BUDT703: Database Management Systems

Submitted to : Professor Adam Lee

Submitted by : Project_0507_13

PeakPerformance Partners

Transforming Teams To Triumph

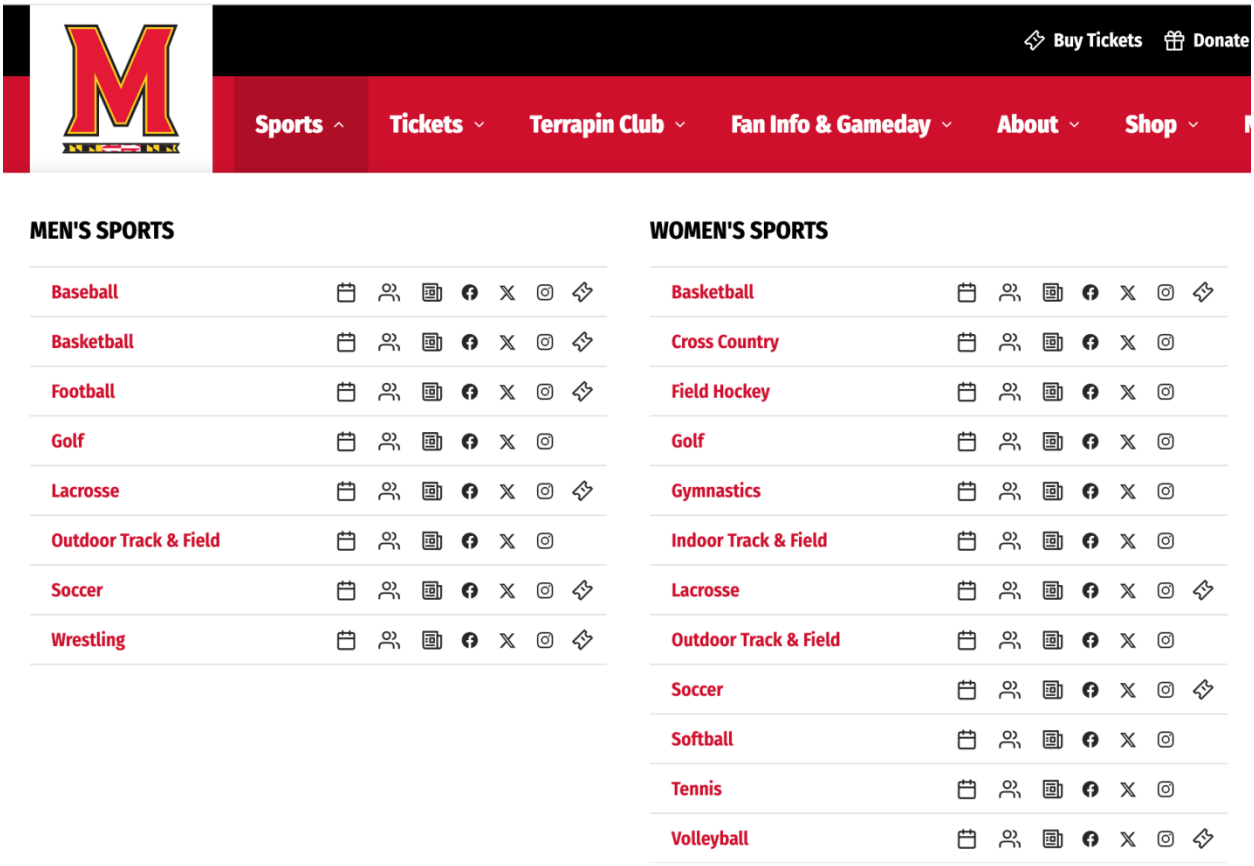
Team members:

Amoghvarsh Kulkarni
Sai Thanmayi Karpurapu
Vedant Kamat
Devni Shah

Step 1: Project Proposal

Collected data from University of Maryland's sports website having information about the schedule of matches which were played from the year 2000 to 2023.

Data Source:
https://umterps.com/sports/softball/schedule



Selected the year for which we wanted to analyze the data-

Selected text only in order to view the data in table format-

Maryland

2014 Softball Schedule

Overall 11–35 .239

Conference 7–16 .304

Streak W3

Home 8–6

Away 3–18

Neutral 0–11

Date	Time	At	Opponent	Location	Tournament	Result
Feb 7 (Fri)	11:00 AM	Neutral	DePaul	Miami, Fla.		L 4–8
Feb 7 (Fri)	4:00 PM	Away	Florida International	Miami, Fla.		L 9–14
Feb 8 (Sat)	10:00 AM	Neutral	Jacksonville	Miami, Fla.		L 2–6
Feb 8 (Sat)	1:00 PM	Neutral	Florida International	Miami, Fla.		L 5–6
Feb 9 (Sun)	11:00 AM	Neutral	Jacksonville	Miami, Fla.		L 0–1
Feb 14 (Fri)	3:00 PM	Away	Mississippi State	Boca Raton, Fla.		L 1–3
Feb 14 (Fri)	5:30 PM	Away	South Carolina	Boca Raton, Fla.		W 8–4

We considered all the important rows such as date of match, location of match, opponent details and the final match result. Analyzed the data on MS Word and MS excel and then formed the mission statement and mission objectives accordingly.

We identified Entities, Relations, Primary and Foreign keys using which we made our ER diagram and Relational schema.

MISSION STATEMENT

As consultants, our mission is to serve as the catalyst for the University of Maryland women's softball team's transformation, to equip the team with actionable recommendations, enabling them to elevate performance standards and establish a legacy of sporting achievement.

