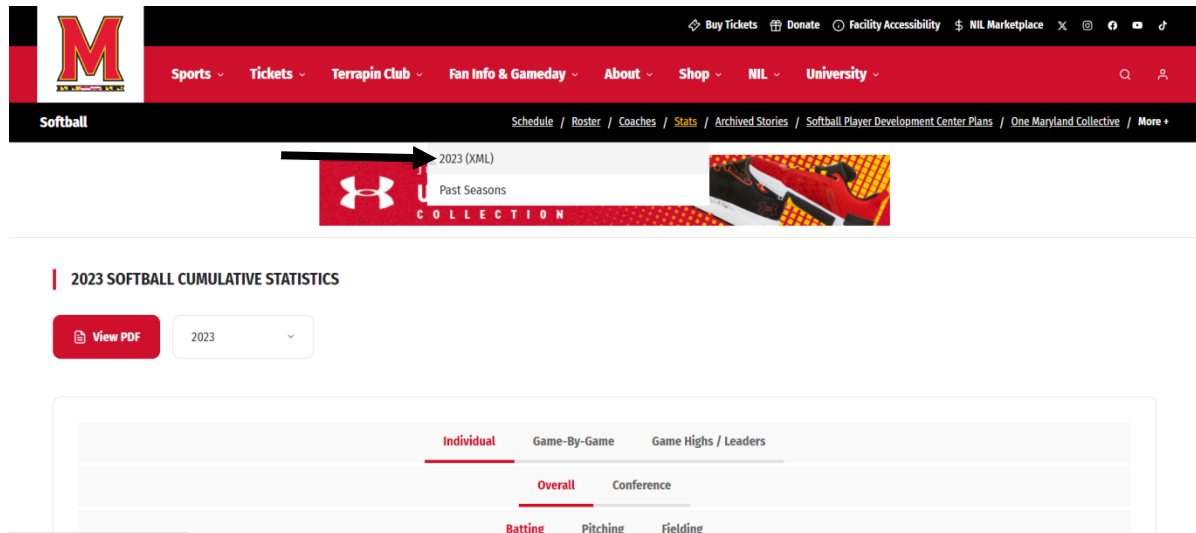


Project Description:

By diving deep into the data of the UMD softball team, we've gained key insights into their overall performance across various metrics. This comprehensive analysis will be a valuable tool for the team to identify areas for improvement and ultimately reach their full potential.

Data Source:

We took last 20 years data from the official website of UMD's softball team website. (<https://umterps.com>) Below screen shot shows exactly where we got our data from.



Data Scrapping:

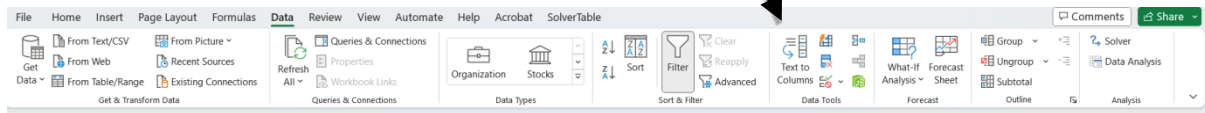
To obtain the data, we employed a manual approach, directly downloading the CSV file from the official UMD softball website (<https://umterps.com>) from the year 2004 to 2023. While utilizing tools like Beautiful Soup in Python could automate the process, we opted for manual download due to potential privacy concerns outlined in the website's policies.

Data Evaluation:

After we scrapped the data, we evaluated the dataset keeping our logical and physical database design in mind which is to follow. We evaluated the redundant fields and data which is not necessary for the client. We have 1052 rows of data when combined for each year starting from year 2004 to 2023.

Data cleaning and Pre-processing:

To prepare the data for analysis, we leveraged Microsoft Excel. Our first step was to utilize the "Text to Columns" feature, separating the data into distinct columns for easier manipulation.



Next, we proceeded to eliminate unnecessary attributes, such as coaches' names, which held no relevance to our investigation. Since the file contained a lot of parathesis, commas and brackets which need to be removed before importing the data into SQL, we used 'find and select' and filtering functions. Below screenshot shows how our data looked before importation into SQL.

