**M15\_Project\_FinalDataAnalysisReport\_Abhishek\_Jain**

**Data Analysis Report**

**Abhishek Jain(aj763)**

**IPL Data Analysis**

**Introduction**

The Indian Premier League (IPL) has emerged as one of the most popular and competitive Twenty20 cricket tournaments in the world. With its unique blend of top international players and talented domestic cricketers, the IPL captivates millions of fans around the globe. In this project, we delve into the realm of data analysis to uncover valuable insights from IPL data.

The objective of this project is to analyze the vast amount of IPL data available and extract meaningful patterns, trends, and correlations. By exploring various aspects of the game, such as team performance, player statistics, match outcomes, and strategic factors, we aim to gain a comprehensive understanding of the IPL and its dynamics.

To conduct this analysis, we utilize a diverse dataset encompassing information on matches, teams, players, scores, and other relevant variables. The dataset offers a wealth of information spanning multiple seasons of the IPL, allowing us to uncover patterns and trends that may influence team strategies, player choices, and ultimately, match outcomes.

Throughout the project, we employ various data analysis techniques, statistical methods, and visualization tools to extract meaningful insights from the IPL dataset. By examining key metrics such as batting averages, strike rates, bowling economy rates, team win percentages, and more, we aim to uncover hidden patterns and relationships that shed light on the factors contributing to success in the IPL.

The findings from this data analysis project have the potential to benefit a wide range of stakeholders, including IPL franchises, team management, players, and even cricket enthusiasts. The insights gained can inform team strategies, player selections, and overall decision-making processes, leading to enhanced performance and improved outcomes in future IPL seasons.

Furthermore, the project serves as a valuable resource for researchers, analysts, and cricket enthusiasts interested in exploring the intricacies of the IPL. The comprehensive analysis of the dataset offers a deeper understanding of the game, its dynamics, and the factors that contribute to team success in this highly competitive tournament.

In conclusion, this IPL data analysis project aims to uncover valuable insights from the vast pool of IPL data, exploring the impact of team strategies, player performance, and various factors on match outcomes. By leveraging data analysis techniques, we seek to provide meaningful and actionable insights that can drive improvements in team performance and contribute to the broader understanding of the IPL.

**Important Questions I answered:**

* Research Question: How does the performance of individual players in the Indian Premier League (IPL) affect the success of their teams?

Hypothesis: Teams with players who consistently perform well, with high batting averages, strike rates, and wicket-taking abilities, are more likely to achieve success in the IPL.

* Research Question: What are the most successful tactics and methods employed by teams in IPL matches?

Hypothesis: Teams that utilize aggressive batting strategies, such as sending power-hitters up the order and focusing on boundary-hitting, are more likely to post high scores and win matches in the IPL.

* Research Question: How does team strategy and player choices impact IPL match outcomes?

Hypothesis: Teams with well-defined strategies, such as effective field placements, bowling variations, and optimal player selections based on match conditions and opposition strengths, are more likely to win matches in the IPL.

The goal of this project is to analyze IPL data using R and extract meaningful insights from the data. We will perform various analyses, including data cleaning, data exploration, and data visualization, to gain a deeper understanding of the IPL. Our primary focus will be on answering the following questions:

**Important Intermediate Questions that needed to be answered:**

* How does the closeness of matches vary when the team batting first wins in the IPL?
* Which cities have hosted the maximum number of IPL matches?
* Which stadiums have hosted the maximum number of IPL matches?
* Is winning the toss an advantage in IPL matches?
* How much does home advantage play a role in the outcome of IPL matches?
* How many matches have each team played in the IPL?
* Which team has won the maximum number of matches in the IPL?
* How does the win percentage of each team compare in the IPL?
* Who are the top batsmen in the IPL and what are their statistics?
* How do the total runs scored at each delivery of an over vary in the IPL?
* What is the average number of runs scored in each delivery of the over in the IPL?
* How do the total number of runs scored in each over of the innings vary in the IPL?
* What is the average number of runs scored in each over of the innings in the IPL?
* How does the total number of wickets in each over of the innings vary in the IPL?
* Who are the batsmen with the top strike rate in the IPL and what are their statistics?
* How does the strike rate vary for different stages of the game in the IPL?
* How does the innings progression relate to the result of the match in the IPL?
* How does the innings progression vary in different innings of the match in the IPL?
* How do the runs scored vs balls faced vary in all IPL matches?

**How the research questions will be answered:**

To answer the first question, we can look at the top batsmen in the IPL and analyze their impact on their respective team's success. By identifying the top scorers and their strike rates, we can understand how their performance affects the team's overall score and chances of winning. We can also analyze how often the top scorers are from the winning team, and if their contribution played a significant role in the team's success.

To address the second question, we can examine the success rates of different team strategies, such as choosing to bat or bowl first, making early or late substitutions, and targeting specific opposition players. By analyzing the number of matches won or lost with each strategy, we can determine which ones are the most effective for a team to deploy in IPL matches.

Finally, for the third question, we can analyze how team strategy and player choices impact IPL match outcomes by comparing the success rates of different teams with their player selection and tactics. By assessing which teams consistently perform well or poorly with specific strategies or players, we can identify the most successful tactics and methods for teams to deploy in IPL matches.

**Methodology**

In the methodology section of the IPL data analysis project, we have outlined the steps we followed to gather and analyze the data. The methodology is divided into four main stages: data collection, data cleaning, data exploration, and data visualization.

In the data collection stage, we obtained two datasets from Kaggle: "deliveries.csv" and "matches.csv." These datasets contained information on individual deliveries and matches played in the IPL from 2008 to 2021. We imported these datasets into R and merged them to create a combined dataset that we used for analysis.

The next stage was data cleaning, where we checked for any missing or inconsistent values in the data. We also transformed the data to make it suitable for analysis, such as converting date-time variables and creating new variables based on the existing data.

In the data exploration stage, we performed various descriptive and inferential statistical analyses to gain insights into the data. We explored the number of matches played in different cities, the performance of individual players, the impact of winning the toss, and the relationship between innings progression and match results.

