

# CSCI-6370 Database Management

## Final Project Report

### “Pro-Kabaddi League Analysis”

Thakkar Vatsal

Deval Aditya

Kandula Punith

The University of Georgia

#### Introduction:

The Pro Kabaddi League, founded in 2014, is a professional Kabaddi league based in India. It has gained immense popularity not just within the country, but also in several other nations. In the first season alone, it attracted 43.5 crores (435 million) viewers, second only to the 2014 Indian Premier League's viewership of 55.2 crores (552 million). Surprisingly, no one has previously worked on analyzing the Kabaddi league data considering all seasons. However, our team has taken the initiative to develop a web application for our final term project that showcases various analyses of Pro-Kabaddi League data for all seasons. Our analysis includes match-specific and player-specific statistics, as well as identifying the factors that influence match outcomes.

#### Dataset & Normalization:

As part of this project, we are gathered data from reliable sources such as Kaggle. This includes valuable information about matches, players, and seasons, as well as specific events that took place during the games. The dataset covers 7 seasons, beginning in 2014 and continuing through 2021, with comprehensive details included for each.

You can find the dataset at,

<https://www.kaggle.com/datasets/sripaadsrinivasan/prokabaddiseason17stats>

It mainly consists of four tables,

1. Ds\_team.csv (34 Columns, ~1400 Rows)
2. Ds\_player.csv (31 Columns, ~17000 Rows)
3. Ds\_match.csv (17 Columns, 684 Rows)
4. Ds\_events.csv (41 Columns, ~67,000 Rows)

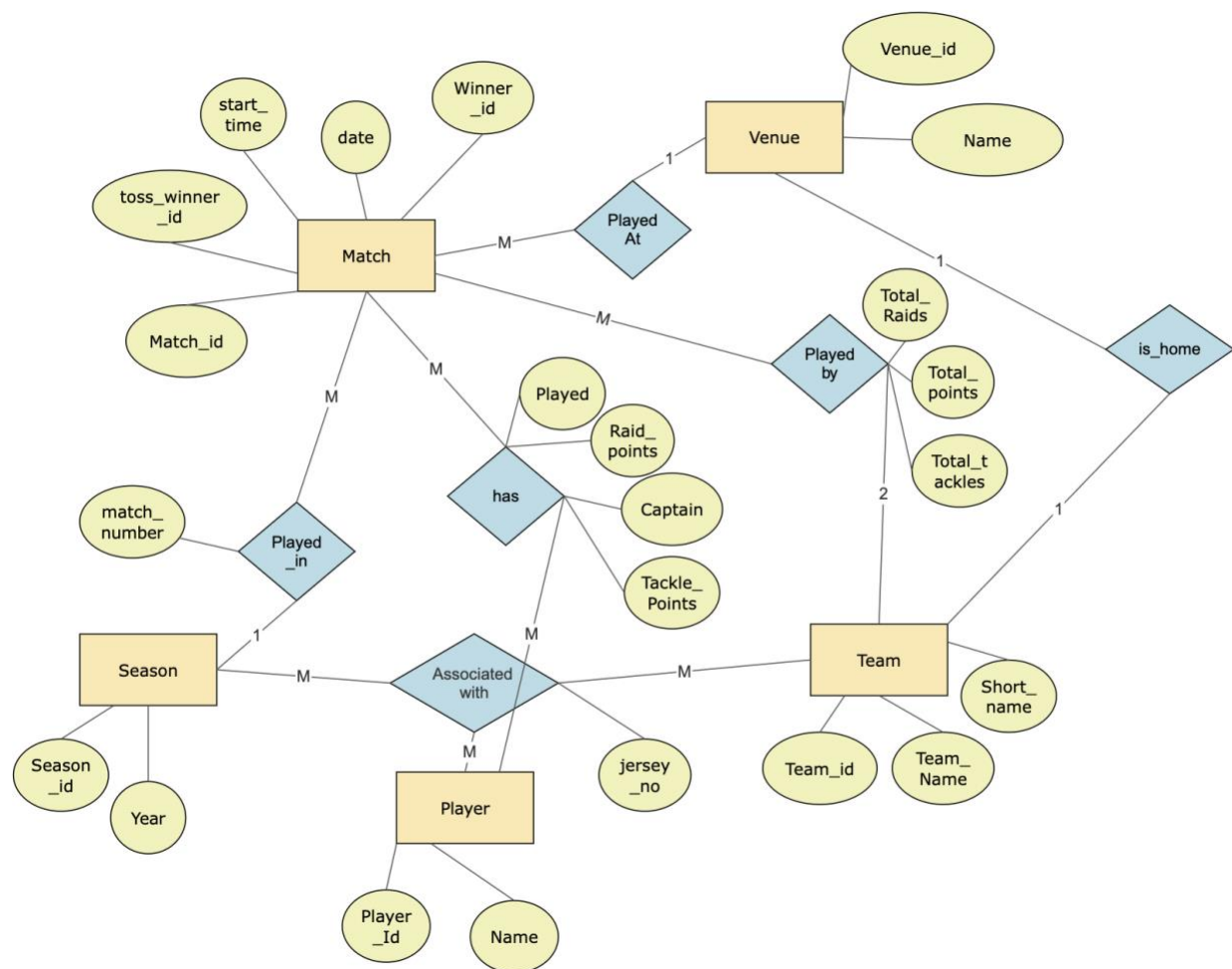
Upon review, we found that the tables were not in a normalized state and contained duplicate data, making it difficult to analyze. To address this, we applied several normalization techniques to the data and separated it into five primary tables and two supporting tables. Additionally, the data set included other data, such as world cup match data, and numerous null values, which we cleaned and preprocessed to make it suitable for our application's specific needs. You can find the code for this and *normalize\_data.ipynb* file.