File Home Insert Page Layout Formulas Data Review View Automate Help Acrobat SolverTable																			Comments		Share																																									
Calibri 11 A A			General Conditional Formatting			Insert			Σ fx Filter Select			Sensitivity			Add-ins			Analyze Data			Create PDF and Share			Create PDF and Outlook																																						
Font			Cell Styles			Delete			Cats			Editing			Sensitivity			Add-ins			Add-ins			Add-ins																																						
Clipboard			Alignment			Number			Styles																																																					
AP14			A B C D E F G H I J K L M N O P Q R S T																																																											
1			open parenthesis			open braces			gid			close braces			comma			open braces			Formatted date			close braces			comma			open braces			Time			close braces			comma			open braces			Season			close braces			comma			open braces			venCategory			close		
2			("			G01			"			,			"			2023-05-20			"			,			"			2 PM			"			,			"			Spring			"			,			"			Neutral			"		
3			("			G02			"			,			"			2023-05-19			"			,			"			2 P.M.			"			,			"			Spring			"			,			"			Neutral			"		
4			("			G03			"			,			"			2023-05-19			"			,			"			5:00 PM			"			,			"			Spring			"			,			"			Neutral			"		
5			("			G04			"			,			"			2023-05-18			"			,			"			12 PM			"			,			"			Spring			"			,			"			Neutral			"		
6			("			G05			"			,			"			2023-05-10			"			,			"			5:30 PM			"			,			"			Spring			"			,			"			Neutral			"		
7			("			G06			"			,			"			2023-05-07			"			,			"			12:00 PM			"			,			"			Spring			"			,			"			Away			"		
8			("			G07			"			,			"			2023-05-06			"			,			"			02:00 PM			"			,			"			Spring			"			,			"			Away			"		
9			("			G08			"			,			"			2023-05-05			"			,			"			05:30 PM			"			,			"			Spring			"			,			"			Away			"		
10			("			G09			"			,			"			2023-05-02			"			,			"			06:00 PM			"			,			"			Spring			"			,			"			Home			"		
11			("			G10			"			,			"			2023-04-29			"			,			"			2 PM			"			,			"			Spring			"			,			"			Home			"		
12			("			G11			"			,			"			2023-04-29			"			,			"			4:30 PM			"			,			"			Spring			"			,			"			Home			"		
13			("			G12			"			,			"			2023-04-23			"			,			"			12 PM			"			,			"			Spring			"			,			"			Home			"		
14			("			G13			"			,			"			2023-04-22			"			,			"			11:30 AM			"			,			"			Spring			"			,			"			Home			"		
15			("			G14			"			,			"			2023-04-21			"			,			"			6:00 PM			"			,			"			Spring			"			,			"			Home			"		
16			("			G15			"			,			"			2023-04-19			"			,			"			6:00 PM			"			,			"			Spring			"			,			"			Home			"		
17			("			G16			"			,			"			2023-04-15			"			,			"			2 PM			"			,			"			Spring			"			,			"			Away			"		
18			("			G17			"			,			"			2023-04-14			"			,			"			4 PM			"			,			"			Spring			"			,			"			Away			"		
19			("			G18			"			,			"			2023-04-14			"			,			"			6:30 PM			"			,			"			Spring			"			,			"			Away			"		
20			("			G19			"			,			"			2023-04-09			"			,			"			12:00 PM			"			,			"			Spring			"			,			"			Home			"		
21			("			G20			"			,			"			2023-04-08			"			,			"			02:00 PM			"			,			"			Spring			"			,			"			Home			"		
22			("			G21			"			,			"			2023-04-07			"			,			"			06:00 PM			"			,			"			Spring			"			,			"			Home			"		
23			("			G22			"			,			"			2023-04-05			"			,			"			01:00 PM			"			,			"			Spring			"			,			"			Away			"		
24			("			G23			"			,			"			2023-04-05			"			,			"			03:00 PM			"			,			"			Spring			"			,			"			Away			"		
25			("			G24			"			,			"			2023-04-02			"			,			"			12:00 PM			"			,			"			Spring			"			,			"			Home			"		
26			("			G25			"			,			"			2023-03-31			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
27			("			G26			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
28			("			G27			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
29			("			G28			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
30			("			G29			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
31			("			G30			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
32			("			G31			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
33			("			G32			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
34			("			G33			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
35			("			G34			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
36			("			G35			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
37			("			G36			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
38			("			G37			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
39			("			G38			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
40			("			G39			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
41			("			G40			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
42			("			G41			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
43			("			G42			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
44			("			G43			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
45			("			G44			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
46			("			G45			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
47			("			G46			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
48			("			G47			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
49			("			G48			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
50			("			G49			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
51			("			G50			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
52			("			G51			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
53			("			G52			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
54			("			G53			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
55			("			G54			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
56			("			G55			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
57			("			G56			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
58			("			G57			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
59			("			G58			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
60			("			G59			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
61			("			G60			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring			"			,			"			Home			"		
62			("			G61			"			,			"			2023-03-24			"			,			"			3:30 PM			"			,			"			Spring																	

Following the data acquisition from the source, extensive cleaning and pre-processing were necessary to ensure compatibility with SQL databases.

- **Date and Time Formatting:** We implemented a consistent date-time format adhering to SQL conventions to facilitate analysis and ensure data integrity.
- **Redundancy Removal:** Duplicate entries were meticulously identified and excluded from the dataset, guaranteeing accurate data representation.
- **Cancellation Handling:** Games marked as "Cancelled" were strategically removed from the dataset to ensure precise analysis and avoid skewed results.
- **Primary Key Assignment:** To enhance data integrity and maintain efficient database management, unique primary keys were generated and assigned to each table.
- **Season Attribute Calculation:** An additional attribute, "Seasons," was calculated and appended to the "IndividualGame" table based on the respective game dates.

Then we split the data into different worksheets according to the number of entities we plan to have in our ERD. A screenshot of which is attached below:

AutoSave OFF latest Insert into table - formatted data - Protected... Saved to this PC

File Home Insert Page Layout Formulas Data Review View Automate Help Acrobat SolverTable

PROTECTED VIEW Be careful—files from the Internet can contain viruses. Unless you need to edit, it's safer to stay in Protected View. Enable Editing

R2 VALUES

	A	B	C	D	E	F	G	H	I
1									
2		open paranthesis	open brace	gld	close brace	comr	open brace	optld	close brace
3		("	G01	"	"	"	O01	"
4		("	G02	"	"	"	O02	"
5		("	G03	"	"	"	O03	"
6		("	G04	"	"	"	O04	"
7		("	G05	"	"	"	O01	"
8		("	G06	"	"	"	O05	"
9		("	G07	"	"	"	O05	"
10		("	G08	"	"	"	O05	"
11		("	G09	"	"	"	O06	"
12		("	G10	"	"	"	O07	"
13		("	G11	"	"	"	O07	"
14		("	G12	"	"	"	O08	"
15		("	G13	"	"	"	O08	"
16		("	G14	"	"	"	O08	"
17		("	G15	"	"	"	O09	"
18		("	G16	"	"	"	O01	"
19		("	G17	"	"	"	O01	"
20		("	G18	"	"	"	O01	"
21		("	G19	"	"	"	O10	"
22		("	G20	"	"	"	O10	"
23		("	G21	"	"	"	O10	"
24		("	G22	"	"	"	O11	"
25		("	G23	"	"	"	O11	"
26		("	G24	"	"	"	O12	"
27		("	G25	"	"	"	O12	"
28		("	G26	"	"	"	O12	"
29		("	G27	"	"	"	O13	"
30		("	G28	"	"	"	O14	"
31		("	G29	"	"	"	O14	"

IndividualGames - formatted GameVenue - formatted OpponentTeam - FORMATTED

Creating the Database:

For creation of the data, we used "USE BUDT703_Project_0507_05" as database.

