**EDA PROJECT - CSM353**

**Exploratory Data Analysis of IPL Matches (2008-2024)**

**Academic Task Number 1**

Submitted to:

***Mr. Ved Prakash Chaubey***

***UID : 63892***



**Dated: 23.10.2024**

**SUBMITTED BY**

**Name of student: *Mahammed Sabeer Ali***

**Registration Number: *12200444***

# Acknowledgement

I would like to express my deepest gratitude to Lovely Professional University for providing me with the opportunity to undertake this project on Exploratory Data Analysis (EDA) in the domain of sports analytics. This project has significantly enhanced my knowledge of data science and its applications in real-world scenarios, specifically in sports performance analysis.

I extend my heartfelt thanks to my mentor, Mr. Ved Prakash Chaubey, for his invaluable guidance, support, and feedback throughout the duration of the project. His insights were crucial in refining the analysis and improving the overall quality of the project.

I am also grateful to my family and friends for their constant support and encouragement, which motivated me to push the boundaries of my knowledge and achieve the objectives of this project.

Finally, I acknowledge UpGrad and other open-source communities that provided access to valuable resources and datasets, enabling the successful completion of this project.

**Table of Contents**

1. **Abstract**

1. **Introduction**

1. **Methodology**

1. **Result and Analysis**

1. **Conclusion**

# Abstract

This project focuses on the Exploratory Data Analysis (EDA) of IPL (Indian Premier League) match data from 2008 to 2024. The primary aim is to identify patterns, trends, and insights within the dataset to enhance understanding of team performance, player statistics, and match outcomes. The analysis employs various data visualization techniques to discover relationships between key variables such as match outcomes, player performances, and team strategies. The findings of this study can provide actionable insights for cricket analysts, teams, and enthusiasts.

# Introduction

The Indian Premier League (IPL) is one of the most popular cricket leagues globally, drawing significant attention from fans, analysts, and teams alike. With over a decade of data available, there is a substantial opportunity to explore and analyze the performance of teams and players, identify trends, and make strategic decisions.

This project aims to perform a detailed Exploratory Data Analysis (EDA) on IPL data from 2008 to 2024. The dataset includes information on match results, team and player performances, and other match-related details. Through EDA, we seek to uncover hidden patterns, trends, and insights that can aid in better decision-making for teams and cricket analysts.

# Methodology

The project follows a structured approach to Exploratory Data Analysis, including:

**Data Collection and Preparation:**

* The dataset was sourced from Kaggle and consists of detailed records of IPL matches from 2008 to 2024. Initial data cleaning steps, such as handling missing values, removing duplicates, and standardizing formats, were performed.

**Univariate Analysis:**

* Analyzing individual features such as total runs scored, wickets taken, player performance metrics, etc., using descriptive statistics and visualizations like histograms, box plots, and bar charts.

**Bivariate and Multivariate Analysis:**

* Exploring the relationships between multiple variables (e.g., team performance vs. player statistics, win/loss ratio vs. home/away matches) using scatter plots, correlation matrices, and pair plots.

**Time Series Analysis:**

* Conducting a time series analysis to examine trends over different IPL seasons, such as changes in average scores, player form, and team performance over time.

**Data Visualization:**

* Utilizing libraries like Matplotlib and Seaborn to create visual representations of the data to identify key patterns and trends.

**Result Interpretation:**

* Interpreting the findings to derive actionable insights and make recommendations based on the analysis.

# RESULT AND ANALYSIS

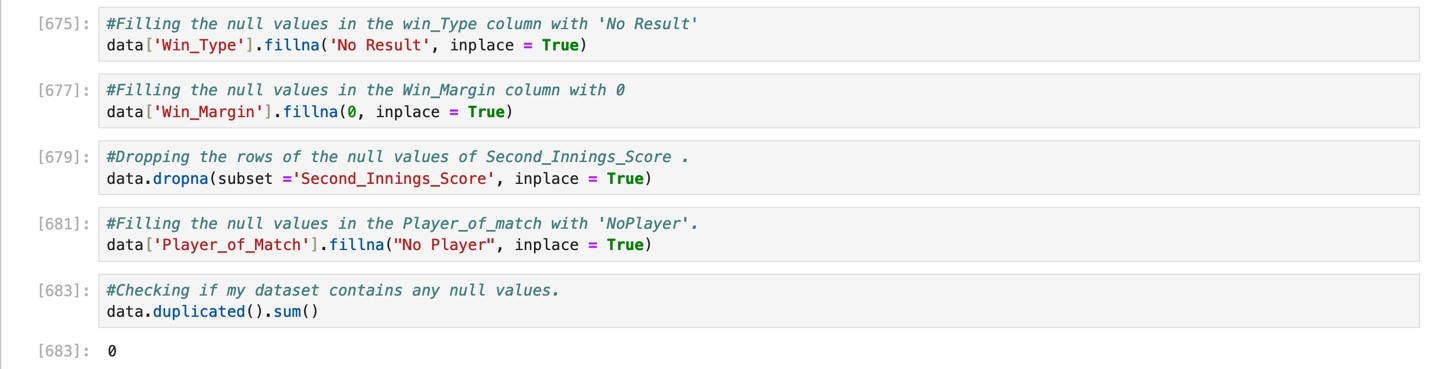
* Importing the libraries
* Loading the dataset
* Counting the columns and rows of the dataset  Seeing the columns of the dataset.



* Finding the null values percentage in each column of the dataset



* In 675 filling the nullvalues in the Win\_Type column with ‘no result’.
* In 677 filling the nullvalues in the Win\_Margin column with 0
* In 679 dropping the rows of the null values in the Second\_Innings\_Score column.
* In 683 finding the duplicate values in the dataset