MISSION OBJECTIVES:

- OBJECTIVE 1: Analyze the top 10 teams where Maryland has lost against the maximum number of times
- OBJECTIVE 2: Analyze the top 10 away locations where Maryland has won the most number of times

- OBJECTIVE 3: Analyze the top 10 matches where Maryland has won with a shutout

- OBJECTIVE 4: Analyze the top 10 of teams against which Maryland has won the maximum number of times

- OBJECTIVE 5: Analyze the top 10 teams against whom Maryland has won by the highest margin

RELATIONAL SCHEMAS

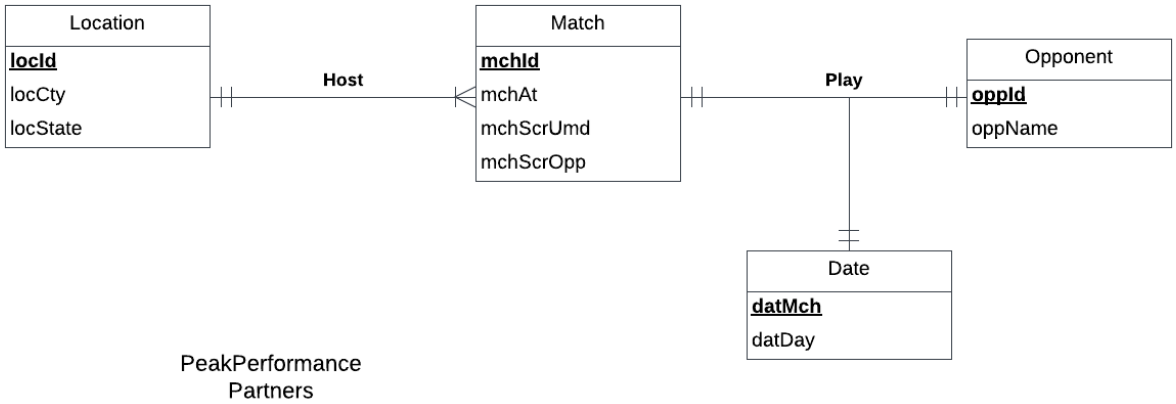
Entities, Attributes and Primary Keys:

Match (**mchId** ,mchAt, mchScrUmd, mchScrOpp,locId)
Date (**datMch** , datDay)
Location (**locId** , locCty ,locState)
Opponent (**oppId** , oppName)

Play (***mchId*** , ***datMch*** , ***oppId***)

Host: Binary Relationship
1 location hosts 1 or more matches
1 match to 1 location

Play: Ternary Relationship
1 match and 1 opponent to 1 date
1 match and 1 date to 1 opponent
1 date and 1 opponent to 1 match



Business rules based on the objectives:

Relation	Foreign Key	Base Relation	Primary Key	Business rule	Constraint: ON DELETE	Business rule	Constraint: ON UPDATE
Host	locId	Location	locId	R1	CASCADE	R2	CASCADE
Host	mchId	Match	mchId	R3	CASCADE	R4	CASCADE
Play	mchId	Match	mchId	R5	CASCADE	R6	CASCADE
Play	datMch	Date	datMch	R7	CASCADE	R8	CASCADE
Play	oppId	Opponent	oppId	R9	CASCADE	R1	CASCADE

Step 2: Data Cleaning

a) Copy pasted the entire columns in Microsoft word in order to remove spaces using find replace function.

```
Apr 19 (Sat)    2:00 PM    Away   Virginia                    Charlottesville, Va.    L 1-5
Apr 2 (Wed)     5:00 PM    Home   George Mason                College Park, Md.       W 11-1
Apr 2 (Wed)     7:00 PM    Home   George Mason                College Park, Md.       W 8-3
Apr 23 (Wed)    5:00 PM    Home   Morgan State                College Park, Md.       W 8-0
Apr 23 (Wed)    7:00 PM    Home   Morgan State                College Park, Md.       W 9-1
Apr 26 (Sat)    3:00 PM    Home   North Carolina              College Park, Md.       W 2-0
Apr 26 (Sat)    5:00 PM    Home   North Carolina              College Park, Md.       L 1-5
Apr 29 (Tue)    5:00 PM    Home   George Washington           College Park, Md.       W 6-0
Apr 29 (Tue)    7:00 PM    Home   George Washington           College Park, Md.       W 6-0
Apr 30 (Wed)    5:00 PM    Home   Towson                      College Park, Md.       W 3-0
Apr 30 (Wed)    7:00 PM    Home   Towson                      College Park, Md.       W 8-0
Apr 6 (Sun)     12:00 PM   Home   Georgia Tech                College Park, Md.       L 1-5
Apr 6 (Sun)     2:00 PM    Home   Georgia Tech                College Park, Md.       L 0-1
```

b) Arranged the data into tabular format using text to table function.
c) Removed the ‘Time result’ column and ‘L’ and ‘W’ from the tournament result.
d) Added quotes and brackets.

(‘Apr 1’	‘(Sat)’	‘TBD’	‘Neutral’	‘UT Martin’	‘Marietta’	‘Ga’	‘1-12’)
(‘Apr 11’	‘(Tue)’	‘TBD’	‘Away’	‘Coppin State’	‘Baltimore’	‘MD’	‘17-2’)
(‘Apr 11’	‘(Tue)’	‘TBD’	‘Away’	‘Coppin State’	‘Baltimore’	‘MD’	‘15-2’)
(‘Apr 12’	‘(Wed)’	‘TBD’	‘Home’	‘Morgan State’	‘College Park’	‘Md.’	‘18-6’)
(‘Apr 12’	‘(Wed)’	‘TBD’	‘Home’	‘Morgan State’	‘College Park’	‘Md.’	‘0-9’)
(‘Apr 15’	‘(Sat)’	‘TBD’	‘Away’	‘UMBC’	‘Baltimore’	‘MD’	‘0-15’)
(‘Apr 15’	‘(Sat)’	‘TBD’	‘Away’	‘UMBC’	‘Baltimore’	‘MD’	‘3-18’)
(‘Apr 17’	‘(Mon)’	‘TBD’	‘Away’	‘Trenton’	‘State Ewing’	‘NJ’	‘2-13’)
(‘Apr 17’	‘(Mon)’	‘TBD’	‘Away’	‘Trenton’	‘State Ewing’	‘NJ’	‘0-11’)
(‘Apr 20’	‘(Thu)’	‘TBD’	‘Neutral’	‘North Carolina’	‘Tallahassee’	‘FL’	‘0-8’)
(‘Apr 21’	‘(Fri)’	‘TBD’	‘Neutral’	‘Virginia’	‘Tallahassee’	‘FL’	‘0-8’)
(‘Apr 24’	‘(Mon)’	‘TBD’	‘Home’	‘Anne Arundel CC’	‘College Park’	‘Md.’	‘6-8’)
(‘Apr 24’	‘(Mon)’	‘TBD’	‘Home’	‘Anne Arundel CC’	‘College Park’	‘Md.’	‘1-2’)
(‘Apr 26’	‘(Wed)’	‘TBD’	‘Away’	‘George Mason’	‘Fairfax’	‘VA’	‘1-9’)

e) Pasted the data into Excel worksheet and removed null and empty cells.

