

# Readme: CurveBall Consultants Softball Project

## Data Sources:

The data for this project is sourced from the University of Maryland Terrapins Softball schedule available at [<https://umterps.com/sports/softball/schedule/>].

## Data Cleanup:

This data from the above sources had to be imported into Microsoft Word and then Excel for cleanup. The extra blank lines between data rows were removed in Word. Excel was used to split the data into columns. Then using Find & Replace, the data points were trimmed, and corrected to match the majority of the existing data. Later this Excel was saved as a comma-separated value document (CSV) to be added to our C# program. Our C# program segregated each row's information into four entities Opponent, Tournament, Location, and Match. By cross-referencing we eliminated duplicates and match records with too many empty cells. We also discarded canceled match records. Finally, the code was utilized to convert data rows into SQL INSERT statements for each entity. The steps to each process mentioned here are given below in the 'How to Test Our Project Section'.

## SQL Database Setup:

1. Database Name: BUDT703\_Project\_0507\_01
2. Tables:
  - `CurveBallConsultants.Opponent`: Stores information about softball opponents.
  - `CurveBallConsultants.Tournament`: Contains data about softball tournaments.
  - `CurveBallConsultants.Location`: Stores details about match locations.
  - `CurveBallConsultants.Match`: Holds information about softball matches, including opponents, locations, scores, and outcomes.
3. Sample Insert Statements:
  - `INSERT INTO [CurveBallConsultants.Opponent] (oppId, oppName) VALUES ('001', 'North Carolina'), ('002', 'George Mason'), ('003', 'Goldey-Beacom');`
  - `INSERT INTO [CurveBallConsultants.Tournament] (trnId, trnName) VALUES ('001', 'Winthrop Tournament'), ('002', 'Buzz Classic'), ('003', 'ACC Tournament');`
  - `INSERT INTO [CurveBallConsultants.Location] (locId, locCity, locState) VALUES ('001', 'Chapel Hill', 'NC'), ('002', 'College Park', 'MD');`
  - `INSERT INTO [CurveBallConsultants.Match] (oppId, locId, trnId, mchDate, mchTime, mchType, runTerps, runOpponent, ssnYear) VALUES ('001', '001', '028', '1999-02-05', '12:00:00', 'Home', 2, 3, '1999');`

## Business Transactions:

1. Top Ten Winning Opponents
  - Query: Identifies the top ten opponents against whom Maryland Softball won the most matches.

- SELECT TOP 5 o.oppName AS 'Opponent', COUNT(\*) as 'Maryland SoftBall Wins'  
 FROM [CurveBallConsultants.Match] m  
 JOIN [CurveBallConsultants.Opponent] o ON m.oppId = o.oppId  
 WHERE m.winLoss = 'W'  
 GROUP BY o.oppName  
 ORDER BY 'Maryland SoftBall Wins' DESC;

	Opponent	Maryland SoftBall Wins
1	UMBC	21
2	Kansas	20
3	Mount St. Marys	18
4	Morgan St	17
5	Rhode Island	17

## 2. Top Ten Losing Opponents

- Query: Identifies the top ten opponents against whom Maryland Softball lost the most matches.
- SELECT TOP 5 o.oppName AS 'Opponent', COUNT(\*) as 'Maryland SoftBall Losses'  
 FROM [CurveBallConsultants.Match] m  
 JOIN [CurveBallConsultants.Opponent] o ON m.oppId = o.oppId  
 WHERE m.winLoss = 'L'  
 GROUP BY o.oppName  
 ORDER BY 'Maryland SoftBall Losses' DESC;

	Opponent	Maryland SoftBall Losses
1	Anne Arundel CC	34
2	Troy	34
3	Connecticut	29
4	Morgan St	28
5	UMBC	22

## 3. Top Five Match Locations

- Query: Identifies the top five locations that hosted the most matches.
- SELECT TOP 5 locCity AS 'City', locState AS 'State', COUNT(\*) AS 'Matches Hosted'  
 FROM [CurveBallConsultants.Match] m  
 JOIN [CurveBallConsultants.Location] l ON m.locId = l.locId  
 WHERE m.ssnYear BETWEEN '1999' AND '2022'  
 GROUP BY l.locCity, l.locState  
 ORDER BY COUNT(\*) DESC;

	City	State	Matches Hosted
1	Arnold	MD	434
2	Greenville	NC	60
3	Charlottesville	VA	52
4	College Park	MD	38
5	Tucson	AZ	33

## 4. Opponent Win Percentage

- Query: Calculates Maryland's all-time win percentage for opponents with more than 5 matches played.
- ```

SELECT o.oppName AS 'Opponent Name',
       CAST(COUNT(CASE WHEN m.winLoss = 'W' THEN 1 END) * 100.0 / COUNT(*) AS
             DECIMAL(6,2)) AS 'Win Percentage (%)'
FROM [CurveBallConsultants.Opponent] o
JOIN [CurveBallConsultants.Match] m ON o.oppId = m.oppId
GROUP BY o.oppName
HAVING COUNT(*) > 5
ORDER BY 'Win Percentage (%)' DESC, o.oppName DESC;
```