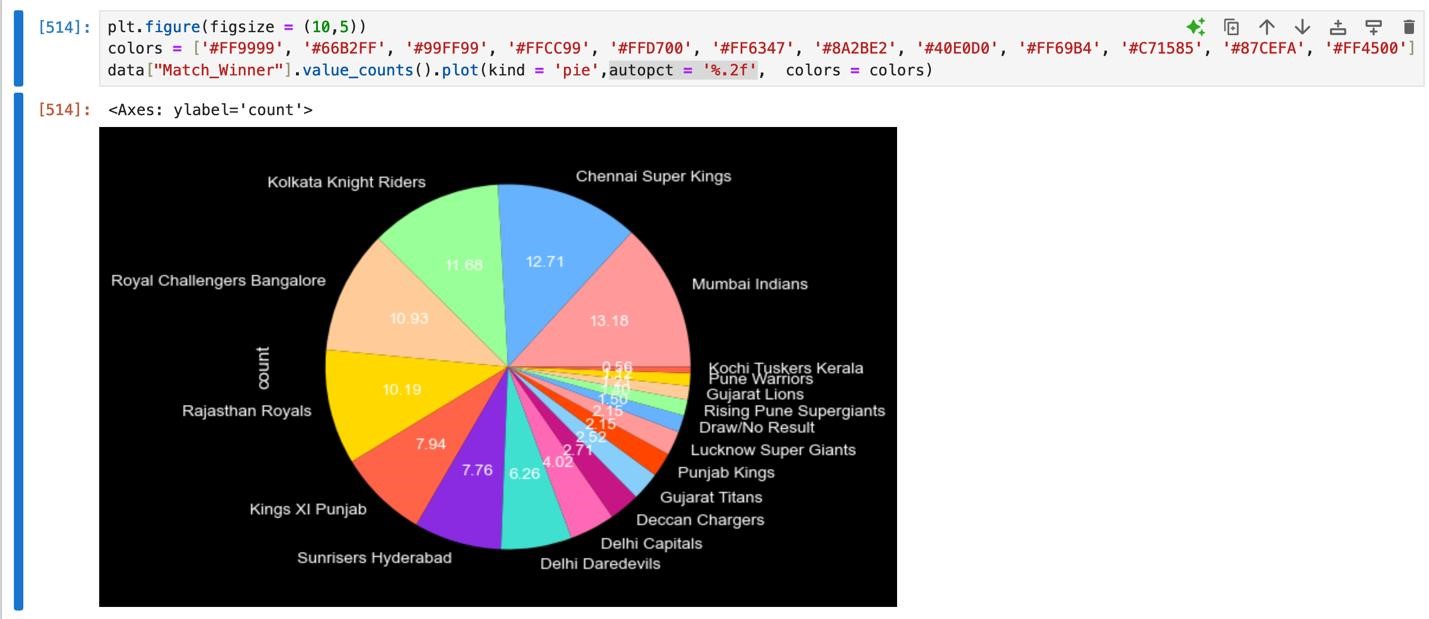
* Converting the date column to datetime
* Creating a new column Year from the Date column.