Our final dataset consists following tables,

1. Team (3 Columns, 12 Rows)
2. Match (8 Columns, ~600 Rows)
3. Player (2 Columns, ~500 Rows)
4. Season (3 Columns, 7 Rows)
5. Venue (4 columns , ~20 Rows)
6. Team\_Match\_Stats (23 Columns, ~1200 Rows)
7. Player\_Match\_stats (20 Columns, ~15000 Rows)

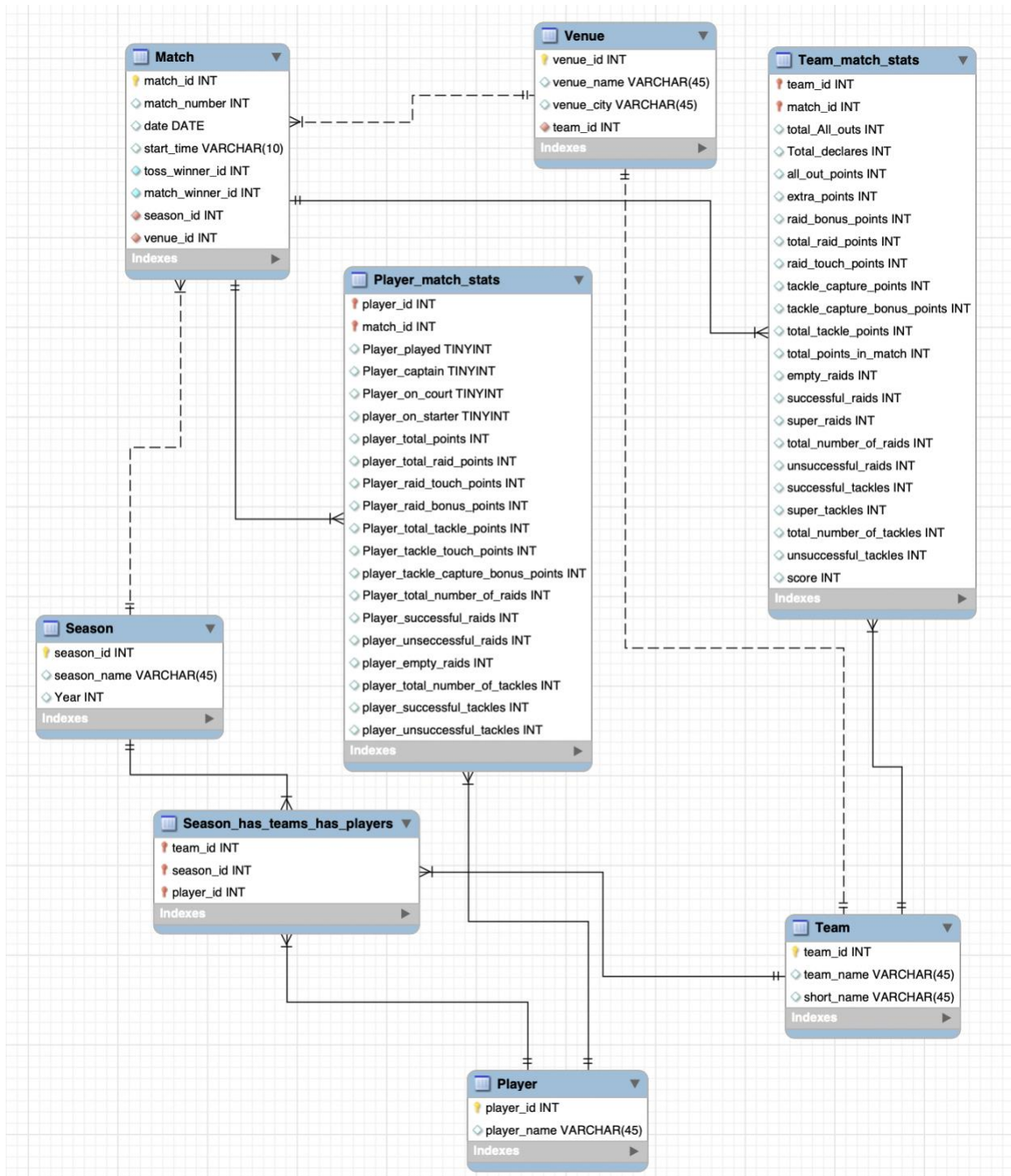
To help us better understand our database, we created a preliminary ER Diagram.

