Terrapin Tactics

Team Members

1. Sriram Ganesh
2. Suvedaranyan Alliarasi Murugaian
3. Taniya Agrawal
4. Jasvinder Singh

Mission Statement

Our mission is to conduct a thorough analysis of the University of Maryland's softball team's historical data in order to generate comprehensive insights about team performance. This analysis serves as a valuable tool to empower the team with actionable information, enabling them to identify key performance metrics, make informed strategic decisions, and ultimately enhance their overall performance on the field.

Mission Objective

- Venue Performance Evaluation: Conduct a comprehensive comparison of the team's performance in various settings—home, away, and neutral venues—to inform strategic venue selection for future seasons and recommend modifications to enhance the home field advantage, thereby improving overall win rates.
- Opponent Performance Analysis: Examine historical game data to pinpoint the top 10 teams where UMD has excelled and the 10 where they have struggled the most. This analysis will be used to derive strategic insights aimed at bolstering team performance against a diverse range of opponents.
- Seasonal Impact Review: Investigate how different seasons and associated weather conditions affect team performance. This review will identify any seasonal patterns that could influence training schedules and match strategies.
- Venue Win Percentage Breakdown: Analyze UMD's win percentages at various venues, identifying the top and bottom 10 to uncover trends and inform future game planning, ensuring the team is positioned to capitalize on venues where they have historically performed well.

Business Tasks and Processes

Data Extraction from Source:
Extracting relevant and actionable data is a critical step in our analytical journey. Here's our systematic approach to data extraction from the specified source: https://umterps.com/sports/softball/schedule:

- ***Web Scraping:*** We plan to employ advanced Excel techniques and formulas for web scraping, which will enable us to systematically navigate web pages, access HTML content, and retrieve essential data. Given the structured nature of the sites with schedules, results, and player statistics, our Excel-based scripts will be specifically designed to parse this information efficiently.
- ***Identifying Relevant Data:*** Not all available data might be relevant to our analysis. We will filter out only the necessary information based on our defined entities and analytical objectives. For instance, while the website might offer a detailed breakdown of every minute of the game, we might be more interested in end-game statistics and specific game-related conditions.
- ***Data Transformation:*** Once extracted, raw data might not be in an immediately usable format. It might require transformation, such as converting strings to numerical values, date-time conversions, or even standardizing certain data points for consistency.
- ***Handling Missing or Incomplete Data:*** Web sources sometimes have incomplete or missing information. We'll implement strategies to handle such scenarios, either by imputation (filling missing values based on certain strategies) or flagging them for manual review.

Entity Definition:
Each entity encapsulates specific facets of the game data, ensuring that we can derive actionable insights. Here's a detailed look at our entity choices and the rationale behind them:

OpponentTeam:

- Definition: This entity represents the teams that are competing against the primary team of interest. Each opponent team has a unique identifier and name.
- Rationale: The OpponentTeam table is chosen to keep a record of all teams that play games against the primary team, allowing for analysis of game outcomes in relation to specific opponents.
- Usage: This table will be used for identifying patterns in wins and losses, comparing performances against different teams, and developing strategies tailored to particular opponents.

GameVenue:

- Definition: This entity captures information about the different locations where games are played, including an identifier and name.
- Rationale: The GameVenue table is included to determine if the location has an effect on the performance of the team, which is essential for location-based analysis.

- Usage: This table will facilitate analysis of team effectiveness in different venues, helping to assess whether the game's location influences the result and thus informing decisions regarding future game locations.

IndividualGame:

- Definition: This entity holds the details of each game played, encompassing identifiers for the game, date, day, venue category (e.g., home, away, neutral) and season in which it occurred.
- Rationale: The IndividualGame table serves as the central hub for collecting all game-related details necessary for comprehensive performance analysis.
- Usage: It will be used for evaluating the consistency of the team's performance based on different time frames and seasons, monitoring the impact of games in various venue categories, and other time-related analyses.

GameResult:

- Definition: This entity is dedicated to storing the outcomes of games, with each record providing a result identifier and the score.
- Rationale: The GameResult table is pivotal for understanding the success of the team by recording detailed game outcomes.
- Usage: This table will be the cornerstone for assessing the team's scoring patterns and overall performance, enabling the analysis of correlations between various factors and game results.

Primary Keys:

- **optId** in OpponentTeam: A unique identifier for each opponent team. This key is systematically generated and is used to distinguish one opponent team from another within the database.
- **venId** in GameVenue: A unique identifier for each game venue. The key is assigned and used to identify and differentiate each venue where games are played. This could be based on a numeric system or a unique code that represents the location.
- **gamId** in IndividualGame: A unique identifier for each individual game. This is generated using a system that may incorporate the date of the game along with a sequential number to ensure that each game is uniquely identifiable.
- **resId** in GameResult: A unique identifier for each game result. This primary key ensures that the results of each game can be tracked and retrieved uniquely, likely using a combination of the game identifier and a unique result number.

