

IPL DATA ANALYSIS REPORT

(GROUP 1)



Manoj Sriram, Vaibhav Negi

25/01/2025
Infotact Data Analyst Internship

GROUP DETAILS

S. No	Name	Class	College
1	Manoj Sriram	CSO - IV Year	Vasireddy Venkatadri Institute of Technology
2	Vaibhav Negi	CSE – III Year	Dev Bhoomi Group of Institutions

TABLE OF CONTENTS

	Page No
1 INTRODUCTION	1
2 DATA EXPLORATION	2
3 DATA PRE-PROCESSING	4
4 DATA ANALYSIS	11
5 CONCLUSION	25

INTRODUCTION

The Indian Premier League (IPL) is one of the most popular and competitive cricket leagues in the world, bringing together players from various countries and showcasing thrilling matches each season. Since its inception in 2008, the league has grown exponentially, becoming a global phenomenon with significant contributions to cricket analytics and sports strategy.

In this project, we delve into the analysis of IPL data to uncover meaningful insights and trends. Our dataset includes detailed records of matches, player performances, team statistics, and venue information. The primary goal is to create a comprehensive data model that provides users with a clear understanding of team dynamics, player contributions, and match outcomes.

Key areas of analysis include:

- Identifying top-performing teams and players.
- Analyzing win probabilities based on factors such as toss decisions and venue conditions.
- Exploring patterns in scoring, bowling, and fielding performances.
- Understanding the impact of home-ground advantage and match venues.

By leveraging data visualization and statistical techniques, this analysis aims to offer actionable insights that can benefit teams, players, and cricket enthusiasts alike, enhancing their understanding of the IPL's complex and dynamic nature.

DATA EXPLORATION

Data Collection

The dataset for this analysis was collected from Kaggle Datasets, a trusted source for publicly available data. This IPL dataset contains detailed records of matches, teams, players, and performance statistics across various seasons of the Indian Premier League. The data includes information such as match results, player contributions, team standings, and venue details.

The dataset was downloaded and stored locally before being imported into a Pandas Dataframe for analysis. Using this framework allowed for efficient data manipulation, cleaning, and preparation, ensuring the dataset was ready for generating meaningful insights and visualizations.

Features Identified for Analysis

The dataset used for this analysis comprises two files: matches.csv and deliveries.csv. These files provide a comprehensive view of the Indian Premier League's match and player statistics.

1. **matches.csv:** This file contains high-level match data such as:
 - **Match details:** ID, season, city, date, match type, and venue.
 - **Teams:** Names of the competing teams, toss winner, toss decision, and the match winner.
 - **Outcomes:** Result type, result margin, and any specific method applied (e.g., DLS).
 - **Other information:** Player of the match, umpires, and details of any super overs.
2. **deliveries.csv:** This file provides ball-by-ball data for all IPL matches, including:
 - **Match and innings information:** Match ID, inning number, batting and bowling teams.
 - **Ball-specific details:** Over, ball, batter, bowler, and non-striker.
 - **Performance metrics:** Runs scored by the batter, extra runs, total runs, and type of extras.
 - **Wicket details:** Information about dismissals, player dismissed, dismissal type, and fielder involved.

These features were chosen to explore various aspects of the IPL, including:

- Match outcomes based on toss decisions and team strategies.

- Player performance metrics such as runs scored, wickets taken, and contributions in key moments.
- Ball-by-ball analysis to study scoring patterns, economy rates, and wicket-taking trends.
- Insights into the impact of venues and match conditions on team performance.

By focusing on these features, we aim to uncover actionable insights into team dynamics, player contributions, and trends that define the IPL's competitive edge.

DATA PRE-PROCESSING

The following pre-processing techniques were employed to prepare the IPL dataset for analysis:

1. Loading the Dataset: After downloading the dataset from Kaggle, the first step was to load the data. Python libraries such as Pandas, Matplotlib, and Seaborn were imported for data manipulation, visualization, and exploratory analysis. The dataset, consisting of two CSV files (matches.csv and deliveries.csv), was loaded into Pandas Dataframes to facilitate further processing.

2. Understanding the Dataset: To ensure accurate analysis, a thorough understanding of the dataset was developed. This included:

- Examining column names and their meanings.
- Identifying data types of each column.
- Displaying the first and last few rows of data to understand its structure.
- Identifying unique values in categorical columns such as teams, venues, and players.
- Generating statistical summaries to gain insights into numerical data such as runs, overs, and margins.

3. Dataset Cleaning: Data cleaning was performed to address inconsistencies and prepare the dataset for analysis:

- Checking for and handling null or missing values appropriately.
- Standardizing column names for easier access and readability.
- Removing or correcting duplicate or irrelevant records.
- Converting data types where necessary to ensure compatibility with analysis methods.
- Filtering and formatting data to align with the analysis objectives.

These pre-processing steps ensured the dataset was clean, well-structured, and ready for analysis, enabling accurate and insightful results in the study of IPL matches and player performances.

The steps followed in the data pre-processing are:

Step 1: Importing libraries

```
import pandas as pd
```

- **Pandas:**

- Purpose: Used for data manipulation and analysis.
- Common Use Cases:
 - Handling structured data in the form of Dataframes.
 - Reading/writing data from CSV, Excel, etc.
 - Cleaning and transforming data.

Step2: Load the Dataset

```
deliveries_path = '/content/deliveries.csv'
matches_path = '/content/matches.csv'

deliveries_data = pd.read_csv(deliveries_path)
matches_data = pd.read_csv(matches_path)
```

- The read_csv function is used to read a CSV (Comma-Separated Values) file into a pandas Dataframe. It is a powerful function provided by the pandas library in Python.

Step 3: Correlation

```
merged_data = deliveries.merge(matches[['id', 'season']], left_on='match_id', right_on='id',
how='left')

numerical_columns = [
    'over', 'ball', 'batsman_runs', 'extra_runs', 'total_runs',
    'is_wicket', 'season']

correlation_data = merged_data[numerical_columns]

correlation_matrix = correlation_data.corr()

print("Correlation Matrix:")
print(correlation_matrix)

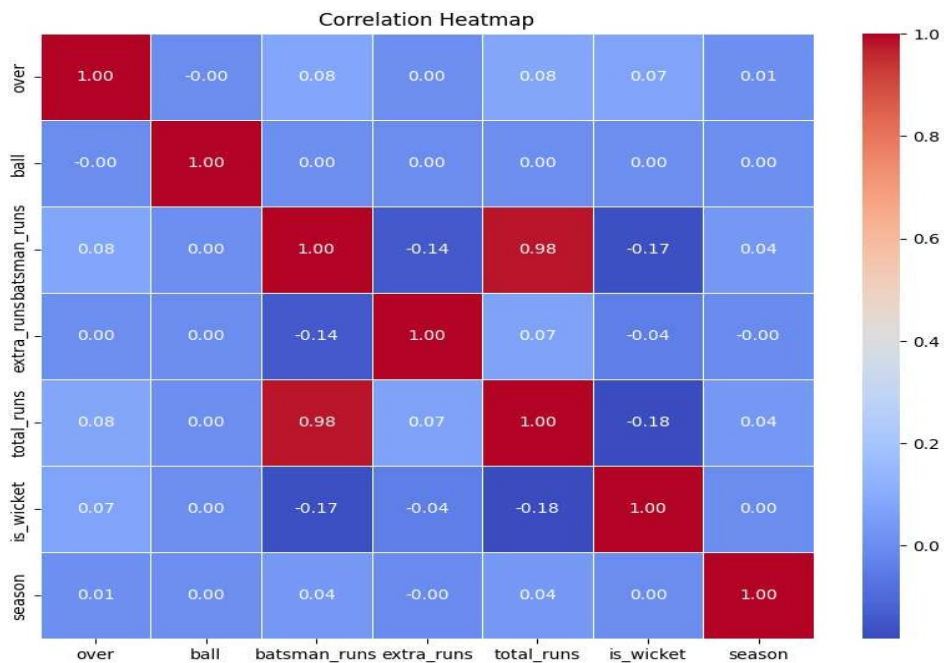
plt.figure(figsize=(10, 8))
sns.heatmap(
    correlation_matrix,annot=True,cmap='coolwarm',fmt='.2f',linewidths=0.5
)
plt.title('Correlation Heatmap')
plt.show()
```


Output:

Correlation Matrix:

```
      over  ball batsman_runs extra_runs total_runs \
over      1.000000 -0.001241  0.080040  0.004437  0.081610
ball     -0.001241  1.000000  0.004391  0.000176  0.004463
batsman_runs 0.080040 0.004391  1.000000 -0.142079  0.977934
extra_runs  0.004437 0.000176 -0.142079  1.000000  0.067852
total_runs  0.081610 0.004463  0.977934  0.067852  1.000000
is_wicket   0.072963 0.004258 -0.171751 -0.041903 -0.181955
season      0.010881 0.004824  0.036320 -0.004712  0.035614
```

```
      is_wicket  season
over      0.072963 0.010881
ball      0.004258 0.004824
batsman_runs -0.171751 0.036320
extra_runs  -0.041903 -0.004712
total_runs  -0.181955 0.035614
is_wicket   1.000000 0.001204
season      0.001204 1.000000
```



The correlation matrix provides a measure of the strength and direction of the relationship between different numerical variables in your dataset. The values range between -1 and 1, where:

- 1: Perfect positive correlation (as one variable increases, the other also increases).
- -1: Perfect negative correlation (as one variable increases, the other decreases).
- 0: No correlation (no relationship between the variables).