### The Preliminary ER Diagram of our database:



Furthermore, we also developed a final EER diagram of our Pro-kabaddi League (PKL) database schema.

## EER Diagram of PKL schema (Created in MySQL workbench):



Note: You can also install this database in your system by running the python script provided with the code.

## Queries and Impact of Project:

To maximize the potential of the data set, we have formulated analytical queries that concentrate on match-specific and player-specific metrics. Additionally, we aim to pinpoint the factors that influence the outcome of a match, such as toss and venue. This information can help us make better decisions and improve overall performance of teams and players. We also strive to identify the factors that affect match outcomes, such as toss and venue, to make better decisions and improve overall team and player performance. By analyzing these factors, teams can uncover patterns and trends to optimize their strategies and gain a competitive advantage. This type of analysis can also assist team staff in developing better strategies for player auctions, ultimately leading to greater success and efficient goal attainment.

Following is the list of queries:

1. What is the success rate of the raiders who have attempted the most raids across all seasons?
2. a season-wise comparison of the maximum and average raid points, tackle points, and total points?
3. How does each team's win percentage compare between home matches, neutral venues, and away matches?
4. What is the win rate of a team when they win the toss compared to when they don't?
5. Who are the defenders with the highest success rate in tackling across all seasons?
6. What are the success and failure rates of teams in knockout or decider matches?

To display the results of this queries we designed a web application using Servlet and JSP. The procedure to run the web application is provided into the README.md file which is provided with the code. There are 4 major components of the web application.

1. ***Index.jsp*** file : which serves as the default home page or landing page for a web application that is built using Java Server Pages (JSP) technology.
2. ***IndexServlet.java*** file: a Java-based server-side technology that runs on a application server and is used to process and respond to client requests over HTTP or HTTPS protocols.
3. ***DBConnector.java*** file: which establishes connection to and interact with databases using SQL queries.
4. ***ProjectQueries.java*** file: Contains all the queries which is going to be executed.

## Queries in SQL:

**Query-1:** What is the success rate of the raiders who have attempted the most raids across all seasons?

```
SELECT player_name AS 'Player Name',
       ROUND(SUM(Player_successful_raids)/SUM(player_total_number_of_raids)
* 100, 2) AS 'Raider Success Rate(%) '
FROM player_match_stats
JOIN Player ON player.player_id = player_match_stats.player_id
WHERE player_match_stats.player_id in
      (SELECT player_id
      FROM
        (SELECT player_id,
                SUM(player_total_number_of_raids) AS total_raids
        FROM player_match_stats
        GROUP BY player_id
        ORDER BY Total_raids DESC
        LIMIT 10) AS T)
GROUP BY player_match_stats.player_id ;
```