K71

	A	B	C	D	E	F	G	H
1								
2	'04/10/2004'	'Sat '	'Away '	' North Carolina State '	'Raleigh '	' N.C. '	0	5
3	'04/10/2004'	'Sat '	'Away '	' North Carolina State '	'Raleigh '	' N.C. '	1	0
4	'04/11/2004'	'Sun '	'Away '	' North Carolina '	'Chapel Hill '	' N.C. '	0	1
5	'04/11/2004'	'Sun '	'Away '	' North Carolina '	'Chapel Hill '	' N.C. '	3	5
6	'04/15/2004'	'Thu '	'Home '	' St Francis Pa '	'College Park '	' Md. '	9	0
7	'04/15/2004'	'Thu '	'Home '	' St Francis Pa '	'College Park '	' Md. '	10	2
8	'04/17/2004'	'Sat '	'Away '	' James Madison '	'Harrisonburg '	' Va. '	5	0
9	'04/17/2004'	'Sat '	'Away '	' James Madison '	'Harrisonburg '	' Va. '	3	0
10	'04/18/2004'	'Sun '	' Home '	' Florida State '	'College Park '	' Md. '	0	7

f) Created primary keys for Match as mchId, date as datMch, Location as locId and for Opponent as oppId.

g) Arranged the data as per the entities and created tables.

	A	B	C	D	E	F	G	H	I	J	K
1	mchId	datMch	oppid								
2	m200401	'04/10/2004'	op72		(	,	)	'	('m200401','04/10/2004','op72'),		
3	m200402	'04/10/2004'	op72		(	,	)	'	('m200402','04/10/2004','op72'),		
4	m200403	'04/11/2004'	op08		(	,	)	'	('m200403','04/11/2004','op08'),		
5	m200404	'04/11/2004'	op08		(	,	)	'	('m200404','04/11/2004','op08'),		
6	m200405	'04/15/2004'	op216		(	,	)	'	('m200405','04/15/2004','op216'),		
7	m200406	'04/15/2004'	op216		(	,	)	'	('m200406','04/15/2004','op216'),		
8	m200407	'04/17/2004'	op126		(	,	)	'	('m200407','04/17/2004','op126'),		
9	m200408	'04/17/2004'	op126		(	,	)	'	('m200408','04/17/2004','op126'),		
10	m200409	'04/18/2004'	op137		(	,	)	'	('m200409','04/18/2004','op137'),		
11	m200410	'04/18/2004'	op137		(	,	)	'	('m200410','04/18/2004','op137'),		
12	m200411	'04/25/2004'	op84		(	,	)	'	('m200411','04/25/2004','op84'),		
13	m200412	'04/25/2004'	op84		(	,	)	'	('m200412','04/25/2004','op84'),		
14	m200413	'04/29/2004'	op62		(	,	)	'	('m200413','04/29/2004','op62'),		
15	m200414	'04/29/2004'	op62		(	,	)	'	('m200414','04/29/2004','op62'),		
16	m200415	'04/06/2004'	op67		(	,	)	'	('m200415','04/06/2004','op67'),		
17	m200416	'04/06/2004'	op67		(	,	)	'	('m200416','04/06/2004','op67'),		
18	m200417	'04/08/2004'	op222		(	,	)	'	('m200417','04/08/2004','op222'),		
19	m200418	'04/08/2004'	op222		(	,	)	'	('m200418','04/08/2004','op222'),		
20	m200419	'02/20/2004'	op232		(		)	'	('m200419','02/20/2004','op232')		

Data | OPPONENET 2004 | PLAY 2004 | MATCH 2004 | DATE 2004 | LOCATION 2004 | +

Step 3: Creating tables in SQL

DROP TABLE:

```
--SQL DROP TABLES:

DROP TABLE IF EXISTS [PPP.Play];
DROP TABLE IF EXISTS [PPP.Opponent];
DROP TABLE IF EXISTS [PPP.Match];
DROP TABLE IF EXISTS [PPP.Location];
DROP TABLE IF EXISTS [PPP.Date];
```

CREATE TABLE:

```
--SQL CREATE TABLES:

CREATE TABLE [PPP.Date] (
    datMch DATE NOT NULL,
    datDay CHAR (3) NOT NULL,
    CONSTRAINT pk_Date_datMch PRIMARY KEY (datMch))


CREATE TABLE [PPP.Location] (
    locId CHAR(5) NOT NULL,
    locCty VARCHAR (20),
    locState VARCHAR (5),
    CONSTRAINT pk_Location_locId PRIMARY KEY (locId)


CREATE TABLE [PPP.Match] (
    mchId CHAR(10) NOT NULL,
    mchAt VARCHAR (10),
    mchScrUmd INT,
    mchScrOpp INT,
    locId CHAR(5) NOT NULL,
    CONSTRAINT pk_Match_mchId PRIMARY KEY (mchId),
    CONSTRAINT fk_Match_locId FOREIGN KEY (locId)
        REFERENCES [PPP.Location] (locId)
        ON DELETE CASCADE ON UPDATE CASCADE)
```

```
CREATE TABLE [PPP.Opponent] (
    oppId CHAR(5) NOT NULL,
    oppName VARCHAR(20),
    CONSTRAINT pk_Opponent_oppId PRIMARY KEY (oppId))


CREATE TABLE [PPP.Play] (
    mchId CHAR(10) NOT NULL,
    datMch DATE NOT NULL,
    oppId CHAR(5) NOT NULL
CONSTRAINT pk_Play_mchId_datMch_oppId PRIMARY KEY (mchId,datMch,oppId),
CONSTRAINT fk_Play_mchId FOREIGN KEY (mchId)
        REFERENCES [PPP.Match] (mchId)
        ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT fk_Play_datMch FOREIGN KEY (datMch)
        REFERENCES [PPP.Date] (datMch)
        ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT fk_Play_oppId FOREIGN KEY (oppId)
        REFERENCES [PPP.Opponent] (oppId)
        ON DELETE CASCADE ON UPDATE CASCADE)
```

INSERT VALUES:

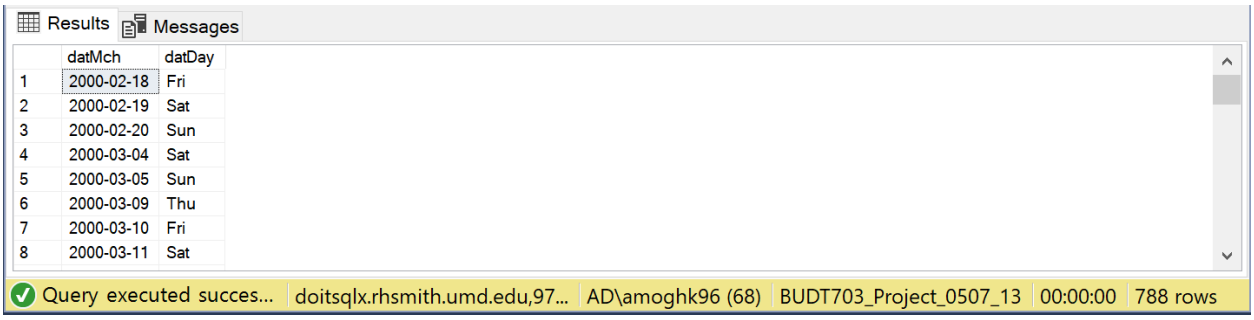
```
--SQL INSERT DATA:
INSERT INTO [PPP.Date] VALUES
('04/01/2000','Sat'),
('04/12/2000','Wed'),
('04/14/2000','Fri'),
('04/16/2000','Sun'),
('04/22/2000','Sat')
INSERT INTO [PPP.Location] VALUES
('lo001','Marietta','GA'),
('lo002','Baltimore','MD'),
('lo003','College Park','MD'),
('lo004','State Ewing','NJ'),
('lo005','Tallahassee','FL')
INSERT INTO [PPP.Match] VALUES
('m200001', 'Home',5,0,'lo003'),
('m200002','Home',6,1,  'lo003'),
('m200003','Home',4,2  ,  'lo003'),
('m200004','Home',4,1  ,  'lo003'),
('m200005','Home',2,3  ,  'lo010')
```

```
INSERT INTO [PPP.Opponent] VALUES
('op01','Robert Morris'),
('op02','Fiu'),
('op03','Binghamton'),
('op04','Mount St. Marys'),
('op05','South Carolina')
INSERT INTO [PPP.Play] VALUES

('m200001','04/01/2000','op212'),
('m200002','04/01/2000','op212'),
('m200003','04/12/2000','op62'),
('m200004','04/12/2000','op62'),
('m200005','04/14/2000','op08')
```

TABLES:

DATE

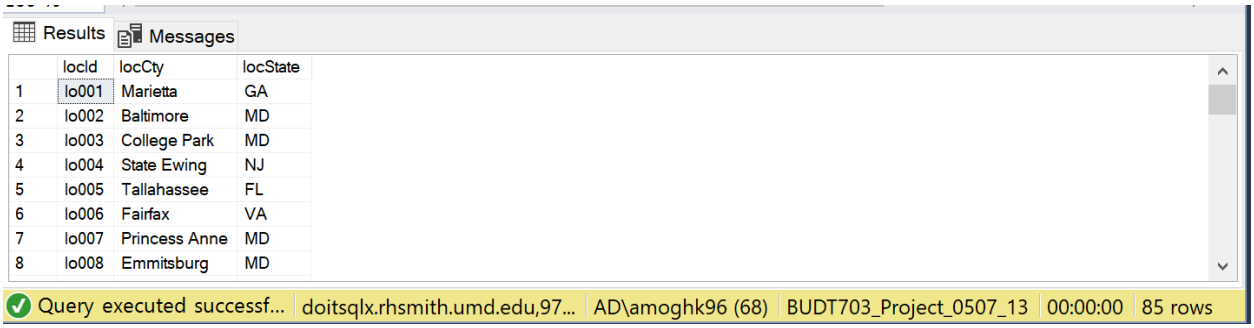


Results | Messages

	datMch	datDay
1	2000-02-18	Fri
2	2000-02-19	Sat
3	2000-02-20	Sun
4	2000-03-04	Sat
5	2000-03-05	Sun
6	2000-03-09	Thu
7	2000-03-10	Fri
8	2000-03-11	Sat

Query executed succes... | doitsqlx.rhsmith.umd.edu,97... | AD\amoghk96 (68) | BUDT703_Project_0507_13 | 00:00:00 | 788 rows

LOCATION



Results | Messages

	locId	locCty	locState
1	lo001	Marietta	GA
2	lo002	Baltimore	MD
3	lo003	College Park	MD
4	lo004	State Ewing	NJ
5	lo005	Tallahassee	FL
6	lo006	Fairfax	VA
7	lo007	Princess Anne	MD
8	lo008	Emmitsburg	MD

Query executed successf... | doitsqlx.rhsmith.umd.edu,97... | AD\amoghk96 (68) | BUDT703_Project_0507_13 | 00:00:00 | 85 rows

MATCH

Results Messages

	mchId	mchAt	mchScrUmd	mchScrOpp	locId
1	m200001	Home	5	0	lo003
2	m200002	Home	6	1	lo003
3	m200003	Home	4	2	lo003
4	m200004	Home	4	1	lo003
5	m200005	Home	2	3	lo010
6	m200006	Away	3	4	lo022
7	m200007	Home	5	3	lo003
8	m200008	Away	2	6	lo005

Query executed succ... | doitsqlx.rhsmith.umd.edu,97... | AD\amoghk96 (68) | BUDT703_Project_0507_13 | 00:00:00 | 1,211 rows

OPPONENT

Results Messages

	oppId	oppName
1	op01	robert morris
2	op02	fiu
3	op03	binghamton
4	op04	mount st. marys
5	op05	south carolina
6	op06	drexel
7	op07	southern illinois
8	op08	North Carolina

Query executed succes... | doitsqlx.rhsmith.umd.edu,97... | AD\amoghk96 (68) | BUDT703_Project_0507_13 | 00:00:00 | 257 rows

PLAY

Results Messages

	mchId	datMch	oppId
1	m200001	2000-04-01	op212
2	m200002	2000-04-01	op212
3	m200003	2000-04-12	op62
4	m200004	2000-04-12	op62
5	m200005	2000-04-14	op08
6	m200006	2000-04-16	op143
7	m200007	2000-04-22	op84
8	m200008	2000-04-29	op137

Query executed succ... | doitsqlx.rhsmith.umd.edu,97... | AD\amoghk96 (68) | BUDT703_Project_0507_13 | 00:00:00 | 1,195 rows

In SQL only 1000 rows can be inserted into one single insert command, due to which we used 2 insert commands for inputting values in Match and Play table.