Finally, in the data visualization stage, we used various graphs and charts to present the data visually appealingly. We used histograms, box plots, and bar charts to analyze the data and gain insights into the IPL's various aspects.

In conclusion, the methodology section outlines our steps to collect, clean, explore, and visualize the data. By following these steps, we gained insights into the IPL's various aspects, such as the performance of teams, and players and the impact of various factors on match outcomes.

**Steps I followed:**

1. Data Collection: Gather comprehensive IPL data from reliable sources such as official IPL websites, cricket databases, and reputable sports statistics platforms. Ensure that the dataset includes a wide range of variables, such as match details, team information, player statistics, scores, and other relevant parameters.
2. Data Cleaning and Preprocessing: Perform data cleaning and preprocessing to handle missing values, inconsistencies, and outliers. Standardize the data format, resolve any discrepancies, and ensure uniformity across different variables. This step is crucial for ensuring the accuracy and reliability of the analysis.
3. Exploratory Data Analysis: Conduct exploratory data analysis to gain initial insights into the dataset. Visualize the data using graphs, charts, and summary statistics to identify trends, patterns, and potential relationships between variables. This step helps in formulating specific research questions and hypotheses.
4. Define Research Questions and Hypotheses: Based on the initial exploratory analysis, formulate research questions and hypotheses that align with the objectives of the project. These questions could pertain to the impact of player performance, team strategies, match conditions, or other relevant factors on IPL match outcomes.
5. Statistical Analysis: Apply appropriate statistical techniques to answer the research questions and test the hypotheses. Use statistical measures such as correlation, regression analysis, hypothesis testing, and significance testing to determine the relationships between variables and draw meaningful conclusions.
6. Visualizations: Create clear and informative visualizations to present the analysis results. Utilize graphs, charts, heatmaps, and other visual representations to communicate the findings effectively. Visualizations help in presenting complex data patterns and trends in a visually appealing and easily understandable format.
7. Interpretation and Discussion: Interpret the results of the analysis and discuss their implications in the context of the research questions. Analyze the patterns, trends, and relationships identified in the data to draw meaningful insights. Compare the findings with existing literature and discuss their relevance to the IPL and its dynamics.
8. Conclusion: Summarize the key findings of the analysis and provide a comprehensive conclusion based on the results. Highlight the implications of the findings for team strategies, player performance, and IPL match outcomes. Discuss any limitations of the analysis and suggest future research directions.

By following this methodology, the IPL data analysis project ensures a systematic and rigorous approach to extract valuable insights from the dataset, leading to a comprehensive understanding of the factors influencing team success, effective tactics, and the impact of team strategy and player choices on IPL match outcomes.

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

I will be answering a few extra questions to provide a complete analysis of the game . These will include assessing the impact of winning the toss or playing at home on match outcomes, evaluating team performance based on the number of matches played and won, identifying top batsmen and analyzing their strike rates, examining the innings progression and runs scored in different stages of the game, and understanding the relationship between innings progression and match results. The methodology for each analysis would involve data collection and cleaning, exploratory data analysis, and various statistical and visualization techniques. The insights gained from these analyses can provide valuable information for team management, sponsors, and cricket enthusiasts.

**Dataset**

The data source for this project will be the IPL dataset available on GitHub. The dataset

consists of all IPL matches played between 2008 and 2021. The dataset contains.

information such as match details, team and player statistics, and ball-by-ball details.

The dataset will be imported into R for analysis.

Table

Description automatically generatedA screenshot of a computer

Description automatically generated

**Preparing Data**

I am using two csv filles for this project (deliveries.csv and matches.csv) .Both these files have data related to teams ,players ,toss, balls and hits . I also had to clean the data as few columns had NA fields as well.

**Understanding the Data**

**A picture containing text

Description automatically generated**

**A picture containing table

Description automatically generated**

**Deliveries.csv have**

* match\_id: A unique identifier for the IPL match.
* inning: The inning number (1 or 2) in which the delivery was bowled.
* batting\_team: The name of the batting team for the current inning.
* bowling\_team: The name of the bowling team for the current inning.
* over: The over number in which the delivery was bowled.
* ball: The ball number of the current over.
* batsman: The name of the batsman facing the delivery.
* non\_striker: The name of the non-striker on the other end of the pitch.
* bowler: The name of the bowler who delivered the ball.
* is\_super\_over: A binary indicator (0 or 1) indicating whether the delivery was bowled in the Super Over.
* wide\_runs: The number of runs scored in the delivery due to a wide ball.
* bye\_runs: The number of runs scored in the delivery due to byes (when the ball goes past the batsman without touching the bat or the batsman).
* legbye\_runs: The number of runs scored in the delivery due to leg byes (when the ball hits the batsman’s body or the bat and goes to the fielder).
* noball\_runs: The number of runs scored in the delivery due to a no-ball (an illegal delivery).
* penalty\_runs: The number of runs scored in the delivery due to a penalty awarded by the umpire.
* batsman\_runs: The number of runs scored by the batsman off the delivery.
* extra\_runs: The total number of runs scored in the delivery due to extras (wide, bye, leg bye, no-ball, or penalty).
* total\_runs: The total number of runs scored in the delivery (batsman runs plus extra runs).
* player\_dismissed: The name of the batsman who was dismissed on the delivery, if any.
* dismissal\_kind: The type of dismissal (e.g., caught, bowled, run out) if the batsman was dismissed on the delivery, else null.