Key Observations in the correlation matrix are:

1. High Positive Correlation

- batsman_runs and total_runs (0.977): This indicates a strong positive correlation, which makes sense since batsman runs are a significant component of total runs.
- over and total_runs (0.081): A weak positive correlation indicates that as the game progresses into later overs, total runs slightly tend to increase.

2. Weak or No Correlation

- ball and other variables (close to 0): Variables like ball have almost no correlation with others, which might be because it is more of a sequence indicator rather than a performance metric.
- season and other variables (close to 0): There is no strong season-based trend observed in most metrics. For example, the correlation between season and batsman_runs is 0.036.

3. Negative Correlation

- is_wicket and batsman_runs (-0.172): A weak negative correlation indicates that matches with more wickets tend to have fewer runs scored by individual batsmen.
- is_wicket and total_runs (-0.182): Slight negative correlation suggests that higher wicket losses might lead to lower total scores, as expected.

4. Self-Correlation

- Diagonal values are 1:
- Each variable has a perfect correlation with itself.
- The strongest relationship in the data is between batsman_runs and total_runs, which aligns with expectations, as these metrics are directly related.
- Wickets (is_wicket) tend to slightly reduce total runs and batsman runs, but the relationship isn't very strong.
- Season does not significantly impact other metrics, meaning performance metrics like runs, wickets, and extras are consistent across seasons.

Step 4: Standardization (Reducing Team Names To Shortcuts And Merging Duplicate Venue Details)

```
team_name_mapping =  
{ 'Chennai Super Kings':  
  'CSK', 'Delhi Daredevils':  
  'DC', 'Delhi Capitals': 'DC',
```

```

'Kings XI Punjab': 'PBKS',
'Punjab Kings': 'PBKS',
'Kolkata Knight Riders': 'KKR',
'Mumbai Indians': 'MI',
'Rajasthan Royals': 'RR',
'Kochi Tuskers Kerala': 'KTK',
'Royal Challengers Bangalore': 'RCB',
'Royal Challengers Bengaluru': 'RCB',
'Sunrisers Hyderabad': 'SRH',
'Deccan Chargers': 'SRH',
'Rising Pune Supergiants': 'RPS',
'Rising Pune Supergiant': 'RPS',
'Gujarat Lions': 'GL',
'Pune Warriors': 'PW',
'Lucknow Super Giants': 'LSG',
'Gujarat Titans': 'GT'
}

venue_mapping = {
'Arun Jaitley Stadium, Delhi': 'Arun Jaitley Stadium',
'Brabourne Stadium, Mumbai': 'Brabourne Stadium',
'Dr DY Patil Sports Academy, Mumbai': 'Dr DY Patil Sports Academy',
'Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Stadium, Visakhapatnam': 'Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Stadium',
'Eden Gardens, Kolkata': 'Eden Gardens',
'Feroz Shah Kotla': 'Arun Jaitley Stadium',
'Himachal Pradesh Cricket Association Stadium, Dharamsala': 'Himachal Pradesh Cricket Association Stadium',
'M Chinnaswamy Stadium, Bengaluru': 'M Chinnaswamy Stadium',
'MA Chidambaram Stadium, Chepauk': 'MA Chidambaram Stadium',
'MA Chidambaram Stadium, Chepauk, Chennai': 'MA Chidambaram Stadium',
'Maharashtra Cricket Association Stadium, Pune': 'Maharashtra Cricket Association Stadium',
'Punjab Cricket Association IS Bindra Stadium, Mohali': 'Punjab Cricket Association Stadium',
'Punjab Cricket Association IS Bindra Stadium': 'Punjab Cricket Association Stadium',
'Rajiv Gandhi International Stadium, Uppal': 'Rajiv Gandhi International Stadium',
'Rajiv Gandhi International Stadium, Uppal, Hyderabad': 'Rajiv Gandhi International Stadium',
'Sawai Mansingh Stadium, Jaipur': 'Sawai Mansingh Stadium',
'Wankhede Stadium, Mumbai': 'Wankhede Stadium',
'Punjab Cricket Association Stadium, Mohali': 'Punjab Cricket Association Stadium',
'Punjab Cricket Association Stadium': 'Punjab Cricket Association Stadium',
'Punjab Cricket Association IS Bindra Stadium, Mohali, Chandigarh': 'Punjab Cricket Association Stadium',
'M.Chinnaswamy Stadium': 'M Chinnaswamy Stadium',
'Sardar Patel Stadium, Motera': 'Narendra Modi Stadium',
'Narendra Modi Stadium, Ahmedabad': 'Narendra Modi Stadium',
}

```

```

matches_data['venue'] = matches_data['venue'].replace(venue_mapping)
matches_data['team1'] = matches_data['team1'].replace(team_name_mapping)
matches_data['team2'] = matches_data['team2'].replace(team_name_mapping)
matches_data['winner'] = matches_data['winner'].replace(team_name_mapping)
matches_data['toss_winner'] = matches_data['toss_winner'].replace(team_name_mapping)
deliveries_data['batting_team'] = deliveries_data['batting_team'].replace(team_name_mapping)
deliveries_data['bowling_team'] = deliveries_data['bowling_team'].replace(team_name_mapping)

list1=matches_data['team1'].unique()
print(list1)

list2=matches_data['venue'].unique()
print(np.sort(list2))
print(len(list2))

list3=matches_data['toss_winner'].unique()
print(list3)

```

Output:

```

['RCB' 'PBKS' 'DC' 'MI' 'KKR' 'RR' 'SRH' 'CSK' 'KTK' 'PW' 'GL' 'RPS' 'LSG'
 'GT']
['Arun Jaitley Stadium' 'Barabati Stadium'
 'Barsapara Cricket Stadium, Guwahati'
 'Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cricket Stadium, Lucknow'
 'Brabourne Stadium' 'Buffalo Park' 'De Beers Diamond Oval'
 'Dr DY Patil Sports Academy'
 'Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Stadium'
 'Dubai International Cricket Stadium' 'Eden Gardens' 'Green Park'
 'Himachal Pradesh Cricket Association Stadium' 'Holkar Cricket Stadium'
 'JSCA International Stadium Complex' 'Kingsmead' 'M Chinnaswamy Stadium'
 'MA Chidambaram Stadium'
 'Maharaja Yadavindra Singh International Cricket Stadium, Mullanpur'
 'Maharashtra Cricket Association Stadium' 'Narendra Modi Stadium'
 'Nehru Stadium' 'New Wanderers Stadium' 'Newlands' 'OUTsurance Oval'
 'Punjab Cricket Association Stadium' 'Rajiv Gandhi International Stadium'
 'Saurashtra Cricket Association Stadium' 'Sawai Mansingh Stadium'
 'Shaheed Veer Narayan Singh International Stadium'
 'Sharjah Cricket Stadium' 'Sheikh Zayed Stadium' 'St George's Park'
 'Subrata Roy Sahara Stadium' 'SuperSport Park'
 'Vidarbha Cricket Association Stadium, Jamtha' 'Wankhede Stadium'
 'Zayed Cricket Stadium, Abu Dhabi']
38
['RCB' 'CSK' 'RR' 'MI' 'SRH' 'PBKS' 'KKR' 'DC' 'KTK' 'PW' 'GL' 'RPS' 'GT'
 'LSG']

```

1. Mappings for Standardization

- **team_name_mapping:** This dictionary maps various names of cricket teams to their standardized abbreviations (e.g., 'Chennai Super Kings' → 'CSK', 'Delhi Daredevils' and 'Delhi Capitals' → 'DC').
- **venue_mapping:** This dictionary maps different versions or formats of venue names to a single standardized version (e.g., 'Feroz Shah Kotla' → 'Arun Jaitley Stadium').

2. Applying Venue Standardization

- **Purpose:** Standardizes the venue names in the `matches_data` DataFrame.
- **How:**
 - The `.replace()` method replaces values in the venue column based on the mappings defined in `venue_mapping`.
 - For example, if a row contains 'Feroz Shah Kotla', it will be replaced with 'Arun Jaitley Stadium'.