First, we drop the tables and then create them in reverse order. Following are the screenshots of the drop and create statements:

.

```
--SQL drop tables:
DROP TABLE IF EXISTS [TerrapinTactics.Play]
DROP TABLE IF EXISTS [TerrapinTactics.OpponentTeam];
DROP TABLE IF EXISTS [TerrapinTactics.GameVenue];
DROP TABLE IF EXISTS [TerrapinTactics.GameResult];
DROP TABLE IF EXISTS [TerrapinTactics.IndividualGame];
```

```
--SQL create tables:
|CREATE TABLE [TerrapinTactics.IndividualGame] (
    gamId CHAR (5) NOT NULL,
    gamDate DATE,
    gamSeason VARCHAR (10),
    gamVenCategory VARCHAR (10),
    CONSTRAINT pk_IndividualGame_gamId PRIMARY KEY (gamId)
)

|CREATE TABLE [TerrapinTactics.GameResult] (
    resId CHAR (5) NOT NULL,
    resUmdScore INT,
    resOptScore INT,
    gamId CHAR (5),
    CONSTRAINT pk_GameResult_resId PRIMARY KEY (resId),
    CONSTRAINT fk_GameResult_gamId FOREIGN KEY (gamId)
        REFERENCES [TerrapinTactics.IndividualGame] (gamId)
        ON DELETE NO ACTION
        ON UPDATE NO ACTION )

|CREATE TABLE [TerrapinTactics.GameVenue] (
    venId CHAR (5) NOT NULL,
    venName VARCHAR (50),
    CONSTRAINT pk_GameVenue_venId PRIMARY KEY (venId)
)

|CREATE TABLE [TerrapinTactics.OpponentTeam] (
    optId CHAR (5) NOT NULL,
    optName VARCHAR (50),
    CONSTRAINT pk_OpponentTeam_optId PRIMARY KEY (optId),
)

|CREATE TABLE [TerrapinTactics.Play] (
    gamId CHAR (5),
    optId CHAR (5),
    venId CHAR (5),
    CONSTRAINT pk_Play_gamId_optId_venId PRIMARY KEY (gamId,optId,venId),
    CONSTRAINT fk_Play_gamId FOREIGN KEY (gamId)
        REFERENCES [TerrapinTactics.IndividualGame] (gamId)
        ON DELETE NO ACTION
        ON UPDATE NO ACTION,
    CONSTRAINT fk_Play_optId FOREIGN KEY (optId)
        REFERENCES [TerrapinTactics.OpponentTeam] (optId)
        ON DELETE NO ACTION
        ON UPDATE NO ACTION,
    CONSTRAINT fk_Play_venId FOREIGN KEY (venId)
        REFERENCES [TerrapinTactics.GameVenue] (venId)
        ON DELETE NO ACTION
        ON UPDATE CASCADE
)

```

Inserting Data:

Subsequently we start inserting the data. The following are the sample of all insert statements.

1. Individual game

```
INSERT INTO [TerrapinTactics.IndividualGame] (gamId, gamDate, gamSeason, gamVenCategory)
VALUES
('G01', '2023-05-20', 'Spring', 'Neutral'),
('G02', '2023-05-19', 'Spring', 'Neutral'),
('G03', '2023-05-19', 'Spring', 'Neutral'),
('G04', '2023-05-18', 'Spring', 'Neutral'),

('G05', '2023-05-10', 'Spring', 'Neutral'),

```

2. Game result

```
INSERT INTO [TerrapinTactics.GameResult] (resId, resUmdScore, resOptScore,gamId)
VALUES
('R01', 3, 4, 'G01'),
('R02', 4, 6, 'G02'),
('R03', 7, 4, 'G03'),
('R04', 8, 0, 'G04'),
('R05', 1, 7, 'G05'),
```

3. Game Venue

```
INSERT INTO [TerrapinTactics.GameVenue] (venId, venName)
VALUES
('V01', 'Fort Collins, CO'),
('V02', 'Urbana-Champaign, IL'),
('V03', 'College Park, MD'),
('V04', 'Iowa City, IA'),
('V05', 'Piscataway, NJ'),
```

4. Opponent Team

```
INSERT INTO [TerrapinTactics.OpponentTeam] (optId, optName)
VALUES
('O01', 'IOWA'),
('O02', 'BYU'),
('O03', 'SAN JOSE STATE'),
('O04', 'CSUN'),
('O05', 'ILLINOIS'),
```

5. Play

```
INSERT INTO [TerrapinTactics.Play] (gamId, optId, venId)
VALUES
('G01', 'O01', 'V01'),
('G02', 'O02', 'V01'),
('G03', 'O03', 'V01'),
('G04', 'O04', 'V01'),
('G05', 'O01', 'V02'),
```

SQL query for use cases:

1. Year on year win percentage of UMD by season

Pitcher Performance:

The advantage given to pitchers during the winter due to atmospheric conditions may not have been fully capitalized upon by UMD's pitchers. It is essential to evaluate whether the pitchers have been trained to adapt to and exploit these conditions. Specialized training programs focusing on ball control and exploiting air movement could enhance pitcher effectiveness.

