**INDIAN PREMIER LEAGUE (2008-2020)**

**A PROJECT REPORT**

**Submitted by**

**Bestha Sahana(1RVU23CSE106)**

**Anushka R Rao(1RVU23CSE068)**

**Ananya M(1RVU23CSE050)**

**Anmol B Rao(1RVU23CSE062)**

*in partial fulfillment for the award of the degree of*

**B.Tech.(Hons) – Computer Science & Engineering**

*in* **Data Analysis with Python**

**to Dr. Sahana Prasad**



**School of Computer Science and Engineering**

**RV University**

**RV Vidyaniketan,8th Mile, Mysore Road, Bengaluru, Karnataka, India – 562112.**

**December 2023**

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
|  | **INDIAN PREMIER LEAGUE(2008-2020)** |  |
| **1.0** | **INTRODUCTION** | **3** |
|  |  |  |
|  |  |  |
| **2.0** | **METHODOLOGY** | **4** |
|  |  |  |
|  |  |  |
| **3.0** | **INFERENCE** | **5** |
|  |  |  |
|  |  |  |
| **4.0** | **RESULT AND DISCUSSION** | **6-22** |
|  |  |  |
|  |  |  |
| **5.0** | **CONCLUSION** | **23** |

INDIAN PREMIER LEAGUE (2008-2020)

**1.INTRODUCTION**

The Indian Premier League (IPL) has not only revolutionized the landscape of cricket but has also generated a vast amount of data that holds valuable insights into player performance, team statistics, and the dynamics of the game. analyzing the IPL dataset presents an exciting opportunity to delve into the numbers behind the on-field action and uncover patterns that can influence decision-making in cricket.

The IPL dataset encompasses a wide array of information, including player statistics, team performance metrics, match outcomes, and more. This wealth of data provides a comprehensive view of the league’s history, allowing analysts and enthusiasts to explore trends, identify key players, and gain a deeper understanding of the factors contributing to success or challenges within the tournament.

The period from 2008-2020 covers the first thirteen seasons of the IPL, spanning a diverse range of cricketing conditions, team dynamics, and player performance. Analyzing the data generated during these years can provide valuable insights into various aspects of the game, including team strategies, player contribution, and overall trends in the tournament.

The primary objective of this data analysis project is to delve into the wealth of data generated by the IPL seasons from 2008-2020 and extract meaningful insights. Through statistical analysis, visualizations, and trend identification, we aim to uncover patterns, correlations, and key performance indicators that define successful teams, players and strategies in the IPL.

INDIAN PREMIER LEAGUE (2008-2020)

**2. METHODOLOGY**

**1.Data Collection:**

* Gathered comprehensive IPL data including match results, player statistics ,team information , venue etc. from a source called **kaggle.com**.

**2.Data cleaning and Preparation:**

* Handled missing data or inconsistent data.
* Identified the null values and dropped it.
* Removed the duplicates in the data.
* Converted the raw data into a format which is suitable for data analysis.

**3.Data Analysis:**

* Used statistical methods to explore the trends in the dataset.
* Identified outliers and anomalies in the dataset.

**4.Hypothesis Testing:**

* Formulated hypotheses based on the initial observations.
* Used statistical tests to validate or reject the hypotheses.

**5. Data Visualization:**

* Created informative and visually appealing charts and graphs to represent the findings.

**6.Interpretation and Conclusions:**

* Summarized key findings.
* Drew conclusions about team and player performance over the specified period.

INDIAN PREMIER LEAGUE (2008-2020)

**3. INFERENCE**

Analyzing the IPL dataset revealed notable trends, with certain players consistently exhibiting outstanding performance, influencing their team’s success. The strategies, player composition and batting orders, emerged as crucial factors affecting overall performance. The impact of winning the toss varied, indicating the importance of adapting strategies based on match conditions. This comprehensive exploration underscore the intricate dynamics of the IPL, highlighting the significance of individual brilliance , strategic management , and adaptability in achieving success in one of the world’s premier cricket leagues.

Investigating the evolution of team strategies over the years, including changes in batting orders, bowling rotations, and fielding tactics. Examining the impact of toss decision on match results and identifying trends in team decision making during critical moments.

Analyzing trends in player actions, including the correlation between player salaries and on-field performance. understanding the financial dynamics of terms, including patterns in team spending and its reflection on on-field success

Identifying and evaluating breakthrough players in each session, assuring their impact on team performance. Comparing the performance of debutantes with established players, providing insights into the emergence of new cricketing talent.

Deriving insights that can be applied for future IPL seasons, aiding terms and players in refining their strategies. Providing the data-driven perspective for cricket enthusiasts, analysts and term management to enhance their understanding of the game.

By addressing these challenges and questions the IPL dataset analysis project aims to contribute valuable insights to the cricketing community, fostering a deeper understanding of the game’s dynamics and providing a foundation for data driven decision making in feature IPL sessions.

INDIAN PREMIER LEAGUE (2008-2020)

**4. RESULT AND DISCUSSION**

* Through the ongoing analysis of the IPL dataset, we ascertain that a total of **816 matches** were played in the Indian Premier League from 2008 to 2020.
* The ongoing analysis of the IPL dataset has provided us with the names of the teams participating in the tournament .

**Teams participating:** 'Royal Challengers Bangalore' 'Kings XI Punjab' 'Delhi Daredevils' 'Mumbai Indians' 'Kolkata Knight Riders' 'Rajasthan Royals' 'Deccan Chargers' 'Chennai Super Kings' 'Kochi Tuskers Kerala' 'Pune Warriors' 'Sunrisers Hyderabad' 'Gujarat Lions' 'Rising Pune Supergiant' 'Rising Pune Supergiant' 'Delhi Capitals' .

