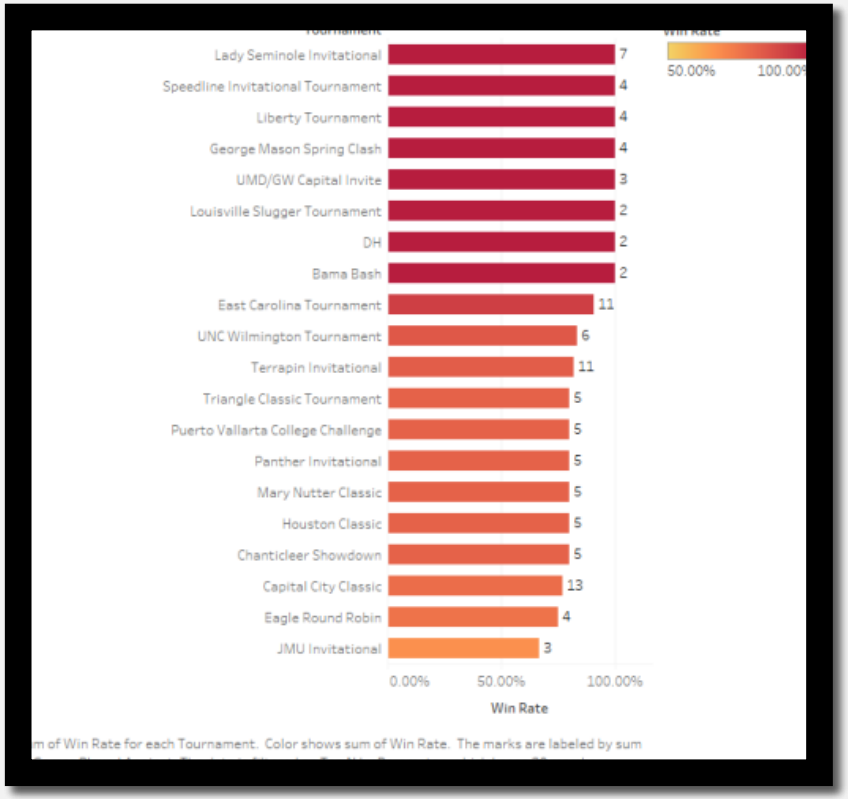
1) What tournaments did the UMD softball team perform best in?

```
ELECT *
ROM TournamentWinRate
RDER BY [Tournament Win Rate] DESC, [Games Played Against] DESC;
```

Results | Messages

	Tournament	Games Played Against	Wins	Losses	Ties	Tournament Win Rate
1	Lady Seminole Invitational	7	7	0	0	1.0
2	Liberty Tournament	4	4	0	0	1.0
3	Speedline Invitational Tournament	4	4	0	0	1.0
4	George Mason Spring Clash	4	4	0	0	1.0
5	UMD/GW Capital Invite	3	3	0	0	1.0
6	Louisville Slugger Tournament	2	2	0	0	1.0
7	DH	2	2	0	0	1.0
8	Bama Bash	2	2	0	0	1.0
9	Arizona State Coca-Cola Classic	1	1	0	0	1.0
10	Georgia Tech / Buzz Classic	1	1	0	0	1.0
11	University of South Florida Tournament	1	1	0	0	1.0
12	East Carolina Tournament	11	10	1	0	0.9

Query executed successfully. | doitsqlx.rhsmith.umd.edu,97... | AD\bhoomi (68) | BUDT703_Project_0507_08 | 00:00:00 | 79 rows