Batter Adaptability:

Batters face additional challenges in winter, such as adjusting to increased ball movement. Their ability to adapt to these conditions is crucial. Batting practice that simulates winter conditions, including the use of machines that can mimic increased ball movement, may help batters better prepare for winter games.

Recruitment and Scouting:

Recruitment strategies might need to focus on attracting pitchers with a proven ability to utilize and control the movement of the ball in cold conditions. Similarly, batters adept at handling swing in such conditions could be a valuable asset for the winter season.

```
DROP VIEW IF EXISTS SeasonalPerformanceTrends
GO
CREATE VIEW SeasonalPerformanceTrends AS
SELECT YEAR(i.gamDate) AS Year, gamSeason,
       COUNT(*) AS TotalGames,
       SUM(CASE WHEN resUmdScore > resOptScore THEN 1 ELSE 0 END) AS Wins,
       SUM(CASE WHEN resUmdScore < resOptScore THEN 1 ELSE 0 END) AS Losses,
       ROUND(
         CASE
           WHEN COUNT(*) = 0 THEN NULL
           ELSE CAST(SUM(CASE WHEN resUmdScore > resOptScore THEN 1 ELSE 0 END) AS FLOAT) / COUNT(*)
         END * 100, 2
       ) AS WinPercentage
FROM [TerrapinTactics.IndividualGame] i
JOIN [TerrapinTactics.GameResult] ON i.gamId = [TerrapinTactics.GameResult].gamId
GROUP BY YEAR(gamDate), gamSeason
GO
SELECT * FROM SeasonalPerformanceTrends
ORDER BY Year
```

2. Year on year win percentage of UMD by venue category

```

DROP VIEW IF EXISTS HomeAwayPerformancePerYear
GO
CREATE VIEW HomeAwayPerformancePerYear AS
SELECT
    YEAR(i.gamDate) AS Year,
    i.gamVenCategory AS 'Venue Category',
    COUNT(*) AS TotalGames,
    SUM(CASE WHEN g.resUmdScore > g.resOptScore THEN 1 ELSE 0 END) AS Wins,
    SUM(CASE WHEN g.resUmdScore < g.resOptScore THEN 1 ELSE 0 END) AS Losses,
    ROUND(
        CAST(SUM(CASE WHEN g.resUmdScore > g.resOptScore THEN 1 ELSE 0 END) AS FLOAT) / COUNT(*) * 100, 2
    ) AS WinPercentage
FROM [TerrapinTactics.IndividualGame] i
JOIN [TerrapinTactics.GameResult] g ON i.gamId = g.gamId
GROUP BY YEAR(i.gamDate), i.gamVenCategory;
GO

SELECT * FROM HomeAwayPerformancePerYear
ORDER BY Year, 'Venue Category';

```

3. UMD Dominance: top 10 teams with highest win percentage

```

DROP VIEW IF EXISTS TopTenOpponentWinPercentages
GO
CREATE VIEW TopTenOpponentWinPercentages AS
WITH WinPercentages AS (
    SELECT
        o.optName AS Opponent, COUNT(*) AS TotalGames,
        SUM(CASE WHEN g.resUmdScore > g.resOptScore THEN 1 ELSE 0 END) AS Wins,
        SUM(CASE WHEN g.resUmdScore < g.resOptScore THEN 1 ELSE 0 END) AS Losses,
        CAST(SUM(CASE WHEN g.resUmdScore > g.resOptScore THEN 1 ELSE 0 END) AS FLOAT) / COUNT(*) * 100 AS WinPercentage
    FROM [TerrapinTactics.IndividualGame] i
    JOIN [TerrapinTactics.GameResult] g ON i.gamId = g.gamId
    JOIN [TerrapinTactics.Play] p ON i.gamId = p.gamId
    JOIN [TerrapinTactics.OpponentTeam] o ON p.optId = o.optId
    GROUP BY o.optName
)
SELECT TOP 10
    Opponent, TotalGames, Wins, Losses, ROUND(WinPercentage, 2) AS WinPercentage
FROM WinPercentages
WHERE TotalGames >= 10
ORDER BY WinPercentage DESC;
GO
SELECT * FROM TopTenOpponentWinPercentages;

```

4. UMD'S toughest opponent: Top 10 teams with the lowest win percentage

```

DROP VIEW IF EXISTS BottomTenOpponentWinPercentages
GO
CREATE VIEW BottomTenOpponentWinPercentages AS
WITH WinPercentages AS (
    SELECT
        o.optName AS Opponent, COUNT(*) AS TotalGames, SUM(CASE WHEN g.resUmdScore > g.resOptScore THEN 1 ELSE 0 END) AS Wins,
        SUM(CASE WHEN g.resUmdScore < g.resOptScore THEN 1 ELSE 0 END) AS Losses,
        CAST(SUM(CASE WHEN g.resUmdScore > g.resOptScore THEN 1 ELSE 0 END) AS FLOAT) / COUNT(*) * 100 AS WinPercentage
    FROM [TerrapinTactics.IndividualGame] i
    JOIN [TerrapinTactics.GameResult] g ON i.gamId = g.gamId
    JOIN [TerrapinTactics.Play] p ON i.gamId = p.gamId
    JOIN [TerrapinTactics.OpponentTeam] o ON p.optId = o.optId
    GROUP BY o.optName
)
SELECT TOP 10
    Opponent, TotalGames, Wins, Losses, ROUND(WinPercentage, 2) AS WinPercentage
FROM WinPercentages
WHERE TotalGames >= 10
ORDER BY WinPercentage ASC; -- Ordering by ascending win percentage to get the bottom 10
GO
SELECT * FROM BottomTenOpponentWinPercentages;