3. Standardizing Team Names in matches_data

- **Purpose:** Ensures team names in `matches_data` are consistent.
- **Columns Affected:**
 - `team1`: The first playing team.
 - `team2`: The second playing team.
 - `winner`: The match-winning team.
 - `toss_winner`: The team that won the toss.
- **How:**
 - The `.replace()` method replaces team names in these columns using the `team_name_mapping` dictionary.
 - Example: 'Delhi Daredevils' or 'Delhi Capitals' will be replaced by 'DC'.

4. Standardizing Team Names in deliveries_data

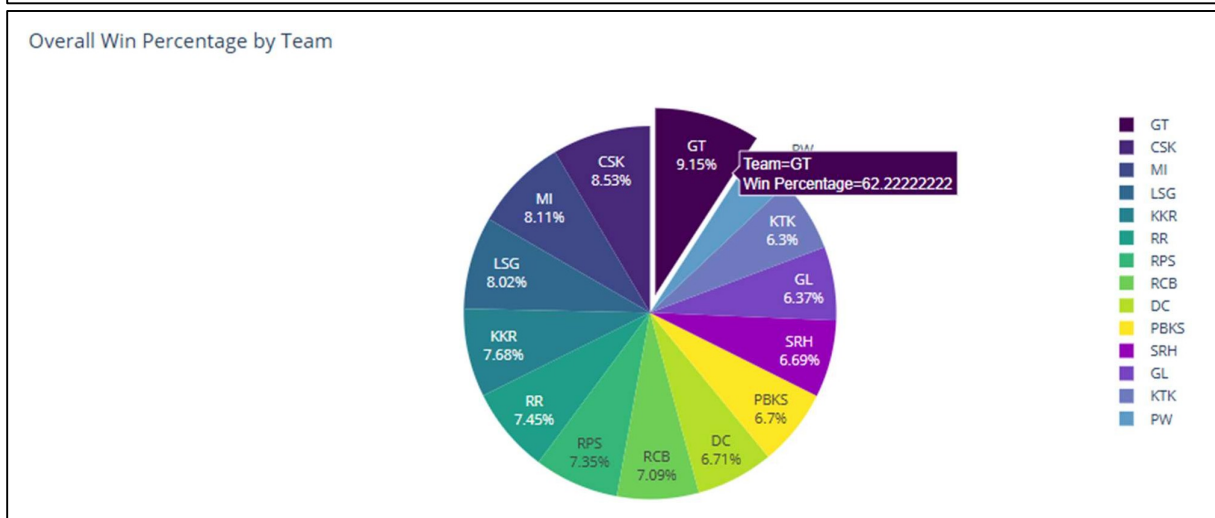
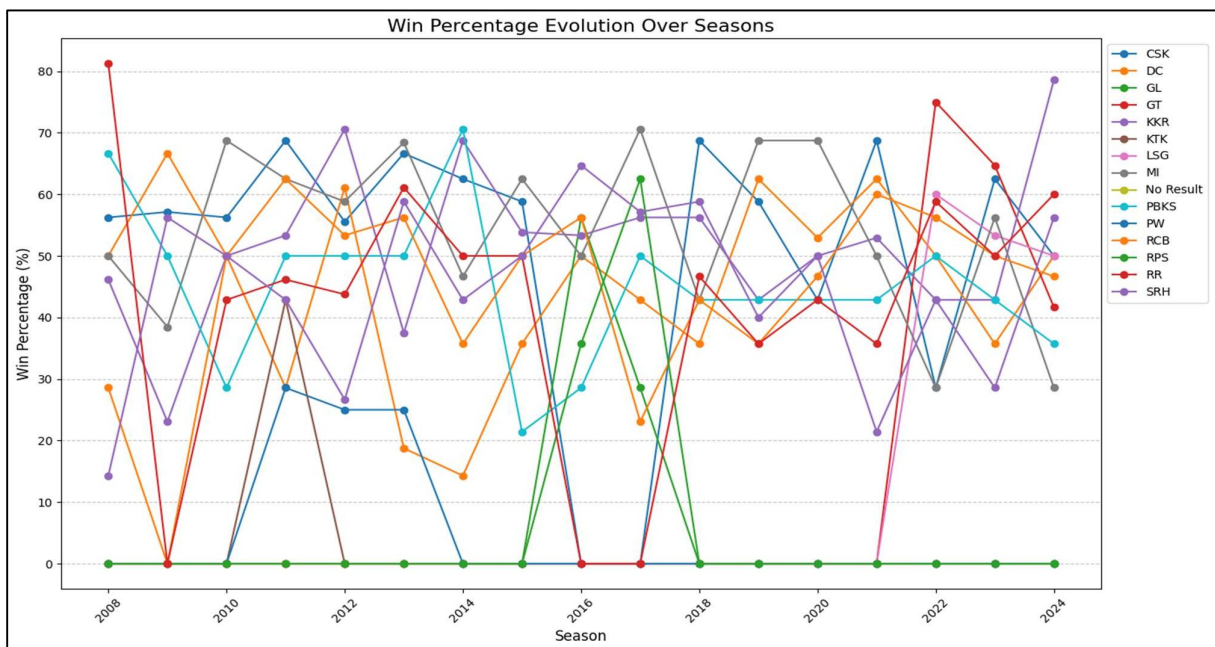
- **Purpose:** Ensures team names in the `deliveries_data` DataFrame are consistent.
- **Columns Affected:**
 - `batting_team`: The team currently batting.
 - `bowling_team`: The team currently bowling.
- **How:**
 - The `.replace()` method applies the `team_name_mapping` dictionary to these columns.
 - Example: 'Kings XI Punjab' will be replaced by 'PBKS'.

DATA ANALYSIS

A. USING MATPLOTLIB.PYPILOT, SEABORN, AND PLOTLY EXPRESS

1. VISUALIZE MATCH OUTCOMES (WIN/LOSS) ACROSS DIFFERENT YEARS

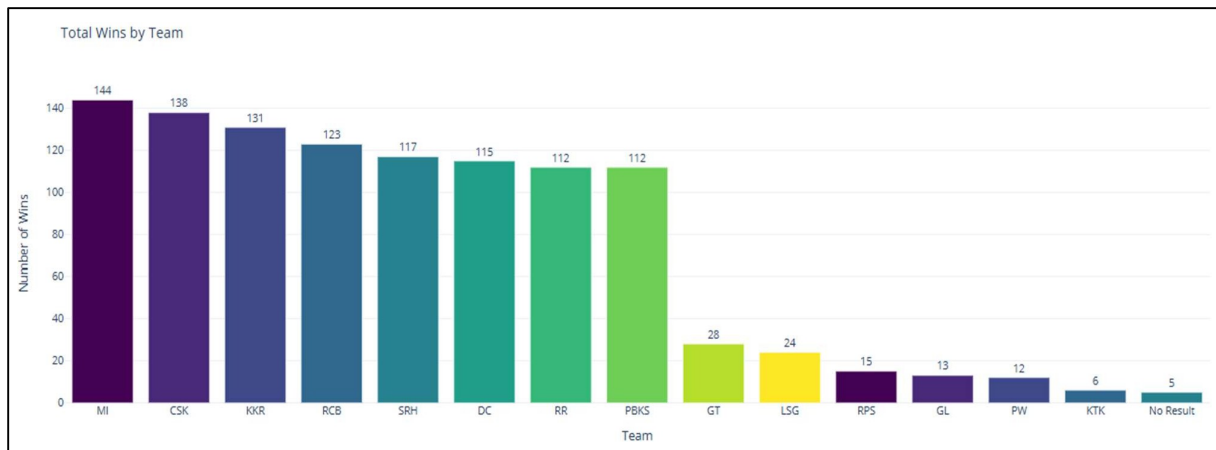
- *Objective:* To analyse match outcomes (wins/losses) across multiple IPL seasons.
- *Insights Expected:*
 - Identify winning patterns for different teams over the years.
 - Highlight consistent performers and fluctuating trends.
- *Visualization:* Line charts and Pie charts showing the percentage of matches won each team per season.



- The trend identified in the above line chart is the most consistent team is “CSK” and boasts a winning percentage of “57.9%”.
- While Coming to the winning percentage, the new team “GT(Gujarat Titans)” boasts a winning percentage of “62.2%”.

2. ANALYZE TEAM PERFORMANCE BASED ON HISTORICAL DATA

- *Objective:* To measure and compare team performances over the IPL's history.
- *Insights Expected:*
 - Which teams have dominated historically.
 - Comparative analysis of count of matches win by each team in the history.
- *Visualization:* Bar chart representing the total wins each team has won throughout all the seasons.



- The trend identified is that “MI(Mumbai Indians)” has won the most number of matches(144), and there lies “CSK(Chennai Super Kings)” claiming the 2nd spot with 138 wins.