Step 4: Executing queries in SQL as per the mission objectives

Query 1

What are the details of the top 10 teams which Maryland has won against keeping a Shutout?

```
SELECT TOP 10  o.oppName AS 'Opponent Name',
COUNT(m.mchId) AS 'Number of Shutouts'
FROM [PPP.Match] m, [PPP.Opponent] o, [PPP.Play] p
WHERE p.oppId=o.oppId
AND m.mchScrOpp=0
AND m.mchId=p.mchId
AND m.mchScrUmd>m.mchScrOpp
GROUP BY p.oppId, o.oppName
ORDER BY 'Number of Shutouts' DESC
```

	Opponent Name	Number of Shutouts
1	George Washington	8
2	Georgetown	7
3	Virginia	7
4	Delaware State	6
5	James Madison	6
6	Howard	6
7	Maryland Estn Shore	6
8	Towson	6
9	George Mason	5
10	Boston College	4

Query execu... | doitsqlx.rhsmith.umd.edu,97... | AD\amoghk96 (68) | BUDT703_Project_0507_13 | 00:00:00 | 10 rows

Query 2

What are the details of the top 10 away locations where Maryland has won the most number of times?

```
SELECT TOP 10  l.locState AS 'Location State',
l.locCty AS 'Location City',
COUNT (CASE WHEN m.mchScrUmd > m.mchScrOpp THEN 1 END) AS 'Match Won'
FROM [PPP.Match] m, [PPP.Location] l
WHERE l.locId=m.locId
AND mchAt='Away'
GROUP BY l.locCty, l.locId,locState
ORDER BY 'Match Won' DESC
```

	Location State	Location City	Match Won
1	FL	Tallahassee	19
2	GA	Marietta	13
3	FL	Tampa	10
4	NC	Raleigh	9
5	VA	Fairfax	9
6	NC	Chapel Hill	8
7	VA	Charlottesville	8
8	FL	Boca Raton	7
9	NY	Brooklyn	6
10	VA	Blacksburg	6

Query execu... | doitsqlx.rhsmith.umd.edu,97... | AD\amoghk96 (68) | BUDT703_Project_0507_13 | 00:00:00 | 10 rows

Query 3

What are the details of the top 10 teams that Maryland has lost against the maximum number of times?

```
SELECT TOP 10  o.oppName AS 'Opponent Name',
COUNT (CASE WHEN m.mchScrUmd < m.mchScrOpp THEN 1 END)
AS 'Number of Times Match Lost'
FROM [PPP.Match] m, [PPP.Opponent] o, [PPP.Play] p
WHERE p.oppId=o.oppId
AND m.mchId=p.mchId
GROUP BY p.oppId, o.oppName
ORDER BY COUNT (CASE WHEN m.mchScrUmd < m.mchScrOpp THEN 1 END) DESC
```

	Opponent Name	Number of Times Match Lost
1	Georgia Tech	34
2	Florida State	33
3	North Carolina	28
4	Michigan	23
5	Virginia Tech	20
6	Virginia	20
7	Ohio State	18
8	Penn State	17
9	Michigan State	17
10	Boston College	15

Query execu... | doitsqlx.rhsmith.umd.edu,97... | AD\amoghk96 (68) | BUDT703_Project_0507_13 | 00:00:00 | 10 rows

Query 4

What are the details of the top 10 teams against which Maryland has won the maximum number of times?

```
SELECT TOP 10 o.oppName AS 'Opponent Name',
COUNT (CASE WHEN m.mchScrUmd > m.mchScrOpp THEN 1 END) AS
'Maryland wins'
FROM [PPP.Match] m, [PPP.Opponent] o, [PPP.Play] p
WHERE p.oppId=o.oppId
AND m.mchId=p.mchId
GROUP BY p.oppId,o.oppName
ORDER BY 'Maryland wins' DESC
```

	Opponent Name	Maryland wins
1	Virginia	22
2	Delaware State	20
3	Towson	19
4	Boston College	17
5	North Carolina	17
6	Virginia Tech	15
7	George Mason	15
8	Penn State	14
9	Michigan State	13
10	Iowa	12

Query executed succe... | doitsqlx.rhsmith.umd.edu,97... | AD\amoghk96 (68) | BUDT703_Project_0507_13 | 00:00:00 | 10 rows

Query 5

What are the details of the top 10 teams against which Maryland has lost by the highest margin?

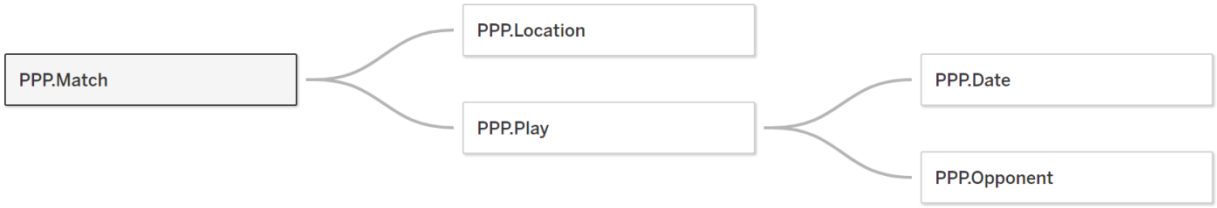
```
SELECT TOP 10 o.oppName AS 'Opponent Name',
(m.mchScrUmd-m.mchScrOpp) AS 'Score Margin'
FROM [PPP.Match] m, [PPP.Opponent] o, [PPP.Play] p
WHERE p.oppId=o.oppId
AND m.mchId=p.mchId
AND m.mchScrUmd<m.mchScrOpp
ORDER BY 'Score Margin' ASC
```

	Opponent Name	Score Margin
1	Nebraska	-22
2	Ohio State	-22
3	Clemson	-19
4	Illinois	-18
5	Alabama	-18
6	Texas	-17
7	Florida	-16
8	Michigan	-16
9	Oregon	-14
10	Arizona State	-14