```

5. Top 10 best performing venues for UMD

```
DROP VIEW IF EXISTS TopVenuesByWinPercentage
GO
CREATE VIEW TopVenuesByWinPercentage AS
WITH VenueWins AS (
    SELECT
        gv.venId, gv.venName, COUNT(CASE WHEN gr.resUmdScore > gr.resOptScore THEN 1 END) AS Wins,
        COUNT(*) AS TotalGames
    FROM [TerrapinTactics.GameVenue] gv
    JOIN [TerrapinTactics.Play] p ON gv.venId = p.venId
    JOIN [TerrapinTactics.GameResult] gr ON p.gamId = gr.gamId
    GROUP BY gv.venId, gv.venName
    HAVING COUNT(*) >= 5 -- Ensures only venues where UMD has played 5 or more games are considered
)
SELECT TOP 10
    venName, Wins, TotalGames, ROUND((CAST(Wins AS FLOAT) / TotalGames) * 100, 2) AS WinPercentage
FROM VenueWins
ORDER BY WinPercentage DESC, venName;
GO
SELECT * FROM TopVenuesByWinPercentage;
```

6. Top 10 worst performing venues for UMD

```
DROP VIEW IF EXISTS BottomVenuesByWinPercentage;
GO
CREATE VIEW BottomVenuesByWinPercentage AS
WITH VenueWins AS (
    SELECT gv.venId, gv.venName, COUNT(CASE WHEN gr.resUmdScore > gr.resOptScore THEN 1 END) AS Wins,
        COUNT(*) AS TotalGames
    FROM [TerrapinTactics.GameVenue] gv
    JOIN [TerrapinTactics.Play] p ON gv.venId = p.venId
    JOIN [TerrapinTactics.GameResult] gr ON p.gamId = gr.gamId
    GROUP BY
        gv.venId, gv.venName
    HAVING
        COUNT(*) >= 5
)
SELECT TOP 10
    venName, Wins, TotalGames, ROUND((CAST(Wins AS FLOAT) / TotalGames) * 100, 2) AS WinPercentage
FROM VenueWins
ORDER BY WinPercentage ASC, venName; -- Order by ascending win percentage
GO
SELECT * FROM BottomVenuesByWinPercentage;
```


How to test the project:

In order to test the project open the MySQL software. Open the file ' Project_0507_05_Create'.

1. Drop the tables: Select the below code and click on execute

```
--SQL drop tables:
DROP TABLE IF EXISTS [TerrapinTactics.Play]
DROP TABLE IF EXISTS [TerrapinTactics.OpponentTeam];
DROP TABLE IF EXISTS [TerrapinTactics.GameVenue];
DROP TABLE IF EXISTS [TerrapinTactics.GameResult];
DROP TABLE IF EXISTS [TerrapinTactics.IndividualGame];
```

2. Run the create statements to create the tables : select the below code and click on execute.
This will create all the tables with required constraints and attributes.

```
--SQL create tables:
|CREATE TABLE [TerrapinTactics.IndividualGame] (
    gamId CHAR (5) NOT NULL,
    gamDate DATE,
    gamSeason VARCHAR (10),
    gamVenCategory VARCHAR (10),
    CONSTRAINT pk_IndividualGame_gamId PRIMARY KEY (gamId)
)

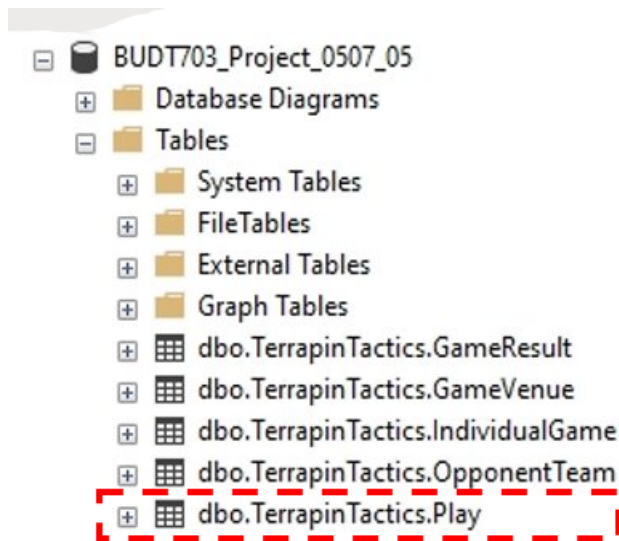
|CREATE TABLE [TerrapinTactics.GameResult] (
    resId CHAR (5) NOT NULL,
    resUmdScore INT,
    resOptScore INT,
    gamId CHAR (5),
    CONSTRAINT pk_GameResult_resId PRIMARY KEY (resId),
    CONSTRAINT fk_GameResult_gamId FOREIGN KEY (gamId)
        REFERENCES [TerrapinTactics.IndividualGame] (gamId)
        ON DELETE NO ACTION
        ON UPDATE NO ACTION )

|CREATE TABLE [TerrapinTactics.GameVenue] (
    venId CHAR (5) NOT NULL,
    venName VARCHAR (50),
    CONSTRAINT pk_GameVenue_venId PRIMARY KEY (venId)
)

|CREATE TABLE [TerrapinTactics.OpponentTeam] (
    optId CHAR (5) NOT NULL,
    optName VARCHAR (50),
    CONSTRAINT pk_OpponentTeam_optId PRIMARY KEY (optId),
)

|CREATE TABLE [TerrapinTactics.Play] (
    gamId CHAR (5),
    optId CHAR (5),
    venId CHAR (5),
    CONSTRAINT pk_Play_gamId_optId_venId PRIMARY KEY (gamId,optId,venId),
    CONSTRAINT fk_Play_gamId FOREIGN KEY (gamId)
        REFERENCES [TerrapinTactics.IndividualGame] (gamId)
        ON DELETE NO ACTION
        ON UPDATE NO ACTION,
    CONSTRAINT fk_Play_optId FOREIGN KEY (optId)
        REFERENCES [TerrapinTactics.OpponentTeam] (optId)
        ON DELETE NO ACTION
        ON UPDATE NO ACTION,
    CONSTRAINT fk_Play_venId FOREIGN KEY (venId)
        REFERENCES [TerrapinTactics.GameVenue] (venId)
        ON DELETE NO ACTION
        ON UPDATE CASCADE
)
```