3. SEASONAL TRENDS FOR ALL TEAMS

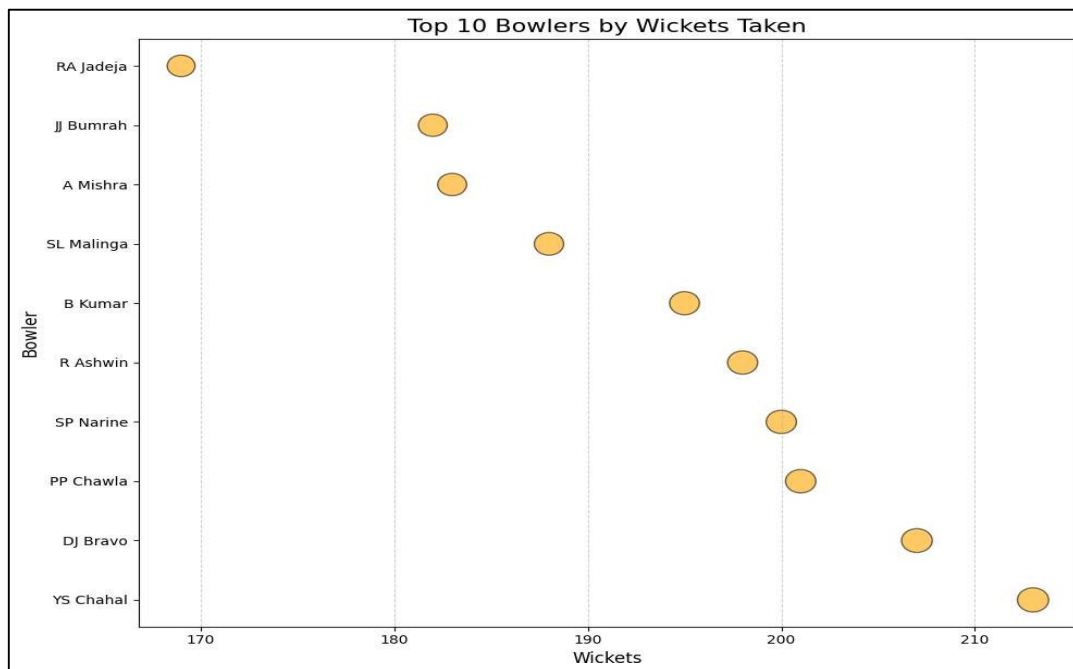
- *Objective:* To identify performance variations for teams during different IPL seasons.
- *Insights Expected:*
 - Seasonal peaks and slumps for each team.
 - Correlation of team performance with external factors like venue or player changes.
- *Visualization:* Line plots showcasing seasonal trends for all teams.



- The trend identified during this analysis is that CSK, MI, SRH perform consistently over the seasons.

4. WICKETS TAKEN PER BOWLER

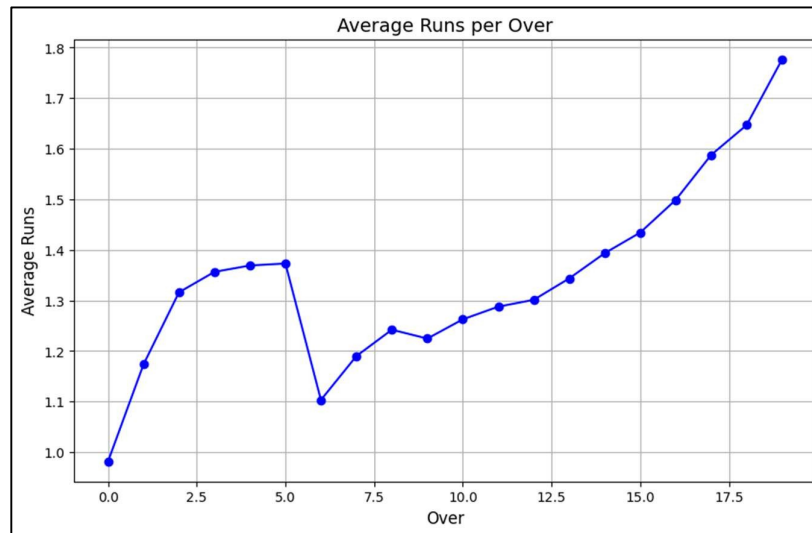
- Objective:** To highlight the top 10 bowlers by the number of wickets taken.
- Insights Expected:**
 - Identifies the most impactful bowlers in the dataset and their wicket-taking ability.
 - A horizontal dot plot visually represents each bowler's contribution, making it easier to compare values at a glance.
 - The use of dot size adds an additional visual cue to emphasise the difference between players.



- This analysis is conducted to find the bowler with most wickets in history of IPL and Yuzvendra Chahal leads the list by taking 211 wickets.

5. AVERAGE RUNS PER OVER ACROSS MATCHES

- *Objective:* To analyze the scoring pattern across different overs in an IPL match.
- *Insights Expected:*
 - Identify which overs contribute the most to the total runs.
 - Determine if there are specific overs where scoring is significantly higher or lower.
- *Visualization:* A line plot showing the average runs scored in each over across all IPL matches.

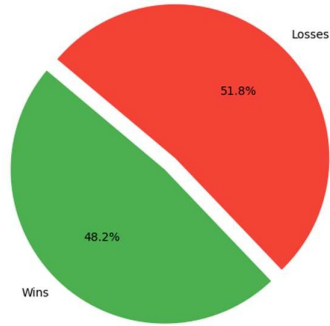


- The analysis indicates that the scoring rate generally increases in the final overs, with a peak towards the death overs (16-20), highlighting aggressive batting strategies in the closing stages of the innings.

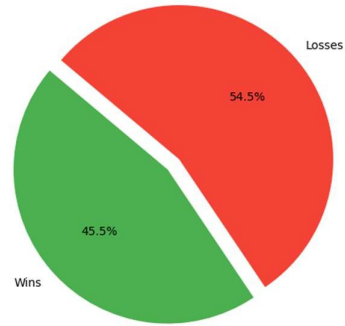
6. COMPARE TEAM PERFORMANCE USING LINE AND PIE CHARTS FOR WIN PERCENTAGE

- *Objective:* To compare the win-loss ratios of IPL teams and analyze their overall performance.
- *Insights Expected:*
 - Identify teams with the highest win percentages over the seasons.
 - Highlight teams that struggle to secure wins compared to their total matches played.
- *Visualization:* Pie charts displaying the proportion of wins and losses for each team.