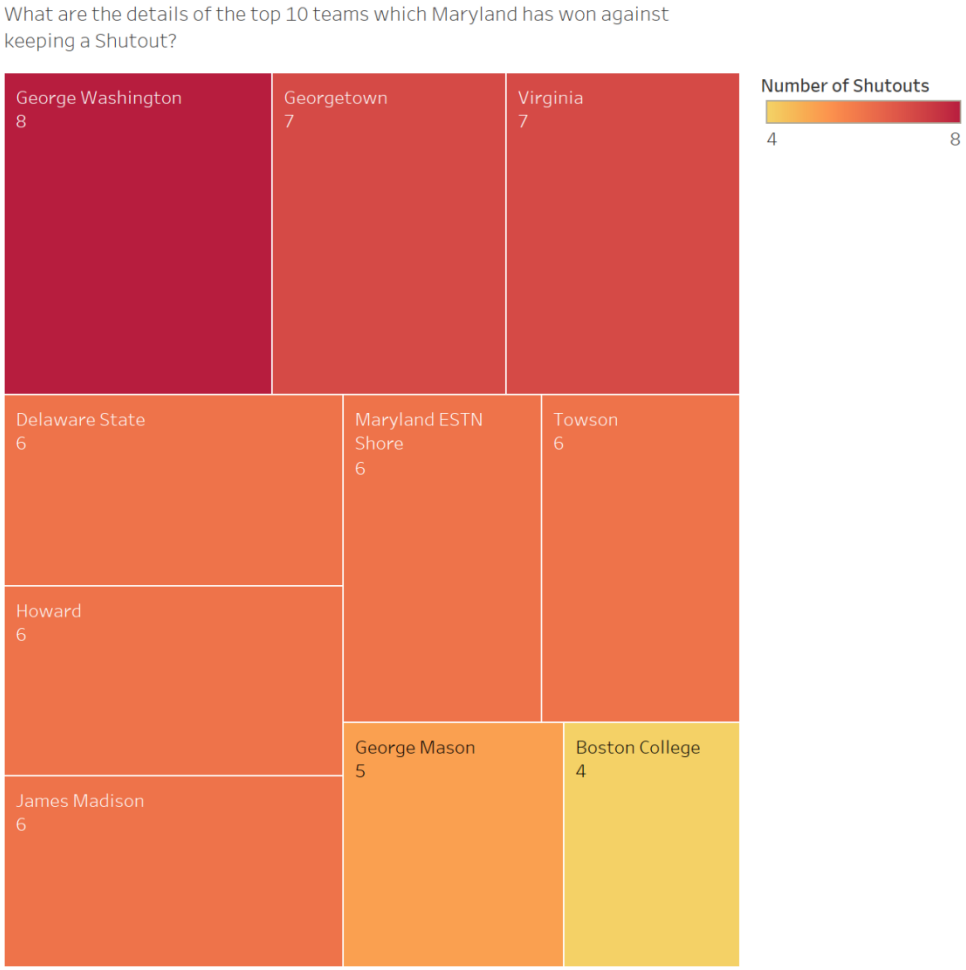
Query executed succe... | doitsqlx.rhsmith.umd.edu,97... | AD\amoghk96 (68) | BUDT703_Project_0507_13 | 00:00:00 | 10 rows

Step 5: Data Visualization using Tableau

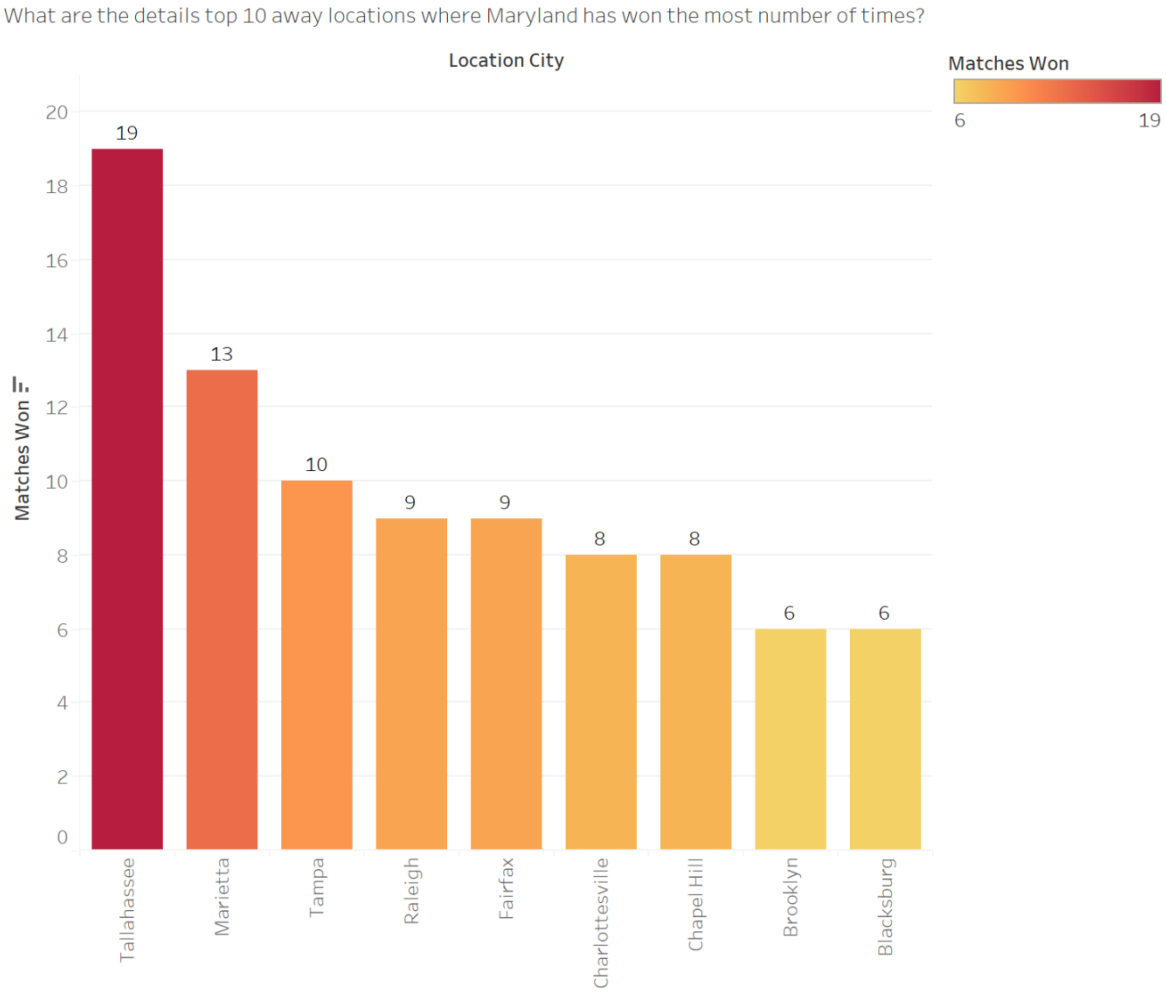
Created connections between all 4 entities.



