Business Tasks and Processes-

1. Data source

   https://umterps.com/sports/softball/schedule
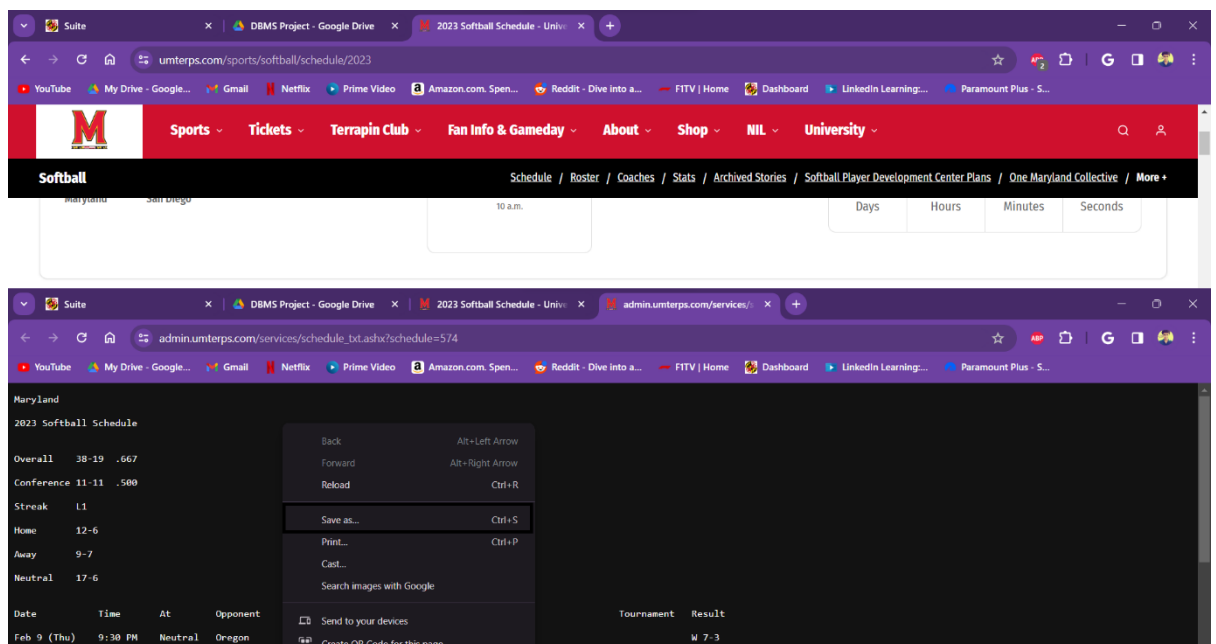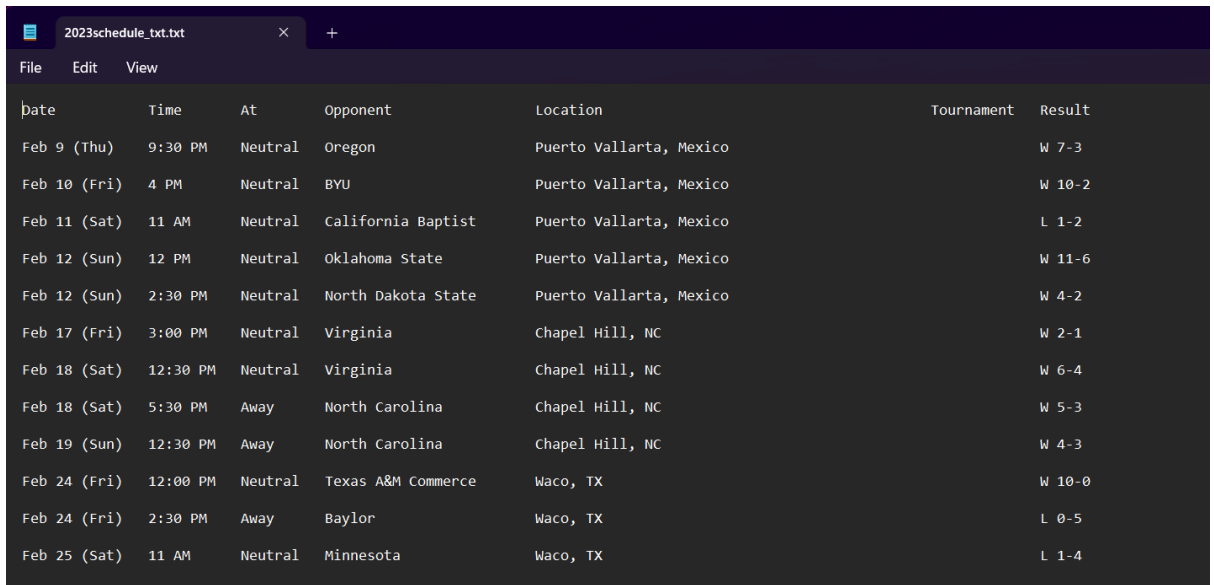
2. Data Download
   - Download data for each year from the above website by selecting only the single years from the dropdown and clicking on text only.
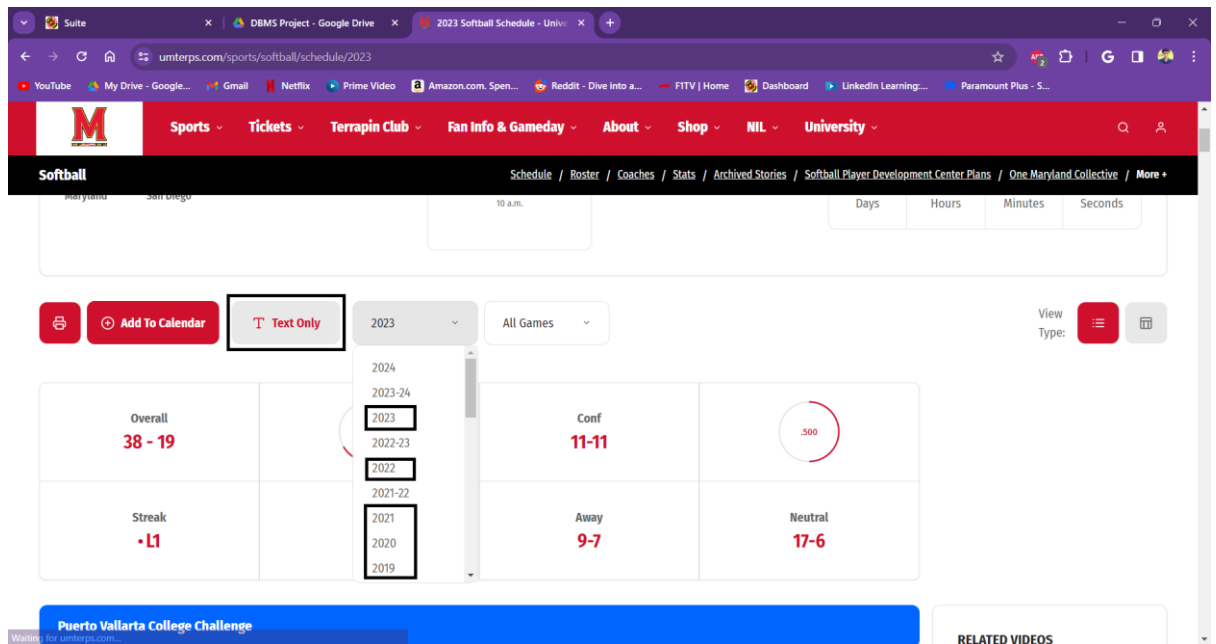
   - This opens the below window, click on save as.