|    | Opponent Name       | Win Percentage (%) |    | Opponent Name       | Win Percentage (%) |    | Opponent Name    | Win Percentage (%) |
|----|---------------------|--------------------|----|---------------------|--------------------|----|------------------|--------------------|
| 1  | Western Michigan    | 100.00             | 16 | Mount St. Marys     | 75.00              | 31 | Penn St          | 50.00              |
| 2  | Seton Hall          | 100.00             | 17 | Michigan St         | 75.00              | 32 | UMBC             | 48.84              |
| 3  | Mercer              | 100.00             | 18 | NorthweStm          | 71.43              | 33 | Georgia Southern | 48.48              |
| 4  | Kansas              | 100.00             | 19 | Charleston South... | 71.43              | 34 | Florida A&M      | 44.44              |
| 5  | Akron               | 100.00             | 20 | Bethune Cookman     | 69.23              | 35 | St Peters        | 42.86              |
| 6  | East Carolina       | 94.12              | 21 | UMES                | 66.67              | 36 | Monmouth III     | 42.86              |
| 7  | UT Martin           | 91.67              | 22 | Oklahoma St         | 66.67              | 37 | Princeton        | 41.67              |
| 8  | Rutgers             | 88.89              | 23 | Nevada              | 66.67              | 38 | Morgan St        | 37.78              |
| 9  | Bucknell University | 88.89              | 24 | Northern Iowa       | 62.50              | 39 | Bucknell         | 35.00              |
| 10 | Washington          | 87.50              | 25 | North Carolina St   | 62.50              | 40 | Pittsburgh       | 33.33              |
| 11 | Harvard             | 83.33              | 26 | Cal Poly            | 62.50              | 41 | George Mason     | 33.33              |
| 12 | Georgia Tech        | 83.33              | 27 | South Florida       | 55.56              | 42 | Troy             | 26.09              |
| 13 | Virginia            | 81.82              | 28 | Illinois-Chicago    | 55.56              | 43 | Norfolk St       | 25.00              |
| 14 | Nebraska            | 78.57              | 29 | North Carolina      | 54.55              | 44 | Saint Josephs    | 23.08              |
| 15 | Tennessee           | 75.00              | 30 | Rhode Island        | 53.13              | 45 | Anne Arundel CC  | 22.73              |

## 5. Home Win Percentage by Season

- Query: Calculates Maryland's win percentage at home across all years.
- ```

SELECT m.ssnYear AS 'Season', CAST(COUNT(CASE WHEN m.winLoss = 'W' THEN 1 END) * 100.0 / COUNT(*) AS DECIMAL(6,2)) AS 'Win Percentage (%)'
FROM [CurveBallConsultants.Match] m
WHERE m.mchType = 'Home'
GROUP BY m.ssnYear;
```

	Season	Win Percentage (%)		Season	Win Percentage (%)
1	1999	76.00	12	2010	67.50
2	2000	78.57	13	2011	66.67
3	2001	80.00	14	2012	66.67
4	2002	66.67	15	2013	69.23
5	2003	66.67	16	2014	57.14
6	2004	56.10	17	2015	42.11
7	2005	62.07	18	2016	31.25
8	2006	76.19	19	2017	29.41
9	2007	55.56	20	2018	38.89
10	2008	77.78	21	2019	35.29
11	2009	54.05	22	2021	56.25
12	2010	67.50	23	2022	58.82
13	2011	66.67			

## How to Test the Project:

1. Data Source: Go to <https://umterps.com/sports/softball/schedule/> and follow the steps mentioned in the screenshot.

The screenshot shows the softball schedule page for the University of Maryland. At the top, there's a navigation bar with links for Sports, Tickets, Terrapin Club, Fan Info & Gameday, About, Shop, NIL, and University. Below the navigation is a black header bar with the word "Softball". The main content area displays the "2024 SOFTBALL SCHEDULE". A game card for the first game on Friday, February 9, at 10 a.m. in Los Angeles, CA, is shown. The game is between Maryland (M logo) and San Diego (SD logo). To the right of the game card, a timer shows "63 Days, 13 Hours, 51 Minutes, 24 Seconds" until the next game. Below the game card, there are buttons for "Add To Calendar" and "Text Only". A red arrow points from the text "Click here to get data in Text Mode which makes it easier to copy for all singular years from 1999 - 2022" to the "Text Only" button. Another red arrow points from the text "Select all singular years from 1999 - 2022 and then select Text Only to get data for each year" to the "Text Only" button. The date dropdown menu shows "2024" and "All Games". The background features Under Armour HOVR Sonic 5 shoes. An advertisement for "The Jaklitsch Law Group" is visible on the right side.

Here is a screenshot of how the Text Only data looks like:

