BUDT 703 DBMS Readme document

1. **Data Source**

* The below link gives the information about the opponent, date, time, tournament, venue, games scores and game result. This data is available for every year starting 1999. And we use the text only option to generate the data in an extractable format.   
  <https://umterps.com/sports/baseball/schedule>

A close-up of a computer screen

Description automatically generated

* Along with this we used google to get the venue name. For example, if Terps play UMBC at an away game in UMBC then the venue name is Baseball Factory.
* Similarly, we added the university name for the respective opponent team.

1. **Data cleaning and pre-processing**

* We paste the data into excel from the page shown above.

A screenshot of a computer

Description automatically generated

* Opponent Data

We assign each unique opponent a unique ID, and then we have the opponent Name and Opponent university. We then use the adjacent cell to concatenate the values as per the format required in the INSERT SQL query as highlighted in the top. While doing this we ensured different variations of the same opponent name is normalized to one name and the ID is allocated only once.

A screenshot of a computer

Description automatically generated

* Tournament Data   
  We assign each unique Tournament a unique ID, and then we have the tournament Name. We then use the adjacent cell to concatenate the values as per the format required in the INSERT SQL query as highlighted in the top.

Since Tournament name couldn’t be extracted from the text page of the UMD baseball schedule website, we manually added the data from the main schedule page for every year.

A screenshot of a computer

Description automatically generated

* Calendar Data   
  Here we have the list of months and year We then use the adjacent cell to concatenate the values as per the format required in the INSERT SQL query as highlighted in the top.

A screenshot of a calendar

Description automatically generated

* Venue Data   
  We assign each unique Venue a unique ID, and then we have the venue Name, venue city and venue state. We then use the adjacent cell to concatenate the values as per the format required in the INSERT SQL query as highlighted in the top. While doing this we ensured different variations of the same venue is normalized to one name and the ID is allocated only once.

A screenshot of a computer

Description automatically generated

* Play Data

The data being reported in the table are as follows, date, time, **month, year, opponent ID, tournament ID, Venue ID**, Game type ( Home or Away), Terps’ score and opponent score. The items bolded are foreign keys and are filled in Excel using VLOOKUP. We then use the adjacent cell to concatenate the values as per the format required in the INSERT SQL query as highlighted in the top.

A screenshot of a computer

Description automatically generated

1. **Data Insertion in SQL**

* Drop Tables

These lines of code is created so that in case any other table with a similar name is present we drop them.

A computer screen shot of a computer code

Description automatically generated

* Create Tables

A screenshot of a computer

Description automatically generated

A white background with text

Description automatically generated

* Insert Tables.

These are the lines of code written to insert the values into the respective table of the database.

A computer screen shot of a computer code

Description automatically generated

A white background with black text

Description automatically generated

A close-up of a computer code

Description automatically generated

A white background with black text

Description automatically generated

A group of people in a row

Description automatically generated with medium confidence



A screen shot of a computer code

Description automatically generated

1. SQL Query and Result Interpretation

* **Transaction 1 What is the ratio of game wins based on home, away, and neutral sites?**

Our query aims to calculate the win ratio for the games based on the location where the game is played, which is the play type: home, neutral or away sites to identify the presence of a home game advantage for the Terps. The query retrieves the win ratio for each play type by counting the number of wins and calculating the ratio over the total number of plays. This is then grouped by the play type and years and ordered by years as well so that the obtained output can be interpreted over the years and trends can be analyzed.

A screenshot of a computer

Description automatically generated

* **Transaction 2 What is the team’s average margin of wins and losses in the previous years?**

For this transaction we calculate the average margin of win or loss. First, we take the difference of the scores in the matches where Terps won and divide it by the number of games won. This will give the average win margin. Similarly, we compute the average loss margin with the numerator being the difference of the scores in the matches where Terps lost and divide it by the games terps lost. This is then grouped by year yielding the trend.

A computer screen shot of a computer

Description automatically generated

* **Transaction 3 What is the ratio of game wins in Big Ten conference games?**

In this transaction we aim to see the superiority of the Terps baseball team at the Big Ten Conference. We start of by calculating the number of wins and divide it by the total number of matches played in the tournament. Subsequently we filter out the results for the BIG Ten tournament and then group by tournament name and year.

A computer screen shot of a computer

Description automatically generated

* **Transaction 4 What is the team’s Winning rate against opponents where atleast 10 games being played?**

In order to analyze the superiority of the terps team against opponents we face regularly (matches played overall >=10) we create a query to obtain the win margin. We start of by calculating the number of wins and divide it by the total number of matches played against that opponent using the data from the Play and Opponent table. We set a condition where the number of matches played is greater than or equal to 10 and we then sort by the descending order of win ratio.

A white background with yellow text

Description automatically generated

* **Transaction 5 What is the win ratio over the years across different States?**

In order to analyze the Identify the states where our team traditionally performed well versus, we create a query to obtain the win margin. We start of by calculating the number of wins and divide it by the total number of matches in a particular state using the data from the Play and Venue table. We group the data by the state and then sort it by the descending order of win ratio.

A computer screen shot of a computer

Description automatically generated

1. **Creating Tableau Views**

**A screenshot of a computer

Description automatically generated**

We first connect the database to the tableau server and then make the connections as shown above, which imitates the quaternary relation we have among the tables.

* Create calculated fields for usage in Tableau views.
  + Games = Count([PlayResult])
  + AvgLossMargin - AVG(IF [Play Result]= 'L' THEN [Play Terp Score] - [Play Opponent Score] ELSE 0 END)
  + AvgWinMargin - AVG(IF [Play Result]= 'W' THEN [Play Terp Score]-[Play Opponent Score] ELSE 0 END)
  + WinRatio - ROUND(SUM(IF [Play Result]= 'W' THEN 1 ELSE 0 END)/COUNT([Play Result]),2) Wins - SUM(IF [Play Result]= 'W' THEN 1 ELSE 0 END
* Tableau Visualizations
* **Transaction 1 What is the ratio of game wins based on home, away, and neutral sites?**

A graph of lines on a white background

Description automatically generated

* **Transaction 2 What is the team’s average margin of wins and losses in the previous years?**

**A screenshot of a graph

Description automatically generated**

* **Transaction 3 What is the ratio of game wins in Big Ten conference games?**

**A screenshot of a computer

Description automatically generated**

* **Transaction 4 What is the team’s Winning rate against opponents where atleast 10 games being played?**

**A screenshot of a graph

Description automatically generated**

* **Transaction 5 What is the win ratio over the years across different States?**

In the Data pane, click the data type icon next to 'Venue State' field under 'Moneryball.Venue', select Geographic Role, and then select the 'State/Province' role to assign to the field.

**A map of the united states

Description automatically generated**

* **Tableau Dashboard**

**A screenshot of a graph

Description automatically generated**