3. Ensure that the tables are created in the database : An example is shown below in the screenshot

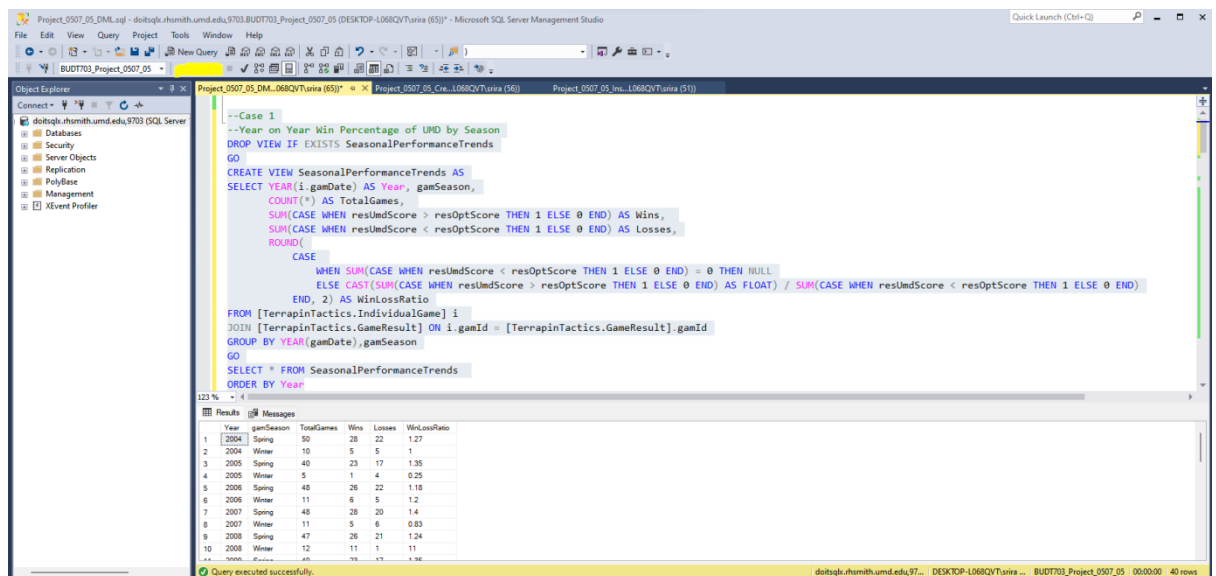


4. Run insert statements: Open the file 'Project_0507_05_Insert '
Select the INSERT query and click on execute. A snippet of the INSERT query is shown below.

```
USE BUDT703_Project_0507_05
```

```
INSERT INTO [TerrapinTactics.IndividualGame] (gamId, gamDate, gamSeason, gamVenCategory)
VALUES
('G01', '2023-05-20', 'Spring', 'Neutral'),
('G02', '2023-05-19', 'Spring', 'Neutral'),
('G03', '2023-05-19', 'Spring', 'Neutral'),
('G04', '2023-05-18', 'Spring', 'Neutral'),
('G05', '2023-05-10', 'Spring', 'Neutral'),
('G06', '2023-05-07', 'Spring', 'Away'),
('G07', '2023-05-06', 'Spring', 'Away'),
('G08', '2023-05-05', 'Spring', 'Away'),
('G09', '2023-05-02', 'Spring', 'Home'),
('G10', '2023-04-29', 'Spring', 'Home'),
```

5. In order to run the use cases open file 'Project_0507_05_DML' and select the required query and click on the execute. A snippet of the same has been attached below. The execute button is highlighted in yellow color.



Data Visualization and interpretation:

We used tableau for visualizing the use cases that we came up with. The following are the tableau outputs and their interpretation.

1. Year on year win percentage of UMD by season

The provided line chart visualizes the University of Maryland's (UMD) softball team's win percentages over a series of years, segmented into two seasons: Winter and Spring. Several insights can be drawn from this visualization to inform strategic decisions and actions:

Pitcher Performance:

The advantage given to pitchers during the winter due to atmospheric conditions may not have been fully capitalized upon by UMD's pitchers. It is essential to evaluate whether the pitchers have been trained to adapt to and exploit these conditions. Specialized training programs focusing on ball control and exploiting air movement could enhance pitcher effectiveness.