Maryland 2022 Softball Schedule						
Overall 29-23 .558						
Conference 14-7 .667						
Streak L1						
Home 10-7						
Away 10-6						
Neutral 9-10						
Date	Time	At	Opponent	Location	Tournament	Result
Feb 11 (Fri)	12:30 PM	Neutral	Tennessee	Miami, FL		L 1-2
Feb 11 (Fri)	5:30	Away	Florida International University	Miami, FL		W 10-2
Feb 12 (Sat)	10:00 AM	Neutral	UNC Greensboro	Miami, FL		L 5-6
Feb 12 (Sat)	4:30 PM	Neutral	UNC Greensboro	Miami, FL		W 5-3
Feb 13 (Sun)	10:00 AM	Neutral	Tennessee	Miami, FL		Canceled
Feb 18 (Fri)	1:30 PM	Neutral	Liberty	Orlando, FL		L 0-9
Feb 18 (Fri)	4:00 PM	Neutral	Stetson	Orlando, FL		L 3-8
Feb 19 (Sat)	11:00 AM	Neutral	FAU	Orlando, FL		L 0-1
Feb 19 (Sat)	1:30 PM	Neutral	Bethune-Cookman	Orlando, FL		W 11-3
Feb 20 (Sun)	11:30 AM	Neutral	Missouri	Orlando, FL		W 1-0
Feb 25 (Fri)	9:00 AM	Neutral	Central Michigan	Myrtle Beach, SC		L 4-5
Feb 25 (Fri)	2:00 PM	Neutral	Monmouth	Myrtle Beach, SC		W 14-4
Feb 26 (Sat)	11:30 AM	Neutral	North Carolina A&T	Myrtle Beach, SC		W 7-3
Feb 26 (Sat)	2:00 PM	Away	Coastal Carolina	Myrtle Beach, SC		W 5-2
Feb 27 (Sun)	9:00 AM	Neutral	Dartmouth	Myrtle Beach, SC		W 5-2
Mar 4 (Fri)	11:00 AM	Neutral	Villanova	Charlotte, NC		W 2-1
Mar 4 (Fri)	4:00 PM	Away	Charlotte	Charlotte, NC		L 2-5
Mar 5 (Sat)	11:00 AM	Neutral	Boston College	Charlotte, NC		L 2-3
Mar 5 (Sat)	1:30 PM	Neutral	Villanova	Charlotte, NC		L 7-8
Mar 8 (Tue)	5:30 PM	Away	Virginia	Charlottesville, VA		L 2-10
Mar 11 (Fri)	12:30 PM	Neutral	Arkansas	Harrisonburg, VA		L 0-6
Mar 11 (Fri)	5:30 PM	Neutral	James Madison	Harrisonburg, VA		W 12-5

## 2. Data Cleanup:

Copy and Paste it into Microsoft Word to get better-sorted data. We got the following output:

Date	Time	At	Opponent	Location	Tournament	Result
Feb 25	(Sat) TBD	Away	North Carolina	Chapel Hill, NC	L	0-8
Feb 25	(Sat) TBD	Away	North Carolina	Chapel Hill, NC	L	0-11
Mar 7	(Tue) TBD	Home	George Mason	College Park, Md.	L	0-6
Mar 7	(Tue) TBD	Home	George Mason	College Park, Md.	L	1-11
Mar 11	(Sat) TBD	Home	<del>GCCCCCCCCC</del>	College Park, Md.	L	8-15
Mar 11	(Sat) TBD	Home	<del>GCCCCCCCCC</del>	College Park, Md.	L	2-5
Mar 12	(Sun) TBD	Home	Bowie State	College Park, Md.	W	20-2
Mar 12	(Sun) TBD	Home	Bowie State	College Park, Md.	W	12-4
Mar 18	(Sat) TBD	Home	UNC Wilmington	College Park, Md.	W	7-6
Mar 18	(Sat) TBD	Home	UNC Wilmington	College Park, Md.	L	5-6
Mar 21	(Tue) TBD	Home	Cornell	College Park, Md.	W	9-3
Mar 21	(Tue) TBD	Home	Cornell	College Park, Md.	W	6-5
Mar 24	(Fri) TBD	Neutral	George Mason	Rock Hill, S.C. (Winthrop Tournament)	L	5-9
Mar 24	(Fri) TBD	Away	Winthrop	Rock Hill, S.C. (Winthrop Tournament)	L	0-15
Mar 25	(Sat) TBD	Neutral	Wright State	Rock Hill, S.C. (Winthrop Tournament)	L	0-12
Mar 25	(Sat) TBD	Neutral	Charleston	Rock Hill, S.C. (Winthrop Tournament)	L	3-4
Mar 28	(Tue) TBD	Home	Saint Francis	College Park, Md.	W	2-0
Mar 28	(Tue) TBD	Home	Saint Francis	College Park, Md.	L	3-11
Mar 30	(Thu) TBD	Away	Georgia Tech	Atlanta, GA	L	2-4
Mar 30	(Thu) TBD	Away	Georgia Tech	Atlanta, GA	L	1-8
Mar 31	(Fri) TBD	Neutral	Troy	Marietta, Ga. (Buzz Classic)	L	1-16
Mar 31	(Fri) TBD	Neutral	Georgia State	Marietta, Ga. (Buzz Classic)	L	3-8
Mar 31	(Fri) TBD	Neutral	Toledo	Marietta, Ga. (Buzz Classic)	L	0-9
Apr 1	(Sat) TBD	Neutral	UT Martin	Marietta, Ga. (Buzz Classic)	L	1-12
Apr 3	(Mon) TBD	Home	Winthrop	College Park, Md.	L	0-10
Apr 3	(Mon) TBD	Home	Winthrop	College Park, Md.	L	1-9
Apr 5	(Wed) TBD	Away	Mount St. Mary's	Emmitsburg, MD	W	15-14
Apr 5	(Wed) TBD	Away	Mount St. Mary's	Emmitsburg, MD	W	11-1
Apr 6	(Thu) TBD	Home	UMES	College Park, Md.	W	9-7
Apr 6	(Thu) TBD	Home	UMES	College Park, Md.	W	14-4
Apr 8	(Sat) TBD	Away	Anne Arundel CC	Arnold, MD	L	7-8
Apr 8	(Sat) TBD	Away	Anne Arundel CC	Arnold, MD	L	6-9
Apr 11	(Tue) TBD	Away	Coppin State	Baltimore, MD	W	17-2
Apr 11	(Tue) TBD	Away	Coppin State	Baltimore, MD	W	15-2
Apr 12	(Wed) TBD	Home	Morgan State	College Park, Md.	W	18-6
Apr 12	(Wed) TBD	Home	Morgan State	College Park, Md.	L	0-9
Apr 15	(Sat) TBD	Away	UMBC	Baltimore, MD	L	0-15
Apr 15	(Sat) TBD	Away	UMBC	Baltimore, MD	L	3-18
Apr 17	(Mon) TBD	Away	Trenton State (TCNJ)	Ewing, NJ	L	2-13
Apr 17	(Mon) TBD	Away	Trenton State (TCNJ)	Ewing, NJ	L	0-11
Apr 20	(Thu) TBD	Neutral	North Carolina	Tallahassee, FL (ACC Tournament)	L	0-8
Apr 21	(Fri) TBD	Neutral	Virginia	Tallahassee, FL (ACC Tournament)	L	0-8
Apr 24	(Mon) TBD	Home	Anne Arundel CC	College Park, Md.	L	6-8
Apr 24	(Mon) TBD	Home	Anne Arundel CC	College Park, Md.	L	1-2
Apr 26	(Wed) TBD	Away	George Mason	Fairfax, VA	L	1-9
Apr 26	(Wed) TBD	Away	George Mason	Fairfax, VA	L	3-7
Apr 29	(Sat) TBD	Away	UMES	Princess Anne, MD	W	15-9
Apr 29	(Sat) TBD	Away	UMES	Princess Anne, MD	W	11-3
Date	Time	At	Opponent	Location	Tournament	Result
Feb 24	(Sat) TBD	Neutral	Minnesota	Raleigh, N.C. (Triangle Classic)	L	0-1
Feb 24	(Sat) TBD	Neutral	Eastern Michigan	Raleigh, N.C. (Triangle Classic)	L	0-2