* Themean of result margin in the Indian Premier League (IPL) across all matches from 2008 to 2020 is 17.32.
* The **median of result margin** by all teams in the Indian Premier League (IPL), considering all matches from 2008 to 2020 is 8.0

* **The mean is a better measure to calculate the average of total runs in the IPL because it considers every run scored, providing a balanced representation of the overall batting performance, whereas the median may be influenced by extreme values from a few exceptional innings.**

### The range is a basic measure of dispersion that provides a quick assessment of how spread out the values in a dataset are. The Number of batsman runs in the IPL Match is 6.

### The variance of total scores in the IPL being 487indicates a notable degree of variability in team performances throughout the season.

### The standard deviationof total runs (batsman runs and extra runs) in an IPL match is 1.59, it suggests a relatively low level of variability or dispersion in the scoring patterns during the matches.

INDIAN PREMIER LEAGUE (2008-2020)

### One-hot encoding is used to create and display the dummy values in the data. The above code converts each category in a categorical variable into a binary vector. It creates new binary columns for Winner and Result columns representing the presence or absence of the data in the category.It simply creates additional features based on the number of unique values in the categorical feature.

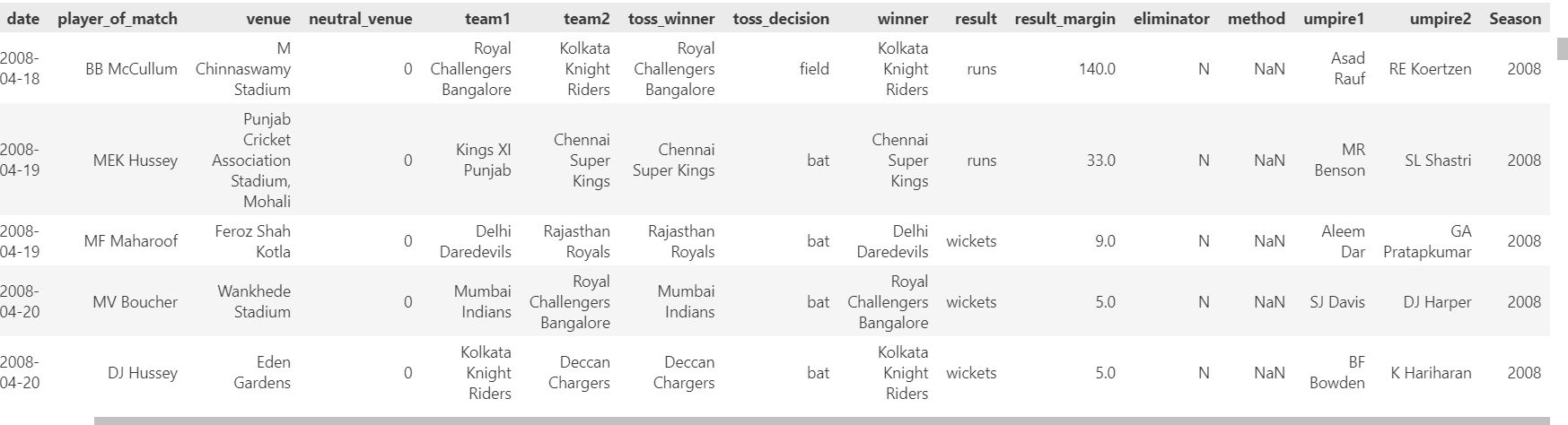
* The given data is pandas. core. frame. DataFrame type**.**
* **ANALYZING THE DATASET BY USING 'KURTOSIS' METHOD:**

### The positive kurtosis value 1.58 suggests that the distribution of total runs has heavier tails and a sharper peak than a normal distribution. This could indicate that there are some matches where teams scored significantly higher than the average, leading to a more concentrated distribution with a higher likelihood of extreme run totals.

* **ANALYZING THE DATASET USING 'SKEWNESS' METHOD:**

### The skewness of a batsman's runs and the total runs (batsman runs plus extra runs) are nearly equal, but the skewness of extra runs is significantly more positive, it suggests that the distribution of additional runs (extras) is driven by occasional high values, potentially indicating instances of exceptional performance in terms of extras.

* In the current analysis, we extract the year values from the date column to gain a temporal perspective on the dataset.



INDIAN PREMIER LEAGUE (2008-2020)

* Using the seasons we can determine the total number of matches played, total runs scored in each year and runs scored per match.



* In the ongoing analysis, the dataset reveals that, out of the total 812 matches,13 are tied. Furthermore, teams fielding first have secured victory in435matches , while teams batting first have won364matches.

### Therefore in present analysis it is evident that Eden Gardens stands out as the optimal stadium for winning the matches by wickers, emphasizing a higher likelihood of success when opting to field first.

### In the ongoing analysis Feroz Shah Kotla emerges as the premier stadium for winning matches by runs, underscoring a higher probability of success when choosing to bat first.

### Chinnaswamy stadium (pretty obvious) is the best stadium for RCB in which they have a high chance of winning the match when they win the toss.