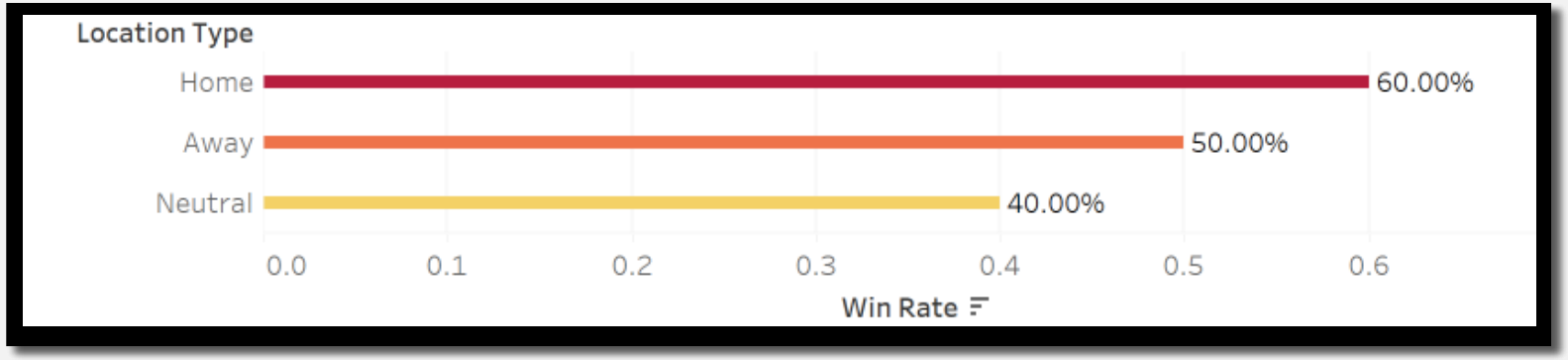
2) What is the win rate of the UMD team with and without home field advantage?

```
SELECT *
FROM LocationWiseScore
ORDER BY [Win Rate] DESC;
```

Results | Messages

	Location Type	Games Played	Wins	Losses	Ties	Win Rate
1	Home	652	395	257	0	0.6
2	Away	533	253	280	0	0.5
3	Neutral	265	117	146	2	0.4

Query executed successfully. | doitsqlx.rhsmith.umd.edu,97... | AD\bhoomi (68) | BUDT703_Project_0507_08 | 00:00:00 | 3 rows



3) Who are the top 10 weakest opponents against UMD?

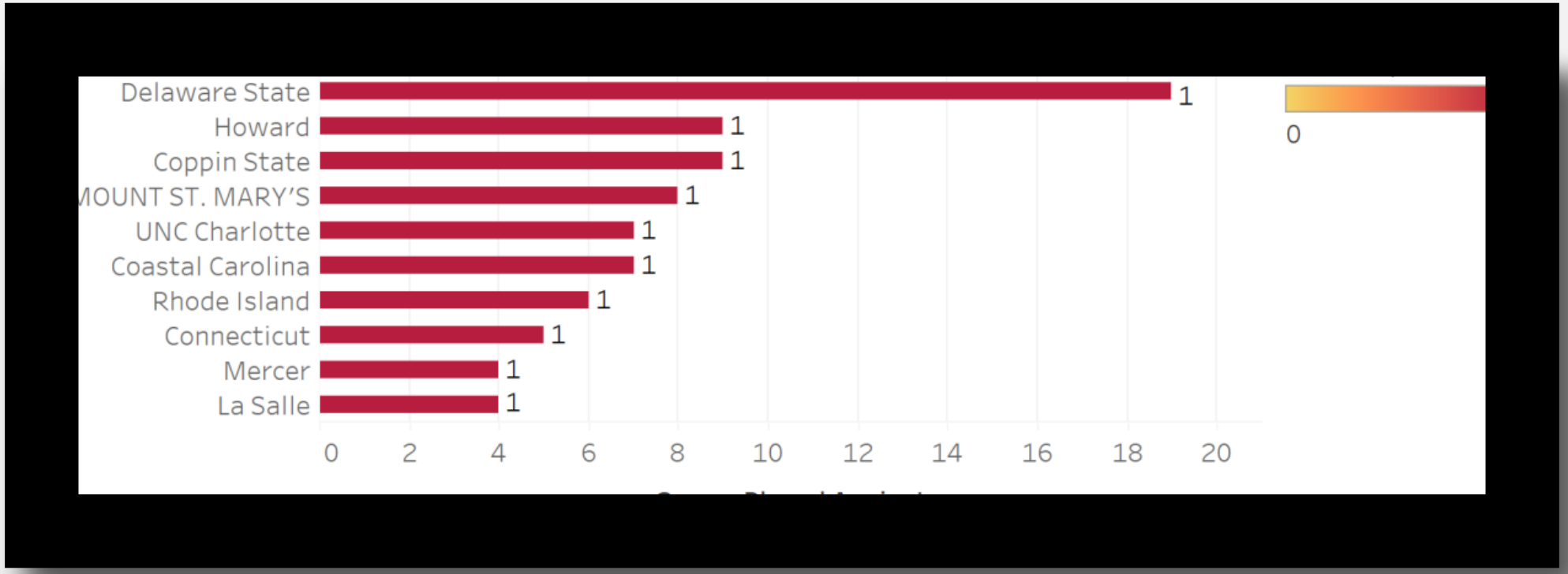
```
SELECT *
FROM OpponentWinRate
ORDER BY [Win Rate Against Opponent] DESC, [Games Played Against] DESC;
```