Copy and paste it into Excel. All the data gets pasted in just one column. The output is as follows:

The screenshot shows the Microsoft Excel ribbon with the 'Data' tab selected. In the 'Text to Columns' section of the ribbon, there is a red circle and an arrow pointing to the 'Text to Columns' button. The main area of the screen displays a table of sports fixtures from February to April. The table includes columns for Date, Time, At, Opponent, Location, Tournament, and Result. Row 11 is highlighted with a green border.

	A	B	C	D	E	F	G	H	I	J	K
1											
2	Date	Time	At	Opponent	Location	Tournament	Result				
3	Feb 25 (Sat)	TBD	Away	North Carolina	Chapel Hill, NC		L 0-8				
4	Feb 25 (Sat)	TBD	Away	North Carolina	Chapel Hill, NC		L 0-11				
5	Mar 7 (Tue)	TBD	Home	George Mason	College Park, Md.		L 0-6				
6	Mar 7 (Tue)	TBD	Home	George Mason	College Park, Md.		L 1-11				
7	Mar 11 (Sat)	TBD	Home	Colgate-Beacon	College Park, Md.		L 8-15				
8	Mar 11 (Sat)	TBD	Home	Colgate-Beacon	College Park, Md.		L 2-5				
9	Mar 12 (Sun)	TBD	Home	Bowie State	College Park, Md.		W 20-2				
10	Mar 12 (Sun)	TBD	Home	Bowie State	College Park, Md.		W 12-4				
11	Mar 18 (Sat)	TBD	Home	UNC Wilmington	College Park, Md.		W 7-6				
12	Mar 18 (Sat)	TBD	Home	UNC Wilmington	College Park, Md.		L 5-6				
13	Mar 21 (Tue)	TBD	Home	Cornell	College Park, Md.		W 9-3				
14	Mar 21 (Tue)	TBD	Home	Cornell	College Park, Md.		W 6-5				
15	Mar 24 (Fri)	TBD	Neutral	George Mason	Rock Hill, S.C. (Winthrop Tournament)		L 5-9				
16	Mar 24 (Fri)	TBD	Away	Winthrop	Rock Hill, S.C. (Winthrop Tournament)		L 0-15				
17	Mar 25 (Sat)	TBD	Neutral	Wright State	Rock Hill, S.C. (Winthrop Tournament)		L 0-12				
18	Mar 25 (Sat)	TBD	Neutral	Charleston	Rock Hill, S.C. (Winthrop Tournament)		L 3-4				
19	Mar 28 (Tue)	TBD	Home	Saint Francis	College Park, Md.		W 2-0				
20	Mar 28 (Tue)	TBD	Home	Saint Francis	College Park, Md.		L 3-11				
21	Mar 30 (Thu)	TBD	Away	Georgia Tech	Atlanta, GA		L 2-4				
22	Mar 30 (Thu)	TBD	Away	Georgia Tech	Atlanta, GA		L 1-8				
23	Mar 31 (Fri)	TBD	Neutral	Troy	Marietta, Ga. (Buzz Classic)		L 1-16				
24	Mar 31 (Fri)	TBD	Neutral	Georgia State	Marietta, Ga. (Buzz Classic)		L 3-8				
25	Mar 31 (Fri)	TBD	Neutral	Toledo	Marietta, Ga. (Buzz Classic)		L 0-9				
26	Apr 1 (Sat)	TBD	Neutral	UT Martin	Marietta, Ga. (Buzz Classic)		L 1-12				
27	Apr 3 (Mon)	TBD	Home	Winthrop	College Park, Md.		L 0-10				
28	Apr 3 (Mon)	TBD	Home	Winthrop	College Park, Md.		L 1-9				
29	Apr 5 (Wed)	TBD	Away	Mount St. Mary's	Emmitsburg, MD		W 15-14				
30	Apr 5 (Wed)	TBD	Away	Mount St. Mary's	Emmitsburg, MD		W 11-1				
31	Apr 6 (Thu)	TBD	Home	UMES	College Park, Md.		W 8-7				

Use Text to Columns (highlighted in the above screenshot) in the Data tab of Microsoft Excel and make the necessary adjustments to split the data into multiple columns. You will need to manually edit some cells to get rid of any inconsistencies.