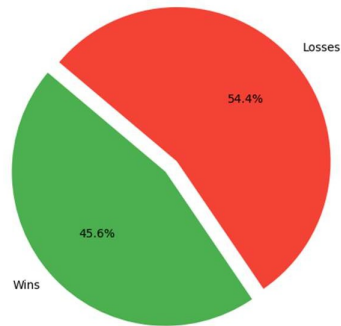
RCB - Wins vs Losses



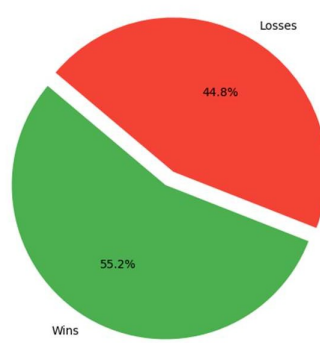
PBKS - Wins vs Losses



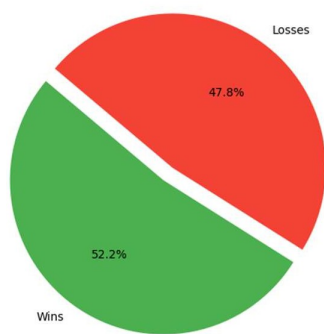
DC - Wins vs Losses



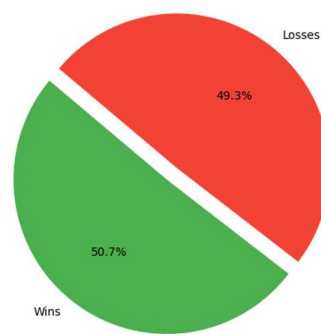
MI - Wins vs Losses



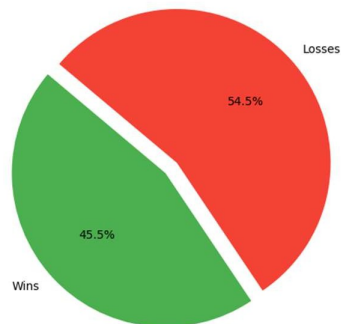
KKR - Wins vs Losses



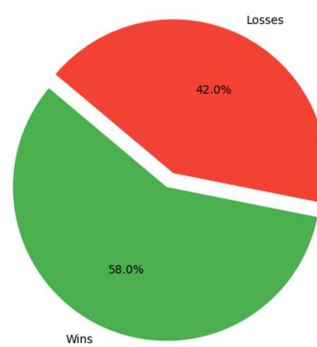
RR - Wins vs Losses



SRH - Wins vs Losses



CSK - Wins vs Losses



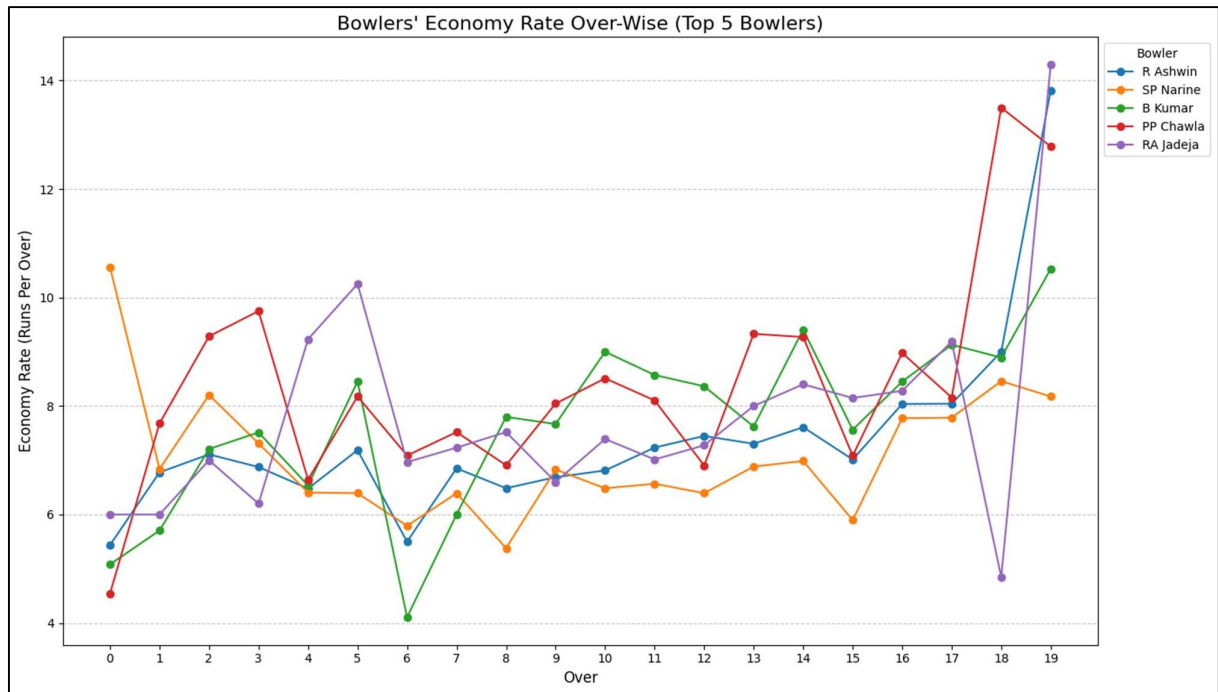


- The analysis reveals that teams like MI, CSK, and KKR have a higher win percentage, indicating strong performances across multiple seasons. On the other hand, newer teams or those with inconsistent squads tend to have a higher loss ratio.

7. BOWLER'S ECONOMY PER OVER

- *Objective:* To analyze the economy rate of the top 5 bowlers across different overs in IPL matches.
- *Insights Expected:*
 - Identify how a bowler's economy rate varies across different overs.

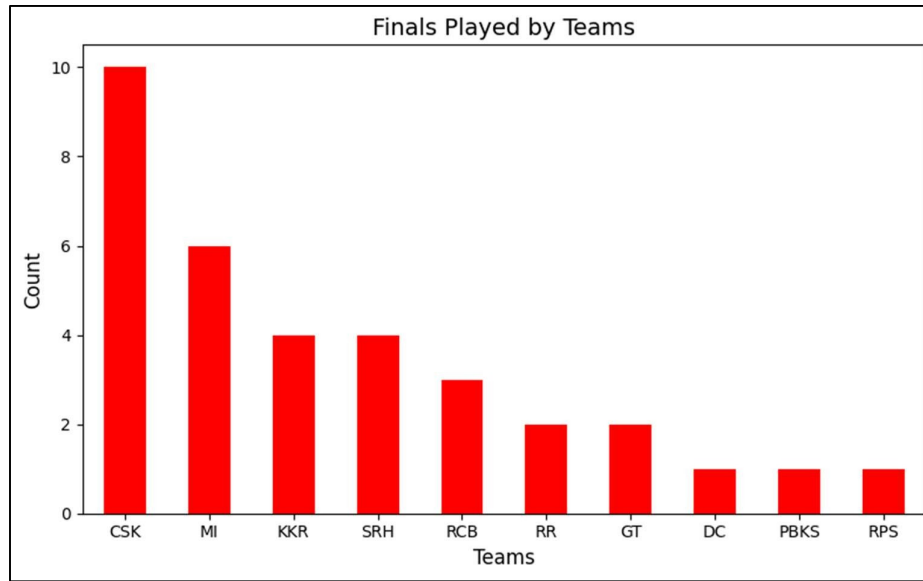
- Determine which bowlers maintain a low economy rate, especially in powerplay and death overs.
- *Visualization:* A line plot showing the economy rate trend for the top 5 bowlers over different overs.



- The analysis shows that economy rates are generally lower in the powerplay overs but tend to increase in the death overs. Some bowlers maintain a consistently low economy rate across all overs, highlighting their effectiveness in restricting runs.