Results | Messages

	Opponent ID	Opponent Team	Games Played Against	Wins	Losses	Ties	Win Rate Against Opponent
1	OPP0167	Delaware State	19	19	0	0	1.0
2	OPP0168	Howard	9	9	0	0	1.0
3	OPP0054	Coppin State	9	9	0	0	1.0
4	OPP0200	MOUNT ST. MARY'S	8	8	0	0	1.0
5	OPP0261	UNC Charlotte	7	7	0	0	1.0
6	OPP0044	Coastal Carolina	7	7	0	0	1.0
7	OPP0088	Rhode Island	6	6	0	0	1.0
8	OPP0194	Connecticut	5	5	0	0	1.0
9	OPP0241	La Salle	4	4	0	0	1.0
10	OPP0082	Nevada	4	4	0	0	1.0
11	OPP0133	Mercer	4	4	0	0	1.0
12	OPP0011	Saint Joseph's	3	3	0	0	1.0
13	OPP0015	Ball State	3	3	0	0	1.0
14	OPP0239	Penn St	3	3	0	0	1.0
15	OPP0227	St. Joseph's	3	3	0	0	1.0

Query executed successfully. | doitsqlx.rhsmith.umd.edu,97... | AD\bhoomi (68) | BUDT703_Project_0507_08 | 00:00:00 | 275 rows

Who are the top 10 weakest opponents facing UMD?



4) Who are the top 10 strongest opponents against UMD?

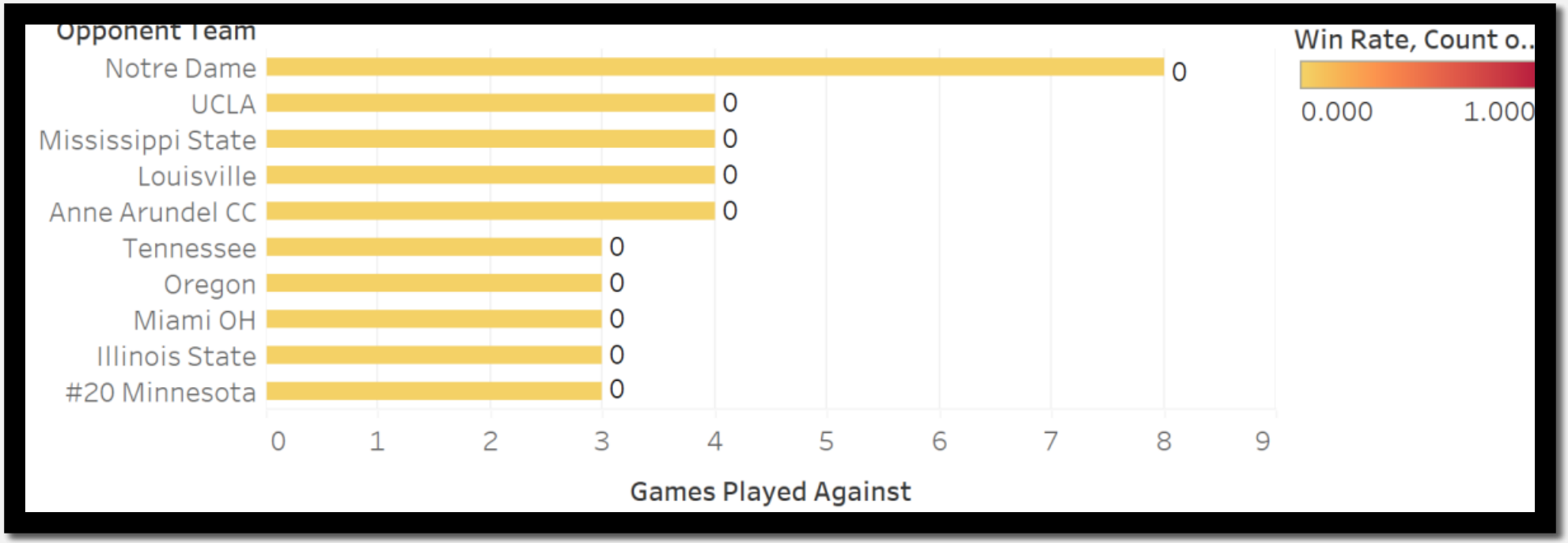
```
SELECT *
FROM OpponentWinRate
ORDER BY [Win Rate Against Opponent] ASC, [Games Played Against] DESC;
```