Below is a screenshot of how the sorted data looks in Excel:

	A	B	C	D	E	F	G	H	I	J	K	L
1	mchDate	mchTime	mchType	oppName	locName	locState	trnName	-=winLoss	-runTerps	-runOpponent		
2	1995-02-25	TBD	Away	North Carolina	Chapel Hill	NC		L	0	8		
3	1995-02-25	TBD	Away	North Carolina	Chapel Hill	NC		L	0	11		
4	1995-03-07	TBD	Home	George Mason	College Park	MD		L	0	6		
5	1995-03-07	TBD	Home	George Mason	College Park	MD		L	1	11		
6	1995-03-11	TBD	Home	Goldey-Beacom	College Park	MD		L	8	15		
7	1995-03-11	TBD	Home	Goldey-Beacom	College Park	MD		L	2	5		
8	1995-03-12	TBD	Home	Bowie State	College Park	MD		W	20	2		
9	1995-03-12	TBD	Home	Bowie State	College Park	MD		W	12	4		
10	1995-03-18	TBD	Home	UNC Wilmington	College Park	MD		W	7	6		
11	1995-03-18	TBD	Home	UNC Wilmington	College Park	MD		L	5	6		
12	1995-03-21	TBD	Home	Cornell	College Park	MD		W	9	3		
13	1995-03-21	TBD	Home	Cornell	College Park	MD		W	6	5		
14	1995-03-24	TBD	Neutral	George Mason	Rock Hill	SC	Winthrop Tournament	L	5	9		
15	1995-03-24	TBD	Away	Winthrop	Rock Hill	SC	Winthrop Tournament	L	0	15		
16	1995-03-25	TBD	Neutral	Wright State	Rock Hill	SC	Winthrop Tournament	L	0	12		
17	1995-03-25	TBD	Neutral	Charleston	Rock Hill	SC	Winthrop Tournament	L	3	4		
18	1995-03-28	TBD	Home	Saint Francis	College Park	MD		W	2	0		
19	1995-03-28	TBD	Home	Saint Francis	College Park	MD		L	3	11		
20	1995-03-30	TBD	Away	Georgia Tech	Atlanta	GA		L	2	4		
21	1995-03-30	TBD	Away	Georgia Tech	Atlanta	GA		L	1	8		
22	1995-03-31	TBD	Neutral	Troy	Marietta	GA	Buzz Classic	L	1	16		
23	1995-03-31	TBD	Neutral	Georgia State	Marietta	GA	Buzz Classic	L	3	8		
24	1995-03-31	TBD	Neutral	Toledo	Marietta	GA	Buzz Classic	L	0	9		
25	1995-04-01	TBD	Neutral	UT Martin	Marietta	GA	Buzz Classic	L	1	12		
26	1995-04-03	TBD	Home	Winthrop	College Park	MD		L	0	10		
27	1995-04-03	TBD	Home	Winthrop	College Park	MD		L	1	9		
28	1995-04-05	TBD	Away	Mount St. Mary's	Emmitsburg	MD		W	15	14		
29	1995-04-05	TBD	Away	Mount St. Mary's	Emmitsburg	MD		W	11	1		
30	1995-04-06	TBD	Home	UMES	College Park	MD		W	9	7		
31	1995-04-06	TBD	Home	UMES	College Park	MD		W	14	4		
32	1995-04-08	TBD	Away	Anne Arundel CC	Arnold	MD		L	7	8		
33	1995-04-08	TBD	Away	Anne Arundel CC	Arnold	MD		L	6	9		
34	1995-04-11	TBD	Away	Coppin State	Baltimore	MD		W	17	2		
35	1995-04-11	TBD	Away	Coppin State	Baltimore	MD		W	15	2		
36	1995-04-12	TBD	Home	Morgan State	College Park	MD		W	18	6		
37	1995-04-12	TBD	Home	Morgan State	College Park	MD		L	0	0		

The above-seen Excel file was then saved as a CSV to be used in Visual Studio to create INSERT statements using the following C# code:

```

using System;
using System.IO;
using System.Security.AccessControl;
using System.Security.Cryptography;
using System.Text;
class Program
{
    static void Main()
    {
        string filePath = "Softball Match Data 99-22.csv"; // Update the file path as needed
        string[] lines = File.ReadAllLines(filePath);

        int oppIdCounter = 1;
        int locIdCounter = 1;
        int trnIdCounter = 1;
        //int mchIdCounter = 1;

        Dictionary<string, string> oppDict = new Dictionary<string, string>();
        Dictionary<string, string> locDict = new Dictionary<string, string>();
        Dictionary<string, string> trnDict = new Dictionary<string, string>();
        Dictionary<int, string> mchDict = new Dictionary<int, string>();

        StringBuilder OppInsertStatements = new StringBuilder();
        StringBuilder LocInsertStatements = new StringBuilder();
        StringBuilder TrnInsertStatements = new StringBuilder();
        StringBuilder MchInsertStatements = new StringBuilder();

        OppInsertStatements.AppendLine($"USE BUDT703_Project_0507_01\n");
        OppInsertStatements.AppendLine($"INSERT INTO [CurveBallConsultants.Opponent] (oppId, oppName) VALUES");

        LocInsertStatements.AppendLine($"$\"USE BUDT703_Project_0507_01\"");
    }
}