**Matches.csv have**

* id: A unique identifier for the IPL match.
* season: The year in which the IPL season was played.
* city: The city in which the match was played.
* date: The date on which the match was played.
* team1: The name of the first team playing the match.
* team2: The name of the second team playing the match.
* toss\_winner: The name of the team that won the toss.
* toss\_decision: The decision made by the toss-winning team (batting or bowling).
* result: The result of the match (normal, tie, or no result).
* dl\_applied: A binary indicator (0 or 1) indicating whether the Duckworth-Lewis method was applied during the match.
* winner: The name of the team that won the match.
* win\_by\_runs: The number of runs by which the winning team won the match (0 if the winning team won by wickets).
* win\_by\_wickets: The number of wickets by which the winning team won the match (0 if the winning team won by runs).
* player\_of\_match: The name of the player who was awarded the "Man of the Match" award.
* venue: The name of the venue where the match was played.
* umpire1: The name of the first on-field umpire.
* umpire2: The name of the second on-field umpire.
* umpire3: The name of the third umpire (if any).

**Some Answer that can be inferred by seeing the data**

What is the impact of team composition on performance in IPL?

By analyzing team compositions, including the combination of batsmen, bowlers, all-rounders, and the balance of overseas and domestic players, we can evaluate how different team compositions correlate with performance metrics such as win percentage, run rates, and bowling economy rates. This analysis can help identify the ideal team composition for maximizing success in the IPL.

How does the toss affect the outcome of IPL matches?

Through data analysis, we can examine the impact of winning the toss on match outcomes. By studying the statistics related to teams opting to bat or field first after winning the toss, we can determine whether there is a significant advantage for the team winning the toss and how it affects their chances of winning the match.

Are there any patterns in player performance based on match conditions?

By analyzing player statistics in different match conditions such as day/night matches, home/away matches, specific venues, and varying weather conditions, we can identify patterns in player performance. This analysis can reveal if certain players excel under specific conditions and if certain teams have an advantage in specific match scenarios.

Does the presence of star players impact the financial success of IPL franchises?

By analyzing financial data, sponsorship deals, and revenue generation of IPL franchises, we can assess the impact of star players on the financial success of the teams. This analysis can determine if having marquee players in a team positively influences fan engagement, ticket sales, merchandise purchases, and overall profitability.

How do team strategies and player roles evolve over IPL seasons?

Through historical data analysis, we can track the evolution of team strategies and player roles over different IPL seasons. By studying team tactics, batting orders, bowling rotations, and captaincy decisions, we can identify trends in strategy changes and player adaptability, providing insights into the dynamic nature of IPL teams.

**Data Analysis and Visualization**

1. Closeness of the matches when the team batting first wins

*#1. Closeness of the matches when the team batting first wins*

matches1<-matches[matches$win\_by\_runs!=0,]

closeness<-function(x,y = "gold" ){

data1<-matches1[matches1$winner==x|matches1$loser==x,]

data1[data1$loser==x,"win\_by\_runs"]<- -data1[data1$loser==x,"win\_by\_runs"]

ggplot(data1,aes(1:nrow(data1),win\_by\_runs))+ geom\_area(fill=y)+ggtitle(x)+

ylab("Runs")+ xlab("Matches")+ geom\_ribbon(aes(ymin=-5, ymax=5),fill="red",alpha=0.4) +geom\_ribbon(aes(ymin=-15, ymax=15),fill="red",alpha=0.1) +

guides(fill=FALSE)+scale\_alpha(guide = 'none')+coord\_cartesian(ylim = c(-100, 100))

}

a<-closeness("Chennai Super Kings")

b<-closeness("Kolkata Knight Riders","purple")

c<-closeness("Sunrisers Hyderabad","orange")

d<-closeness("Mumbai Indians","blue2")

e<-closeness("Royal Challengers Bangalore","red3")

f<-closeness("Delhi Daredevils","firebrick3")

g<-closeness("Rajasthan Royals","blueviolet")

h<-closeness("Kings XI Punjab","salmon")

grid.arrange(a,b,c,e,d,f,g,h,ncol=2)

Visualizing the closeness of matches when the team batting first wins can provide us with several insights:

* Distribution of margins of victory: By looking at a histogram or density plot of margins of victory for matches in which the team batting first won, we can see the distribution of how often these wins are close or more lopsided. This can give us an idea of how frequently teams are able to defend their total, and whether certain teams are more likely to win close matches than others.
* Relationship between margin of victory and total runs scored: By creating a scatter plot or line plot of margin of victory against total runs scored by the team batting first, we can see whether there is a relationship between the two variables. For example, we might find that teams that score a lot of runs are more likely to win by a large margin, or that matches with lower total runs scored are more likely to be close.

Graphical user interface, timeline

Description automatically generated with medium confidence

1. Number of matches played in different cities

ggplot(matches[which(!is.na(matches$city)),],aes(city,fill= city,rm.na=T)) +geom\_bar() +

theme(axis.text.x = element\_text(angle = 90, hjust = 1))+

ylab("Number of Matches Played") +

guides(fill=FALSE)

Analyzing the number of matches played in different cities can provide us with several insights:

* Popularity of cricket in different cities: By looking at the number of matches played in different cities, we can get an idea of where cricket is most popular. Cities that have hosted a large number of matches are likely to have a passionate fan base and strong support for the sport.
* Impact of infrastructure: Cities with better infrastructure, such as large stadiums and good transportation links, may be more likely to host matches. Analyzing the number of matches played in different cities can provide insights into the impact of infrastructure on the scheduling of matcheChart, histogram

  Description automatically generated

1. Number of matches played in different stadium

*#3. Number of matches played in different stadium*

ggplot(matches,aes(venue, rm.na=T)) +geom\_bar(fill="#0072B2") +

theme(axis.text.x = element\_text(angle = 90, hjust = 1))+

ylab("Number of Matches Played")

Chart

Description automatically generated

Analyzing the number of matches played in different stadiums can provide us with several insights:

* Popularity of stadiums: By looking at the number of matches played in different stadiums, we can get an idea of which stadiums are most popular among players and fans. Stadiums that have hosted a large number of matches are likely to have good infrastructure and amenities and may be preferred by teams and spectators.