- The below text file is downloaded. Repeat these steps for all years from 1995 to 2023.


Date	Time	At	Opponent	Location	Tournament	Result
Feb 9 (Thu)	9:30 PM	Neutral	Oregon	Puerto Vallarta, Mexico		W 7-3
Feb 10 (Fri)	4 PM	Neutral	BYU	Puerto Vallarta, Mexico		W 10-2
Feb 11 (Sat)	11 AM	Neutral	California Baptist	Puerto Vallarta, Mexico		L 1-2
Feb 12 (Sun)	12 PM	Neutral	Oklahoma State	Puerto Vallarta, Mexico		W 11-6
Feb 12 (Sun)	2:30 PM	Neutral	North Dakota State	Puerto Vallarta, Mexico		W 4-2
Feb 17 (Fri)	3:00 PM	Neutral	Virginia	Chapel Hill, NC		W 2-1
Feb 18 (Sat)	12:30 PM	Neutral	Virginia	Chapel Hill, NC		W 6-4
Feb 18 (Sat)	5:30 PM	Away	North Carolina	Chapel Hill, NC		W 5-3
Feb 19 (Sun)	12:30 PM	Away	North Carolina	Chapel Hill, NC		W 4-3
Feb 24 (Fri)	12:00 PM	Neutral	Texas A&M Commerce	Waco, TX		W 10-0
Feb 24 (Fri)	2:30 PM	Away	Baylor	Waco, TX		L 0-5
Feb 25 (Sat)	11 AM	Neutral	Minnesota	Waco, TX		L 1-4