Results | Messages

	Opponent ID	Opponent Team	Games Played Against	Wins	Losses	Ties	Win Rate Against Opponent
1	OPP0147	Notre Dame	8	0	8	0	0.0
2	OPP0101	Louisville	4	0	4	0	0.0
3	OPP0150	Mississippi State	4	0	4	0	0.0
4	OPP0161	UCLA	4	0	4	0	0.0
5	OPP0274	Anne Arundel CC	4	0	4	0	0.0
6	OPP0099	Illinois State	3	0	3	0	0.0
7	OPP0103	Miami OH	3	0	3	0	0.0
8	OPP0107	Tennessee	3	0	3	0	0.0
9	OPP0090	#20 Minnesota	3	0	3	0	0.0
10	OPP0120	Oregon	3	0	3	0	0.0
11	OPP0123	UT Martin	2	0	2	0	0.0
12	OPP0141	Oklahoma	2	0	2	0	0.0
13	OPP0066	Clemson	2	0	2	0	0.0
14	OPP0072	Arizona State	2	0	2	0	0.0
15	OPP0275	Trenton State	2	0	2	0	0.0

Query executed successfully. | doitsqlx.rhsmith.umd.edu,97... | AD\bhoomi (68) | BUDT703_Project_0507_08 | 00:00:00 | 275 rows

Who are the top 10 strongest opponents facing UMD ?