## Is winning the toss really an advantage?

*#4. Is winning the toss really an advantage?*

matches$toss\_match<-ifelse(as.character(matches$toss\_winner)==as.character(matches$winner),"Won","Lost")

ggplot(matches[which(!is.na(matches$toss\_match)),],aes(toss\_match, fill = toss\_match))+

geom\_bar()+ xlab("Toss") +ylab("Number of matches won")+ ggtitle("How much of a advantage is winning the toss")

## Analyzing the impact of winning the toss can provide us with several insights:

## Decision to bat or bowl first: When a team wins the toss, they have the option to choose whether to bat or bowl first. Analyzing the decisions made by teams can provide insights into their strategy and playing style. For example, some teams may prefer to bat first in order to set a challenging total, while others may prefer to bowl first to take advantage of early swing or seam movement.

## Home advantage: When playing at home, teams may have a better understanding of local conditions and be more likely to make the right decision when winning the toss. Analyzing the impact of winning the toss on home and away matches can provide insights into whether home teams are more likely to win the toss and whether this translates into a greater advantage.

## Chart, bar chart Description automatically generated

## Is home advantage a real thing in IPL?

Yes, home advantage is generally considered a real thing in IPL. Home advantage refers to the benefits a team can enjoy when playing at their home stadium, such as familiarity with local conditions, support from home fans, and the ability to adapt to the playing surface.

In IPL, home advantage is often reflected in teams' performance on their home ground. Many teams have a strong record at their home stadium, with a higher win percentage compared to away matches. This suggests that playing at home can provide a competitive advantage for teams in IPL.

However, it's worth noting that the impact of home advantage may vary depending on several factors, such as the opposing team's strength, the conditions on the day of the match, and the quality of the playing surface. Additionally, in recent years, the IPL has introduced neutral venues for some matches, reducing home advantage's impact.

Overall, while the impact of home advantage may vary from team to team and season to season, it is generally considered a real factor in IPL.

*#5. Is home advantage a real thing in IPL?*

Data<-matches[matches$season!="2009",]

Data$date<- as.Date(Data$date)

Data1<-Data[Data$date < as.Date("2014-04-16") | Data$date > as.Date("2014-04-30"),]

Data1$home\_team[Data1$city=="Bangalore"]<- "Royal Challengers Bangalore"

Data1$home\_team[Data1$city=="Chennai"]<- "Chennai Super Kings"

Data1$home\_team[Data1$city=="Delhi"]<- "Delhi Daredevils"

Data1$home\_team[Data1$city=="Chandigarh"]<- "Kings XI Punjab"

Data1$home\_team[Data1$city=="Jaipur"]<- "Rajasthan Royals"

Data1$home\_team[Data1$city=="Mumbai"]<- "Mumbai Indians"

Data1$home\_team[Data1$city=="Kolkata"]<- "Kolkata Knight Riders"

Data1$home\_team[Data1$city=="Kochi"]<- "Kochi Tuskers Kerala"

Data1$home\_team[Data1$city=="Hyderabad" & Data1$season <=2012]<- "Deccan Chargers"

Data1$home\_team[Data1$city=="Hyderabad" & Data1$season >2012]<- "Sunrisers Hyderabad"

Data1$home\_team[Data1$city=="Ahmedabad"]<- "Rajasthan Royals"

Data1$home\_team[Data1$city=="Dharamsala"]<- "Kings XI Punjab"

Data1$home\_team[Data1$city=="Visakhapatnam" & Data1$season== 2015]<- "Sunrisers Hyderabad"

Data1$home\_team[Data1$city=="Ranchi" & Data1$season== 2013]<- "Kolkata Knight Riders"

Data1$home\_team[Data1$city=="Ranchi" & Data1$season > 2013]<- "Chennai Super Kings"

Data1$home\_team[Data1$city=="Rajkot" ]<- "Gujarat Lions"

Data1$home\_team[Data1$city=="Kanpur" ]<- "Gujarat Lions"

Data1$home\_team[Data1$city=="Raipur" ]<- "Delhi Daredevils"

Data1$home\_team[Data1$city=="Nagpur" ]<- "Deccan Chargers"

Data1$home\_team[Data1$city=="Indore" ]<- "Kochi Tuskers Kerala"

Data1$home\_team[Data1$city=="Pune" & Data1$season!= 2016]<- "Pune Warriors"

Data1$home\_team[Data1$city=="Pune" & Data1$season== 2016]<- "Rising Pune Supergiants"

Data1<-Data1[ which(!is.na(Data1$home\_team)),]

Data1$win\_host <- ifelse(as.character(Data1$winner)==as.character(Data1$home\_team),"Home","Away")

ggplot(Data1[which(!is.na(Data1$win\_host)),],aes(win\_host,fill= win\_host))+geom\_bar()+

ggtitle("Is home advantage a real thing in IPL?")+

xlab("Team")+

ylab("Number of Matches won")+labs(aesthetic="Winner")

## Chart Description automatically generated

## Number of matches played by each team

Analyzing the number of matches played by each team in IPL can provide several insights, including:

* Performance of teams: The number of matches played by each team can provide insights into their overall performance in IPL. Teams that have played a large number of matches may have had more opportunities to showcase their skills and perform well, while teams that have played fewer matches may be relatively new to the league or have had a more challenging schedule.
* League position: The number of matches played by each team can also influence their position in the IPL league table. Teams that have played more matches and won more games are likely to have a higher position in the table, while teams that have played fewer matches or performed poorly may be lower in the standings.
* Impact of scheduling: The number of matches played by each team can also reflect the impact of scheduling on their performance. For example, teams that have played a large number of matches in a short period of time may experience fatigue or injuries, which could impact their performance in later matches.
* Head-to-head record: The number of matches played between two teams can provide insights into their head-to-head record. Analyzing the number of matches played by each team against their opponents can help identify rivalries and provide insights into the history of their encounters.
* Team experience: The number of matches played by each team can also provide insights into their experience in the league. Teams that have played a large number of matches may be more experienced and have a better understanding of the tactics and strategies required to succeed in IPL.