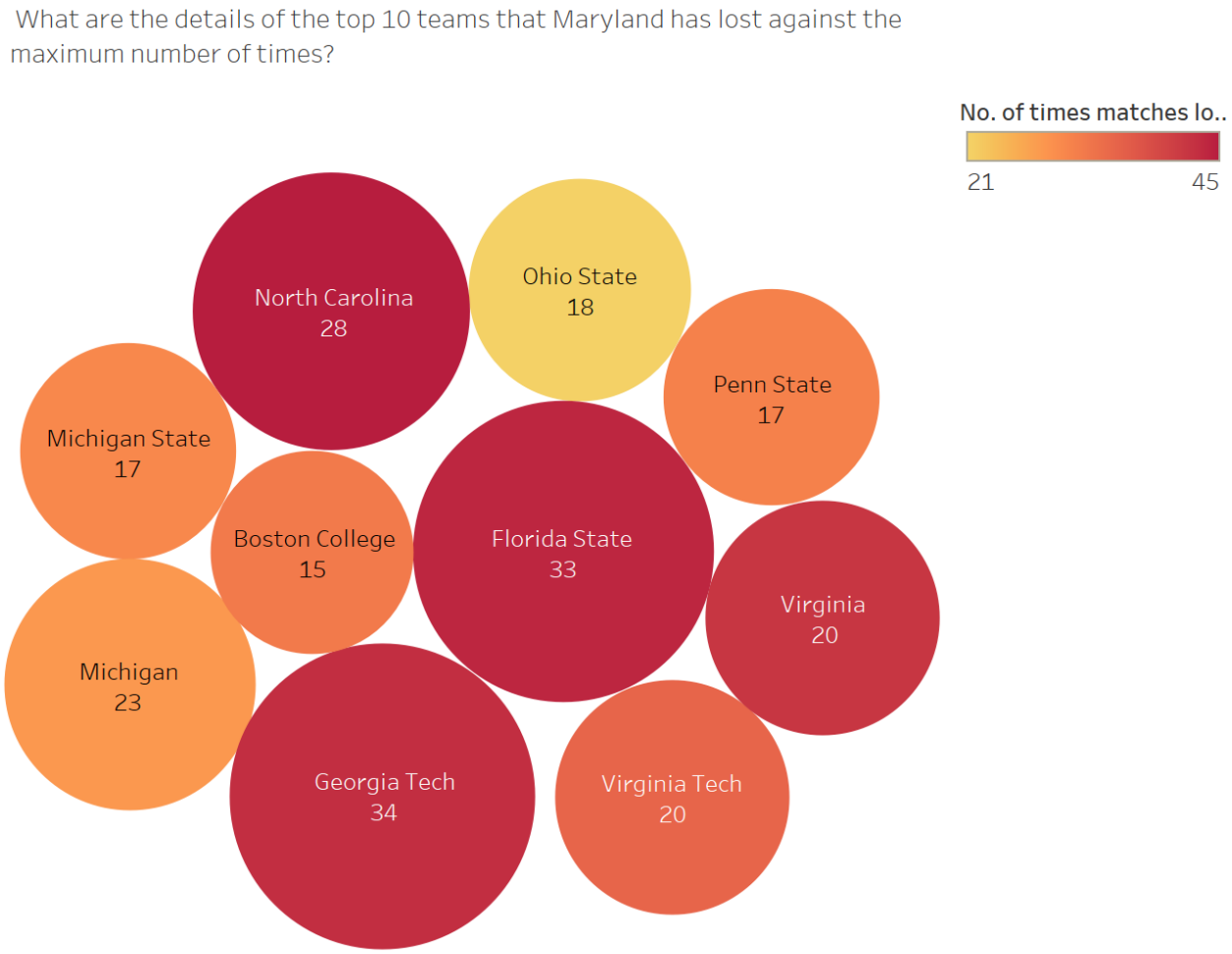
We created a calculated field for calculating Shutout and visualized it using Tree map as seen below.



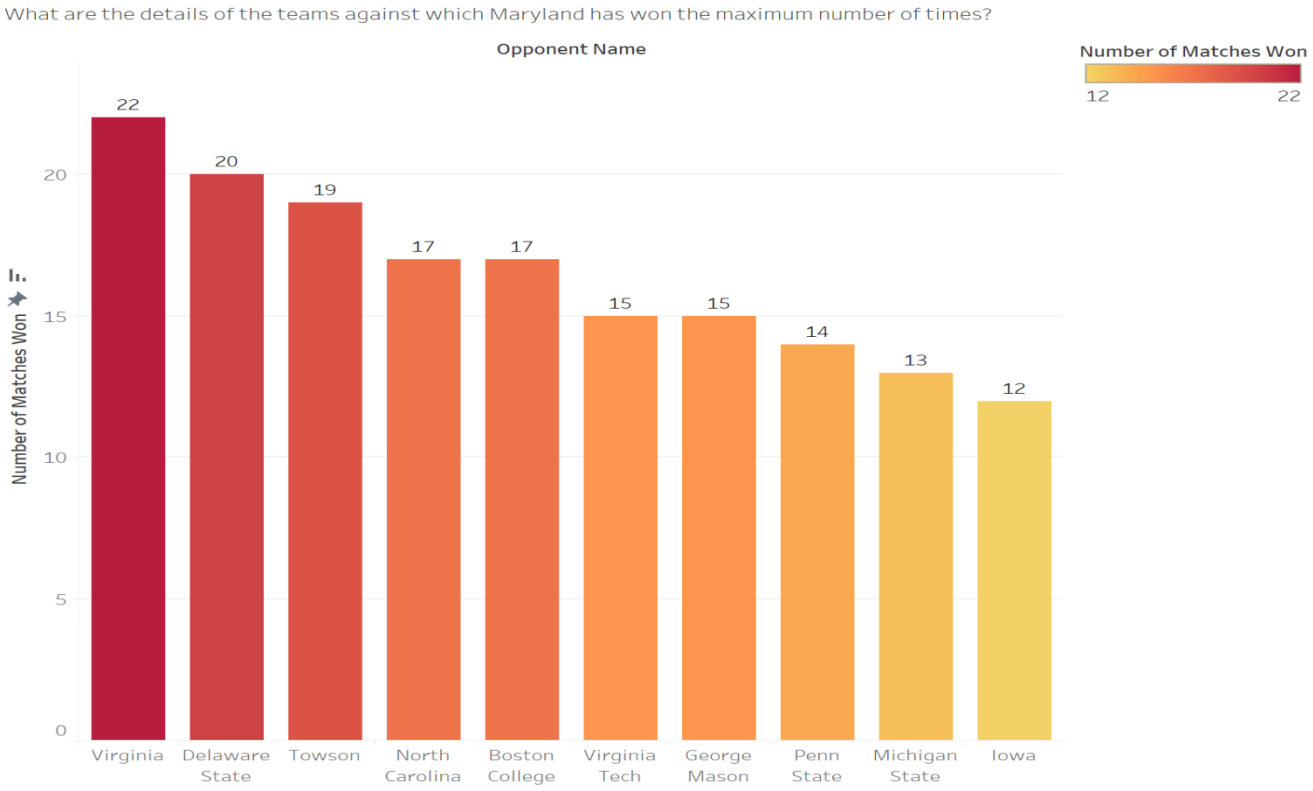
We created a calculated field for when Maryland's score was greater than Opponent's score. We also specified the location to be 'away' and visualized it using a Bar chart.