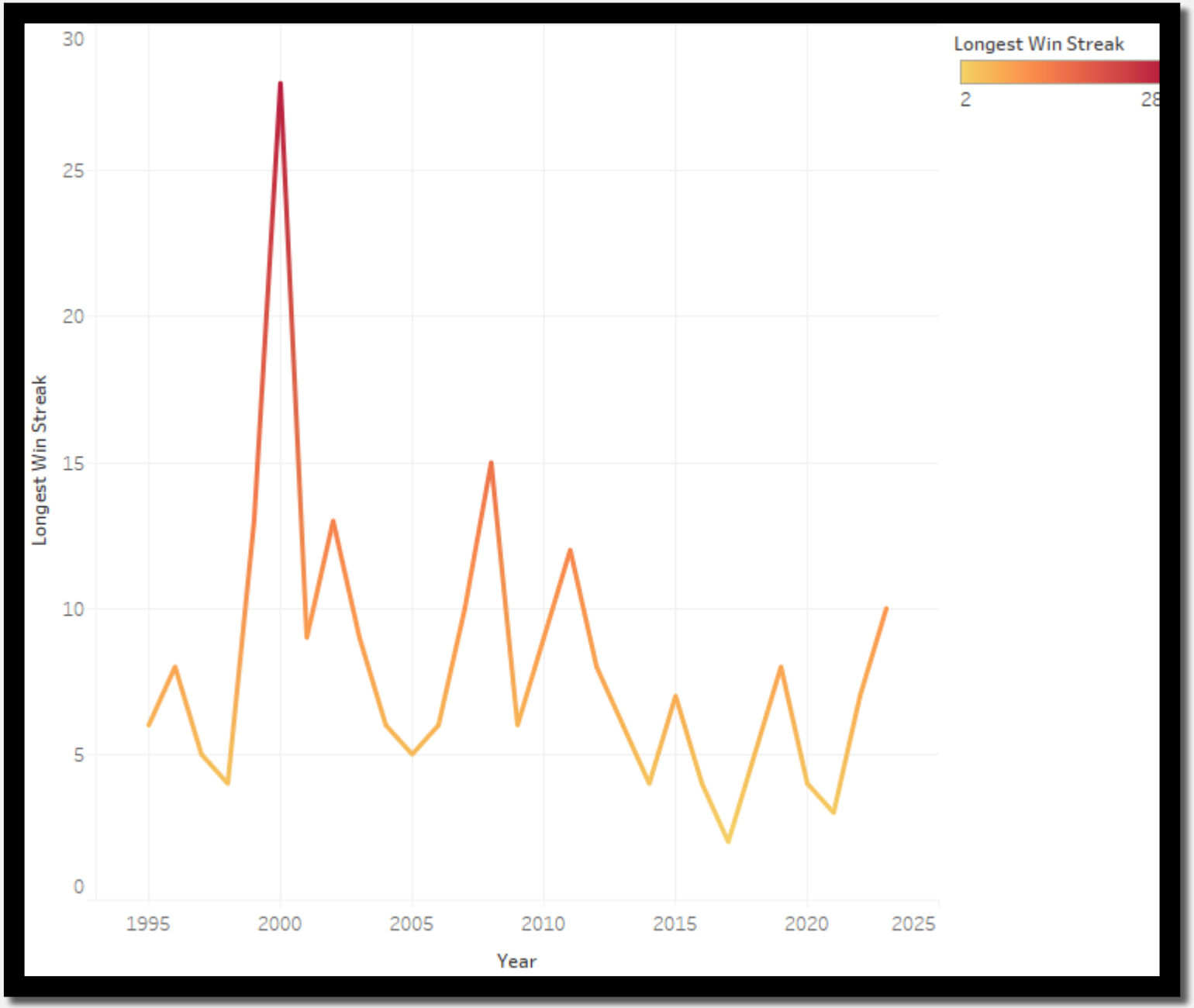


5) What is the longest winning streak?

```
WITH RankedGames AS (
    SELECT
        g.gmeId, g.gmeDate, g.gmeScore, g.trnName,
        ROW_NUMBER() OVER (ORDER BY gmeDate DESC) -
        ROW_NUMBER() OVER
        (PARTITION BY YEAR(gmeDate), CASE WHEN SUBSTRING(g.gmeScore, 2, 1) > SUBSTRING(g.gmeScore, 4, 1) THEN 'W' ELSE 'L' END
        ORDER BY gmeDate DESC) AS GroupNum
    FROM
        [Moneyball.Game] g

ConsecutiveWinGroups AS (
    SELECT
        YEAR(gmeDate) AS GameYear,
        GroupNum,
        COUNT(*) AS ConsecutiveGamesWon
    FROM
        RankedGames
    WHERE
        CASE WHEN SUBSTRING(gmeScore, 2, 1) > SUBSTRING(gmeScore, 4, 1) THEN 'W' ELSE 'L' END = 'W'
    GROUP BY YEAR(gmeDate), GroupNum

LongestWinStreaks AS (
    SELECT
        GameYear,
        MAX(ConsecutiveGamesWon) AS LongestWinStreak
    FROM ConsecutiveWinGroups
    GROUP BY GameYear

ELECT GameYear AS 'Year', LongestWinStreak AS 'Longest Winning Streak'
ROM LongestWinStreaks
RDER BY 'Year' DESC;
```

	Year	Longest Winning Streak
1	2023	10
2	2022	7
3	2021	3
4	2020	4
5	2019	8
6	2018	5
7	2017	2
8	2016	4

What are the longest win streaks by year?



6) What is the longest losing streak?