3. Data Consolidation
   - Once all the relevant data text files have been downloaded, open one of the files and select and copy all the data present in this text file.
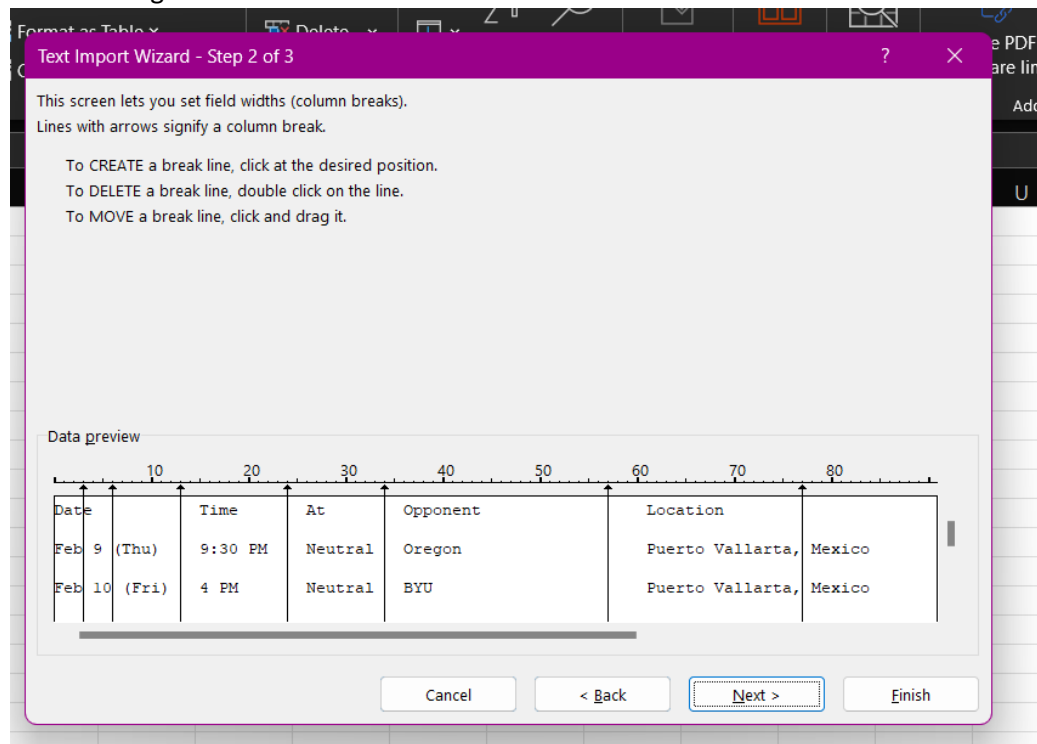   - Open an Excel file and click on 'Use Text Import Wizard'.
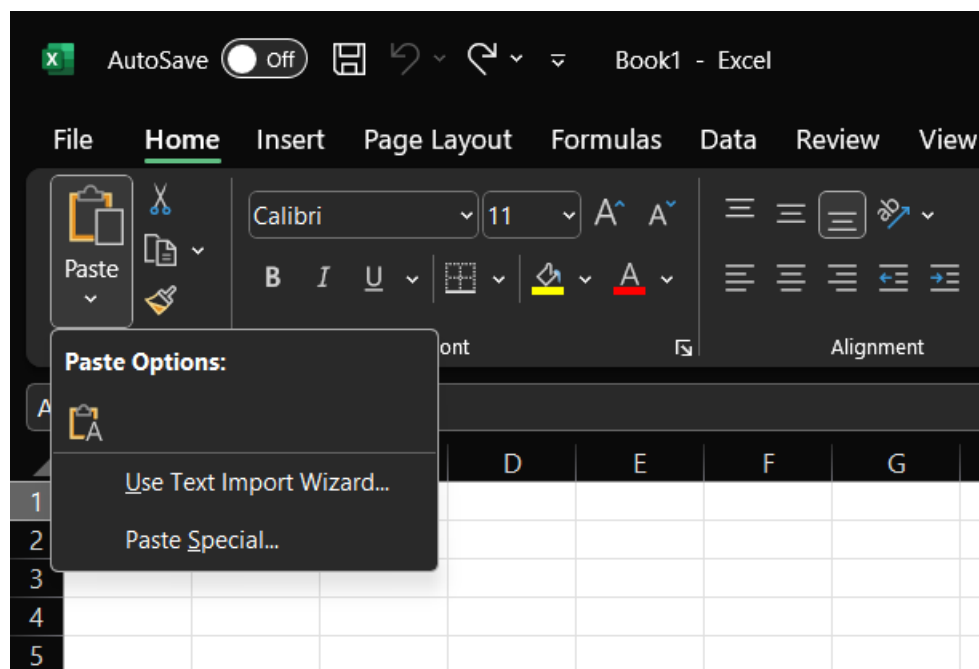
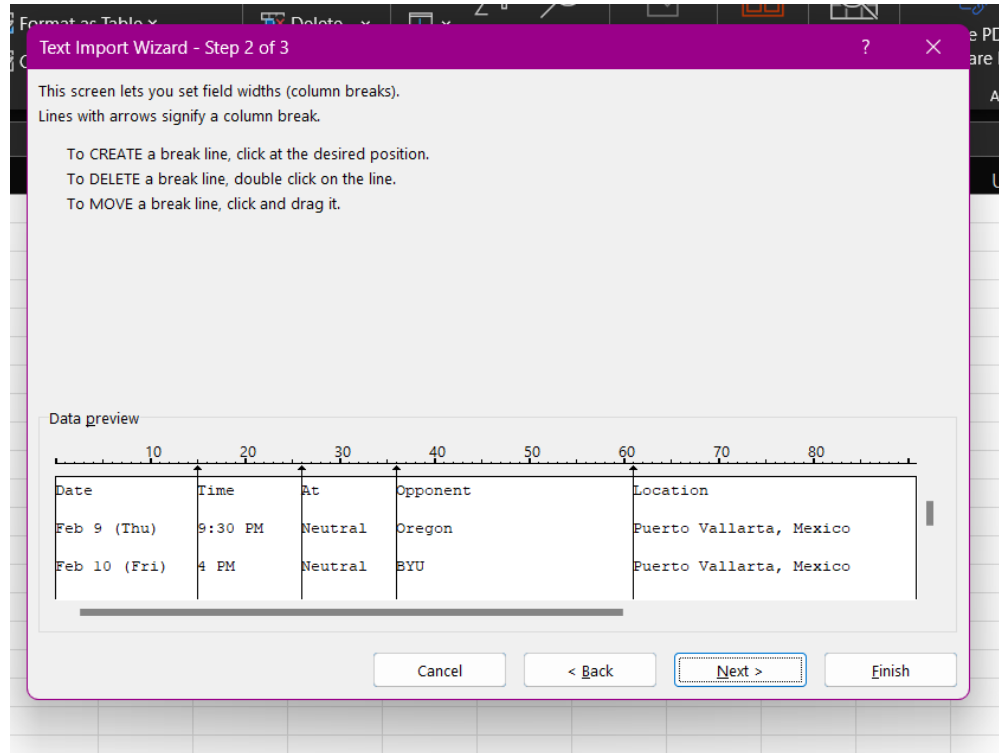- Select 'Fixed Width' in the popup that opens and click on next.
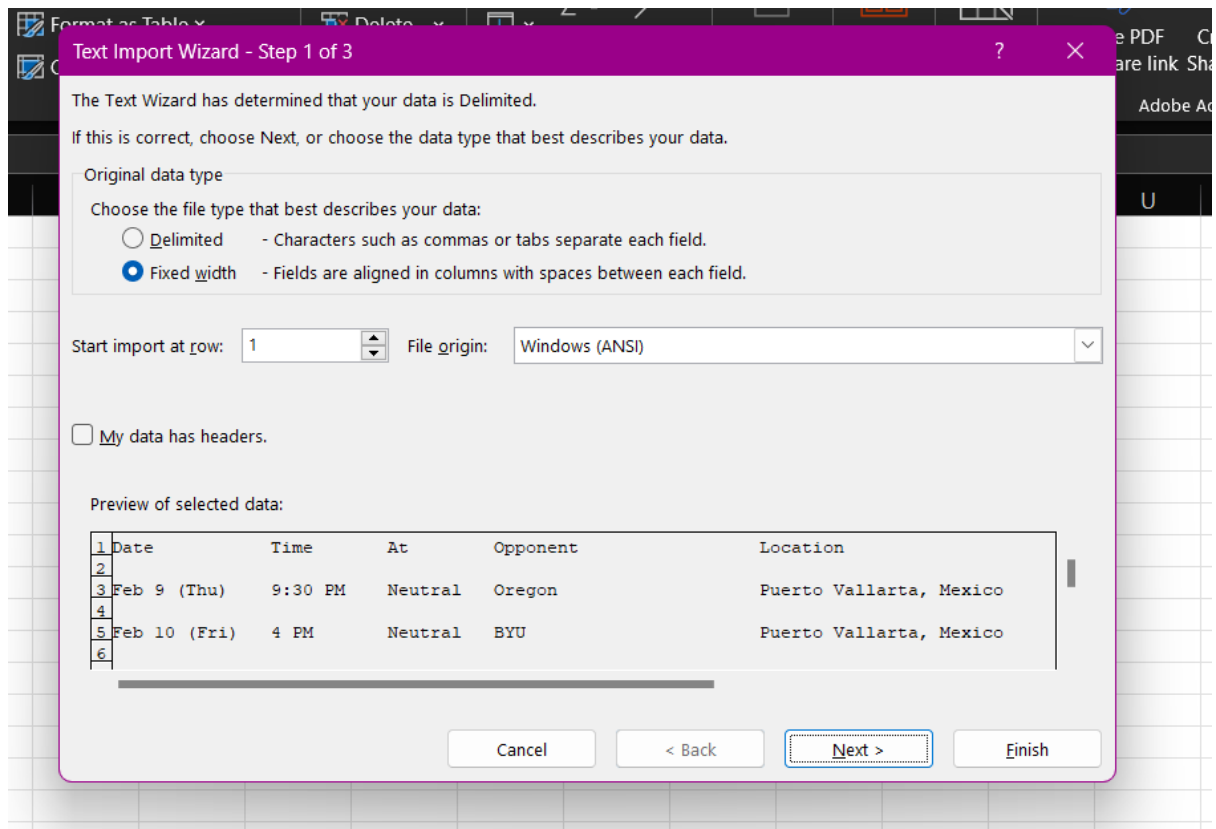
- Drag and move around the lines that appear in the popup to the starting location of each column.
  Initial setting-

Final Setting-

- Click on 'Next' and 'Finish' in the upcoming window. Data will be imported in Excel in the following format –

	A	B	C	D	E	F	G	H
1	Date	Time	At	Opponent	Location	Tourname	Result	
2								
3	Feb 9 (Thu	9:30 PM	Neutral	Oregon	Puerto Vallarta, Mexi		W 7-3	
4								
5	Feb 10 (Fri	4:00 PM	Neutral	BYU	Puerto Vallarta, Mexi		W 10-2	
6								
7	Feb 11 (Sa	11:00 AM	Neutral	California	Puerto Vallarta, Mexi		L 1-2	
8								
9	Feb 12 (Su	12:00 PM	Neutral	Oklahoma	Puerto Vallarta, Mexi		W 11-6	
10								
11	Feb 12 (Su	2:30 PM	Neutral	North Dak	Puerto Vallarta, Mexi		W 4-2	
12								
13	Feb 17 (Fri	3:00 PM	Neutral	Virginia	Chapel Hill, NC		W 2-1	
14								
15	Feb 18 (Sa	12:30 PM	Neutral	Virginia	Chapel Hill, NC		W 6-4	
16								
17	Feb 18 (Sa	5:30 PM	Away	North Car	Chapel Hill, NC		W 5-3	
18								
19	Feb 19 (Su	12:30 PM	Away	North Car	Chapel Hill, NC		W 4-3	
20								
21	Feb 24 (Fri	12:00 PM	Neutral	Texas A&N	Waco, TX		W 10-0	
22								
23	Feb 24 (Fri	2:30 PM	Away	Baylor	Waco, TX		L 0-5	
24								

- Use Excel functionality to create this type of an output and add an additional column showing the year. Copy this to a separate tab, 'Master Data'. Repeat this process for

each year that the data has been downloaded for and keep pasting the yearly data in the 'Master Data' tab one below the other.