INDIAN PREMIER LEAGUE (2008-2020)

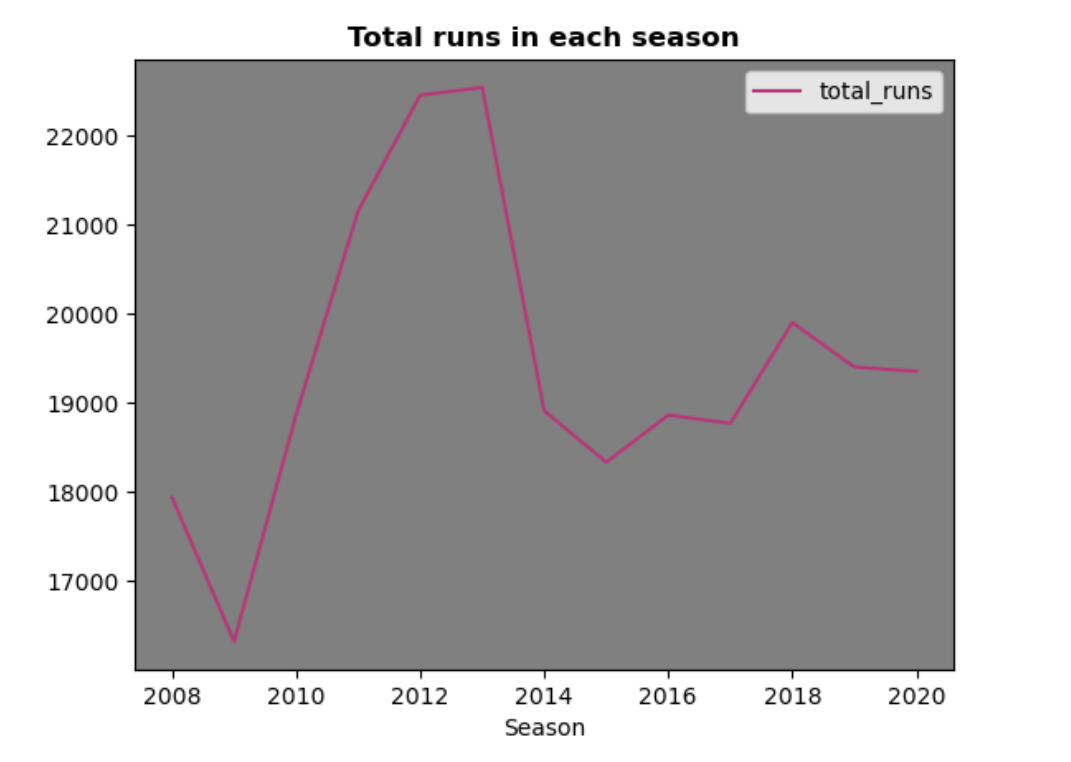
### The Orange Cap is awarded to the highest run-scorer in a particular IPL season. Examining the list of Orange Cap winners over multiple seasons helps identify players who consistently perform well with the bat. It highlights those who have maintained high levels of run-scoring consistency.

### So the Orange cap goes to VIRAT KOHLI !!!!



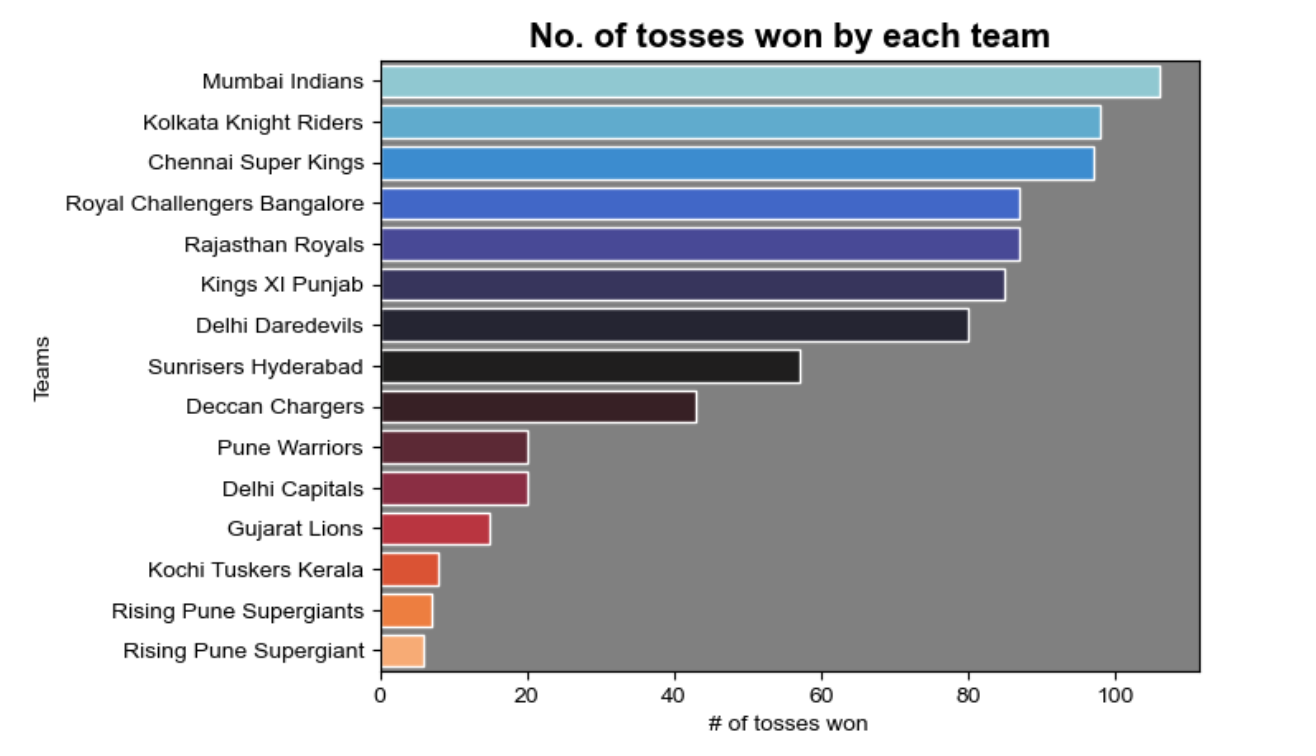
INDIAN PREMIER LEAGUE (2008-2020)

* **By employing graphs in the ongoing visual analysis, we vividly depict the data and discern trends within the dataset:**