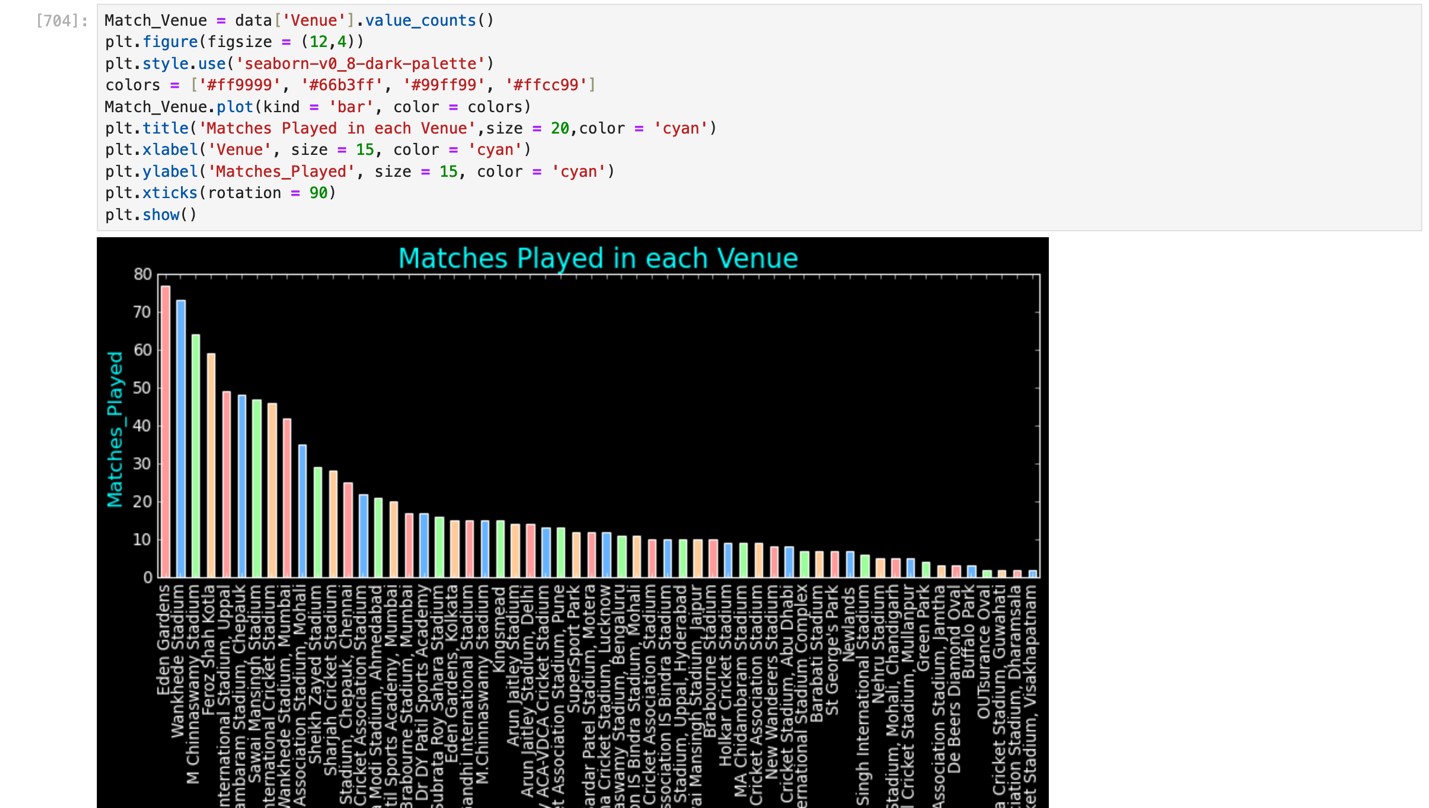
A screen shot of a computer

Description automatically generated

* Plotting a PieChart



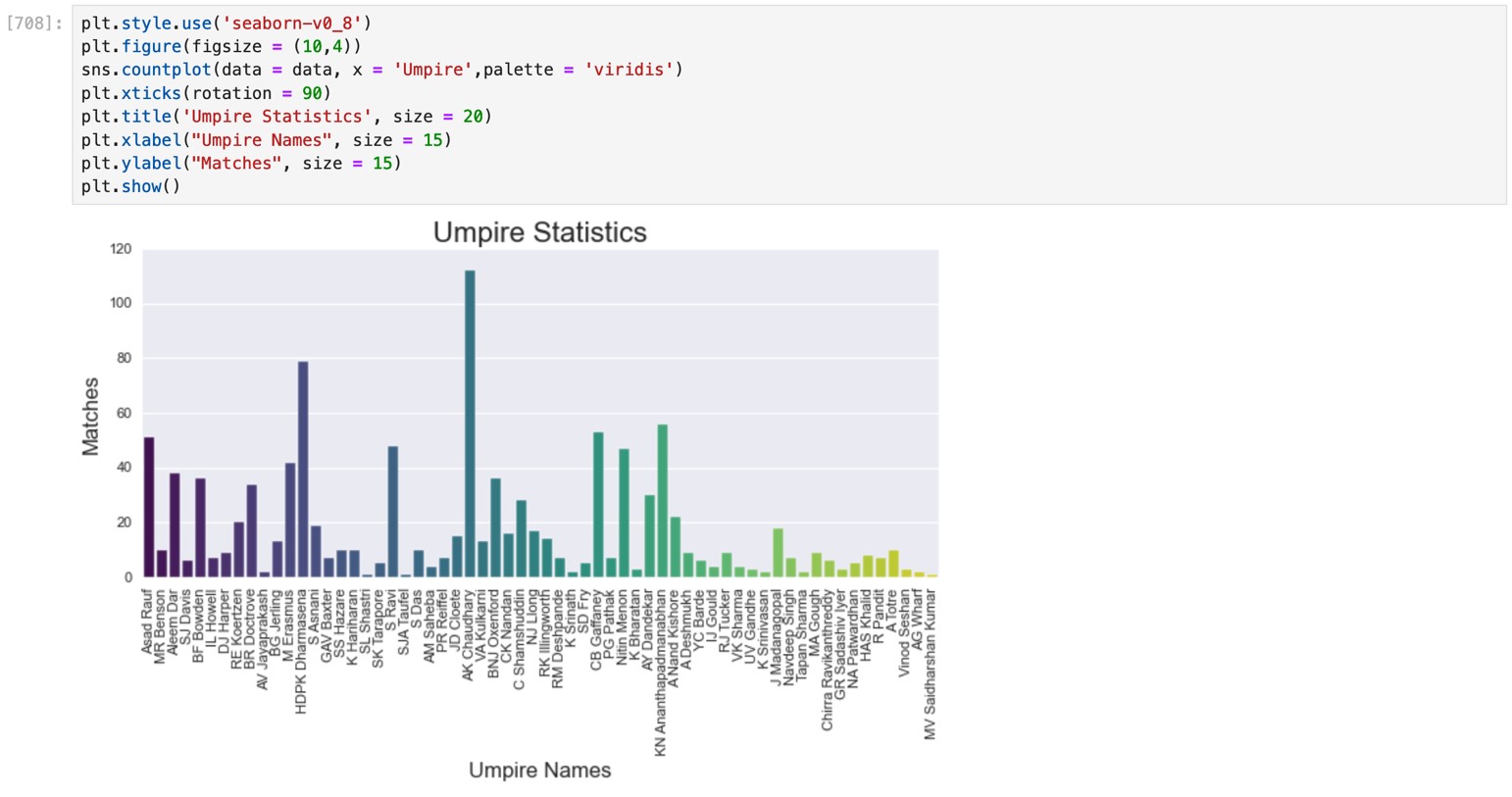
* Plotting a Bar plot to find the number of matches playes in each venue.



* Plotting histograms for Score distribution in Powerplay, Middle\_overs, Death\_overs



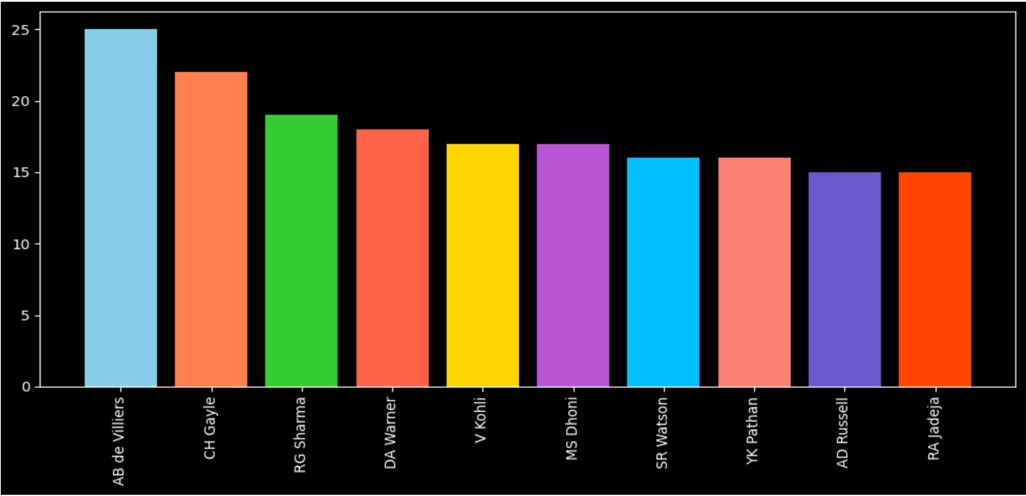
* Plotting Countplot to check the matches umpired by each umpire.



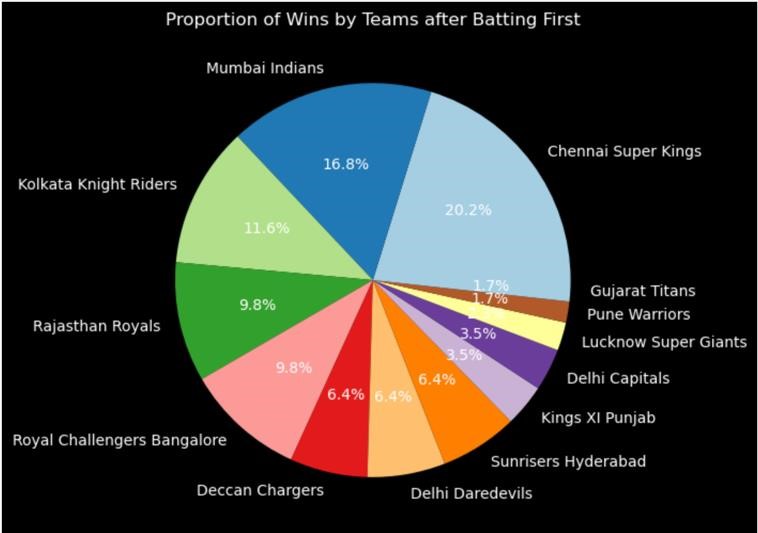
* Stacked bar plot to get Toss Decision by Team after Winning Toss