8. MOST FINALIST OF IPL

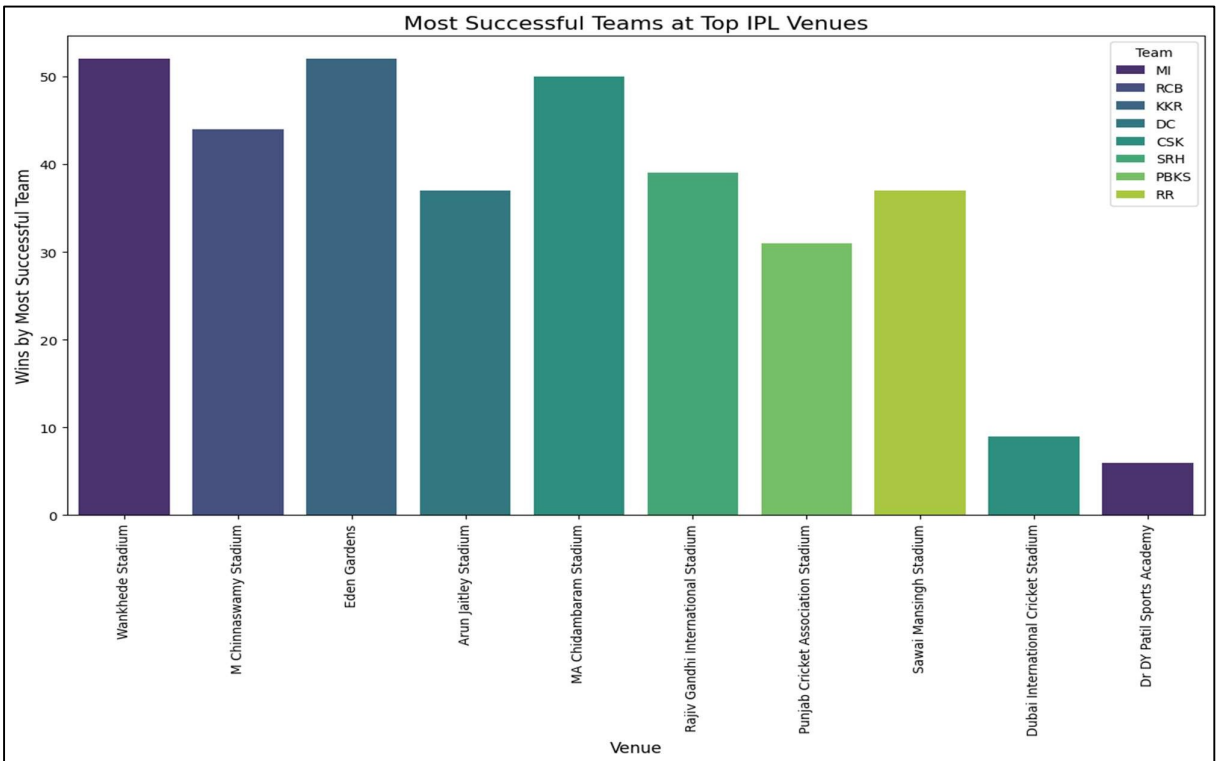
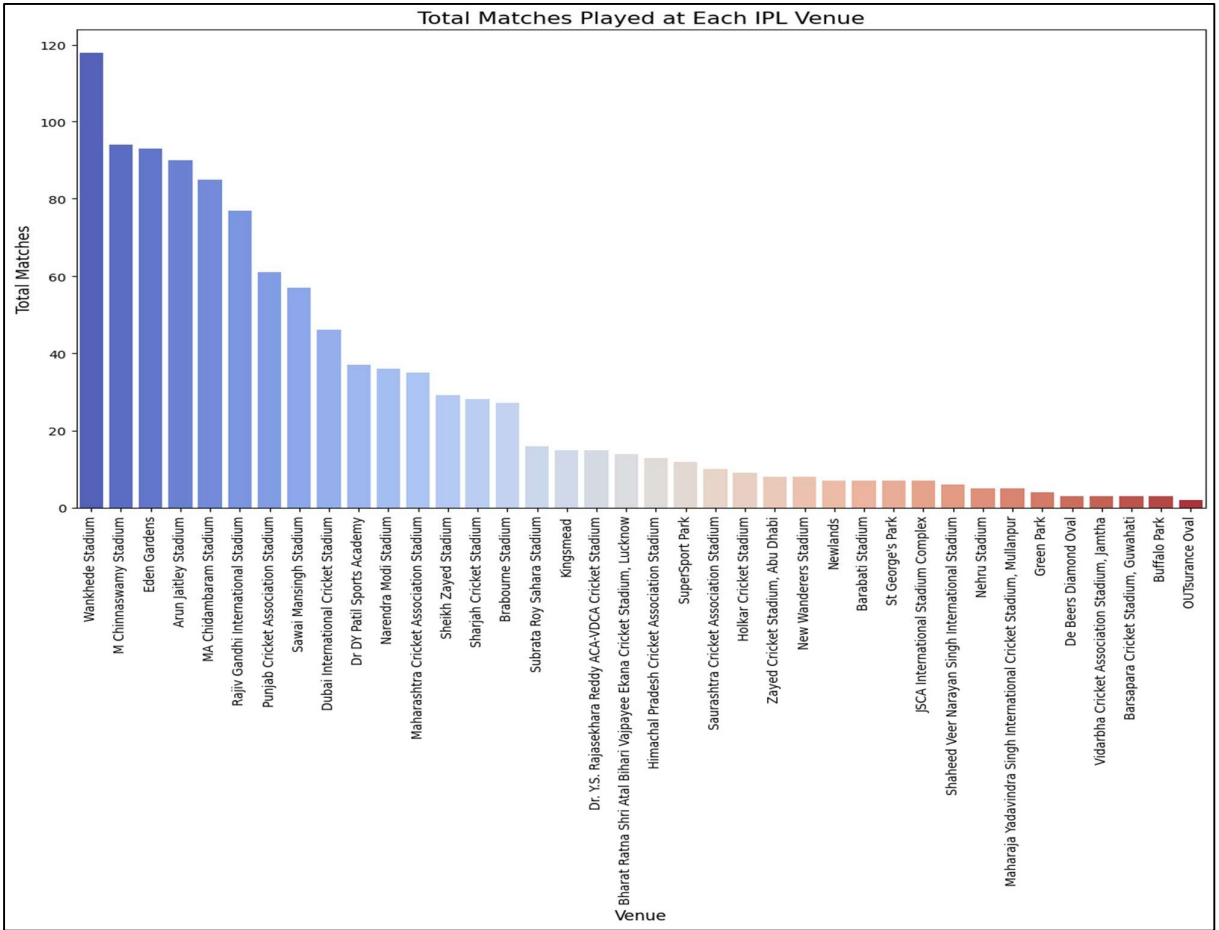
- *Objective:* To analyze the number of IPL final matches each team has participated in.
- *Insights Expected:*
 - Identify teams that have consistently reached the IPL finals.
 - Highlight dominant teams in the tournament's history.
- *Visualization:* A bar chart displaying the number of times each team has played in the IPL final.



- The analysis reveals that teams like CSK and MI have appeared in the most finals, showcasing their dominance in the league. Other teams have reached the finals occasionally but lack consistency compared to these top-performing franchises.

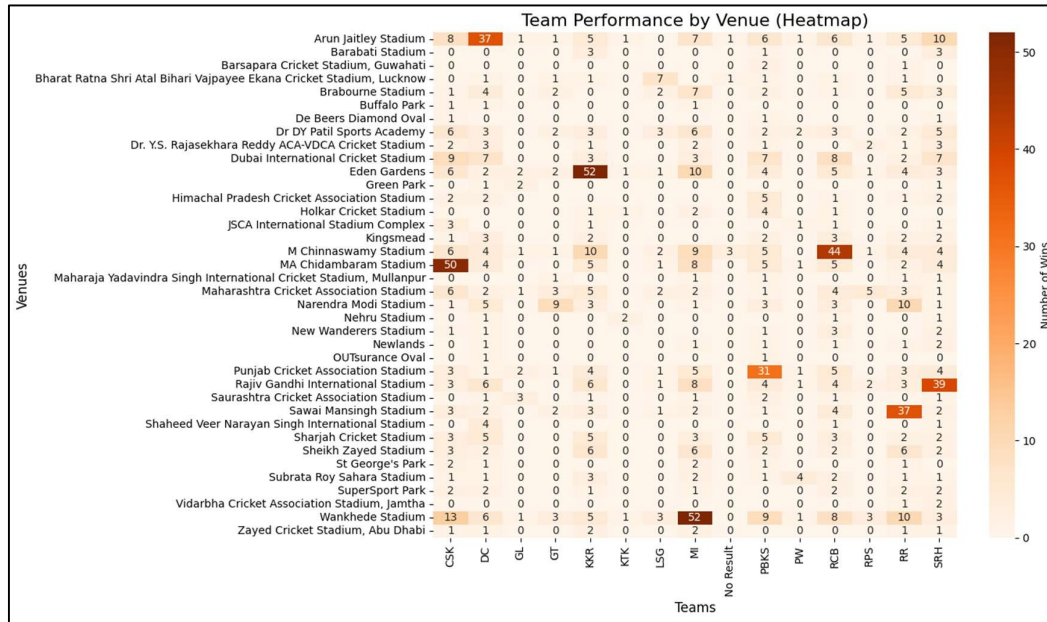
9. TOTAL MATCHES PLAYED AT A VENUE AND WIN COUNT AT EACH VENUE

- *Objective:* To analyze the impact of different IPL venues on match outcomes and identify the most successful teams at each location.
- *Insights Expected:*
 - Identify venues that have hosted the highest number of matches.
 - Determine which teams perform best at specific venues.
- *Visualization:*
 - A bar chart showing the total number of matches played at each venue.
 - A bar chart highlighting the most successful teams at the top 10 IPL venues.
- Wankhede Stadium, M. Chinnaswamy Stadium, and Eden Gardens have hosted the most IPL matches.
- Certain teams have strong home-ground advantages, with MI at Wankhede Stadium and CSK at MA Chidambaram Stadium dominating matches at their respective venues.



10. PERFORMANCE OF TEAMS AT EACH VENUE IN HEATMAP

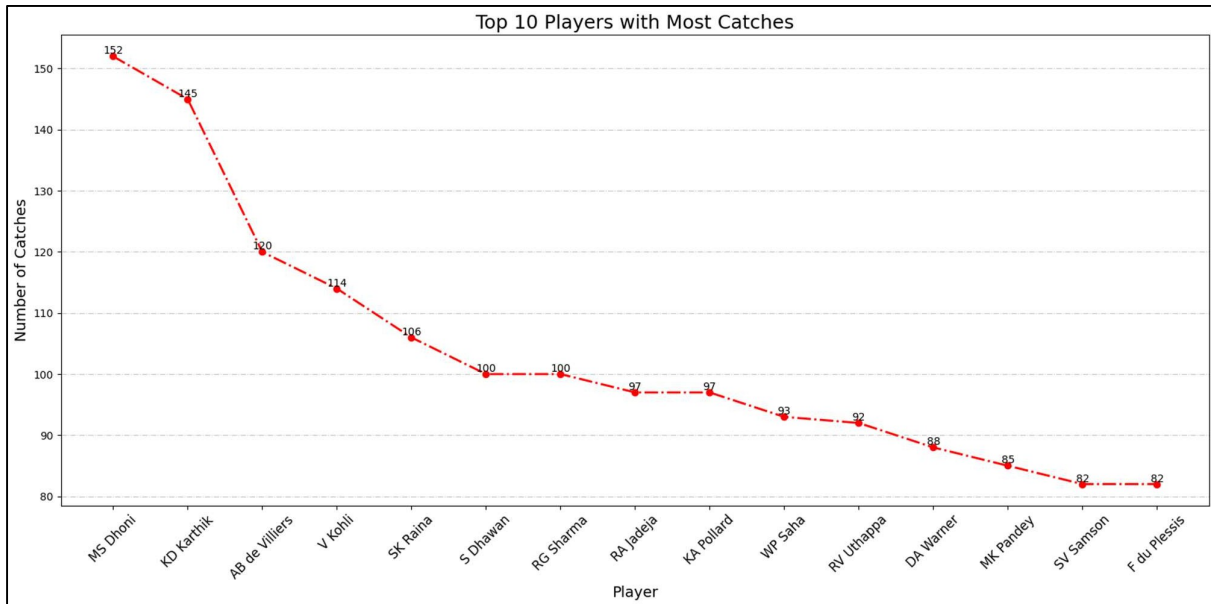
- **Objective:** To analyze how different IPL teams perform across various venues.
- **Insights Expected:**
 - Identify which teams dominate at specific venues.
 - Detect any venue-based advantages for particular teams.
- **Visualization:** A heatmap showing the number of wins for each team at different IPL venues.