****

* In current analysis, a line graph is employed to visually represent and observe the highest runs scored in each season, providing a clear depiction of scoring trends over the time.
* Generally upward-sloping line indicates an increase in total runs scored over the years, suggesting a rise in the overall scoring rate in IPL matches. fluctuations or variations in the line, might indicate changes in scoring patterns in specific seasons. Plateau or a downward trend, it could suggest stability or a decrease in total runs scored over the years. **From this line graph we can see that the highest runs scored is in the 2013 season.**

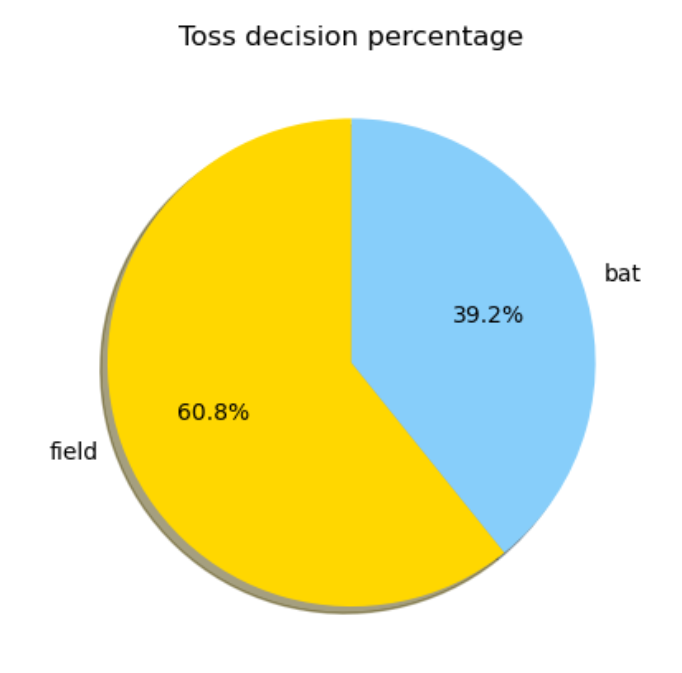
INDIAN PREMIER LEAGUE (2008-2020)



* Analyzing the bar graph can help stakeholders, fans, and analysts understand the toss-winning dynamics and potential strategies adopted by different IPL teams. It can also reveal whether toss wins are evenly distributed among teams or if certain teams consistently win more tosses than others.

### **MUMBAI INDIANS had won the most number of tosses.**

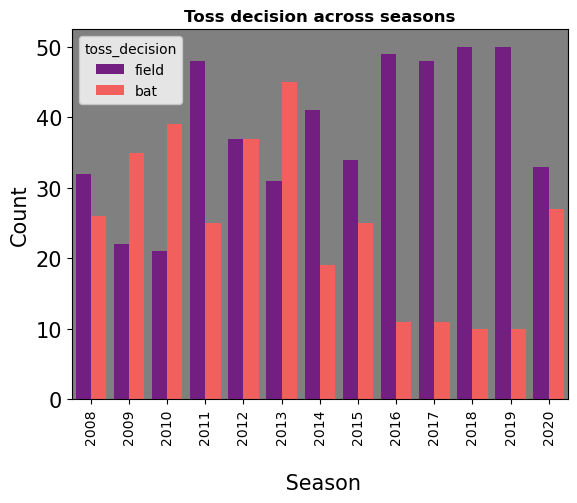
INDIAN PREMIER LEAGUE (2008-2020)

****

### A pie chart representing the toss decision (whether to field or bat) in an IPL provides a visual summary of the distribution of decisions made by teams after winning the toss. Comparing the sizes of the slices allows you to see if certain decisions are more common than others. The size of each slice corresponds to the percentage of times that decision was made. A pie chart helps understanding the overall pattern of toss decisions in the IPL and identify any trends or preferences among teams.

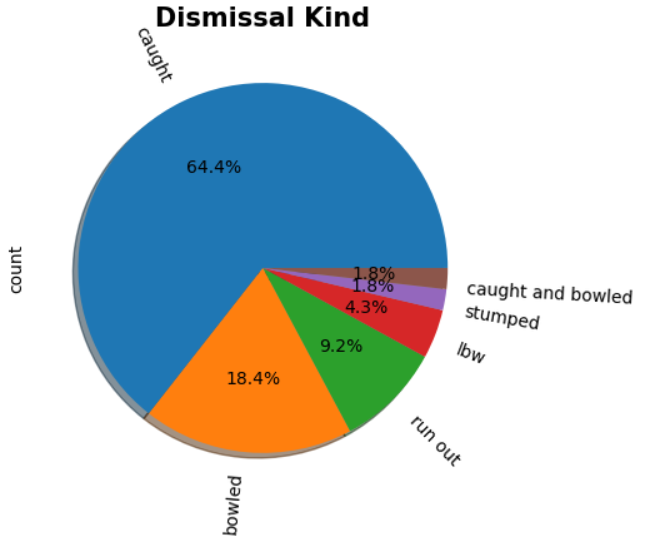
### **Almost 61 % of the toss decisions are made on the field first.**

INDIAN PREMIER LEAGUE (2008-2020)