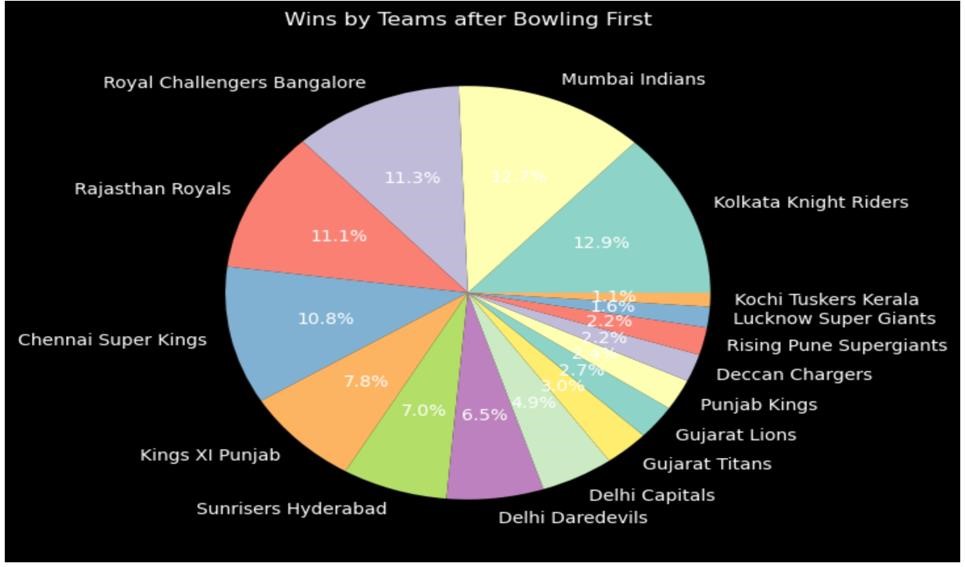
* **Top 10 players who have won most of the M.O.M awards**



* **Proportion of Wins by Teams after Batting First**



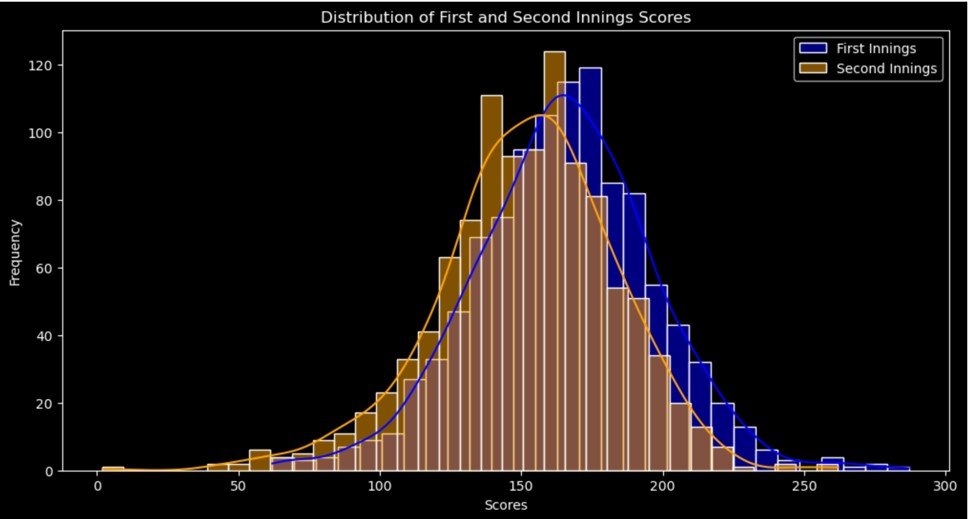
* **Proportion of Wins by Teams after Bowling**



* **Finding the correlation between first\_Innings\_score, Second\_Innings\_Score, Powerplay\_scores, middle\_over\_score, death\_over\_scores, win\_margin**