**Query-2:** A season-wise comparison of the maximum and average raid points, tackle points, and total points?

```
SELECT season_name AS 'Season Name',
       ROUND(AVG(total_raid_points), 0) AS 'Average Raid Points',
       MAX(total_raid_points) AS 'Maximum Raid Points in a Match',
       ROUND(AVG(total_tackle_points), 0) AS 'Average Tackle Points',
       MAX(total_tackle_points) AS 'Maximum Tackle Points in a Match',
       ROUND(AVG(score), 0) AS 'Average Total Points',
       MAX(score) AS 'Maximum Total Points in a match'
FROM team_match_stats AS TMS
JOIN pkl.match AS M ON TMS.match_id = M.match_id
JOIN season AS S ON M.season_id = S.season_id
GROUP BY M.season_id;
```

**Query-3:** How does each team's win percentage compare between home matches, neutral venues, and away matches?

```
SELECT team_name AS 'Team Name',
       ROUND(Numeber_of_home_matches_won/Total_home_matches * 100, 2) AS
'Win Percentage Home',
       ROUND(number_of_neutral_matches_won/Total_neutral_matches * 100, 2)
'Win Percentage Neutral',
       ROUND(Number_of_away_matches_won/Total_away_matches * 100, 2) 'Win
Percentage Away'
FROM
  (SELECT team_id,
         COUNT(CASE
               WHEN team_id = venue_type
                 AND team_id = match_winner_id THEN 1
               END) AS Numeber_of_home_matches_won,
         COUNT(CASE
               WHEN team_id = venue_type THEN 1
               END) AS Total_home_matches,
         COUNT(CASE
               WHEN venue_type='Neutral'
                 AND team_id = match_winner_id THEN 1
               END) AS number_of_neutral_matches_won,
         COUNT(CASE
               WHEN venue_type='Neutral' THEN 1
               END) Total_neutral_matches,
         COUNT(CASE
               WHEN team_id != venue_type
                 AND venue_type!='Neutral'
                 AND team_id = match_winner_id THEN 1
               END) Number_of_away_matches_won,
         COUNT(CASE
               WHEN team_id != venue_type
                 AND venue_type!='Neutral' THEN 1
               END) Total_away_matches
    FROM team_match_stats AS TMS3
   JOIN
     (SELECT M.match_id,
            match_winner_id,
```

```

        team_1,
        team_2,
        CASE
            WHEN team_1 = v.team_id THEN team_1
            WHEN team_2 = v.team_id THEN team_2
            ELSE 'Neutral'
        END AS venue_type
FROM pkl.match AS M
JOIN
    (SELECT TMS.team_id AS team_1,
        TMS2.team_id AS team_2,
        TMS.match_id
    FROM team_match_stats AS TMS
    JOIN
        (SELECT *
        FROM team_match_stats AS T1
        ORDER BY team_id DESC) AS TMS2 ON TMS.match_id = TMS2.match_id
        AND TMS.team_id <> TMS2.team_id
        AND TMS.team_id < TMS2.team_id) AS T ON M.match_id = T.match_id
    JOIN venue AS V ON M.venue_id = V.venue_id) AS T3 ON T3.match_id =
TMS3.match_id
    GROUP BY tms3.team_id) AS T4
JOIN team ON T4.team_id = team.team_id;

```

**Query-4:** What is the win rate of a team when they win the toss compared to when they don't?

```

SELECT team_name AS 'Team Name',
    Round(COUNT(CASE
        WHEN TMS.team_id = match_winner_id
            AND TMS.team_id = toss_winner_id THEN 1
        END) / COUNT(CASE
            WHEN TMS.team_id = toss_winner_id THEN
1
            END)*100, 2) AS 'Win Rate (Toss Won)(%)',
    Round(COUNT(CASE
        WHEN TMS.team_id = match_winner_id
            AND TMS.team_id != toss_winner_id THEN 1
        END) /COUNT(CASE

```

```

                                WHEN TMS.team_id != toss_winner_id THEN
1
                                END) *100, 2) AS 'Win Rate (Toss Not
Won) (%) '
FROM team_match_stats AS TMS
JOIN pkl.match AS M ON TMS.match_id = M.match_id
JOIN team AS T ON TMS.team_id = T.team_id
GROUP BY TMS.team_id;