****

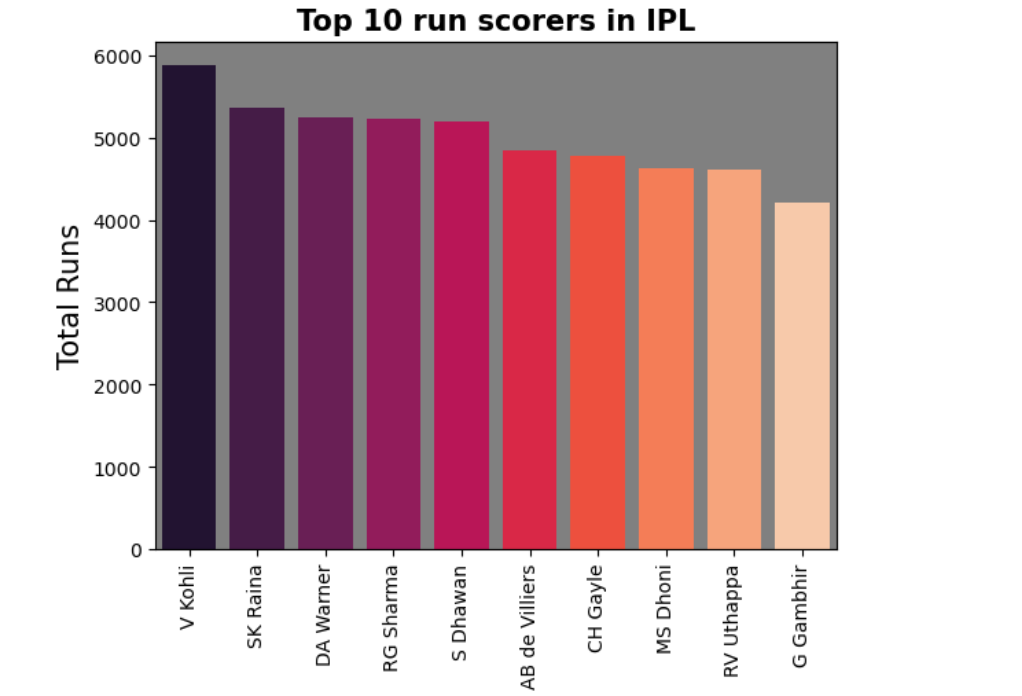
* The bar graph can reveal whether teams tend to choose a particular toss decision consistently across multiple seasons or if there are variations. A consistent trend over seasons may indicate a strategic preference or a team's approach in different conditions. Teams may adjust their toss decisions based on the pitch conditions prevalent during each season. Teams may adjust their toss decisions based on the pitch conditions prevalent during each season.
* **Here the orange indicate batting and purple indicates fielding. So as an example in 2016 , we can see that majority of times, teams prefer to field first after winning the toss.**

INDIAN PREMIER LEAGUE (2008-2020)

****

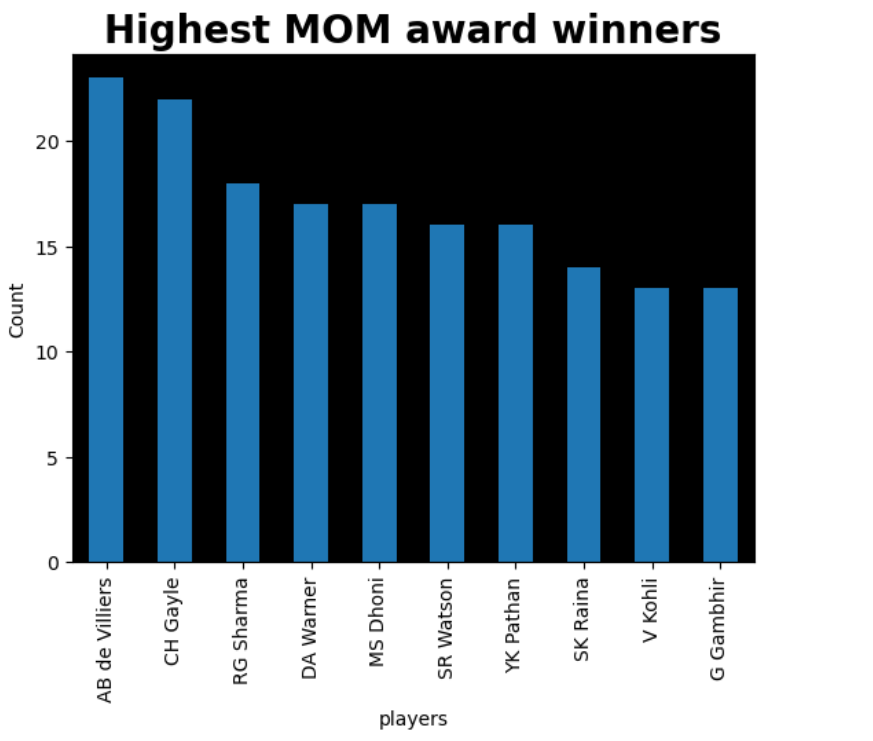
* In ongoing visual analysis , a pie chart is utilized to illustrate the distribution of dismissal types for Virat Kohli in IPL.The visual representation offers insights into the predominant ways in which he has been dismissed throughout the tournament.
* **This tells us that VK has been dismissed by catch majority of times.This is very useful for a player to work on their weakness.**

INDIAN PREMIER LEAGUE (2008-2020)