- The analysis highlights strong home-ground advantages for teams like MI at Wankhede Stadium, CSK at MA Chidambaram Stadium, and KKR at Eden Gardens. Some teams perform consistently well across multiple venues, while others struggle outside their home grounds.

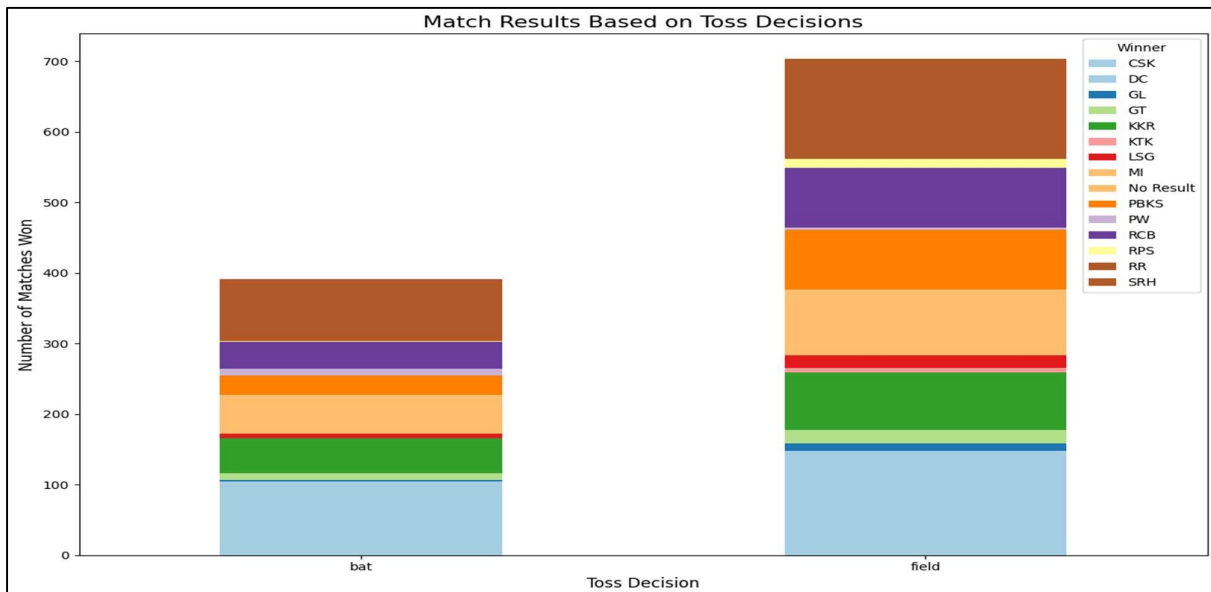
11. MOST CATCHES BY A PLAYER

- **Objective:** To analyze the top fielders in IPL history based on the number of catches taken.
- **Insights Expected:**
 - Identify the most reliable fielders in terms of catches.
 - Highlight players with exceptional fielding abilities and consistency.
- **Visualization:** A line plot displaying the top 10 players with the most catches in IPL history.
- The analysis reveals that players like MS Dhoni, Dinesh Karthik, Suresh Raina and AB de Villiers have consistently been among the top fielders, showcasing their agility and contribution to their teams' success through fielding.



12. MATCH RESULTS BASED ON TOSS DECISIONS

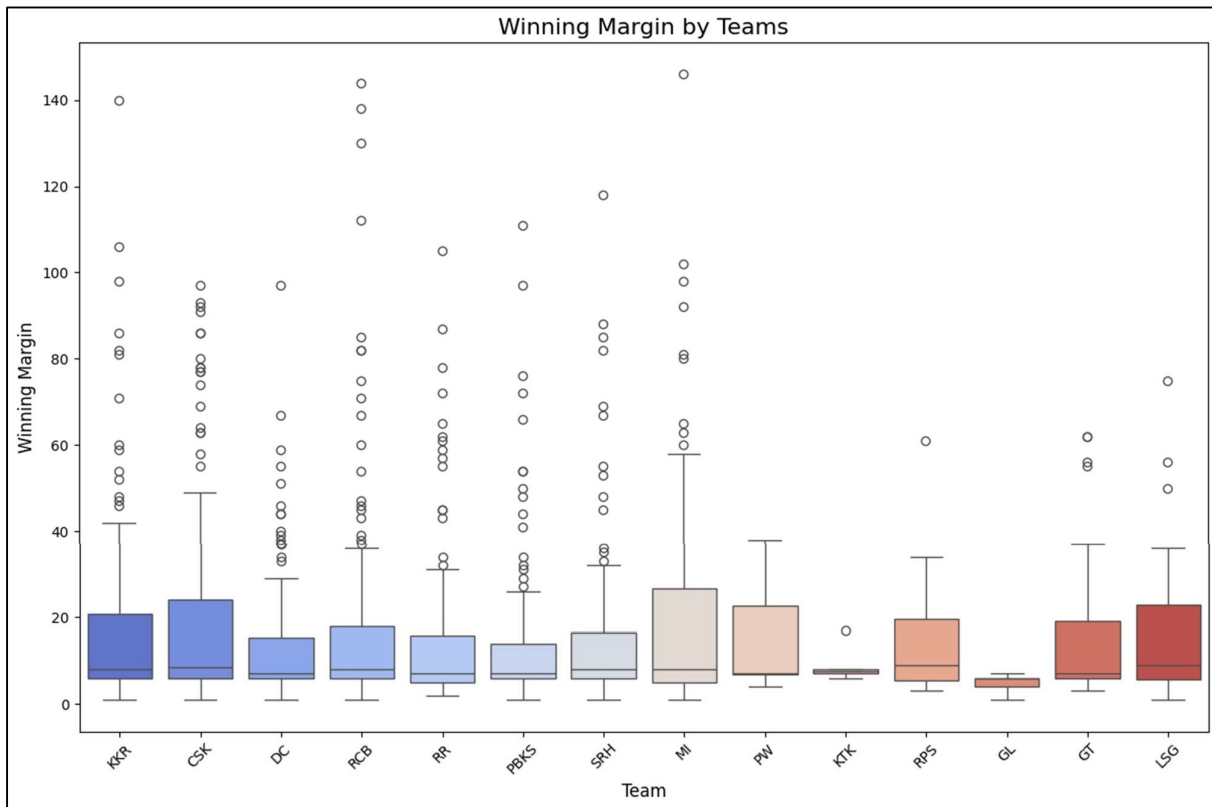
- **Objective:** To analyze the impact of toss decisions (batting or bowling first) on match outcomes.
- **Insights Expected:**
 - Determine whether teams winning the toss prefer batting or bowling.
 - Identify which toss decision leads to more match victories.
- **Visualization:** A stacked bar chart displaying match wins based on toss decisions.



- The analysis shows that teams generally prefer to bowl first, and in many cases, chasing teams have a higher success rate. However, certain teams, depending on pitch and match conditions, have excelled while setting a target as well.

13. WINNING MARGIN ANALYSIS

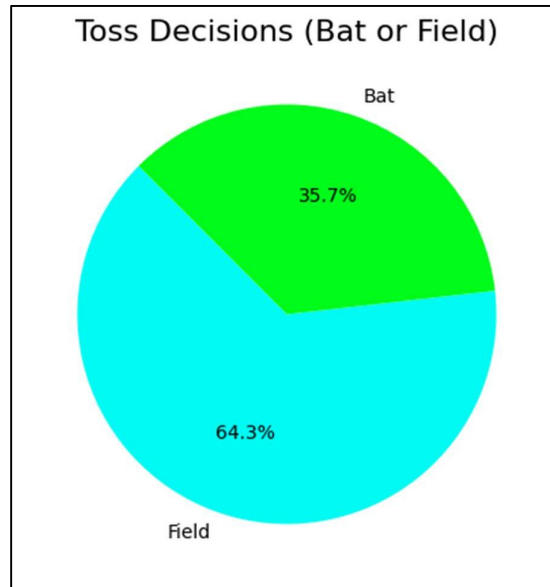
- *Objective:* To analyze the margin of victory for different IPL teams and identify dominant performances.
- *Insights Expected:*
 - Identify teams that consistently win by large margins.
 - Determine whether certain teams win more frequently by runs or wickets.
- *Visualization:* A box plot displaying the distribution of winning margins for each team.