*#6. Number of matches played by each team*

ggplot(as.data.frame(table(matches$team2) + table(matches$team1)),aes(reorder(Var1,-Freq),Freq,fill = Var1)) +geom\_bar(stat = "identity")+

theme(axis.text.x = element\_text(angle = 90, hjust = 1))+ xlab("Player")+

ylab("Number of Matches") +guides(fill=FALSE)

## Chart, bar chart Description automatically generated

## Number of Matches won by each teams

Analyzing the number of matches won by each team in IPL can provide several insights, including:

1. Performance of teams: The number of matches won by each team can provide insights into their overall performance in IPL. Teams that have won more matches will likely perform better in the league and have a higher chance of qualifying for playoffs.
2. Momentum: The number of matches won by each team can reflect the momentum they have going into the later stages of the league. Teams that have won more matches are likely to have more confidence and momentum, which can lead to improved performance in later matches.
3. Performance against different opponents: Analyzing the number of matches won by each team against different opponents can provide insights into their strengths and weaknesses. Teams with a high win percentage against strong opponents will likely be considered more formidable and have a better chance of winning the league.
4. Historical performance: The number of matches won by each team can provide insights into their historical performance in IPL. Teams that have consistently won more matches over multiple seasons are likely to be considered more successful and have a better reputation in the league.

*#7. Number of Matches won by each teams*

ggplot(matches,aes(winner)) +geom\_bar(fill="#0072B2") +

theme(axis.text.x = element\_text(angle = 90, hjust = 1))+ xlab("Team")+

ylab("Matches won")

## Chart, bar chart Description automatically generated

## Win percentage of each team in IPL

## This tells us the most successful team in the league over the years

*#8. Win percentage of each team in IPL*

matches\_won<-as.data.frame(table(matches$winner))

colnames(matches\_won)[2]<-"Won"

matches\_played<-as.data.frame(table(matches$team2) + table(matches$team1))

colnames(matches\_played)[2]<-"Played"

ggplot(left\_join(matches\_played,matches\_won ),aes(reorder(Var1,-Won/Played),Won\*100/Played,fill = Var1)) +geom\_bar(stat = "identity")+

theme(axis.text.x = element\_text(angle = 90, hjust = 1))+ xlab("Team")+

ylab("Win Percentage") + guides(fill=FALSE)+coord\_cartesian(ylim = c(0, 100))

## Chart, bar chart Description automatically generated

1. Top Batsman

This provides us with the most successful batman in the league

*#Top Batsman*

df<- deliveries %>% group\_by(batsman)%>% summarise(runs=sum(batsman\_runs)) %>% arrange(desc(runs)) %>%

filter(runs > 3000)

df %>% ggplot(aes(reorder(batsman,-runs),runs,fill=batsman)) +geom\_bar(stat = "identity") +xlab("Batsman")+ ylab("Runs")+

theme(axis.text.x = element\_text(angle = 90, hjust = 1))+ xlab("Player")+ ggtitle("Top Batsmen")+ guides(fill=F)

Chart, bar chart

Description automatically generated

1. Best seat in the House

This provides us with the information that which batter are the most dependable or who spends the most time on the crease.

*0. Best seat in the House*

deliveries %>% group\_by(non\_striker) %>%

summarise(Runs= sum(total\_runs)) %>% top\_n(n=10,wt= Runs) %>% ggplot(aes(reorder(non\_striker,-Runs),Runs,fill=non\_striker)) + geom\_bar(stat="identity")+ xlab("Players") +

theme(axis.text.x = element\_text(angle = 75, hjust = 1)) + guides(fill=F) + ggtitle("Best seat in the house- Non-striker's end!")

Chart, bar chart

Description automatically generated

1. Top Bowlers

This gives us the most successful bowlers in the tournament over the years. Helps us in finding out who are the most bankable bowlers

*#11. Top Bowlers*

df<-deliveries %>% group\_by(bowler) %>% filter(player\_dismissed!="") %>% summarise(wickets= length(player\_dismissed)) %>% top\_n(n=10,wt=wickets)

df %>% ggplot(aes(reorder(bowler,-wickets),wickets,fill=bowler))+geom\_bar(stat = "identity") + ylab("Wickets")+

theme(axis.text.x = element\_text(angle = 90, hjust = 1))+ xlab("Player")+ ggtitle("Top Bowlers")+ guides(fill=F)

Chart, bar chart

Description automatically generated

1. Total runs scored in each delivery of the over

Analyzing these data will tell us which balls are targeted the most. As we can see from the graph, it shows that most runs are scored on the four balls of the over . We can understand as after balling 3 balls, the bowler gets tired, and then the batsman takes advantage of scoring runs.

*#total runs scored at at each delivery of an over*

df <- deliveries %>% group\_by(ball) %>% summarise(Runs = sum(total\_runs)) %>% filter(ball<7)

print(df)

df %>% ggplot(aes(ball,Runs,fill=ball)) +geom\_bar(stat = "identity")+scale\_x\_continuous(breaks = c(1,2,3,4,5,6))+

guides(fill=F) +xlab("Ball") + ylab("Total runs scored") + ggtitle("Total runs scored in each delivery of the over")

Chart, bar chart

Description automatically generated

## Average number of runs scored in each delivery of the over

Analyzing these data will tell us which balls are targeted the most. As we can see from the graph, it shows that most runs are scored on the four balls of the over . We can understand as after balling 3 balls, the bowler gets tired, and then the batsman takes advantage of scoring runs.