****

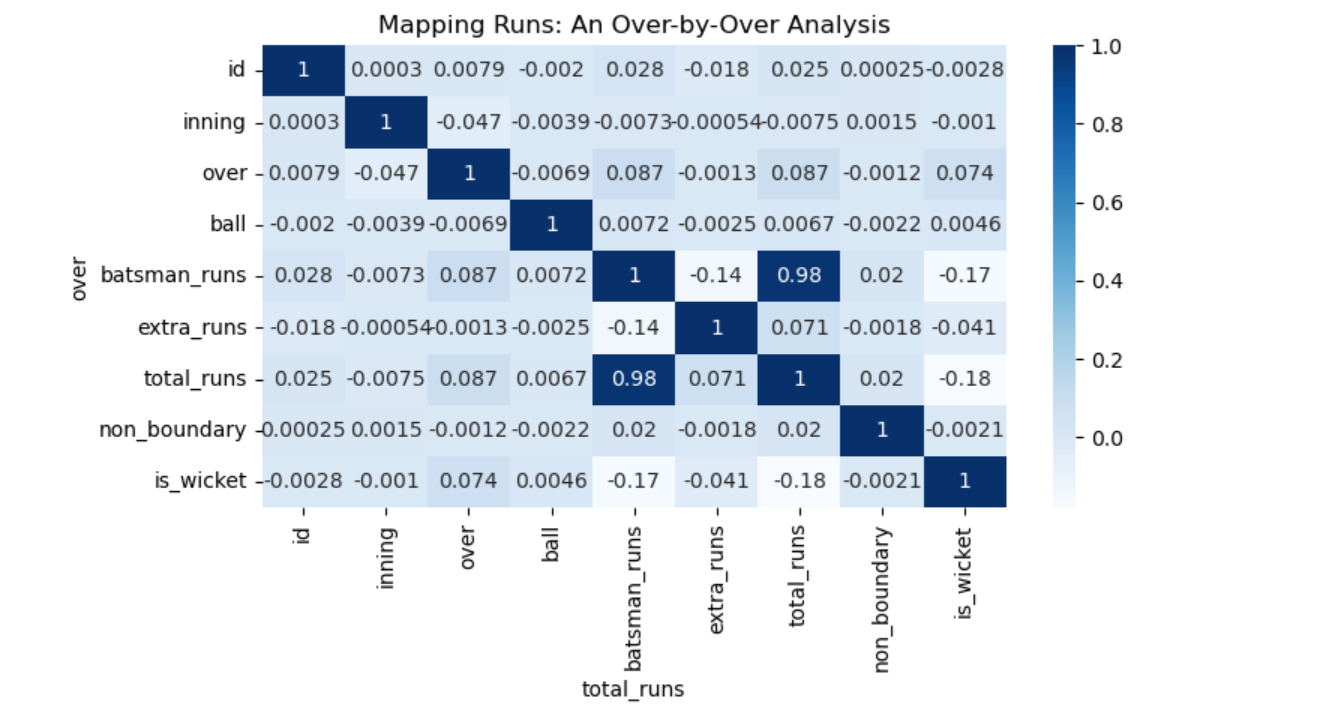
* The bar graph allows you to easily compare the run-scoring performances of the top 10 batsmen in the IPL.Observing the heights of the bars can indicate the consistency and dominance of certain players. A higher bar suggests a player who consistently scores a significant number of runs.The graph can indicate the impact of foreign players among the top run-scorers, highlighting the contribution of overseas players to the league.
* The graph is likely to capture the attention of fans, as they can easily identify their favorite players and track their performances in comparison to other top run-scorers.
* **Virat Kohli is the top most run scorer in the IPL seasons.**

INDIAN PREMIER LEAGUE (2008-2020)

****

* The bar graph allows you to identify players who have consistently made significant contributions to their team's success, as reflected by winning the Man of the Match award.Players who frequently win the Man of the Match award are likely to be fan favorites. The graph can engage fans and spark discussions about player performances.
* **AB de Villiers has won most "MAN OF THE MATCH" awards.**

INDIAN PREMIER LEAGUE (2008-2020)

****

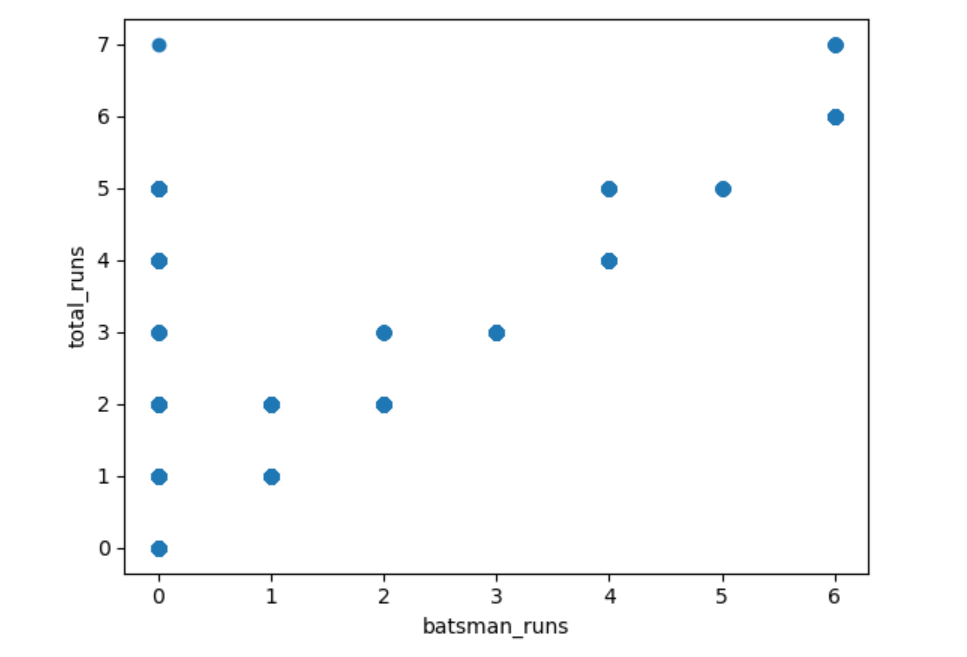
### The purpose of a heatmap is to provide a visual representation of the relative intensity or magnitude of values in a matrix, where colors are used to convey the magnitude of each value.

### **Darker cells in the heatmap indicate overs where a higher number of total runs were scored.** These overs can be considered as scoring hotspots or periods of aggressive batting.

### **Lighter colors may represent overs where the team is scoring at a slower rate (tight bowling).**

### Changes in color patterns can reveal shifts in the team's scoring rate. A transition from lighter to darker colors might signify an acceleration in scoring.