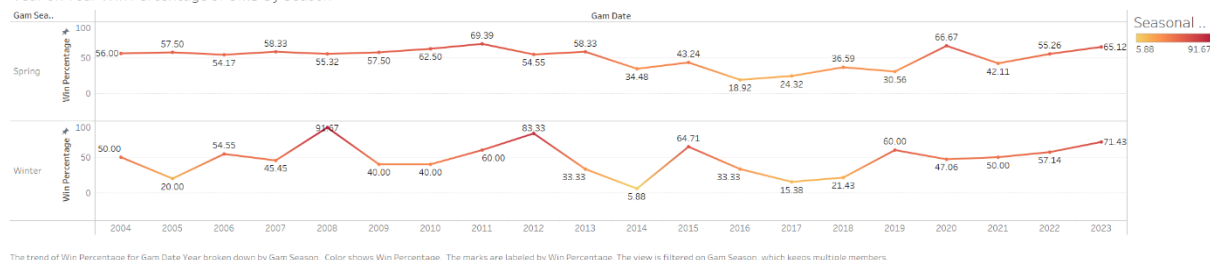
Batter Adaptability:

Batters face additional challenges in winter, such as adjusting to increased ball movement. Their ability to adapt to these conditions is crucial. Batting practice that simulates winter conditions, including the use of machines that can mimic increased ball movement, may help batters better prepare for winter games.

Recruitment and Scouting:

Recruitment strategies might need to focus on attracting pitchers with a proven ability to utilize and control the movement of the ball in wintry conditions. Similarly, batter's adept at handling swing in such conditions could be an asset for the winter season.

Year on Year Win Percentage of UMD by Season



2. Year on year win percentage of UMD by venue category

The bar chart analysis reveals a shift in the University of Maryland's (UMD) softball team performance by venue type over the years. UMD has historically shown a higher win percentage in home games, but recent data points to a significant improvement in both away and neutral venues. Key insights and strategic recommendations include:

Chasing Advantage: UMD's increased away wins suggest they excel in batting second, capitalizing on the chance to chase the target. This indicates a solid batting lineup and effective game pacing.

Away Game Strategies: Leveraging the chasing strength, UMD should continue to cultivate a resilient batting team adept at strategizing according to the game's flow and opponent's target.

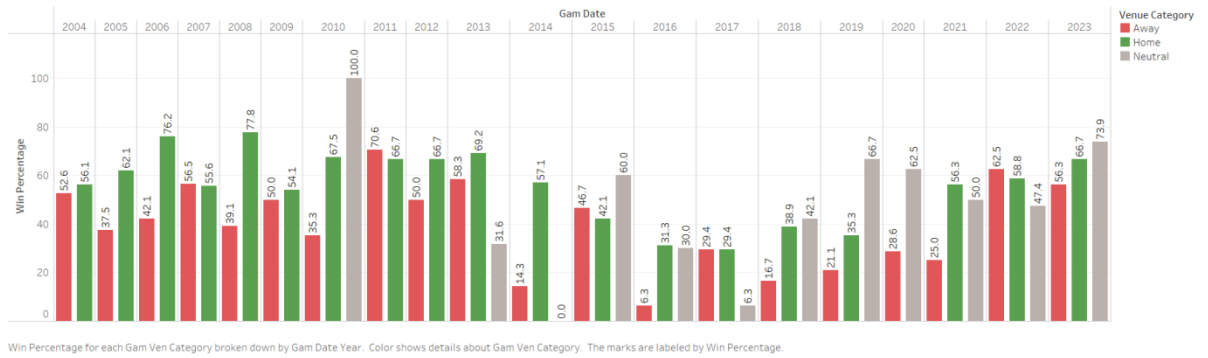
Neutral Venue Success: The rise in neutral venue wins suggests UMD's ability to adapt and possibly benefit from batting second, often a privilege of the higher-ranked team. This points to improved rankings and adaptability.

Home Venue Tactics: With a relatively smaller increase in home wins, UMD should reassess their approach when batting first. This could involve aggressive run-scoring strategies and fostering a strong opening partnership.

Scheduling for Strength: Future scheduling could focus on selecting away games where UMD has historically succeeded and maintaining a high ranking to bat second in neutral games.

Home Ground Focus: To bolster home win rates, UMD may look to aggressive batting tactics that build early pressure, ensuring a defensible score is set from the outset.

Year on Year Win Percentage of UMD by Venue Category



3. UMD Dominance: top 10 teams with highest win percentage

The bubble chart provided illustrates UMD's win percentage against opponents where they have played at least five games. The size of the bubble represents the win percentage, providing a visual indicator of UMD's dominance over certain teams. Here are actionable insights from the chart:

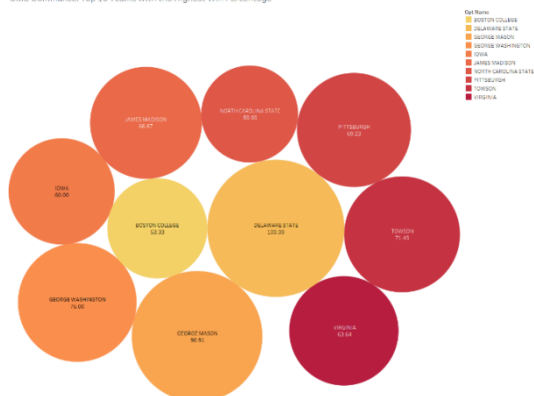
Targeted Strategy for High Win Percentage Opponents:

UMD has achieved high win percentages against teams such as George Mason and Delaware State. This indicates a successful strategy or a favorable matchup. UMD should analyze what factors have contributed to this success, such as specific game plans, player matchups, or psychological advantages, and look to replicate these strategies in future games.