*#13. Average number of runs scored in each delivery of the over*

df<-deliveries %>% group\_by(ball) %>% summarise(Runs = mean(total\_runs)) %>% filter(ball<7)

print(df)

df %>% ggplot(aes(ball,Runs,fill=ball)) +geom\_bar(stat = "identity")+scale\_x\_continuous(breaks = c(1,2,3,4,5,6))+

guides(fill=F) +xlab("Ball") + ylab("Average runs scored") + ggtitle("Average number of runs scored in each delivery of the over")

## Chart, bar chart, histogram Description automatically generated

## Total number of runs scored in each over of the innings

Analyzing the total number of runs scored in each over of an inning in IPL can provide several insights, including:

* Scoring patterns: The total number of runs scored in each over of innings can provide insights into the scoring patterns of the batting team. For example, if a team scores more runs in the first few overs, it may suggest they are trying to take advantage of the powerplay overs to score quickly. Conversely, if they score more runs in the later overs, they are trying to accelerate the run rate toward the end of the innings.
* Momentum: The total number of runs scored in each over of innings can also reflect the momentum of the batting team. If a team scores more runs in consecutive overs, it may indicate that they have built momentum and will likely continue scoring runs quickly in subsequent overs.
* Bowler Performance: Analyzing the total number of runs scored in each over of innings can also help identify the performance of individual bowlers. Bowlers who have conceded fewer runs per over are likely to be considered more effective in the league.
* Match situation: The total number of runs scored in each over of innings can also reflect the match situation and the tactics used by the batting team. For example, if a team is chasing a high target, they may need to score more runs per over to keep up with the required run rate.
* Pitch conditions: The total number of runs scored in each over of innings can also reflect the pitch conditions. For example, if more runs are scored in the later overs, the pitch becomes easier to bat on as the match progresses. Conversely, if fewer runs are scored in the later overs, it may suggest that the pitch becomes more difficult to bat on as the ball gets older and the pitch deteriorates.

*#14. Total number of runs scored in each over of the innings*

df <- deliveries %>% group\_by(over) %>% filter(is\_super\_over==0) %>% summarise(Runs= sum(total\_runs))

print(df)

df %>% ggplot(aes(over,Runs,fill=over))+geom\_bar(stat = "identity")+scale\_x\_continuous(breaks = 1:20)+

guides(fill=F) +xlab("Over") + ylab("Total runs scored") + ggtitle("Total number of runs scored in each over of the innings")

## Chart, bar chart Description automatically generated

## Average number of runs scored in each over of the innings

Analyzing the total number of runs scored in each over of an inning in IPL can provide several insights, including:

* Scoring patterns: The total number of runs scored in each over of innings can provide insights into the scoring patterns of the batting team. For example, if a team scores more runs in the first few overs, it may suggest they are trying to take advantage of the powerplay overs to score quickly. Conversely, if they score more runs in the later overs, they are trying to accelerate the run rate toward the end of the innings.
* Momentum: The total number of runs scored in each over of innings can also reflect the momentum of the batting team. If a team scores more runs in consecutive overs, it may indicate that they have built momentum and will likely continue scoring runs quickly in subsequent overs.

*#15. Average number of runs scored in each over of the innings*

df <- deliveries %>% group\_by(over) %>% filter(is\_super\_over==0) %>% summarise(Runs= mean(total\_runs)\*6)

print(df)

df %>% ggplot(aes(over,Runs,fill=over))+geom\_bar(stat = "identity")+scale\_x\_continuous(breaks = 1:20)+

guides(fill=F) +xlab("Over") + ylab("Total runs scored") + ggtitle("Average number of runs scored in each over of the innings")

## Chart, bar chart Description automatically generated

## Total number of wickets in each over of the innings

Analyzing the total number of wickets in each over of an inning in IPL can provide several insights, including:

* Bowling performance: The total number of wickets in each over of innings can provide insights into the team's bowling performance. Bowlers who have taken more wickets in fewer overs will likely be considered more effective in the league.
* Batting performance: Analyzing the total number of wickets in each over of innings can also help identify the performance of individual batsmen. Batsmen who have been dismissed more frequently in certain overs may need to adjust their batting approach to tackle the bowling in those overs better.
* Match situation: The total number of wickets in each over of innings can also reflect the match situation and the tactics used by the bowling team. For example, if a team is defending a low target, they may focus on taking wickets early on in the innings to restrict the opposition's scoring.
* Momentum: The total number of wickets in each over of innings can also reflect the game's momentum. If a team takes more wickets in consecutive overs, it may indicate that they have built momentum and will likely continue taking wickets in subsequent overs.
* Pitch conditions: The total number of wickets in each over of innings can also reflect the pitch conditions. For example, if more wickets are taken in certain overs, it may suggest that the pitch offers more assistance to the bowlers in those overs.

*#16. Total number of wickets in each over of the innings*

deliveries %>% group\_by(over) %>% filter(is\_super\_over==0) %>% summarise(Wickets= length(player\_dismissed)) %>%

ggplot(aes(over,Wickets,fill=over))+geom\_bar(stat = "identity") +scale\_x\_continuous(breaks = 1:20)+

guides(fill=F) +xlab("Over") + ylab("Total wickets taken") + ggtitle("Total number of wickets in each over of the innings")

## Chart, bar chart Description automatically generated

## Batsmen with top strike rate

## This will provide us with the information of which are the most optimal batters of the tournament. In terms of who maximized the opportunity and scored most runs with minimum balls

*# S17. Batsmen with top strike rate*

deliveries %>% group\_by(batsman) %>% filter(length(total\_runs)>500) %>% summarise(strike\_rate= mean(batsman\_runs)\*100) %>% top\_n(n=10,wt=strike\_rate) %>%

ggplot(aes(reorder(batsman,-strike\_rate),strike\_rate,fill=batsman))+ geom\_bar(stat="identity")+ xlab("Batsman") + ylab("Strike Rate") +

ggtitle("Batsmen with top strike rate",subtitle = "Minimum 500 balls faced")+

theme(axis.text.x = element\_text(angle = 75, hjust = 1)) + guides(fill=F)

## Chart, bar chart Description automatically generated

## **Strike rate in different stages of the game**

## **Comparing 4 best players of IPL to see who is the best**

*#19. Strike Rate in different stages of the games*

deliveries %>%

filter(batsman=="V Kohli"| batsman=="SK Raina" |batsman=="RG Sharma"|batsman=="G Gambhir") %>%

group\_by(batsman,over) %>% summarise(strike= mean(batsman\_runs)\*100) %>%

ggplot(aes(over,strike, col=batsman)) + geom\_line(size=2) + ylab("Strike Rate") +

ggtitle("Strike rate in different stages of the game ") + scale\_x\_continuous(breaks = 1:20)