ER Schema:

Entities, Attributes, and Primary Keys

OpponentTeam (**optId**, optName)
IndividualGame (**gamId**, gamDate, gamDay, gamTime, gamSeason)
GameVenue (**venId**, venName, venCategory)
GameResult (**resId**, resScore, resOutcome)

Relationships, Attributes, Degrees, Participating Entities, and Constraints

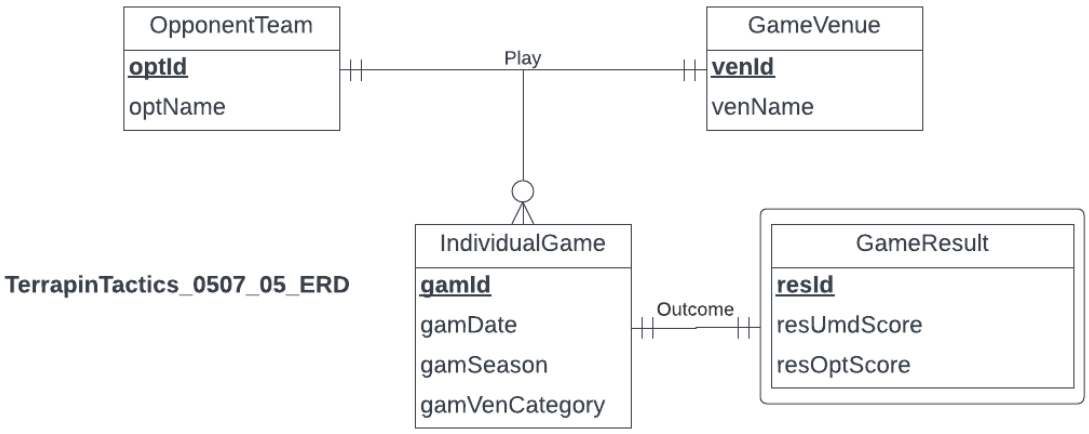
Play: ternary relationship

1 OpponentTeam and 1 IndividualGame to 1 GameVenue
1 IndividualGame and 1 GameVenue to 1 OpponentTeam
1 OpponentTeam and 1 GameVenue to 0 or more IndividualGame

Outcome: binary relationship

1 IndividualGame to 1 GameResult
1 GameResult to 1 IndividualGame

ER Diagram



Relational Schema

OpponentTeam (**optId**, optName)
IndividualGame (**gamId**, gamDate, gamSeason, gamVenCategory)
GameVenue (**venId** venName, venCategory)
GameResult (**resId,** resUmdScore, resOptScore, *gamId*)
Play (***gamId***, ***optId***, ***venId***)

Business Rules

[R1] When a game has been completed, the details of the IndividualGame and GameResult cannot be updated or deleted.

[R2] When a GameVenue details have been deleted, the corresponding details of the OpponentTeam and Individual Game cannot be deleted.

[R3] When a GameVenue details have been updated, it should cascade to update the the corresponding details of the OpponentTeam and Individual Game.

[R4] When a game has been played, the corresponding details of the OpponentTeam and Individual Game cannot be updated or deleted.

Referential Integrities

Relation	Foreign Key	Base Relation	Primary Key	Business Rule	Business Constraint: ON DELETE	Business Rule	Business Constraint: ON UPDATE
Outcome	gamId	IndividualGame	gamId	R1	NO ACTION	R1	NO ACTION
Outcome	resId	GameResult	resId	R1	NO ACTION	R1	NO ACTION
Play	venId	GameVenue	venId	R2	NO ACTION	R3	CASCADE
Play	gamId	IndividualGame	gamId	R4	NO ACTION	R4	NO ACTION
Play	optId	OpponentTeam	optId	R4	NO ACTION	R4	NO ACTION

Sample Data

OpponentTeam

- optId (Integer): 101
- optName (String): "City Raiders"

GameVenue

- venId (Integer): 301
- venName (String): "Downtown Arena"
- venCategory (String): "Stadium"

IndividualGame

- gamId (Integer): 501
- gamDate (Date): "2023-06-12" (YYYY-MM-DD)
- gamDay (String): "Saturday"
- gamTime (Time): "15:00" (HH:MM)
- gamSeason (String): "Summer 2023"

GameResult

- resId (Integer): 701
- resScore (String): "3-2"
- resOutcome (String): "Win"

SQL 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'),
```