- The analysis shows that teams like MI, CSK, and KKR have recorded some of the largest winning margins in IPL history. Some teams tend to win by small margins, indicating closely contested matches, while others have dominant victories.

14. TOSS BASED DECISIONS

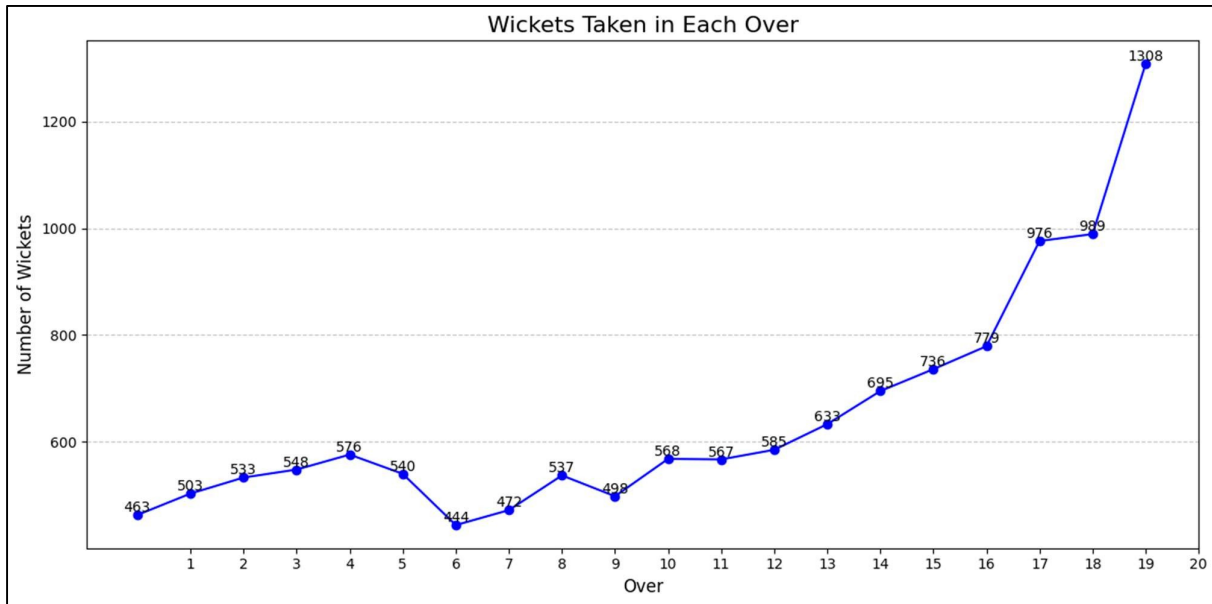
- *Objective:* To analyze the preference of teams when winning the toss in IPL matches.
- *Insights Expected:*
 - Identify whether teams prefer to bat first or field first after winning the toss.
 - Determine if there is a dominant trend in toss decisions across seasons.
- *Visualization:* A pie chart displaying the proportion of toss decisions (batting vs. fielding).



- The analysis reveals that a majority of teams prefer to field first after winning the toss, likely due to factors like dew conditions and better chasing success rates in T20 matches.

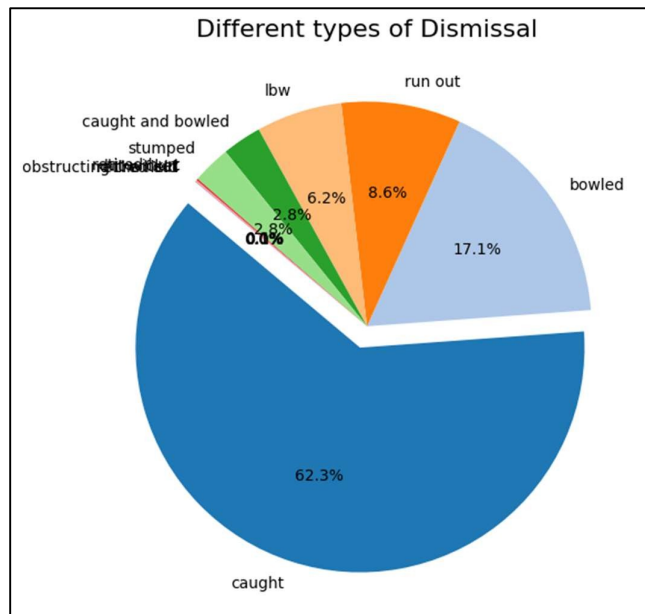
15. WICKETS TAKEN IN EACH OVER

- *Objective:* To analyze the distribution of wickets across different overs in an IPL match.
- *Insights Expected:*
 - Identify which overs see the most dismissals.
 - Determine if there are specific phases in the match where wickets tend to fall more frequently.
- *Visualization:* A line plot displaying the number of wickets taken in each over.
- The analysis reveals that wickets fall more frequently in the powerplay (overs 1-6) and death overs (16-20). The pressure of scoring quickly in the final overs often leads to more dismissals, while early overs see wickets due to the movement of the ball and aggressive field placements.



16. DIFFERENT TYPES OF DISMISSALS

- *Objective:* To analyze the various ways in which batsmen get out in IPL matches.
- *Insights Expected:*
 - Identify the most common modes of dismissal.
 - Understand the distribution of different types of dismissals in T20 cricket.
- *Visualization:* A pie chart displaying the proportion of each type of dismissal.



- The analysis shows that "Caught" is the most frequent mode of dismissal, highlighting the impact of fielding in T20 cricket. Other common dismissals include bowled, leg before wicket (LBW), and run out, while rare dismissals like hit wicket and obstructing the field occur occasionally.

CONCLUSION

The Indian Premier League (IPL) has evolved into one of the most dynamic and competitive T20 leagues in the world. Through this comprehensive data analysis, we have extracted valuable insights into team performance, player contributions, match strategies, and venue influences. The findings provide a deep understanding of how various factors impact the outcomes of IPL matches and offer data-driven perspectives on decision-making in cricket.

Our analysis of match outcomes and win-loss ratios has highlighted that teams like Mumbai Indians (MI), Chennai Super Kings (CSK), and Kolkata Knight Riders (KKR) have consistently performed well over multiple seasons. These teams have shown resilience in both setting and chasing targets, contributing to their dominance in the league. Newer teams like Gujarat Titans (GT) have also emerged as strong contenders, boasting an impressive win percentage in their initial seasons.

Examining scoring trends across overs, we found that the final overs (16-20) witness a significant increase in runs, primarily due to aggressive batting strategies in the death overs. This aligns with modern T20 tactics where teams focus on accelerating the run rate in the latter part of the innings. Similarly, wickets tend to fall more frequently in the powerplay (overs 1-6) and death overs (16-20), indicating that bowlers play a crucial role in disrupting the batting flow during these phases.

The economy rate analysis of top bowlers showed that maintaining a low economy rate throughout the innings is challenging, with economy rates generally increasing in the final overs. However, elite bowlers manage to control the run flow, particularly in the powerplay and middle overs. Our findings suggest that bowlers with better control and variation tend to excel in containing batsmen.

From a venue perspective, stadiums like Wankhede Stadium, M. Chinnaswamy Stadium, and Eden Gardens have hosted the most IPL matches. Some teams leverage their home-ground advantage, as seen with MI at Wankhede and CSK at MA Chidambaram Stadium, where they have secured the most wins. Additionally, toss analysis indicates that teams often prefer to field first, likely due to the advantage of chasing under dew conditions in night matches.

Our findings on fielding and dismissals revealed that "Caught" is the most common mode of dismissal, emphasizing the importance of sharp fielding and catching in the T20 format. Players like MS Dhoni, Dinesh Karthik, and Suresh Raina have proven to be exceptional fielders, taking a high number of catches over multiple seasons.

In conclusion, IPL data analysis has provided crucial insights into the league's patterns and trends. Understanding these dynamics helps teams refine their strategies, whether it be optimizing batting lineups, making tactical bowling changes, or leveraging home-ground advantages. The integration of statistical methods and visualizations allows for a more data-driven approach to cricket strategy, benefiting players, coaches, analysts, and fans alike. The ever-evolving nature of the IPL ensures that future seasons will continue to present new challenges and trends, making data analytics an essential tool for success in the modern game.