* **Distribution and Frequency of First and Second Innings Scores**

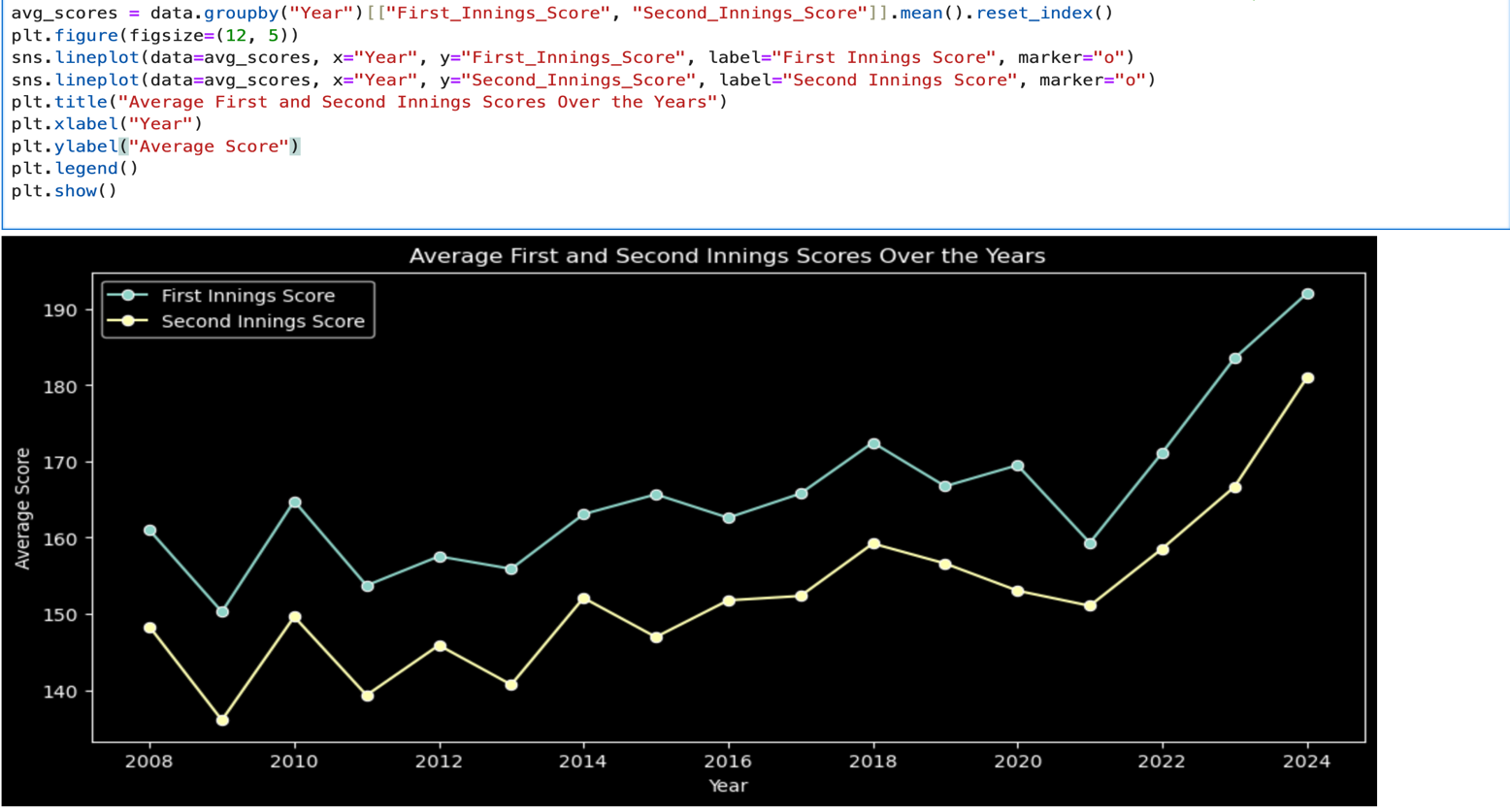


### Countplot to find the Toss Decision depeding on the Venue.

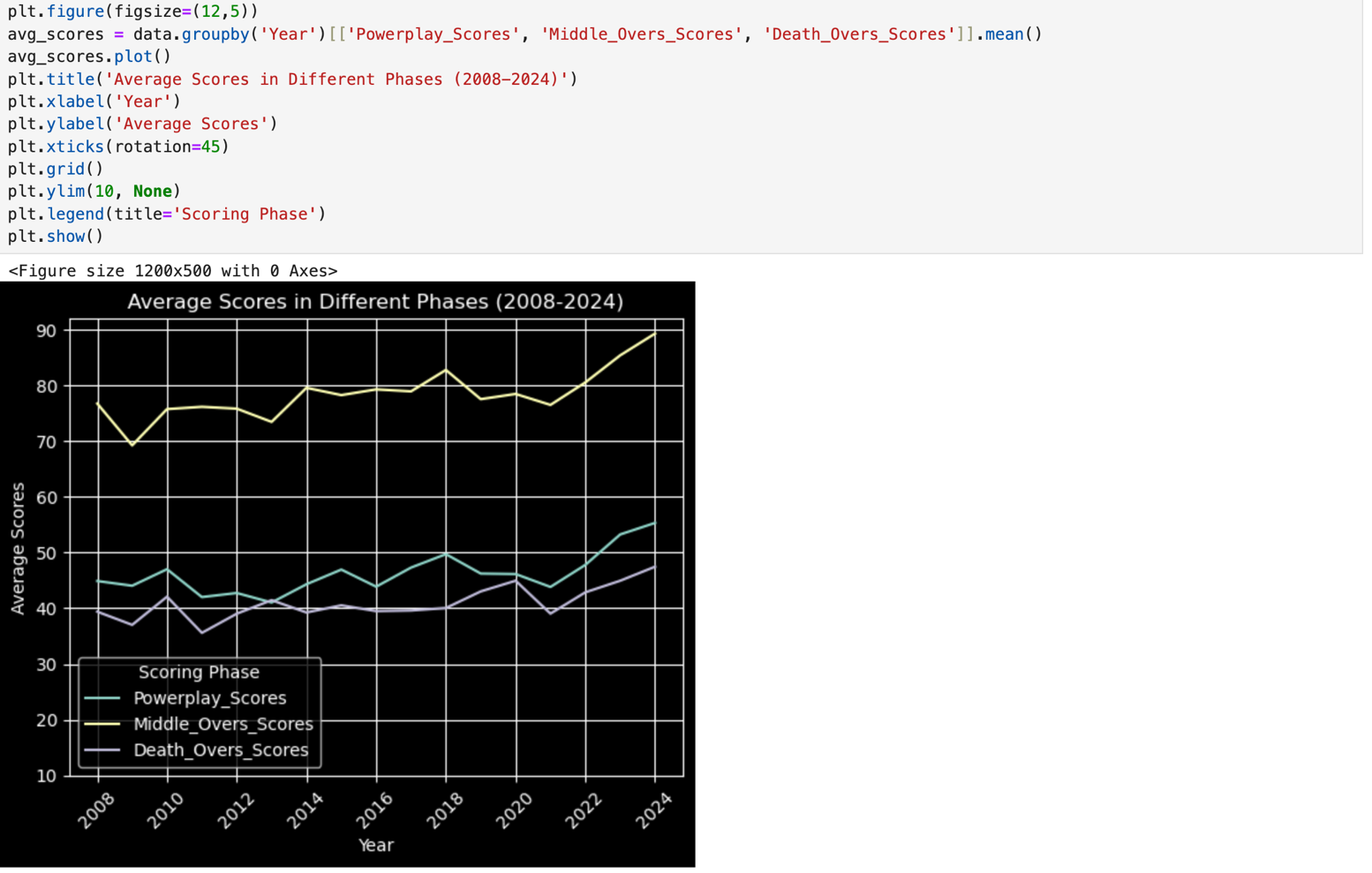
A screen shot of a computer

Description automatically generated

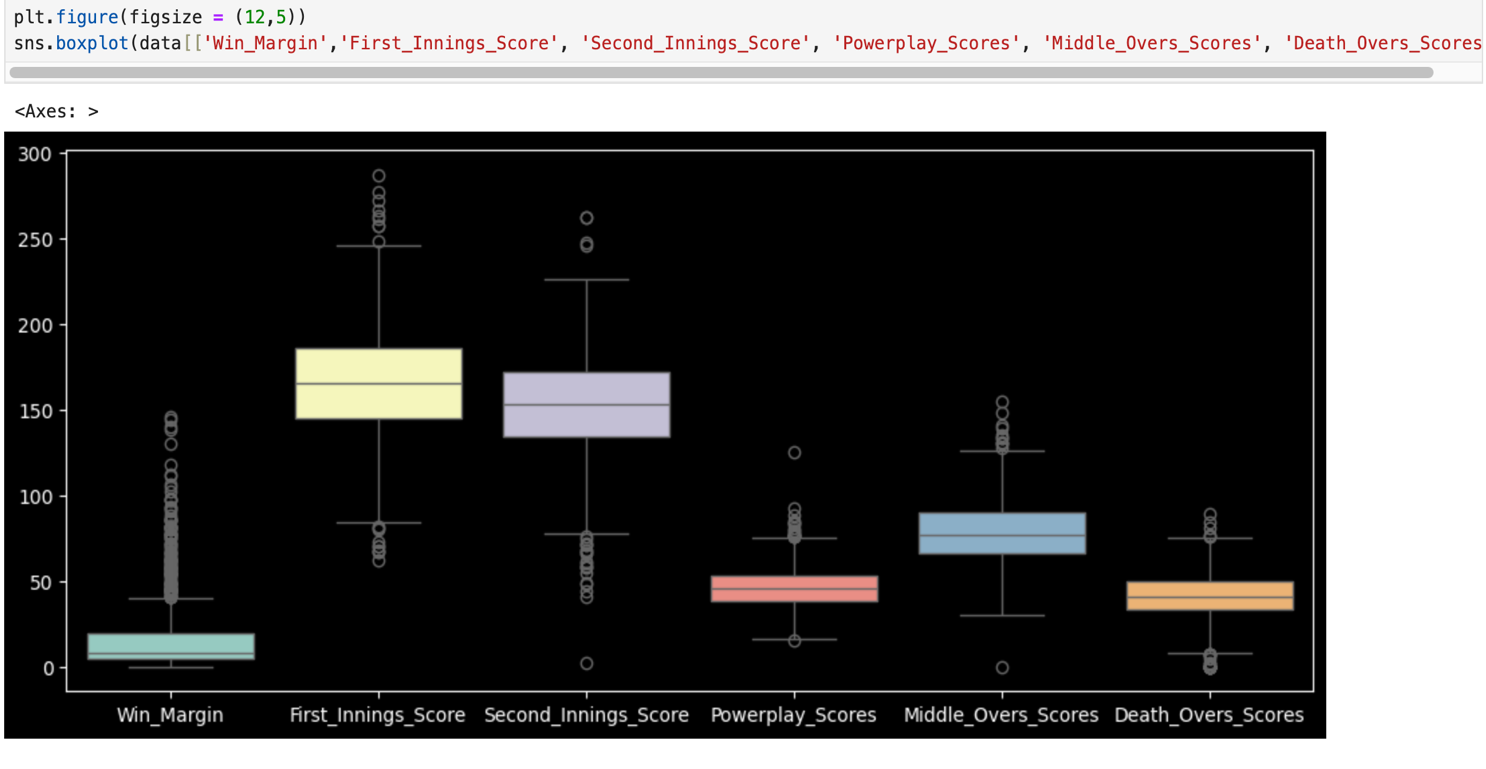
### Line Plot of Average First and Second Innings Scores Over the Years



### Line Plot of Powerplay\_scores, Middle\_over\_scores, Death\_over\_scores over years.



* Box Plot for finding the Outliers



1. **How many matches were played each year from 2008 to 2024?**

Sol:Each season typically features around 60-90 matches, with the number increasing as new teams were added.

1. **What are the top venues in terms of the number of matches hosted?**

Sol: Popular venues include Wankhede Stadium (Mumbai), Eden Gardens (Kolkata), and M. Chinnaswamy Stadium (Bangalore), which host numerous matches each season.

1. **How does the win percentage of teams vary across different venues?**

* Teams tend to perform better at home venues, and some venues have a bias favoring teams that bat first.

4. **What are the most common toss decisions (bat/field) over the years?**

* Initially, teams preferred batting first, but in recent seasons, chasing has become more popular due to improvements in chasing strategies.

5. **How often does the toss winner end up winning the match?**

* Around 55-60% of toss winners also win the match, with some variability depending on the venue and match conditions.

6. **How has the number of matches per season changed over time?**

* The number of matches has increased over time as more teams were added, especially from 2011 onwards with the introduction of additional teams.

7. **What are the most frequent match outcomes (win by runs or win by wickets)?**

* Wins by wickets (chasing) are more common, particularly in recent seasons, where teams have become adept at chasing down targets.

8. **What is the distribution of win margins (both in terms of runs and wickets)?**

* Common win margins range from 5-10 wickets for chases, and 20-50 runs for defending teams.

9. **Which teams have been the most successful in IPL history?**

* Mumbai Indians and Chennai Super Kings are historically the most successful IPL teams, winning multiple titles.

1. **Are there any trends in home vs away performance for teams?**