```

**Query-5:** Who are the defenders with the highest success rate in tackling across all seasons?

```

SELECT player_name AS 'Player Name',

ROUND(SUM(player_successful_tackles)/SUM(player_total_number_of_tackles) *
100, 2) AS 'Tackle Success Rate(%) '
FROM player_match_stats
JOIN Player ON player.player_id = player_match_stats.player_id
WHERE player_match_stats.player_id in
    (SELECT player_id
    FROM
        (SELECT player_id,
            SUM(player_total_number_of_tackles) AS total_tackles
        FROM player_match_stats
        GROUP BY player_id
        ORDER BY total_tackles DESC
        LIMIT 10) AS T)
GROUP BY player_match_stats.player_id ;

```



**Query-6:** What are the success and failure rates of teams in knockout or decider matches?

```
SELECT team_name AS 'Team Name',
       ROUND(matches_lost/matches_played*100, 2) 'Losing Rate(%)',
       ROUND(matches_won/matches_played*100, 2) 'Winning Rate(%)',
       matches_played AS 'Total Deciders Played'
FROM
  (SELECT team_id,
         COUNT(CASE
                WHEN team_id != match_winner_id THEN 1
                END) matches_lost,
         COUNT(CASE
                WHEN team_id= match_winner_id THEN 1
                END) matches_won,
         COUNT(team_id) matches_played
  FROM pkl.match AS M
  JOIN team_match_stats AS TMS ON M.match_id = TMS.match_id
  WHERE match_number NOT LIKE 'MATCH%'
  GROUP BY team_id) AS T1
JOIN team ON T1.team_id = team.team_id;
```

## Results of query and Snapshot of web application:

### Home Page:

### CSCI-6370 Database Management - Term Project

#### Pro-Kabaddi League Analysis

**Note:** Click on the Buttons to execute respected Queries. Also, there is a button at the bottom of the table to display the result of recently executed query in case if you close the modal window.

|    |   |         |
|----|---|---------|
| 1. | The success rate of raiders who have made the most attempted raids across all seasons   | Query-1 |
| 2. | Season-wise comparison between the maximum and average raid points, tackle points, and total points.  | Query-2 |
| 3. | For each team, the comparison between their win percentage at home matches, their win percentage at neutral venues, and their win percentage at away matches. | Query-3 |
| 4. | The win rate of a team is being compared between the scenarios where they won the toss and where they did not win the toss                                    | Query-4 |
| 5. | The success rate of defenders who have attempted the highest number of tackles across all seasons.  | Query-5 |
| 6. | The success and failure rates of teams in knockout or decider matches .   | Query-6 |

Show Result

© Copyright by Vatsal Thakkar

### Result of Query-1:

### CSCI-6370 Database Management - Term Project

#### Pro-Kabaddi League Analysis

**Note:** Click on the Buttons to execute respected Queries. Also, there is a button at the bottom of the table to display the result of recently executed query in case if you close the modal window.

|    |   |         |
|----|---|---------|
| 1. | The success rate of raiders who have made the most attempted raids across all seasons   | Query-1 |
| 2. | Season-wise comparison between the maximum and average raid points, tackle points, and total points.  | Query-2 |
| 3. | For each team, the comparison between their win percentage at home matches, their win percentage at neutral venues, and their win percentage at away matches. | Query-3 |
| 4. | The win rate of a team is being compared between the scenarios where they won the toss and where they did not win the toss                                    | Query-4 |
| 5. | The success rate of defenders who have attempted the highest number of tackles across all seasons.  | Query-5 |
| 6. | The success and failure rates of teams in knockout or decider matches .   | Query-6 |

Show Result

© Copyright by Vatsal Thakkar

#### Result of Query

| Player Name        | Raider Success Rate(%) |
|--------------------|------------------------|
| Rahul Chaudhari    | 40.03                  |
| Pardeep Narwal     | 45.79                  |
| Ajay Thakur        | 38.92                  |
| Deepak Nivas Hudda | 40.44                  |
| Rohit Kumar        | 39.78                  |
| Anup Kumar         | 31.43                  |
| Rishank Devadiga   | 37.56                  |
| Kashiling Adake    | 36.13                  |
| Maninder Singh     | 46.79                  |
| Jasvir Singh       | 30.33                  |