4. Split Data into relevant tables

- The 'Master Data' tab containing data for all the years needs to be split into different tables based on the entities. This is done using basic excel functionality.

  Date – An Id column is created for each unique match date.
  Columns – dteId, dteMatchDate

  Opponent - An Id column is created for each unique opponent.
  Columns – oppId, oppName

  Location – An Id column is created for each unique city in a state.
  Columns – locId, locCity, locState

  Score – An Id column is created for each unique score combination.
  Columns – scrId, scrUMD, scrOpp

  Record – Created using the unique opponent + date + score combinations from the 'Master Data'.
  Columns – oppId, dteId, scrId

5. Integrate Data into SQL

- Using Excel formulae, create a concatenated column having all the relevant columns of the above tables. E.g.

	Clipboard ⌐			Font		⌐		Alignment	

F2		∨	⋮	✕ ✓ *fx*	=CONCAT("('",A2,"', ",B2,", ",C2,", '",D2,"'),")			

	A	B	C	D	E	F	G	H
1	scrId	scrUMD	scrOpp	scrOutcome				
2	S0001	7	8	L		('S0001', 7, 8, 'L'),		
3	S0002	6	9	L		('S0002', 6, 9, 'L'),		
4	S0003	6	8	L		('S0003', 6, 8, 'L'),		
5	S0004	1	2	L		('S0004', 1, 2, 'L'),		
6	S0005	20	2	W		('S0005', 20, 2, 'W'),		
7	S0006	12	4	W		('S0006', 12, 4, 'W'),		
8	S0007	3	4	L		('S0007', 3, 4, 'L'),		
9	S0008	17	2	W		('S0008', 17, 2, 'W'),		
10	S0009	15	2	W		('S0009', 15, 2, 'W'),		
11	S0010	9	3	W		('S0010', 9, 3, 'W'),		
12	S0011	6	5	W		('S0011', 6, 5, 'W'),		
13	S0012	0	6	L		('S0012', 0, 6, 'L'),		
14	S0013	1	11	L		('S0013', 1, 11, 'L'),		
15	S0014	5	9	L		('S0014', 5, 9, 'L'),		
16	S0015	1	9	L		('S0015', 1, 9, 'L'),		
17	S0016	3	7	L		('S0016', 3, 7, 'L'),		
18	S0017	3	8	L		('S0017', 3, 8, 'L'),		
19	S0018	2	4	L		('S0018', 2, 4, 'L'),		
20	S0019	1	8	L		('S0019', 1, 8, 'L'),		
21	S0020	8	15	L		('S0020', 8, 15, 'L'),		
22	S0021	2	5	L		('S0021', 2, 5, 'L'),		
23	S0022	18	6	W		('S0022', 18, 6, 'W'),		
24	S0023	0	9	L		('S0023', 0, 9, 'L'),		
25	S0024	15	14	W		('S0024', 15, 14, 'W'),		
26	S0025	11	1	W		('S0025', 11, 1, 'W'),		
27	S0026	0	8	L		('S0026', 0, 8, 'L'),		
28	S0027	0	11	L		('S0027', 0, 11, 'L'),		
29	S0028	0	8	L		('S0028', 0, 8, 'L')		

‹ ›	Play	Record	Date	Opponent	Location	Score

- Copy paste this column into the insert statement query in SQL for the respective tables.

6. Create tables in SQL:

```sql
USE BUDT703_Project_0507_04

DROP TABLE IF EXISTS Play
DROP TABLE IF EXISTS Record;
DROP TABLE IF EXISTS Score;
DROP TABLE IF EXISTS Location;
DROP TABLE IF EXISTS Date;
DROP TABLE IF EXISTS Opponent;

CREATE TABLE Opponent (
 oppId CHAR(5) NOT NULL,
 oppName VARCHAR(100),
 CONSTRAINT pk_Opponent_oppId PRIMARY KEY(oppId)
)

CREATE TABLE Date (
 dteId CHAR(5) NOT NULL,
 dteMatchDate Date,
 CONSTRAINT pk_Date_dteId PRIMARY KEY(dteId)
)

CREATE TABLE Location  (
 locId CHAR(5) NOT NULL,
 locCity VARCHAR(50),
 locState VARCHAR(50),
 CONSTRAINT pk_Location_locId PRIMARY KEY(locId)
)

CREATE TABLE Score (
 scrId CHAR(5) NOT NULL,
 scrUMD int,
 scrOpp int,
 CONSTRAINT pk_Score_scrId PRIMARY KEY(scrId)
)


CREATE TABLE Record (
 oppId CHAR(5) NOT NULL,
 dteId CHAR(5) NOT NULL,
 scrId CHAR(5) NOT NULL,
 CONSTRAINT pk_Record_oppId_dteId_scrId PRIMARY KEY(oppId, dteId,scrId),
 CONSTRAINT fk_Record_oppId FOREIGN KEY(oppId)
   REFERENCES [Opponent](oppId)
   ON DELETE NO ACTION ON UPDATE CASCADE,
 CONSTRAINT fk_Record_dteId FOREIGN KEY(dteId)
   REFERENCES[Date](dteId)
   ON DELETE NO ACTION ON UPDATE CASCADE,
 CONSTRAINT fk_Record_scrId FOREIGN KEY(scrId)
   REFERENCES [Score] (scrId)
   ON DELETE NO ACTION ON UPDATE CASCADE,
)

CREATE TABLE Play  (
 playId CHAR(5) NOT NULL,
 oppId CHAR(5) NOT NULL,
 locId CHAR(5) NOT NULL,
 ground VARCHAR(10),
 CONSTRAINT pk_Play_playId PRIMARY KEY(playId),
 CONSTRAINT fk_Play_oppId FOREIGN KEY(oppId)
   REFERENCES [Opponent](oppId)
```

```
      ON DELETE NO ACTION ON UPDATE CASCADE,
  CONSTRAINT fk_Play_locId FOREIGN KEY(locId)
     REFERENCES [Location](locId)
     ON DELETE NO ACTION ON UPDATE CASCADE
)

SELECT * FROM Opponent
SELECT * FROM Score
SELECT * FROM Record, DATE where Record.dteId = Date.dteId
SELECT * FROM Date
SELECT * FROM Play
```