* Teams generally perform better at home venues, although successful teams like Mumbai and Chennai have consistent away records too.

Team Performance Analysis:

1. **Which teams have the best performance in the powerplay overs?**

* Teams like Chennai Super Kings and Kolkata Knight Riders are known for strong powerplay performances due to their aggressive openers.

12. **How do team scores vary between the first innings and second innings?**

* First-innings scores are often higher as teams tend to be cautious while chasing, leading to lower second-innings scores on average.

13. **What is the average winning margin for each team?**

* Winning margins vary; for teams like Mumbai Indians, the average winning margin while defending is typically around 30-40 runs.

14. **Which team has the highest average first innings score?**

* Teams like Mumbai Indians and Chennai Super Kings consistently post high first innings scores.

15. **How does the powerplay performance of teams correlate with match outcomes?**

* Teams with high powerplay scores are more likely to win, particularly if they score above 50 runs in the first six overs.

16. **How do teams perform in the death overs compared to the middle overs?**

* Teams that excel in death overs, such as Sunrisers Hyderabad, often make up for slow middle overs, especially in close matches.

17. **Which team has been the most consistent in setting high scores in death overs?**

* Mumbai Indians are known for explosive finishes in death overs, often adding 5070 runs in the last 5 overs.

18. **What are the highest-scoring matches across all seasons?**

* High-scoring games, such as the 2010 match between Chennai Super Kings and Kings XI Punjab, often exceed 400 combined runs.

19. **Which team performs better while chasing compared to defending a score?**

* Chennai Super Kings are known for their chasing abilities, with MS Dhoni excelling in finishing matches.

20. **How do different teams perform under pressure in knockout games?**

* Teams like Mumbai Indians have a solid record in knockout games, with their experienced players handling pressure well.

Toss and Match Outcomes:

21. **How frequently do teams choose to field after winning the toss?**

* In recent years, fielding first has become a more common choice, with about 6070% of toss winners opting to chase.

22. **Is there a correlation between toss decisions and match results?**

* Teams that choose to field first tend to win more often, particularly in day-night matches.