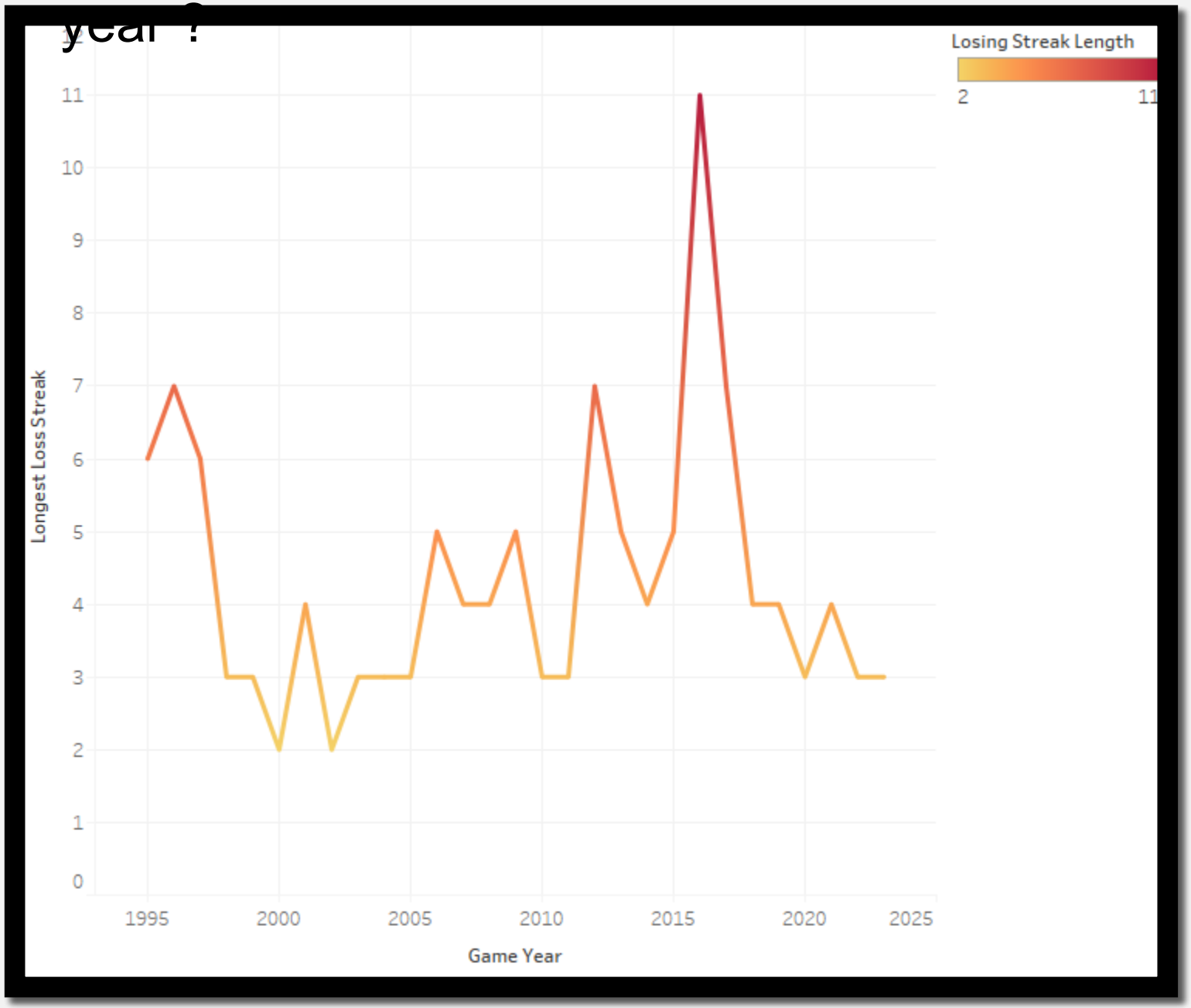
```
ITH RankedGames AS (
  SELECT
      gmeId, gmeDate, gmeScore, trnName,
      ROW_NUMBER() OVER (ORDER BY g.gmeDate DESC) -
      ROW_NUMBER() OVER
      (PARTITION BY YEAR(g.gmeDate), CASE WHEN SUBSTRING(g.gmeScore, 2, 1) > SUBSTRING(g.gmeScore, 4, 1) THEN 'W' ELSE 'L' END
      ORDER BY g.gmeDate DESC) AS GroupNum
  FROM
      [Moneyball.Game] g

ConsecutiveLossGroups AS (
  SELECT
      YEAR(r.gmeDate) AS GameYear,
      r.GroupNum,
      COUNT(*) AS ConsecutiveGamesLost
  FROM
      RankedGames r
  WHERE
      CASE WHEN SUBSTRING(r.gmeScore, 2, 1) < SUBSTRING(r.gmeScore, 4, 1) THEN 'L' ELSE 'W' END = 'L'
  GROUP BY
      YEAR(r.gmeDate),
      r.GroupNum

LongestLossStreaks AS (
  SELECT
      GameYear,
      MAX(ConsecutiveGamesLost) AS LongestLossStreak
  FROM
      ConsecutiveLossGroups
  GROUP BY
      GameYear

ELECT
  GameYear AS [Year],
  LongestLossStreak AS [Longest Losing Streak]
ROM
  LongestLossStreaks
RDER BY
  GameYear DESC;
```