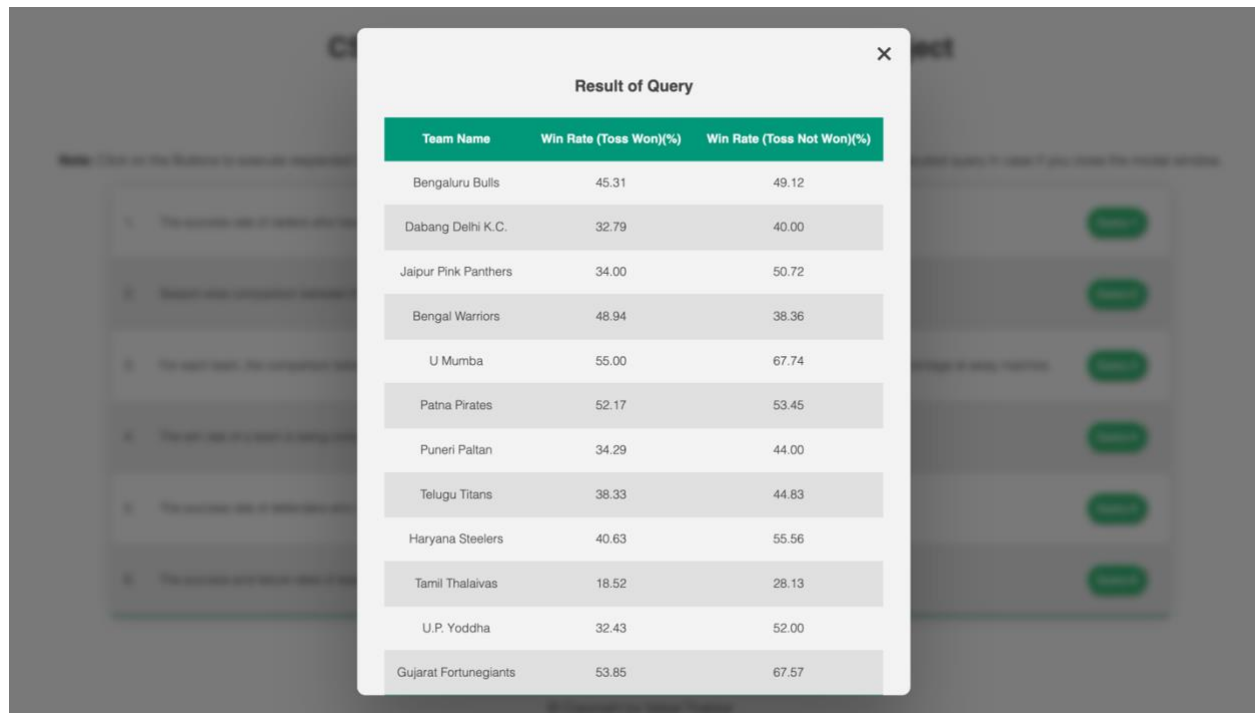
## Result of Query-2:

| Result of Query             |                     |                                |                       |                                  |                      |                                 |
|-----------------------------|---------------------|--------------------------------|-----------------------|----------------------------------|----------------------|---------------------------------|
| Season Name                 | Average Raid Points | Maximum Raid Points in a Match | Average Tackle Points | Maximum Tackle Points in a Match | Average Total Points | Maximum Total Points in a match |
| Pro Kabaddi League Season 1 | 21                  | 35                             | 9                     | 18                               | 33                   | 58                              |
| Pro Kabaddi League Season 2 | 16                  | 35                             | 10                    | 19                               | 30                   | 54                              |
| Pro Kabaddi League Season 3 | 16                  | 34                             | 9                     | 20                               | 30                   | 67                              |
| Pro Kabaddi League Season 4 | 16                  | 27                             | 9                     | 18                               | 29                   | 51                              |