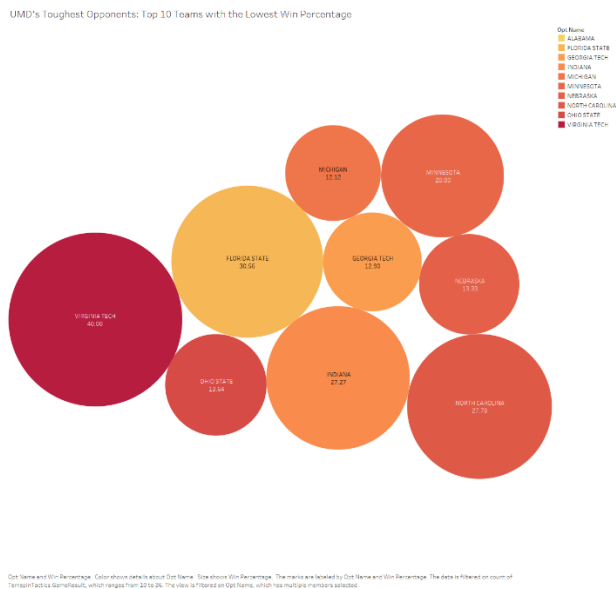
Contextual Performance Evaluation:

Evaluate the context of past performances, considering variables such as weather conditions, player availability, and game significance, to understand the true nature of each victory and loss.

UMD Dominance: Top 10 Teams with the Highest Win Percentage



4. UMD'S toughest opponent: Top 10 teams with the lowest win percentage



5. Top 10 best performing venues for UMD

The bar chart illustrates the top 10 venues where the University of Maryland (UMD) has the highest win percentages. From the chart, we can infer several meaningful insights:

Venue-Specific Performance:

UMD has varying levels of success at different venues, with some locations like Greenville, NC, and Puerto Vallarta, MX, showing significantly higher win rates. This suggests that the team may be particularly well-adapted to the conditions or atmosphere at these venues.

Environmental and Regional Considerations:

The venues listed are geographically diverse, indicating that UMD's team is capable of performing well under a variety of environmental conditions. This adaptability is a strength that can be further leveraged.

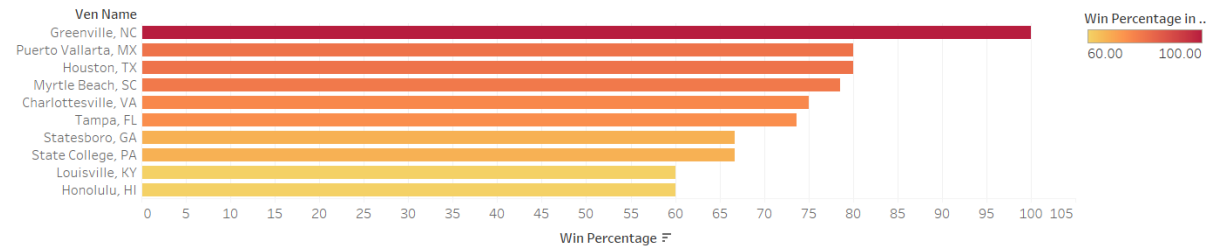
Strategic Scheduling:

When possible, UMD might consider scheduling more games at these top-performing venues to capitalize on their historical success rates.

Analyzing Venue Commonalities:

A deeper analysis of the common factors shared by these venues—such as field type, weather patterns, or travel logistics—could provide insights that help UMD replicate success across other venues.

Top 10 best performing venues for UMD



Win Percentage for each Ven Name. Color shows Win Percentage. The data is filtered on Distinct no of games played in venue, which ranges from 5 to 408. The view is filtered on Ven Name, which has multiple members selected.

6. Top 10 worst performing venues for UMD

The provided bar chart represents the top 10 venues where the University of Maryland (UMD) has experienced the lowest win percentages, which can provide valuable insights for the team’s strategic development:

Venue-Specific Challenges:

UMD should conduct a thorough review of past performances at these venues to identify any common challenges, such as environmental factors, venue-specific conditions, or psychological barriers that may impact their play.

Customized Training Regimens:

Tailoring training regimens to simulate the conditions of these challenging venues could help the team adapt better. For instance, if a venue like Ann Arbor, MI, presents specific weather challenges, practicing under similar conditions could prove beneficial.

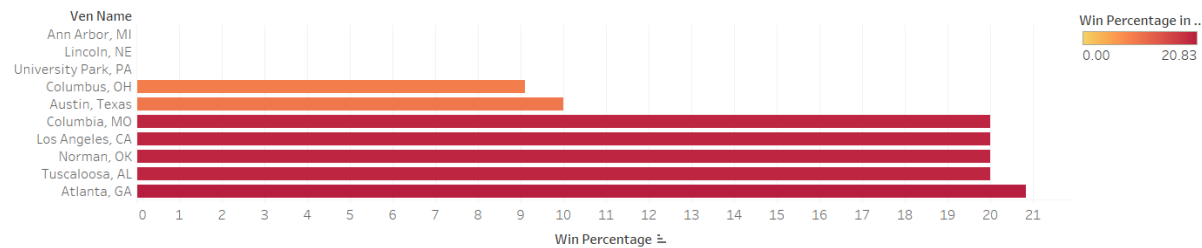
Data-Driven Tactical Adjustments:

Analyzing game data to understand the tactical adjustments required when playing at these venues will be crucial. For example, if certain pitches or strategies are less effective, coaches can revise the game plan accordingly.

Strategic Scheduling:

While not always within the team’s control, strategically scheduling games to minimize consecutive challenging away matches could help prevent fatigue and maintain morale.

Top 10 worst performing venues for UMD



Win Percentage for each Ven Name. Color shows Win Percentage. The data is filtered on Distinct no of games played in venue, which ranges from 5 to 408. The view is filtered on Ven Name, which has multiple members selected.

DASHBOARD