7. Insert data into tables: Use excel columns created in Step 5

```
USE BUDT703_Project_0507_04

DELETE FROM Record
DELETE FROM Play
DELETE FROM Opponent
DELETE FROM Location
DELETE FROM Date
DELETE FROM Score
```

For eg: Insert into Score table:

```
INSERT INTO Score VALUES
('S0001', 7, 8),
('S0002', 6, 9),
('S0003', 6, 8),
('S0004', 1, 2),
('S0005', 20, 2),
('S0006', 12, 4),
('S0007', 3, 4),
('S0008', 17, 2),
('S0009', 15, 2),
('S0010', 9, 3),
('S0011', 6, 5),
('S0012', 0, 6),
('S0013', 1, 11),
('S0014', 5, 9),
('S0015', 1, 9),
('S0016', 3, 7),
('S0017', 3, 8),
('S0018', 2, 4),
('S0019', 1, 8),
('S0020', 8, 15),
('S0021', 2, 5),
('S0022', 18, 6),
('S0023', 0, 9),
```

8. Connect SQL to Tableau server and create required data model:

9. Visualizing key metrics:

1) What are the total matches played per year?

```sql
SELECT YEAR(d.dteMatchDate) AS 'Year of Game', COUNT(r.scrId) AS 'Total matches played
each year'
FROM Date d, Record r
WHERE d.dteId = r.dteId
GROUP BY YEAR(d.dteMatchDate)
ORDER BY YEAR(d.dteMatchDate)
```

	Year of Game	Total matches played each year
1	1995	48
2	1996	57
3	1997	48
4	1998	50
5	1999	33
6	2000	42
7	2001	33
8	2002	45
9	2003	44
10	2004	59
11	2005	45
12	2006	58
13	2007	59
14	2008	59
15	2009	54

	Year of Game	Total matches played each year
16	2010	57
17	2011	57
18	2012	56
19	2013	57
20	2014	46
21	2015	54
22	2016	52
23	2017	51
24	2018	55
25	2019	51
26	2020	23
27	2021	44
28	2022	52
29	2023	48

AD\ronak99 (97) | BUDT703_Project_0507_04 | 00:00:00 | 29 rows

AD\ronak99 (97) | BUDT703_Project_0507_04 | 00:00:00 | 29 rows

Total matches pla..
23   59

2) What is the total percentage of wins in each year?

```sql
WITH CalNum AS (
SELECT YEAR(d.dteMatchDate) AS 'Year of Game' , COUNT(r.scrId) AS 'Total Wins'
FROM Record r, Date d, Score s
WHERE d.dteId = r.dteId AND s.scrUMD>s.scrOpp AND r.scrId = s.scrId
GROUP BY YEAR(d.dteMatchDate)
),
CalDen AS (
SELECT YEAR(d.dteMatchDate) AS 'Year of Game', COUNT(r.scrId) AS 'Total Matches
Played'
FROM Record r, Date d
WHERE d.dteId = r.dteId
GROUP BY YEAR(d.dteMatchDate)
)
SELECT c1.[Year of Game], ROUND(100*CONVERT(FLOAT ,c1.[Total Wins])/CONVERT(FLOAT,
c2.[Total Matches Played]),2) AS 'Win percentage'
FROM CalNum c1, CalDen c2
WHERE c1.[Year of Game] = c2.[Year of Game]
ORDER BY  c1.[Year of Game]
```
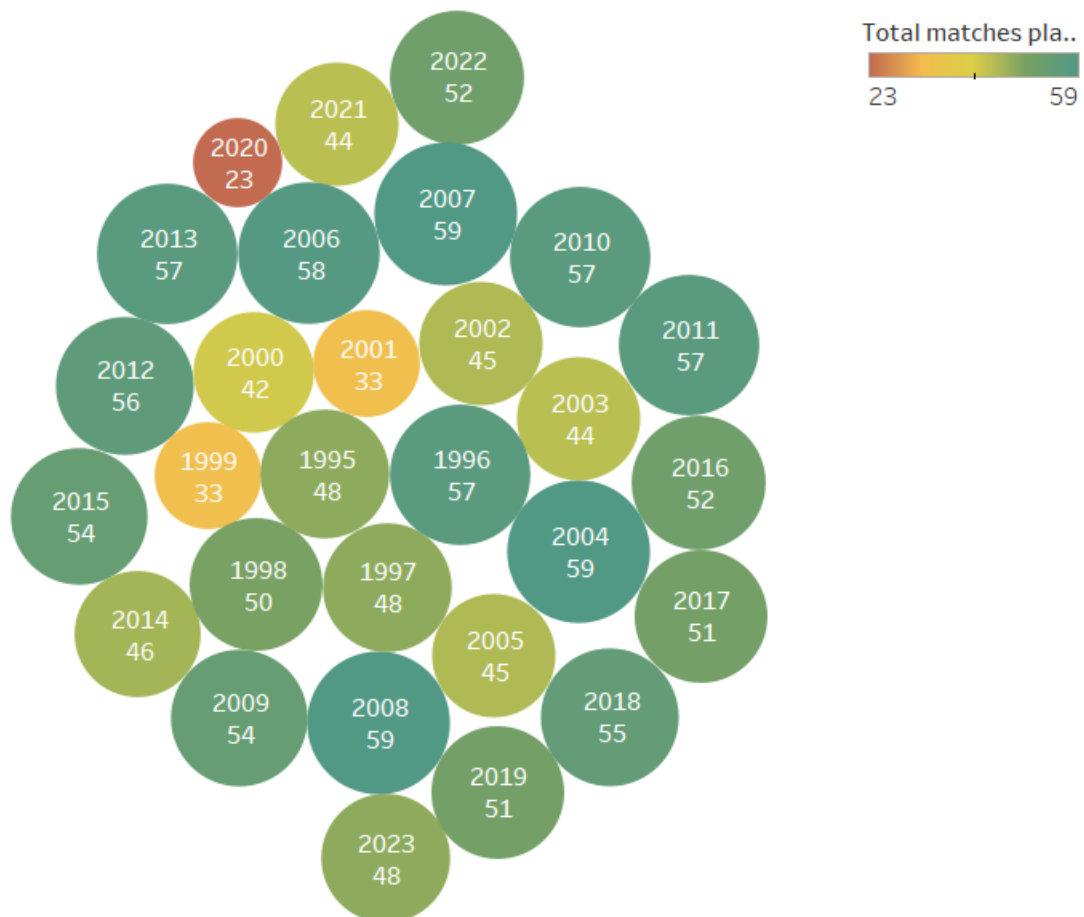
	Year of Game	Win percentage
1	1995	31.25
2	1996	57.89
3	1997	52.08
4	1998	56
5	1999	75.76
6	2000	83.33
7	2001	60.61
8	2002	75.56
9	2003	59.09
10	2004	54.24
11	2005	53.33
12	2006	55.17
13	2007	55.93
14	2008	62.71
15	2009	51.85