INDIAN PREMIER LEAGUE (2008-2020)

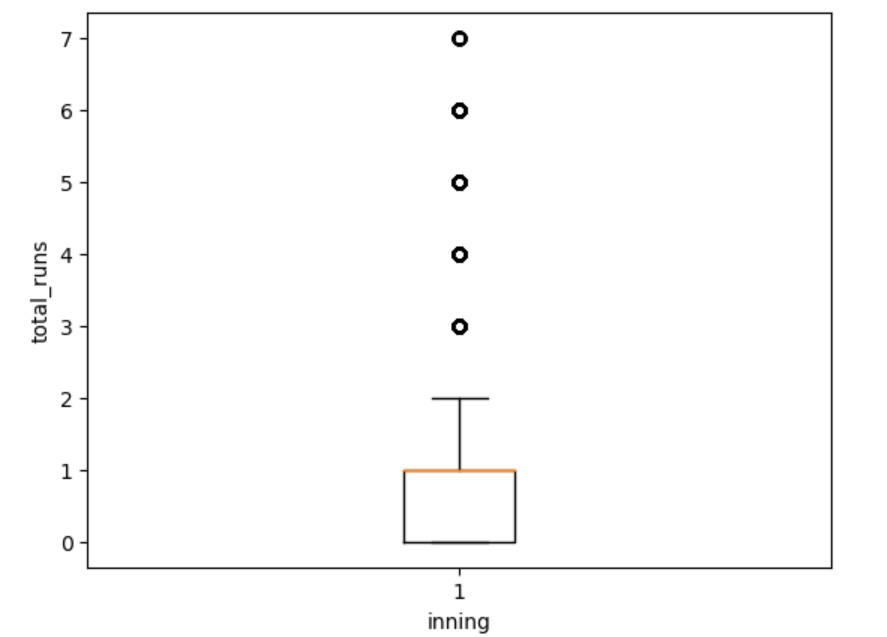
****

### A scatter plot representing the relationship between total runs and batsmen runs in an IPL dataset provides a visual representation of how individual batsmen contribute to their team's overall run total.

### **Batsmen whose points are both higher and to the right on the plot are considered highly impactful,** contributing significantly to both their individual scores and the team's total runs.The given scatter plot is positive.**Positive correlation suggests that as a batsman scores more, the team tends to score more as well**

### Outliers on the plot may represent exceptional performances where a batsman has significantly outperformed in terms of individual runs or where the team has achieved an exceptionally high total

INDIAN PREMIER LEAGUE (2008-2020)

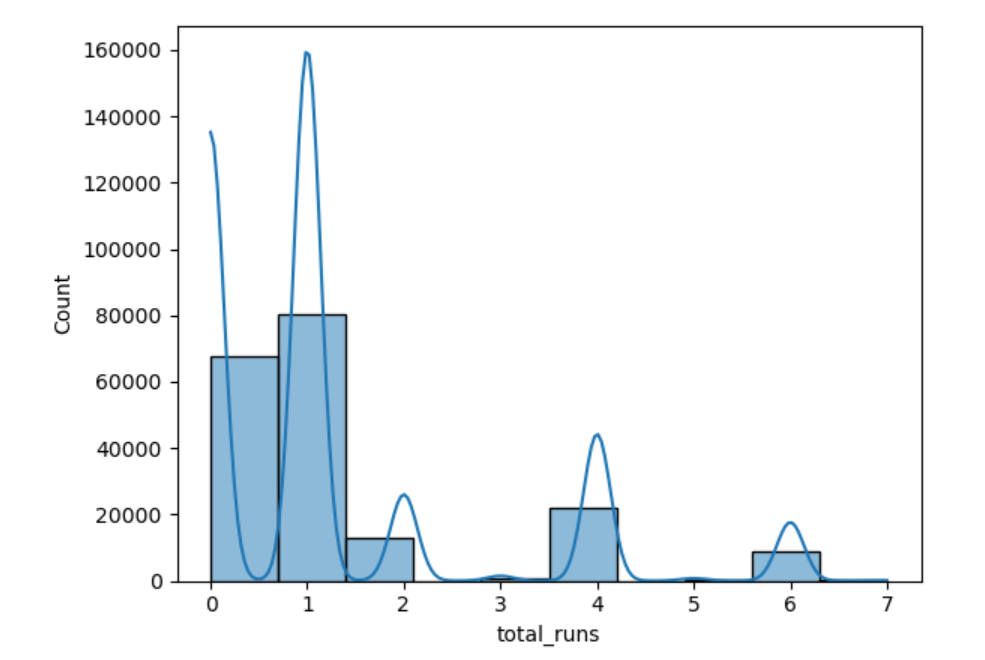
****

### The box plot of the given data tells us that the median is closer to the upper quartile. There is no box above the median which indicates that the distribution of variability in total runs in the given data is highly negatively skewed.

### **This means that a significant portion of the total runs is dispersed below the median** and the lower side is stretched as there are more data points with lower values leading to a negatively skewed distribution.