23. **Which toss decision (batting or fielding) has resulted in more wins over time?**

* Fielding first (chasing) has yielded more wins in recent seasons, especially under favorable conditions.

24. **Does the impact of toss decisions differ between teams?**

* Some teams, like Chennai Super Kings, tend to perform well regardless of the toss, while others rely more on winning the toss.

25. **Which venues show a bias towards a particular toss decision?**

* Venues like Wankhede Stadium in Mumbai often favor teams that chase, as conditions tend to improve for batting under lights.

26. **Are teams that choose to bat first more likely to win by runs?**

* Yes, teams that bat first are more likely to win by larger run margins when they post high scores.

27. **What is the win/loss record for teams that chose to field first?**

* Teams fielding first have a higher win rate, especially when chasing under lights.

Player Analysis:

28. **Who are the most frequently awarded ‘Player of the Match’ across seasons?**

* Players like AB de Villiers and Chris Gayle have frequently won 'Player of the Match' awards due to their match-winning performances.

29. **Which players have had the most impact on their teams’ success?**

* Players like MS Dhoni, Rohit Sharma, and Virat Kohli have been crucial to their teams’ success.

30. **How do individual performances vary across different venues?**

* Players tend to perform better at venues like Eden Gardens and Wankhede, which are known for batting-friendly conditions.

31. **Which players consistently perform well under pressure (knockout stages)?**

* MS Dhoni and Rohit Sharma are known for consistently strong performances in knockout matches.

32. **What are the top all-round performances in a match?**

* Players like Shane Watson and Andre Russell have delivered exceptional allround performances in both batting and bowling.

33. **How do the performance metrics differ for players batting first vs. chasing?**

* Players like Virat Kohli and MS Dhoni tend to have better stats while chasing compared to setting a target.

Umpire and Decision Analysis:

34. **Which umpires have officiated the most matches?**

* Umpires like Simon Taufel and Sundaram Ravi have officiated in many IPL matches.

35. **Is there a correlation between umpire decisions and match outcomes?**

* No significant correlation has been observed, though certain umpires may be associated with particular high-pressure games.

36. **How often do certain umpires officiate key knockout matches?**

* Experienced umpires like Simon Taufel are often assigned to key knockout matches due to their consistency and reputation.

37. **Do certain umpires show any bias towards teams or certain venues?**

* There is no clear evidence of umpire bias, although fans sometimes speculate on favoritism.

38. **What is the distribution of umpire presence across different IPL seasons?**

* A few senior umpires consistently appear across multiple seasons, while new umpires are rotated in occasionally.

Time Series Analysis:

39. **How have total scores in IPL matches evolved from 2008 to 2024?**

* Scores have generally increased as batting techniques and pitch conditions have improved.

40. **Are there any clear trends in winning margins over time?**

* Winning margins have become smaller as teams have become more skilled at chasing, leading to closer finishes.

41. **What is the trend of first innings scores over the years?**

* First-innings scores have gradually increased, with many matches now seeing 180+ runs scored in the first innings.

42. **Has the average run rate during powerplay overs increased or decreased over the seasons?**

* The average run rate in powerplay overs has increased, particularly with more aggressive opening batsmen.

43. **How have the strategies related to toss decisions changed over time?**

* Teams now prefer chasing more often due to better analysis of second-innings conditions.

1. **Is there any seasonality or trend in the number of matches won by wickets?**

* There is a slight increase in matches won by wickets in the latter half of the tournament due to chasing teams.

1. **Are there any specific years with unusually high or low win margins?**

* Some seasons have had closer matches, particularly in 2014 and 2021, where many games were decided by small margins.

Inning and Score Analysis:

46. **What are the typical winning first innings scores across seasons?**

* A typical winning first innings score is around 170-180 runs, with scores above 200 being more likely to result in wins.

47. **How does the margin of victory differ between low-scoring and highscoring matches?**

* Low-scoring matches often result in small victory margins, while high-scoring matches can have wider win margins.

48. **What is the distribution of scores in the powerplay, middle overs, and death overs?**

* Powerplay scores average around 45-55 runs, middle overs around 70-80, and death overs around 50-70 runs.

49. **What is the correlation between powerplay scores and total match scores?**

* A strong powerplay often leads to a high total score, especially if teams maintain their momentum through the middle overs.

50. **How often do teams chase down high targets successfully?**

* Teams successfully chase high targets (180+) around 40-50% of the time, with the success rate increasing in recent years.

Venue and Geographic Analysis:

51. **Which venues tend to favor high-scoring matches?**

* Venues like Wankhede Stadium and M. Chinnaswamy Stadium are known for high-scoring matches due to flat pitches.

1. **What are the top venues for close finishes (wins by a small margin)?**

* Eden Gardens and Feroz Shah Kotla have witnessed several close finishes with teams winning by narrow margins.

1. **Are certain venues more likely to produce matches decided by a significant win margin?**

* Venues like Rajiv Gandhi Stadium (Hyderabad) often produce larger win margins due to challenging pitch conditions.

54. **Which venues have a bias towards teams winning after batting first?**

* Venues like Chepauk Stadium (Chennai) often see teams win by defending totals due to spin-friendly conditions.

55. **Which venues have the lowest average win margins?**

* Small grounds like Sharjah have seen low win margins due to the ease of chasing high scores.

Miscellaneous:

56. **What are the largest margins of victory in IPL history?**

* Some of the largest margins include 100+ runs in matches where teams posted massive scores and bowled the opposition out cheaply.

57. **Which teams tend to perform better in high-pressure matches (finals/semifinals)?**

* Teams like Mumbai Indians and Chennai Super Kings have a history of performing well in high-pressure matches.

58. **What are the records of individual players for the most 'Player of the Match' awards across different venues?**

* AB de Villiers and Chris Gayle have consistently won 'Player of the Match' awards at venues like Bangalore.

59. **Are there any patterns in match outcomes when certain umpires are officiating?**

* No specific umpire-related patterns are consistently observed, though experienced umpires tend to officiate crucial matches.

60. **How do team performances change based on match location (home vs away games)?**

* Teams generally perform better at home, but top teams like Mumbai Indians and Chennai Super Kings have strong away records as well.

1. **How does a team's performance change when playing under lights compared to day matches?**

* Teams often perform better under lights due to cooler conditions and pitch behavior, favoring batting.

1. **Which team has the best record in the IPL playoffs?**

* Mumbai Indians have the best playoff record, consistently making deep runs and winning titles.

1. **How do teams perform when chasing a target versus setting one?**

* Chasing is generally more successful in recent years, with teams adapting better to pressure situations.