AD\ronak99 (97) | BUDT703_Project_0507_04 | 00:00:00 | 29 rows

	Year of Game	Win percentage
16	2010	57.89
17	2011	66.67
18	2012	60.71
19	2013	54.39
20	2014	23.91
21	2015	50
22	2016	23.08
23	2017	21.57
24	2018	32.73
25	2019	39.22
26	2020	52.17
27	2021	43.18
28	2022	55.77
29	2023	31.25

AD\ronak99 (97) | BUDT703_Project_0507_04 | 00:00:00 | 29 rows

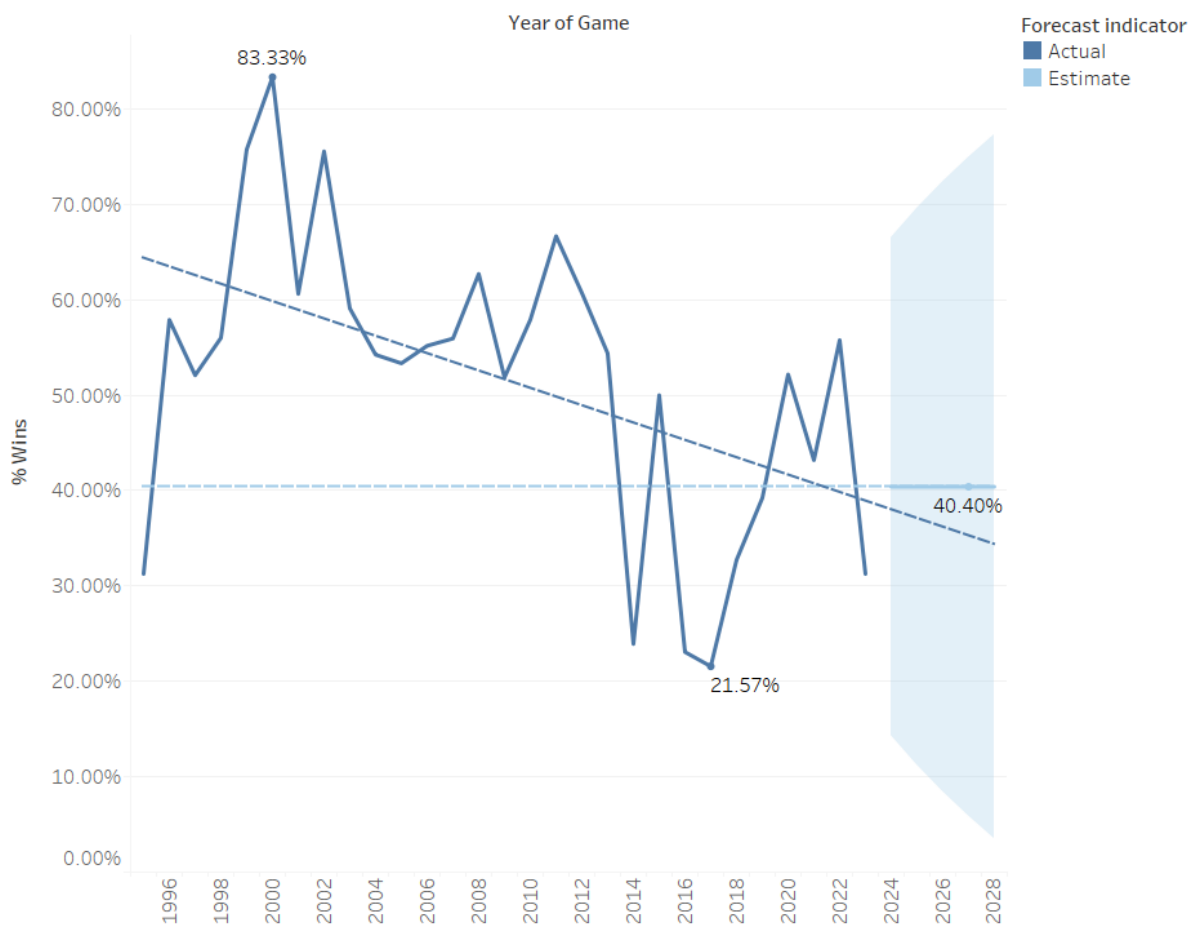3) What is the total percentage of losses in each year?

```sql
WITH CalNum AS (
SELECT YEAR(d.dteMatchDate) AS 'Year of Game' , COUNT(r.scrId) AS 'Total Losses'
FROM Record r, Date d, Score s
WHERE d.dteId = r.dteId AND s.scrUMD<s.scrOpp AND r.scrId = s.scrId
GROUP BY YEAR(d.dteMatchDate)
),
CalDen AS (
SELECT YEAR(d.dteMatchDate) AS 'Year of Game', COUNT(r.scrId) AS 'Total Matches
Played'
FROM Record r, Date d
WHERE d.dteId = r.dteId
GROUP BY YEAR(d.dteMatchDate)
)
SELECT c1.[Year of Game], ROUND(100*CONVERT(FLOAT ,c1.[Total Losses])/CONVERT(FLOAT,
c2.[Total Matches Played]),2) AS 'Loss percentage'
FROM CalNum c1, CalDen c2
WHERE c1.[Year of Game] = c2.[Year of Game]
ORDER BY c1.[Year of Game]
```
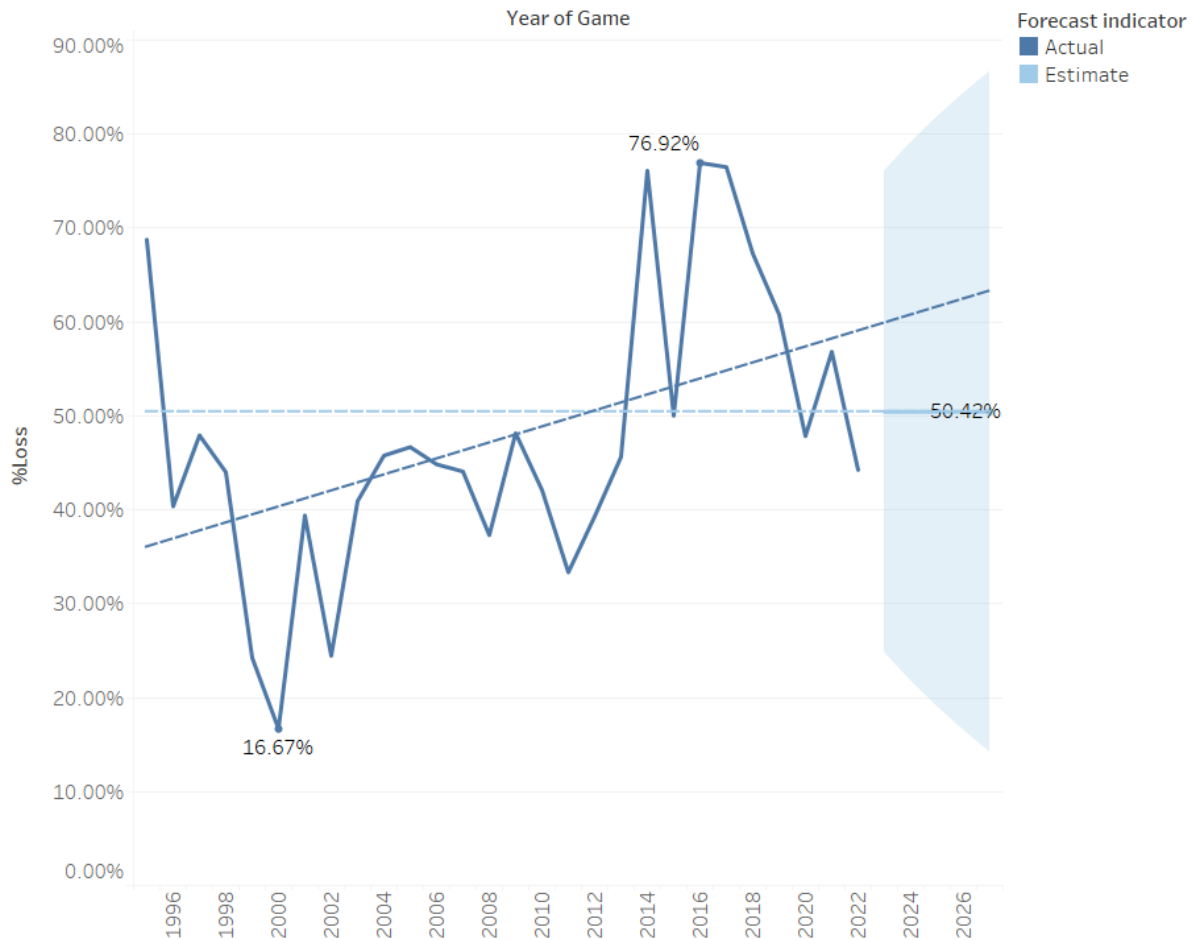
	Year of Game	Loss percentage
1	1995	68.75
2	1996	40.35
3	1997	47.92
4	1998	44
5	1999	24.24
6	2000	16.67
7	2001	39.39
8	2002	24.44
9	2003	40.91
10	2004	45.76
11	2005	46.67
12	2006	44.83
13	2007	44.07
14	2008	37.29
15	2009	48.15