### The given dataset has a significant number of outliers.

INDIAN PREMIER LEAGUE (2008-2020)

****

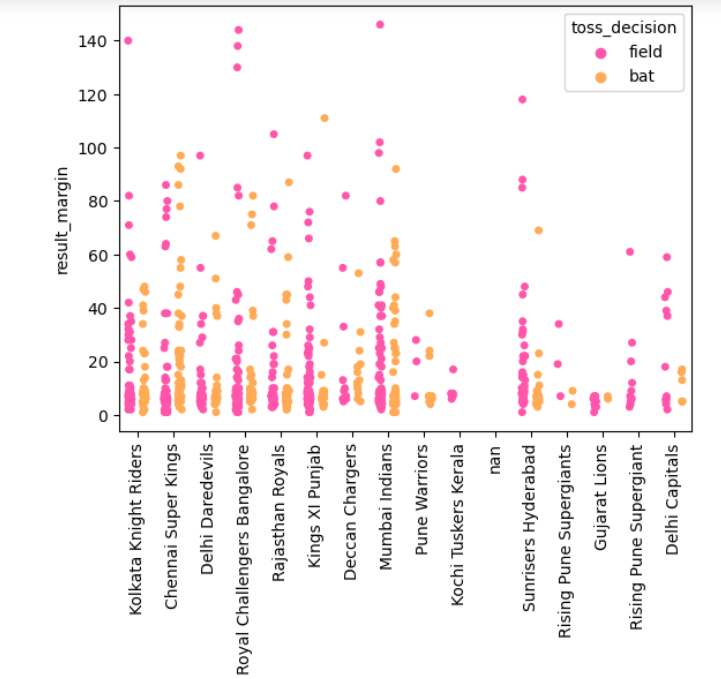
### The histogram displays the distribution of total runs scored, showing the frequency of different ranges or bins of total runs in the dataset.

### Peaks or high bars in the histogram indicate the most common or frequently occurring total runs scored. These are the values with the highest count.

### You can observe how often total runs fall within specific ranges or intervals. Extreme values or outliers in the dataset may be visible as tall bars at the edges of the histogram.

### **These bars represent instances of exceptionally high or low total runs.**

INDIAN PREMIER LEAGUE (2008-2020)

****

### The strip plot shows the distribution of winners in the dataset, indicating which teams have won matches. The spread and distribution of points along the x-axis provide insights into the competitiveness of matches.

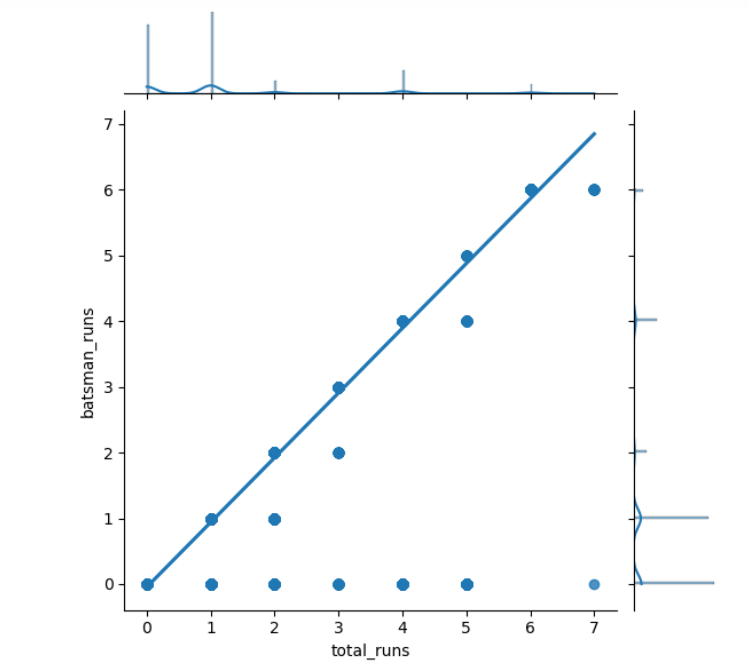
### Color or marker style differentiation on the strip plot can represent the toss decision (bat or field). **This allows you to see if there's a correlation between the toss decision and the winning team or result margin.**

### Outliers in the strip plot may indicate matches where the result margin was unexpectedly high or low compared to typical matches. **These points can represent upsets or exceptional performances.**

### The strip plot allows you to compare the performance of different teams in terms of winning matches, result margins, and toss decisions.

### 

### INDIAN PREMIER LEAGUE (2008-2020)

****

### The joint plot typically includes a scatter plot where each point represents a combination of a batsman's runs and the corresponding total runs in a match. A regression line may be included to show the trend in the data.

### The scatter plot helps you assess the correlation or relationship between the runs scored by an individual batsman and the total runs scored by the team. Positive correlation suggests that as individual batsmen score more, the team tends to score more as well.

### If certain **batsmen consistently appear in matches where the team scores high, it suggests that the team is dependent on these key batsmen for significant contributions.**

### The plot allows you to visually evaluate the effectiveness of the overall batting lineup by examining the distribution of runs across different batsmen.

INDIAN PREMIER LEAGUE (2008-2020)

**5. CONCLUSION**

The results of the IPL dataset analysis illuminate key patterns and insights. Player performance metrics, including batting averages and bowling economy, reveal consistent standout performers across seasons. Team strategies, reflected in composition and decision-making after winning the toss, significantly influence match outcomes. Comparative team analyses demonstrate the evolving dynamics of team success. These findings contribute to a nuanced understanding of the IPL, emphasizing the interplay of individual brilliance, team dynamics, and strategic acumen. The discussion delves into actionable implications for team management, offering a holistic perspective on factors shaping success in the highly competitive and dynamic landscape of the IPL.

In essence, the results derived from the comprehensive analysis of the IPL dataset from

2008 to 2020 shed light on intricate patterns and strategic nuances that define the league’s narrative.The consistent excellence of individual players, as evidenced by metrics like batting averages and bowling economy, underscores the enduring impact of cricketing talent. Additionally, the observed influence of team strategies on match outcomes highlights the critical role of decision-making in the fast-paced T20 format.

As we compare team performances across seasons, a nuanced understanding of the league’s evolving dynamics emerges, emphasizing the multifaceted nature of success in

the IPL. These insights offer actionable implications for team management, providing a roadmap to navigate the complexities of team composition, strategic choices, and player utilization. In essence, the analysis encapsulates the essence of the IPL, where skill, strategy, and teamwork converge in a captivating blend, defining an era of cricket that transcends the boundaries of sport and entertainment.