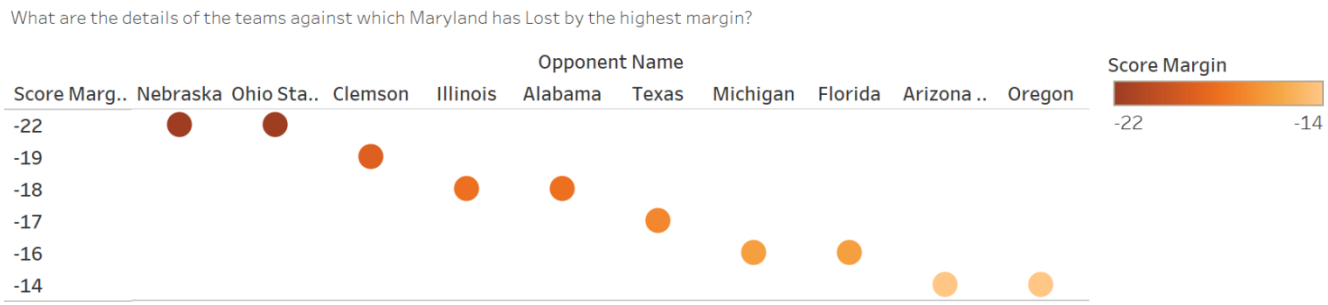
We created a calculated field for when Opponent's score was greater than Maryland's score and visualized its total count using a Bubble chart.



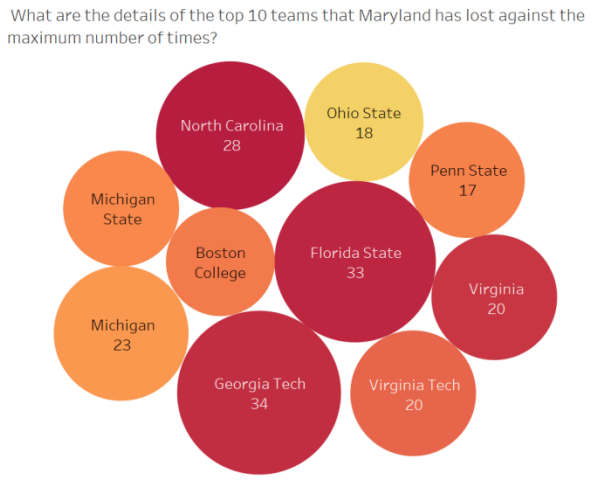
We created a calculated field for when Maryland's score was greater than Opponent's score and visualized their maximum wins using a Bar chart.

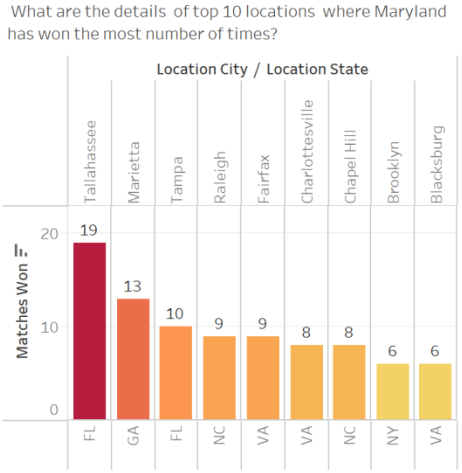


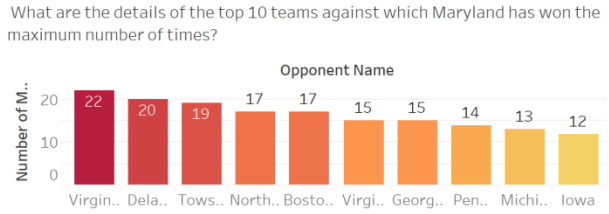
We created a calculated field for getting the Score Margin (Maryland Score-Opponent Score) and visualized it using a dot chart.

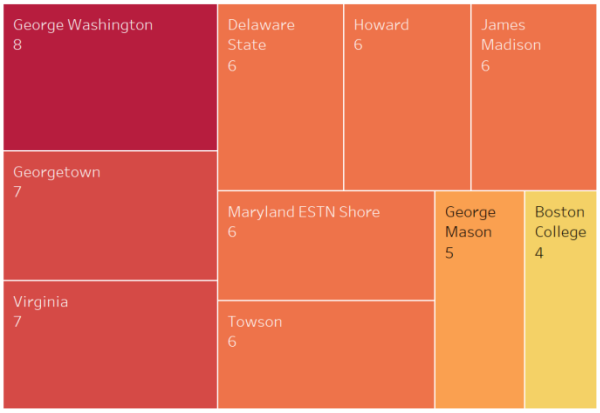


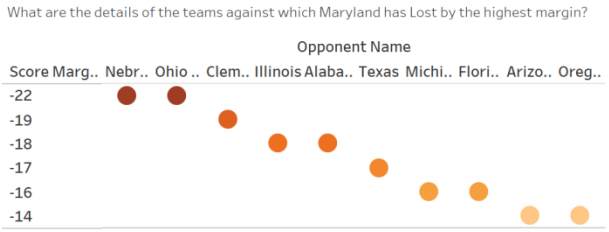
Final Tableau Dashboard-











INSIGHTS:

- Number of “Shutouts” indicate impressive defensive play from the team, as it reflects a strong performance in preventing the opposing team's hitters from reaching home plate. It signifies effective pitching, solid fielding, and teamwork to keep the opponent from scoring. Maryland’s team showed impressive defensive play against George Washington 8 times, Georgetown 7 times etc.
- In contrast, Maryland’s defeat with the highest margins against teams like Nebraska, Ohio State, Clemson etc. show where their pitching, fielding and overall defense was weak. They should focus on defensive aspects of their game against these opponents.

- Maryland has dominated the most against teams like Virginia, Delaware State, Towson etc. They should be confident when playing against these teams.
- The maximum number of defeats against teams like Georgia Tech, Florida State, North Carolina etc. shed light on whom Maryland should practice more, get more fan support and improve their overall performance.