AD\ronak99 (97) | BUDT703_Project_0507_04 | 00:00:00 | 29 rows

	Year of Game	Loss percentage
16	2010	42.11
17	2011	33.33
18	2012	39.29
19	2013	45.61
20	2014	76.09
21	2015	50
22	2016	76.92
23	2017	76.47
24	2018	67.27
25	2019	60.78
26	2020	47.83
27	2021	56.82
28	2022	44.23
29	2023	68.75

AD\ronak99 (97) | BUDT703_Project_0507_04 | 00:00:00 | 29 rows

4) What are the Top 5 teams that beat UMD the maximum time over the years?

```sql
WITH TotalMatches AS (
SELECT o.oppName AS 'Opponent Name', COUNT(r.scrId) AS 'Total Matches Played'
FROM Record r, Opponent o
WHERE o.oppId = r.oppId
GROUP BY o.oppName
),
MatchesLost AS (
SELECT o.oppName AS 'Opponent Name', COUNT(r.scrId) AS 'Total Losses'
FROM Record r, Opponent o, Score s
WHERE o.oppId = r.oppId AND s.scrUMD<s.scrOpp AND r.scrId = s.scrId
GROUP BY o.oppName
)

SELECT TOP (5) m.[Opponent Name],m.[Total Losses], t.[Total Matches Played] AS 'Total
Matches Played'
FROM TotalMatches t,  MatchesLost m
WHERE m.[Opponent Name]=t.[Opponent Name]
ORDER BY m.[Total Losses] DESC
```

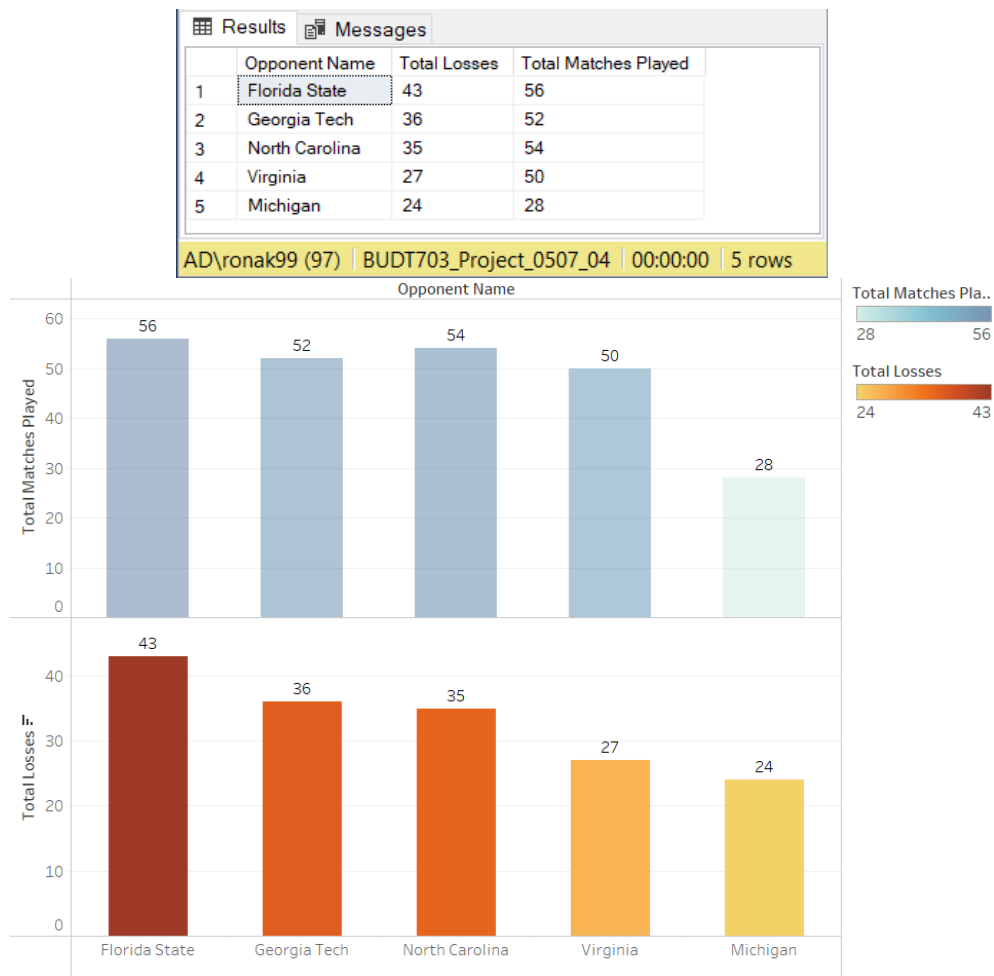|	Opponent Name	Total Losses	Total Matches Played
1	Florida State	43	56
2	Georgia Tech	36	52
3	North Carolina	35	54
4	Virginia	27	50
5	Michigan	24	28