## Result of Query-3:

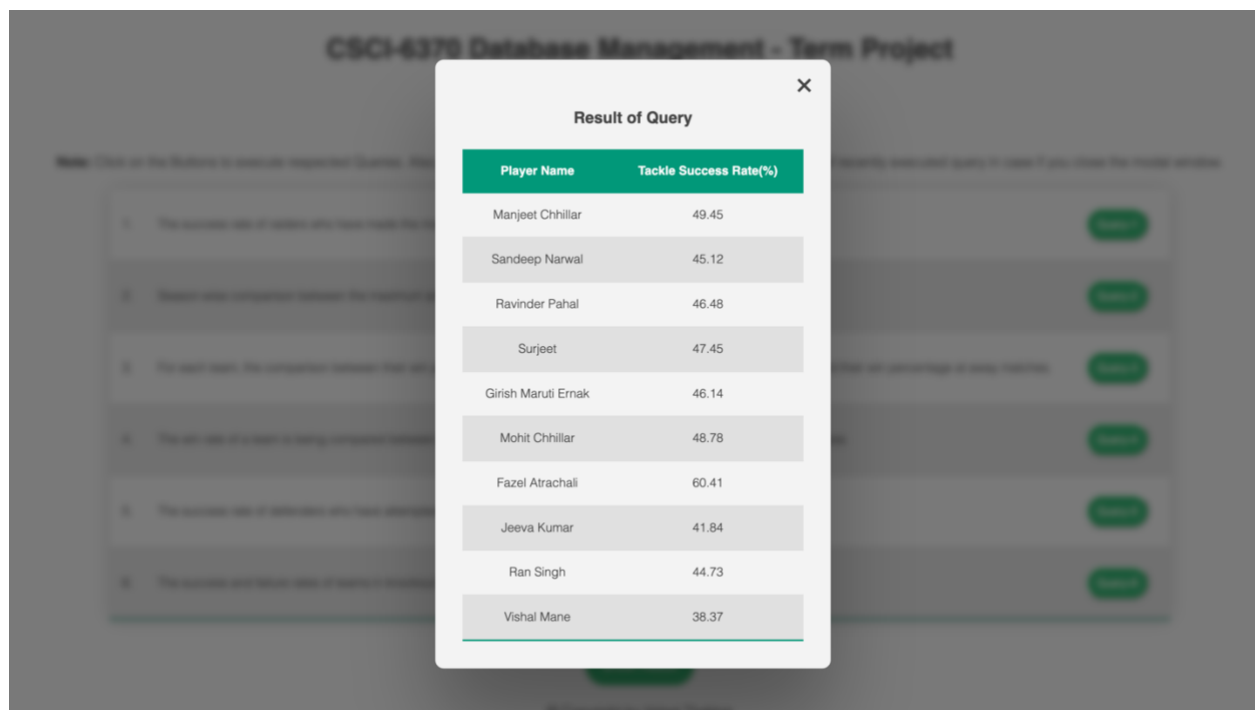
| Result of Query      |                     |                        |                     |
|----------------------|---------------------|------------------------|---------------------|
| Team Name            | Win Percentage Home | Win Percentage Neutral | Win Percentage Away |
| Bengaluru Bulls      | 42.31               | 51.67                  | 42.86               |
| Dabang Delhi K.C.    | 37.50               | 35.59                  | 36.00               |
| Jaipur Pink Panthers | 35.71               | 50.00                  | 40.54               |
| Bengal Warriors      | 48.39               | 41.94                  | 37.04               |
| U Mumba              | 64.86               | 60.00                  | 60.00               |
| Patna Pirates        | 43.75               | 54.69                  | 58.06               |
| Puneri Paltan        | 37.04               | 37.50                  | 41.38               |
| Telugu Titans        | 38.24               | 40.74                  | 46.67               |
| Haryana Steelers     | 33.33               | 47.06                  | 61.54               |
| Tamil Thalaivas      | 6.67                | 26.67                  | 35.71               |
| U.P. Yoddha          | 9.09                | 38.24                  | 64.71               |