```

```

LocInsertStatements.AppendLine($"INSERT INTO [CurveBallConsultants.Location] (locId, locCity, locState) VALUES");

TrnInsertStatements.AppendLine($"USE BUDT703_Project_0507_01\n");
TrnInsertStatements.AppendLine($"INSERT INTO [CurveBallConsultants.Tournament] (trnId, trnName) VALUES");

MchInsertStatements.AppendLine($"USE BUDT703_Project_0507_01\n");
MchInsertStatements.AppendLine($"INSERT INTO [CurveBallConsultants.Match] (oppId, locId, trnId, mchDate, mchTime, mchType,
runTerps, runOpponent, ssnYear) VALUES");

//FOR OPPONENT, LOCATION, TOURNAMENT, SEASON INSERT STATEMENTS
foreach (var line in lines.Skip(1)) // Start from 1 to skip the header line
{
    string[] columns = line.Split(',');
    if (columns[7] != "Cancelled" && columns[7] != "Postponed")
    {
        string oppId = oppIdCounter.ToString("D3"); // Assuming oppId is generated incrementally
        string oppName = columns[3].Trim();

        // Generate SQL INSERT statement
        if (!oppDict.ContainsKey(oppId) && !oppDict.ContainsValue(oppName))
        {
            oppDict.Add(oppId, oppName);
            OppInsertStatements.AppendLine($"({{oppId}}', '{oppName}');");
            oppIdCounter++;
        }

        string locId = locIdCounter.ToString("D3"); // Assuming locId is generated incrementally
        string locCity = columns[4].Trim();
        string locState = columns[5].Trim();

        // Generate SQL INSERT statement
        if (!locDict.ContainsKey(locId) && !locDict.ContainsValue($"{locCity} {locState}") && locCity != "TBD")
        {
            locDict.Add(locId, $"{locCity} {locState}");
            LocInsertStatements.AppendLine($"({{locId}}', '{locCity}', '{locState}');");
            locIdCounter++;
        }

        string trnId = trnIdCounter.ToString("D3"); // Assuming trnId is generated incrementally
        string trnName = columns[6].Trim();

        if (!trnDict.ContainsKey(trnId) && !trnDict.ContainsValue(trnName) && trnName != "")
        {
            trnDict.Add(trnId, trnName);
            TrnInsertStatements.AppendLine($"({{trnId}}', '{trnName}');");
            trnIdCounter++;
        }

        string[] dateItems = columns[0].Trim().Split('-');

    }
}

// Print or save the SQL insert statements
Console.WriteLine(OppInsertStatements.ToString());

// You can also save the SQL insert statements to a file if needed
File.WriteAllText("Project_0507_01_Opponent_INSERT.sql", OppInsertStatements.ToString());

//****************************************************************************
*****// Print or save the SQL insert statements
Console.WriteLine(LocInsertStatements.ToString());

```

```

// You can also save the SQL insert statements to a file if needed
File.WriteAllText("Project_0507_01_Location_INSERT.sql", LocInsertStatements.ToString());

//****************************************************************************
***** /



// Print or save the SQL insert statements
Console.WriteLine(TrnInsertStatements.ToString());

// You can also save the SQL insert statements to a file if needed
File.WriteAllText("Project_0507_01_Tournament_INSERT.sql", TrnInsertStatements.ToString());

//****************************************************************************
***** /



foreach (var line in lines.Skip(1)) // Start from 1 to skip the header line
{
    string[] columns = line.Split(',');
    if (columns[7] != "Cancelled" && columns[7] != "Postponed" && columns[4] != "TBD")
    {
        string[] dateItems = columns[0].Trim().Split('-');
        string ssnYear = dateItems[0];
        string mchDate = columns[0].Trim();
        string mchTime = columns[1].Trim();
        string mchType = columns[2].Trim();
        string oppName = columns[3].Trim();
        string locCity = columns[4].Trim();
        string runTerps = columns[8];
        string runOpponents = columns[9];
        string trnName = columns[6].Trim();

        string oppId = "", locId = "", trnId = "";
        //string mchId = mchIdCounter.ToString("D4");

        foreach (var item in oppDict.Keys)
        {
            if (oppDict[item] == oppName)
            {
                oppId = item;
            }
        }

        foreach (var item in locDict.Keys)
        {
            if (locDict[item].Contains(locCity))
            {
                locId = item;
            }
        }

        foreach (var item in trnDict.Keys)
        {
            if (trnDict[item].Contains(trnName))
            {
                trnId = item;
            }
        }

        MchInsertStatements.AppendLine($"({oppId},{locId},{trnId},{mchDate},{mchTime},{mchType},{runTerps},{runOpponents}, {ssnYear})");
        //mchIdCounter++;
    }
}

```

```

// Print or save the SQL insert statements
Console.WriteLine(MchInsertStatements.ToString());

// You can also save the SQL insert statements to a file if needed
File.WriteAllText("Project_0507_01_Match_INSERT.sql", MchInsertStatements.ToString());
}
}