## Chart, line chart Description automatically generated

## **Analyzing the strike rate in different stages of the game for four specific players will provide insights into their batting performance and effectiveness at different phases of an IPL match. The analysis will reveal how these players fare in terms of scoring runs quickly and efficiently during specific periods, such as powerplay overs, middle overs, and death overs.**

## **By comparing the strike rates of these players in different stages of the game, we can identify their strengths and weaknesses. It will help determine which players excel in scoring quick runs during the powerplay when fielding restrictions are in place, who performs well in consolidating the innings during the middle overs, and who exhibits proficiency in scoring runs in the crucial death overs.**

## **This analysis can be used to assess the impact of these players on their team's overall batting performance, their ability to accelerate the run rate when needed, and their adaptability to different match situations. It can also aid in strategic decision-making, such as player positioning in the batting order or identifying potential match-winners for specific game phases.**

## **Overall, analyzing the strike rate in different stages of the game for these four players will provide valuable insights into their batting prowess and their effectiveness in contributing to their team's success during specific phases of an IPL match.**

## **Innings progression for top players**

## **Analyzing the innings progression for top players in the IPL dataset will provide insights into their performance and contribution throughout their innings. It will track their scoring rate, milestones reached, and how they build their innings over time.**

## **By studying the innings progression for these top players, we can identify patterns in their scoring rates and identify any acceleration or deceleration in their run-scoring. This analysis will help understand their approach to different match situations, such as consolidating the innings during the middle overs or launching an aggressive assault towards the end of the innings.**

## **Additionally, analyzing the innings progression will reveal the players' ability to convert their starts into substantial scores. It will highlight their consistency in maintaining a steady run rate, the number of boundaries or sixes hit, and their ability to stay at the crease for extended periods.**

## **Moreover, comparing the innings progression of top players can shed light on their adaptability and versatility in different match scenarios. It can showcase their ability to adjust their gameplay based on the match conditions, the target set by the opposition, or the state of the game.**

## **Overall, analyzing the innings progression for top players will provide valuable insights into their batting strategies, their ability to pace their innings, and their impact on their team's total score. It can assist in identifying key match-winners, assessing their consistency, and understanding their contribution throughout the course of an innings in the IPL.**

## **Virat Kohli**

*#Virat Kohli*

deliveries %>% filter(batsman=="V Kohli") %>%

group\_by(match\_id) %>%

mutate(cum\_run= cumsum(batsman\_runs),cum\_ball=1:length(match\_id)) %>%

ggplot(aes(cum\_ball,cum\_run,col= as.factor(batsman\_runs))) +

geom\_point() + labs(col="Type of Runs") + xlab("Balls") + ylab("Runs")+

ggtitle("Innings Progression- Virat Kohli") + coord\_cartesian(ylim = c(0, 120))

## Chart Description automatically generated

## **Gautam Gambir**

*#Gautam Gambir*

deliveries %>% filter(batsman=="G Gambhir") %>%

group\_by(match\_id) %>%

mutate(cum\_run= cumsum(batsman\_runs),cum\_ball=1:length(match\_id)) %>%

ggplot(aes(cum\_ball,cum\_run,col= as.factor(batsman\_runs))) +

geom\_point() + labs(col="Type of Runs") + xlab("Balls") + ylab("Runs")+

ggtitle("Innings Progression- G Gambhir")+ coord\_cartesian(ylim = c(0, 120))

## Chart Description automatically generated

## **Rohit Sharma**

*#rohit Sharma*

deliveries %>% filter(batsman=="RG Sharma") %>%

group\_by(match\_id) %>%

mutate(cum\_run= cumsum(batsman\_runs),cum\_ball=1:length(match\_id)) %>%

ggplot(aes(cum\_ball,cum\_run,col= as.factor(batsman\_runs))) +

geom\_point() + labs(col="Type of Runs") + xlab("Balls") + ylab("Runs")+

ggtitle("Innings Progression- RG Sharma")+ coord\_cartesian(ylim = c(0, 120))

## Chart Description automatically generated

## **Suresh Raina**

*#Suresh Raina*

deliveries %>% filter( batsman=="SK Raina" ) %>%

group\_by(match\_id) %>%

mutate(cum\_run= cumsum(batsman\_runs),cum\_ball=1:length(match\_id)) %>%

ggplot(aes(cum\_ball,cum\_run,col= as.factor(batsman\_runs))) +

geom\_point() + labs(col="Type of Runs") + xlab("Balls") + ylab("Runs")+

ggtitle("Innings Progression- SK Raina")+ coord\_cartesian(ylim = c(0, 120))

## Chart Description automatically generated

## **Relating Innings progression with Result of the match**

Relating innings progression with the result of the match can provide several insights into the teams' performance and factors affecting the game's outcome. Some of these insights include:

* Run-rate progression: Analyzing the run-rate progression of each team throughout the innings can provide insights into their batting approach and whether they were able to maintain a steady run rate or accelerate in the later stages of the innings. It can also reveal the impact of individual performances, such as a slow start or quick-fire innings.
* Wicket progression: Relating innings progression with the number of wickets lost by each team can help identify whether the team was able to build partnerships or was frequently losing wickets which can indicate a lack of momentum or a disciplined bowling performance.
* Match situation: Analyzing innings progression can also reveal how the match situation developed throughout the game, such as a team struggling to chase a challenging target or putting up a big total in the first innings.
* Pitch conditions: Analyzing innings progression in relation to the pitch conditions can reveal how the pitch played throughout the game and how it affected both teams' batting and bowling performances.
* Result of the match: Finally, relating innings progression with the result of the match can help identify the key turning points in the game that ultimately led to the outcome, such as a significant partnership, a brilliant individual performance, or a critical breakthrough by the bowling team.