Results Messages

	Year	Longest Losing Streak
1	2023	3
2	2022	3
3	2021	4
4	2020	3
5	2019	4
6	2018	4
7	2017	7
8	2016	11

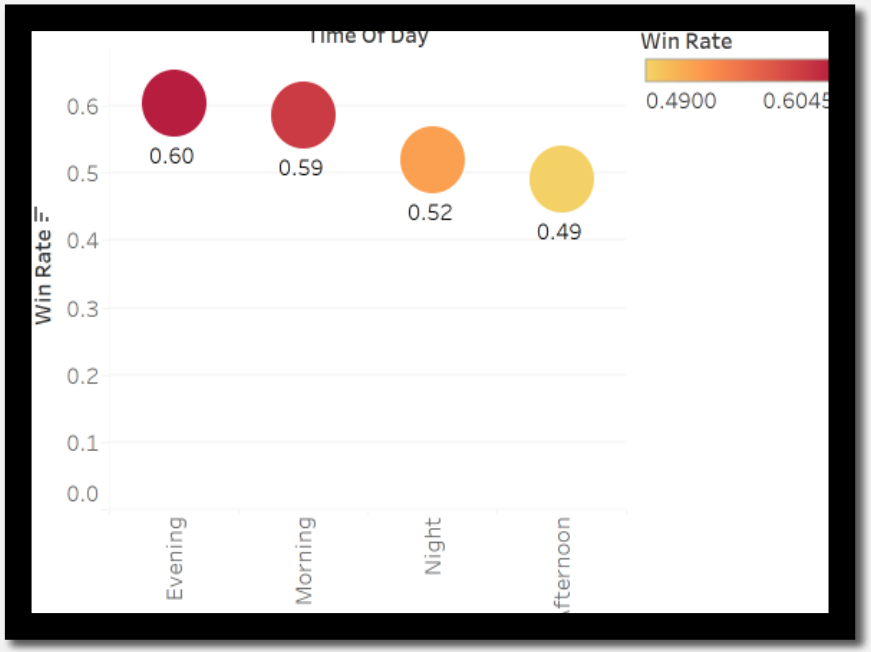
Query executed successfully. doitsqlx.rhsmith.umd.edu,97... AD\neil30 (51) BUDT703_Project_0507_08 00:00:00 29 rows

What is the longest losing streak for each year?



7) What time of day does the UMD softball team perform best?

```
SELECT
    gmeTimeOfDay AS 'Time of Day',
    SUM(CASE
        WHEN CAST(SUBSTRING(gmeScore, 1, CHARINDEX('-', gmeScore) - 1) AS INT) >
             CAST(SUBSTRING(gmeScore, CHARINDEX('-', gmeScore) + 1, LEN(gmeScore)) AS INT) THEN
        ELSE 0
    END) * 1.0 / COUNT(*) AS 'Win Rate'
FROM
    [Moneyball.Game]
WHERE
    gmeTimeOfDay IS NOT NULL
GROUP BY
    gmeTimeOfDay
ORDER BY
    'Win Rate' DESC;
```



Results | Messages

	Time of Day	Win Rate
1	Evening	0.6
2	Morning	0.6
3	Afternoon	0.5
4	Night	0.5

Query executed successfully. | doitsqlx.rbsmith.umd.edu,97 | AD\bhoomi (68) | BUDT703_Project_0507_08 | 00:00:00 | 4 rows

CONCLUSION

- This project has provided valuable insights into the world of sports analytics through the lens of a baseball database.
- By leveraging SQL queries and Tableau visualizations, we have successfully extracted meaningful information from the dataset, ranging from team performance metrics to opponent win rates and tournament statistics.
- The use of SQL views, such as 'OpponentWinRate' and 'TournamentWinRate,' has allowed for a more organized and efficient analysis of key metrics.
- Throughout the project, we have demonstrated the power of combining SQL and Tableau for comprehensive sports analytics, showcasing the potential for data-driven decision-making in the realm of baseball and beyond.