```

3. Database Setup: Execute the SQL script given in ‘Project\_0507\_01\_CREATE.sql’ to create the database and tables. Run the INSERT statements in ‘Project\_0507\_01\_INSERT.sql’ to populate the tables with cleaned data. There are annotations for each DROP TABLE, CREATE TABLE, INSERT, and SELECT statements for the entity in the given file.

```

USE BUDT703_Project_0507_01
-- Opponent CREATE --
CREATE TABLE [CurveBallConsultants.Opponent] (
    oppId CHAR (3) NOT NULL,
    oppName VARCHAR (40),
    CONSTRAINT pk_Opponent_oppId PRIMARY KEY (oppId)
)

-- Tournament CREATE --
CREATE TABLE [CurveBallConsultants.Tournament] (
    trnId CHAR(3) NOT NULL,
    trnName CHAR (50),
    CONSTRAINT pk_Tournament_trnId PRIMARY KEY (trnId)
)

-- Location CREATE --
CREATE TABLE [CurveBallConsultants.Location] (
    locId CHAR(3) NOT NULL,
    locCity CHAR(50),
    locState VARCHAR (20),
    CONSTRAINT pk_Location_locId PRIMARY KEY (locId)
)

-- Match CREATE --
CREATE TABLE [CurveBallConsultants.Match] (
    oppId CHAR(3) NOT NULL,
    locId CHAR(3) NOT NULL,
    trnId CHAR(3),
    mchDate DATE NOT NULL,
    mchTime TIME NOT NULL,
    mchType VARCHAR (7),
    runTerps INT,
    runOpponent INT,
    winLoss AS (
        CASE
            WHEN runTerps - runOpponent > 0 THEN 'W'
            WHEN runTerps - runOpponent < 0 THEN 'L'
            ELSE ''
        END
    )
)

```

Messages

100 %

Query executed successfully.

doitsqlx.rhsmith.umd.edu,97... AD\carter (127) BUDT703\_Project\_0507\_01 00:00:00 0 rows

Ready

Ln 19 Col 35 Ch 35 INS

The screenshot shows the Microsoft SQL Server Management Studio interface. The title bar reads "Project\_0507\_01\_INSERT.sql - doitsqlx.rhsmith.umd.edu,9703.BUDT703\_Project\_0507\_01 (AD\carter (131)) - Microsoft SQL Server Management Studio". The main window displays a query window with the following SQL code:

```
-- Opponent INSERT --
INSERT INTO [CurveballConsultants.Opponent] (oppId, oppName) VALUES
('001', 'North Carolina'),
('002', 'George Mason'),
('003', 'Goldey-Beacom'),
('004', 'Bowling St'),
('005', 'UNC Wilmington'),
('006', 'Cornell'),
('007', 'Widener'),
('008', 'Wright St'),
('009', 'Charleston'),
('010', 'Saint Francis'),
('011', 'Georgia Tech'),
('012', 'Troy'),
('013', 'Georgia St'),
('014', 'Toledo'),
('015', 'UT Martin'),
('016', 'Mount St. Marys'),
('017', 'UMES'),
('018', 'Anne Arundel CC'),
('019', 'Coppin St'),
('020', 'Morgan St'),
('021', 'UMBC'),
('022', 'Trenton St (TCNJ)'),
('023', 'Virginia'),
('024', 'Minnesota'),
('025', 'Eastern Michigan'),
```

The "Messages" pane below the query window shows the results of the execution:

```
(251 rows affected)
(35 rows affected)
(96 rows affected)
(611 rows affected)
(382 rows affected)
```

A status bar at the bottom right indicates "Query executed successfully." and provides details about the session: "doitsqlx.rhsmith.umd.edu,9703 | AD\carter (131) | BUDT703\_Project\_0507\_01 | 00:00:01 | 0 rows".

Run Business Transactions Queries: Execute the provided SQL queries in 'Project\_0507\_01\_SELECT.sql' for each business transaction to retrieve the desired information. There are annotations for each business transaction to specify the question we intend to solve.

In case you need to DROP tables to start over or reset the database use the statements in ‘Project 0507 01 DROP.sql’.

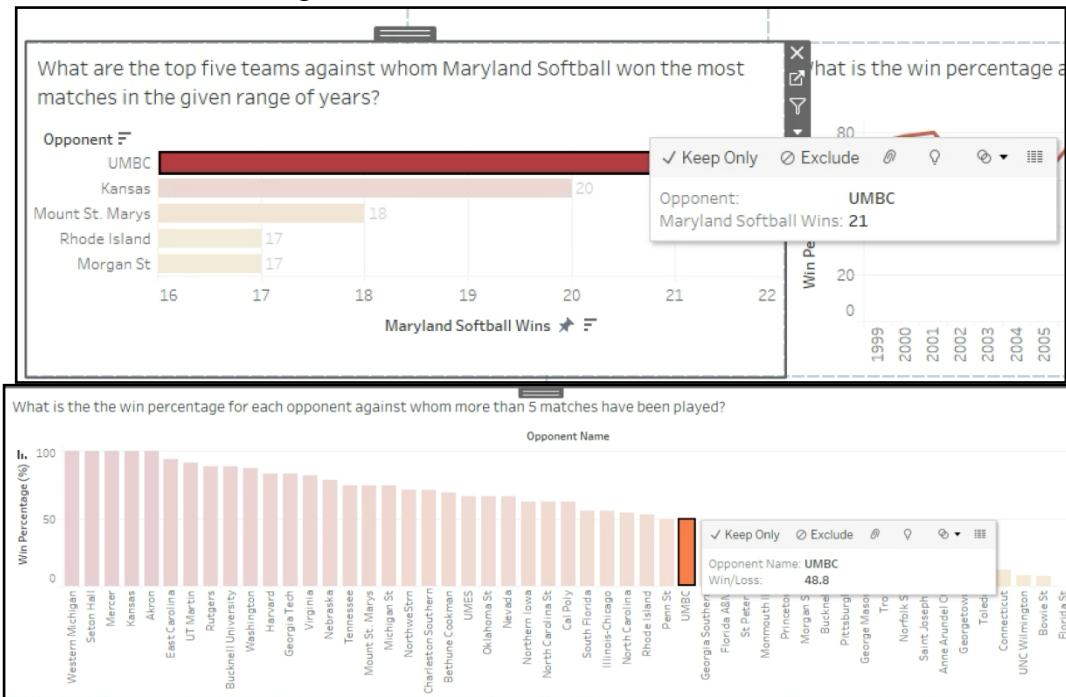
The screenshot shows the Microsoft SQL Server Management Studio (SSMS) interface. The top menu bar includes File, Edit, View, Query, Project, Tools, Window, and Help. The toolbar contains various icons for file operations like New Query, Save, Print, and Execute. The title bar displays the connection information: BUDT03\_Project\_0507\_01, Execute, and the date 2029. The Object Explorer on the left shows the database structure for doitsqlx.rhsmith.umd.edu, port 9703, including Databases, Security, Server Objects, Replication, PolyBase, and Management. The main central pane shows a query window with the following T-SQL code:

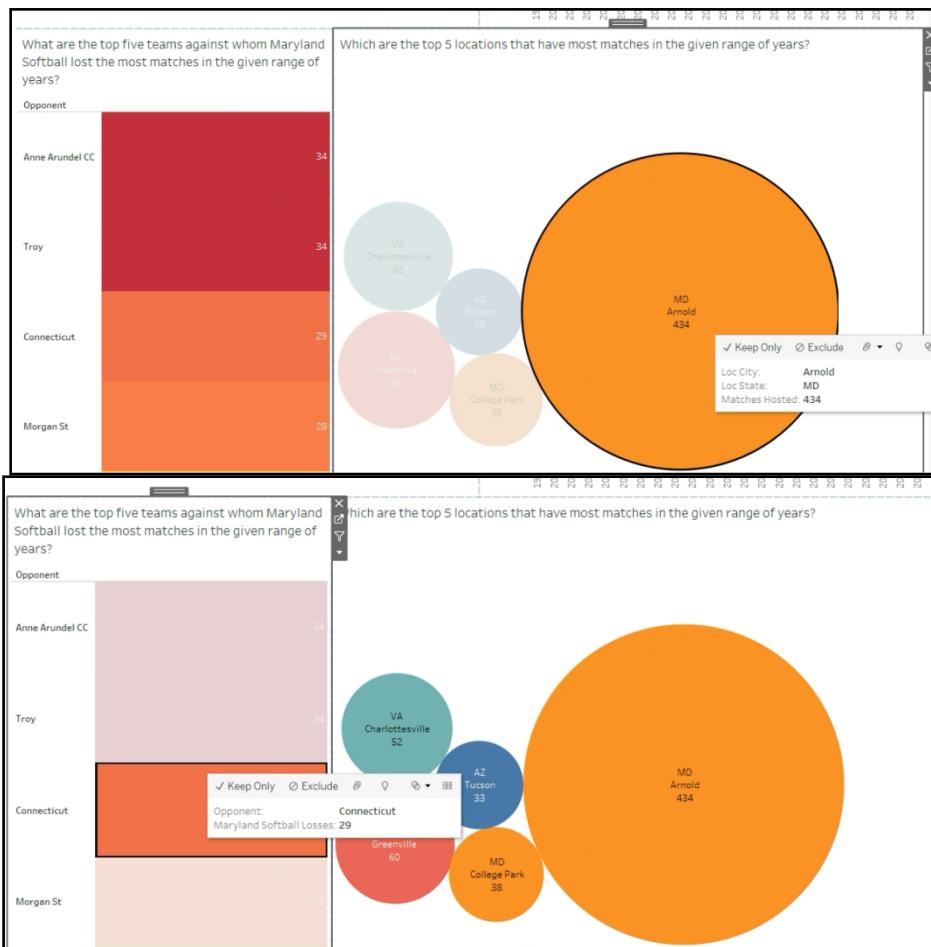
```
USE BUDT03_Project_0507_01

DROP TABLE IF EXISTS [CurveBallConsultants.Match]
DROP TABLE IF EXISTS [CurveBallConsultants.Location]
DROP TABLE IF EXISTS [CurveBallConsultants.Tournament]
DROP TABLE IF EXISTS [CurveBallConsultants.Opponent]
```

The status bar at the bottom indicates the query was executed successfully with 0 rows affected.

- Review Results: Check the output of each query to verify that the results match the expected outcomes. These outcomes have corresponding Tableau visualizations as well. In the given presentation the outputs for the said transactions are mentioned.
- Tableau Dashboard Explanation: For the Tableau dashboard containing each visualization, if you were to click on any bar, bubble, or block of the given visualization, Tableau would highlight the data point associated with that value, as given below.





6. **Modify and Experiment:** Feel free to modify the queries or create new ones to explore different aspects of the softball data.
7. **Adjust Date Range:** Some queries involve date ranges. Adjust the range in the WHERE clause as needed.

By following these steps, you can test and explore the Maryland Softball database using the provided SQL queries.