### Result of Query-4:



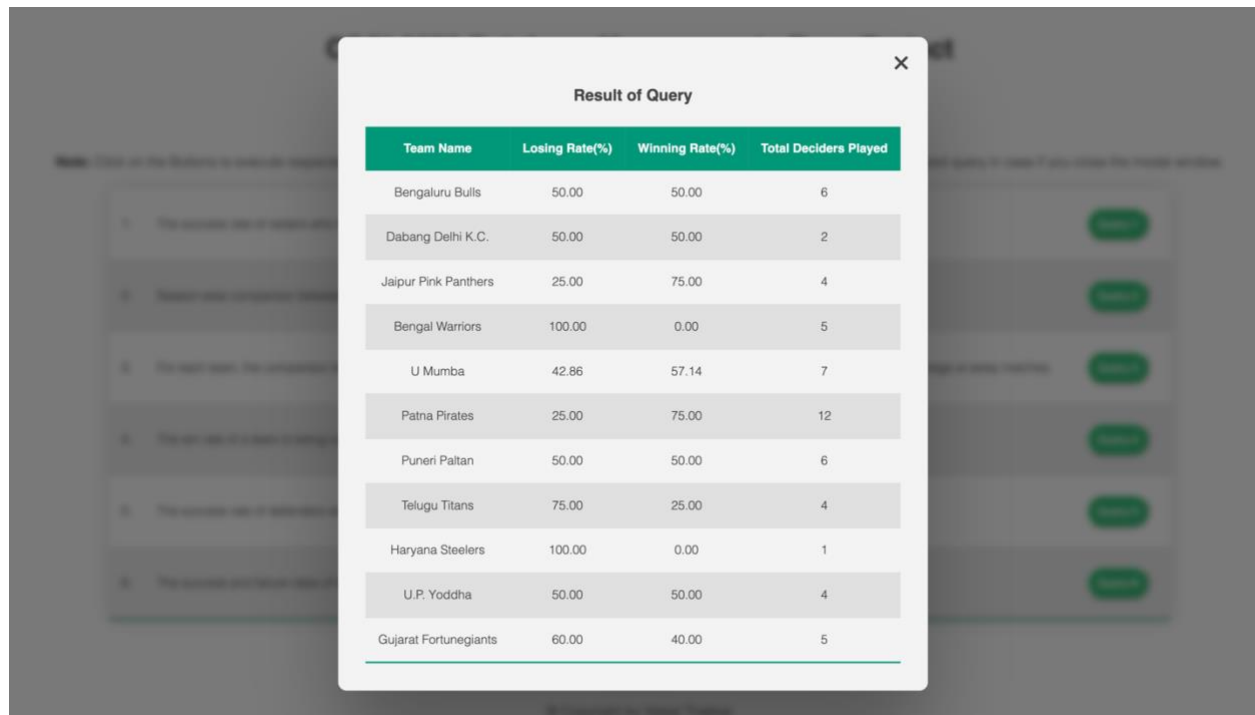
| Team Name             | Win Rate (Toss Won)(%) | Win Rate (Toss Not Won)(%) |
|-----------------------|------------------------|----------------------------|
| Bengaluru Bulls       | 45.31                  | 49.12                      |
| Dabang Delhi K.C.     | 32.79                  | 40.00                      |
| Jaipur Pink Panthers  | 34.00                  | 50.72                      |
| Bengal Warriors       | 48.94                  | 38.36                      |
| U Mumba               | 55.00                  | 67.74                      |
| Patna Pirates         | 52.17                  | 53.45                      |
| Puneri Paltan         | 34.29                  | 44.00                      |
| Telugu Titans         | 38.33                  | 44.83                      |
| Haryana Steelers      | 40.63                  | 55.56                      |
| Tamil Thalaivas       | 18.52                  | 28.13                      |
| U.P. Yoddha           | 32.43                  | 52.00                      |
| Gujarat Fortunegiants | 53.85                  | 67.57                      |

### Result of Query-5:



| Player Name         | Tackle Success Rate(%) |
|---------------------|------------------------|
| Manjeet Chhillar    | 49.45                  |
| Sandeep Narwal      | 45.12                  |
| Ravinder Pahal      | 46.48                  |
| Surjeet             | 47.45                  |
| Girish Maruti Ernak | 46.14                  |
| Mohit Chhillar      | 48.78                  |
| Fazel Atrachali     | 60.41                  |
| Jeeva Kumar         | 41.84                  |
| Ran Singh           | 44.73                  |
| Vishal Mane         | 38.37                  |

### Result of Query-6:



| Team Name             | Losing Rate(%) | Winning Rate(%) | Total Deciders Played |
|-----------------------|----------------|-----------------|-----------------------|
| Bengaluru Bulls       | 50.00          | 50.00           | 6                     |
| Dabang Delhi K.C.     | 50.00          | 50.00           | 2                     |
| Jaipur Pink Panthers  | 25.00          | 75.00           | 4                     |
| Bengal Warriors       | 100.00         | 0.00            | 5                     |
| U Mumba               | 42.86          | 57.14           | 7                     |
| Patna Pirates         | 25.00          | 75.00           | 12                    |
| Puneri Paltan         | 50.00          | 50.00           | 6                     |
| Telugu Titans         | 75.00          | 25.00           | 4                     |
| Haryana Steelers      | 100.00         | 0.00            | 1                     |
| U.P. Yoddha           | 50.00          | 50.00           | 4                     |
| Gujarat Fortunegiants | 60.00          | 40.00           | 5                     |

### Conclusion:

In conclusion, analyzing trends and patterns within the data set can unlock new opportunities for success and help teams achieve their objectives more effectively. This information can also assist team personnel in devising better strategies for player auctions.