AD\ronak99 (97) | BUDT703_Project_0507_04 | 00:00:00 | 5 rows

5) What are the Top 5 teams that UMD beat the most over the years?

```
WITH TotalMatches AS (
SELECT o.oppName AS 'Opponent Name', COUNT(r.scrId) AS 'Total Matches Played'
FROM Record r, Opponent o
WHERE o.oppId = r.oppId
GROUP BY o.oppName
),
MatchesWon AS (
SELECT o.oppName AS 'Opponent Name', COUNT(r.scrId) AS 'Total Wins'
FROM Record r, Opponent o, Score s
WHERE o.oppId = r.oppId AND s.scrUMD>s.scrOpp AND r.scrId = s.scrId
GROUP BY o.oppName
)

SELECT TOP (5) m.[Opponent Name],m.[Total Wins], t.[Total Matches Played] AS 'Total
Matches Played'
FROM TotalMatches t,  MatchesWon m
WHERE m.[Opponent Name]=t.[Opponent Name]
ORDER BY m.[Total Wins] DESC
```
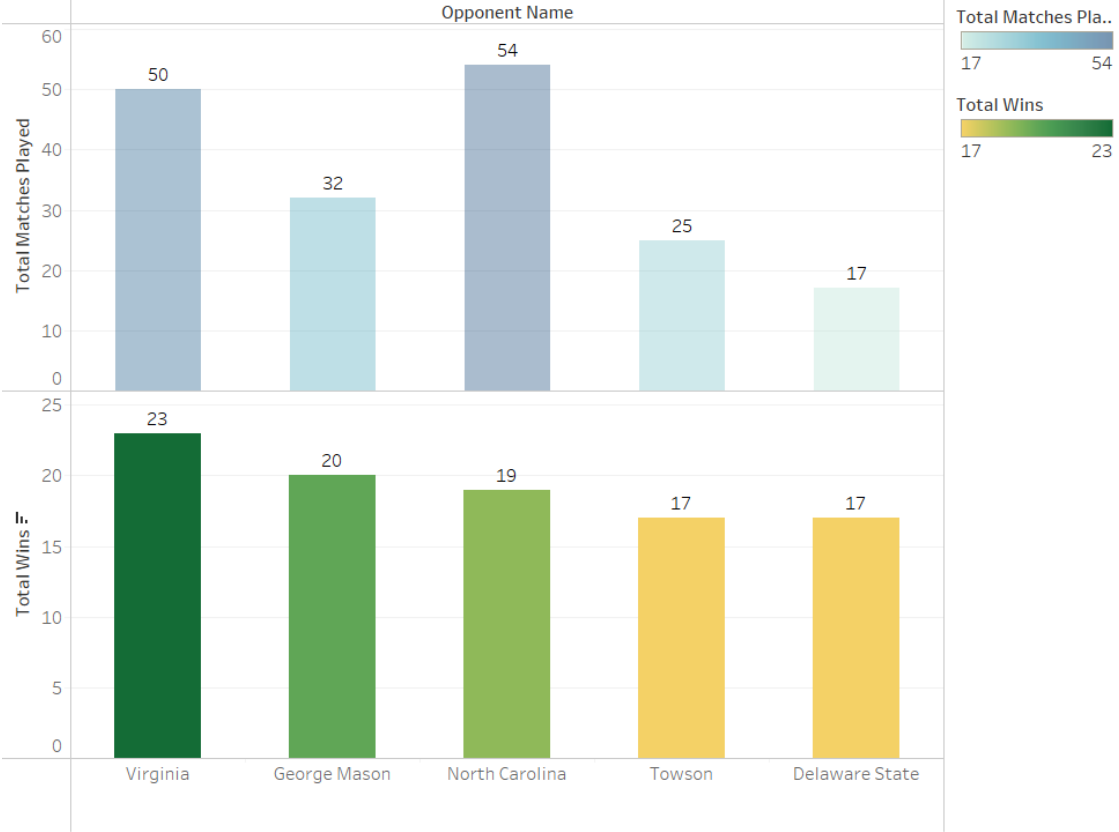
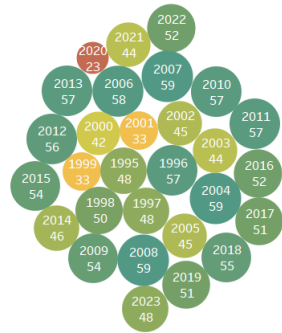|	Opponent Name	Total Wins	Total Matches Played
1	Virginia	23	50
2	George Mason	20	32
3	North Carolina	19	54
4	Delaware State	17	17
5	Towson	17	25

Opponent Name

Total Matches Pla..
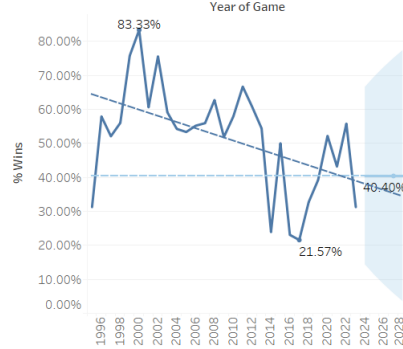17   54

Total Wins
17   23

10.

What are the total matches played per year?

Total matches played
23 — 59

Total Losses
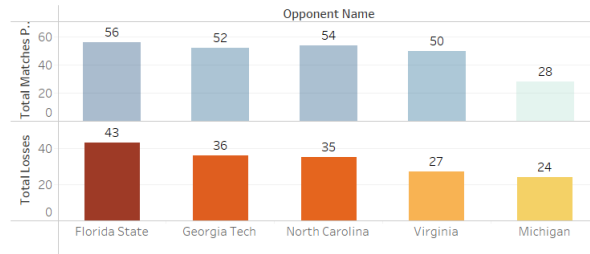24 — 43

What is the total percentage of wins in each year?

83.33%

21.57%

40-40%

% Wins

Year of Game

What is the total percentage of losses in each year?

76.92%

16.67%

50.42%

%Loss

Year of Game
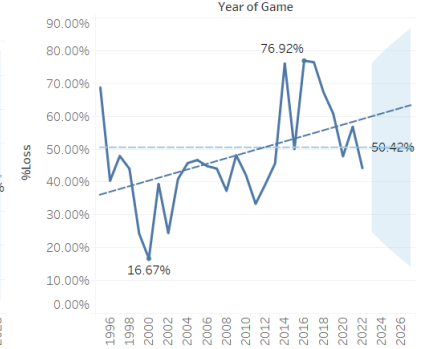
Forecast indicator
■ Actual
■ Estimate

Total Wins
17 — 23

Which are top 5 teams that beat UMD the maximum time over the years?

Opponent Name

Total Matches P..
56   52   54   50   28

Total Losses
43   36   35   27   24

Florida State   Georgia Tech   North Carolina   Virginia   Michigan

Which are top 15 teams that UMD beat the most over the years?

Opponent Name

Total Matches ..
50   32   54   25   17

Total Wins ..
23   20   19   17   17

Virginia   George Mason   North Carolina   Towson   Delaware State