1. **How do teams with the highest number of sixes in a season perform in the tournament?**

* Teams hitting the most sixes tend to have strong batting line-ups but need balance in bowling and fielding to succeed.

1. **Which teams are the most successful in the last five overs of a match?**

* Teams like Mumbai Indians and Chennai Super Kings excel in the last five overs, particularly due to experienced finishers.

1. **How does the batting order influence a team’s chances of winning?**

* A solid batting order with aggressive openers and capable finishers gives teams a better chance of setting or chasing high totals.

1. **What impact does the presence of foreign players have on a team’s performance?**

* Foreign players often provide valuable expertise, especially in crucial roles like finishing, bowling, or spin attack.

1. **Which teams have the highest bowling economy rates in the IPL?**

* Chennai Super Kings and Mumbai Indians often maintain low bowling economy rates due to effective line and length strategies.

1. **How do teams with strong bowlers compare in matches held in high-scoring venues?**

* Strong bowlers struggle in high-scoring venues but can still influence games with key wickets during critical phases.

1. **How often do teams win after posting a score below 150?**

* Winning after posting under 150 is rare but still possible with strong bowling attacks and tight fielding.

1. **Which team is most likely to make a comeback after a poor start in a match?**

* Teams like Kolkata Knight Riders and Sunrisers Hyderabad have a history of bouncing back due to strong middle-order performances.

1. **What is the average number of wickets taken by the top bowlers in each season?**

* Top bowlers usually take around 20-25 wickets in a season, with some exceeding that in standout performances.

### Player Analysis:

1. **Which players have the most IPL centuries?**

* Chris Gayle, AB de Villiers, and David Warner are among the players with the most centuries in IPL history.

1. **How do players like AB de Villiers and Virat Kohli compare in terms of batting average over the years?**

* Both players maintain high batting averages, with AB de Villiers being a more aggressive scorer and Kohli consistent across seasons.

1. **What is the average strike rate for top-order batsmen across seasons?**

* Top-order batsmen typically have strike rates above 130, with some aggressive openers going beyond 140.

1. **Which bowlers have the most 4-wicket hauls in IPL history?**

* Lasith Malinga, Sunil Narine, and Kagiso Rabada are known for their multiple 4-wicket hauls in IPL.

1. **What are the most effective partnerships between batsmen in IPL history?**

* Partnerships like Virat Kohli and AB de Villiers for Royal Challengers Bangalore have set records for consistent run-making.

1. **Which players have the most runs in knockout matches?**

* Players like MS Dhoni and Rohit Sharma shine in knockout matches, leading their teams with match-winning knocks.

1. **How does the performance of young players (under 25) compare to experienced players in the IPL?**

* Young players often bring energy and surprise performances, but experienced players tend to handle pressure better in crucial moments.

1. **Who has the highest batting average in IPL history?**

* Virat Kohli holds one of the highest career batting averages in the IPL due to consistent performances across seasons.

1. **Which bowlers have the best average and strike rate in death overs?**

* Death over specialists like Dwayne Bravo and Jasprit Bumrah maintain exceptional averages and strike rates in the final overs.

1. **How often do players perform better in the second innings compared to the first?**

* Players often perform better in the second innings due to a clearer understanding of pitch conditions and match context.

1. **Which players have the highest number of sixes in IPL history?**

* Chris Gayle and AB de Villiers lead in sixes, known for their aggressive hitting and match-changing capabilities.

1. **Who are the most consistent performers in IPL playoffs and finals?**

* Players like Rohit Sharma and MS Dhoni consistently deliver in IPL playoffs and finals due to their experience.

1. **Do teams with a higher win percentage in the regular season also perform well in the playoffs?**

* Teams with strong regular-season records, like Mumbai Indians and Chennai Super Kings, often carry that form into the playoffs.

1. **What is the impact of choosing to bat first in a high-pressure knockout match?**

* Batting first can put pressure on chasing teams, but it depends on the conditions and the target set.

1. **How often do teams opt to field first in high-humidity conditions?**

* In high-humidity conditions, teams tend to field first to take advantage of the dew factor in the second innings.

1. **Which teams are more successful when they lose the toss?**

* Teams like Kolkata Knight Riders and Mumbai Indians have a strong record even when they lose the toss.

1. **Do teams that win the toss and field first tend to have a higher win rate in rainy conditions?**

* Fielding first in rainy conditions helps as the pitch behavior remains more predictable during the chase.

1. **What are the most common strategies employed by teams that consistently win the toss?**

* Teams often choose to field first to assess the target, especially in day-night matches where dew can be a factor.

### Venue & Geographic Analysis:

1. **Which venues tend to produce the most one-sided matches?**

* Wankhede Stadium and M. Chinnaswamy Stadium often host high-scoring matches, which can result in one-sided contests.

1. **How does the home ground advantage affect teams' performances at specific venues?**

* Teams generally perform better at their home ground due to familiarity with pitch conditions and crowd support.

1. **Which venues see the most runs scored in the death overs?**

* Wankhede and Eden Gardens are known for high-scoring death overs due to their flat pitches and boundaries.

1. **How does the IPL's match schedule affect venue attendance and fan engagement?**

* Matches on weekends and holidays tend to have higher attendance, especially in cities with a strong fan base.

1. **How do teams perform at neutral venues compared to home venues?**

* Teams generally perform worse at neutral venues, as they lack the advantage of home crowd and pitch familiarity.

1. **Which venues are known for producing the most exciting finishes in IPL history?**

* Eden Gardens and Feroz Shah Kotla are famous for their thrilling finishes and last-over finishes.

1. **What is the average win rate for teams playing away matches in cities with high humidity?**

* Teams tend to struggle in high-humidity cities unless they are well-prepared for dew and pitch conditions.

### Umpire & Decision Analysis:

1. **How often do close decisions by umpires affect the outcome of a match?**

* Close umpire decisions can influence match outcomes but are generally neutralized by team performance.

1. **Are there patterns in the number of no-balls and wide deliveries called by certain umpires?**

* There are no significant patterns in no-balls or wide calls, but experienced umpires are less likely to make errors in high-pressure situations.

1. **How do decisions related to the DRS impact match outcomes in IPL?** - DRS decisions can influence match outcomes, especially in tight situations where the outcome may hinge on close LBW or caught behind calls.

# Conclusion

The EDA project successfully uncovered several valuable insights into IPL matches, teams, and players. By analyzing match data from 2008 to 2024, we identified critical factors that influence match outcomes, such as team strategies, player performance, and the impact of winning the toss. The findings provide a deeper understanding of the game dynamics and can assist teams and analysts in making informed decisions.