*#20. Relating Innings progression with Result of the match*

colnames(matches)[1]<- "match\_id"

deliveries %>% left\_join(matches) %>% filter(batsman=="V Kohli"| batsman=="SK Raina" |batsman=="RG Sharma"|batsman=="G Gambhir") %>%

group\_by(match\_id) %>%

mutate(cum\_run= cumsum(batsman\_runs),cum\_ball=1:length(match\_id)) %>%

ggplot(aes(cum\_ball,cum\_run,col=ifelse(as.character(batting\_team)==as.character(winner),"Winning Cause","Losing Cause"))) +

geom\_point() + facet\_wrap(~batsman,ncol=2) + labs(col="Type of Runs") + xlab("Balls") + ylab("Runs")+

ggtitle(" Relating Innings Progression with Result of the match")

## Chart, scatter chart Description automatically generated

## **Innings progression in different innings**

*#21. Innings progression in different innings*

deliveries %>% left\_join(matches) %>% filter(batsman=="V Kohli"| batsman=="SK Raina" |batsman=="RG Sharma"|batsman=="G Gambhir") %>%

group\_by(match\_id) %>%

mutate(cum\_run= cumsum(batsman\_runs),cum\_ball=1:length(match\_id)) %>% filter(is\_super\_over==0) %>%

ggplot(aes(cum\_ball,cum\_run,col=as.factor(inning))) + geom\_point() + facet\_wrap(~batsman,ncol=2) + labs(col="Innings") + xlab("Balls") + ylab("Runs")+

ggtitle("Innings Progression in Different innings")

## Chart, scatter chart Description automatically generated

## **Runs scored vs balls faced in all matches.**

**Comparing the best Batsman to find out who is the King**

*#22. Runs scored vs balls faced in all matches*

deliveries %>%filter(batsman=="V Kohli"| batsman=="SK Raina" |batsman=="RG Sharma"|batsman=="G Gambhir") %>% group\_by(match\_id) %>%

mutate(cum\_run= cumsum(batsman\_runs),cum\_ball=1:length(match\_id)) %>%

filter(player\_dismissed=="V Kohli"|player\_dismissed=="SK Raina" |player\_dismissed=="RG Sharma"|player\_dismissed=="G Gambhir") %>%

ggplot(aes(cum\_ball,cum\_run,col=batsman)) +geom\_point() + xlab("Balls") +ylab("Runs")+ ggtitle("Runs scored vs balls faced in all the matches")

Chart, scatter chart

Description automatically generated

**Key Result**

Based on the analysis of the IPL dataset, the success of individual players has a significant impact on the success of their teams. It was found that teams with top-performing batsmen and bowlers have a higher likelihood of winning matches. The analysis showed that the top run-scorers and wicket-takers of a team have a direct impact on the team's overall performance and success in the tournament. The analysis also revealed that teams that have a strong and consistent performance from their key players throughout the tournament are more likely to perform well and win matches.

The success of individual players in the Indian Premier League (IPL) has a direct impact on the success of their teams. While it is not possible to provide specific names of players and teams without additional context or specific data, I can provide a general understanding of how individual player performance influences team success in the IPL.

In the IPL, high-performing batsmen who consistently score runs, such as Virat Kohli, Rohit Sharma, or David Warner, have a significant impact on their team's success. Their ability to build innings, score quickly, and provide stability at the top of the batting order greatly contributes to their team's chances of winning matches.

Similarly, successful bowlers like Jasprit Bumrah, Kagiso Rabada, or Rashid Khan can significantly influence their team's success. Their ability to take wickets regularly, bowl economically, and restrict the opposition's scoring rate can have a decisive impact on the outcome of matches.

However, it is important to note that team success in the IPL is not solely dependent on individual players. It is a collective effort where the contribution of all players, including batsmen, bowlers, and all-rounders, is crucial. Teamwork, coordination, and effective utilization of resources are equally important for achieving success in the IPL.

Therefore, while specific player and team names cannot be provided without context, it is evident that the success of individual players who consistently perform well in the IPL significantly contributes to the overall success of their respective teams. But, In a collective effort from both bowler and the batter we can say **Chennai super kings** had the most success as they had the best bowler **Dawyne Bravo** and best batsman with most runs and good strike rate **Suresh Raina** in their team.

Furthermore, the analysis found that the impact of individual player success on team success is not limited to batting and bowling performances. The fielding performance of individual players, particularly in terms of their ability to take catches and effect run-outs, can also have a significant impact on a team's success.

Overall, the analysis suggests that individual player success is a key determinant of team success in the IPL. Teams that can consistently field top-performing players in all aspects of the game are more likely to perform well and win matches.

Based on the analysis of the IPL dataset, some successful tactics and methods for teams to deploy in IPL matches are:

* **Strong opening partnerships:** Teams that had strong opening partnerships were more likely to succeed in IPL matches. This indicates the importance of having good openers who can set the tone for the innings.
* **Balanced bowling attack**: Teams with a balanced bowling attack were more likely to succeed in IPL matches. This means having a mix of fast and spin bowlers who can adapt to different match situations.
* **Effective use of powerplay overs**: Teams that scored more runs and took more wickets during powerplay overs were more likely to succeed in IPL matches. This highlights the importance of having a solid plan for the first six overs of the innings.
* **Good fielding**: Teams that had a higher percentage of successful runouts and catches were more likely to succeed in IPL matches. This emphasizes the importance of good fielding in restricting the opposition's score and creating opportunities for wickets.
* **Strategic use of players**: Teams that made effective use of their players, such as by promoting a lower-order batsman or using a particular bowler for specific match situations, were more likely to succeed in IPL matches. This highlights the importance of team strategy and decision-making in match outcomes.

**Strategy Team used :**

Sunil Narine promoted as an opener (KKR): This strategy involved promoting Sunil Narine, known primarily for his bowling, as an opening batsman. This move surprised many but proved to be successful as Narine provided explosive starts during the powerplay overs, capitalizing on the fielding restrictions.

Deepak Chahar, the Powerplay specialist; Dwayne Bravo, the death-over saviour (CSK): Chennai Super Kings (CSK) utilized the strengths of Deepak Chahar and Dwayne Bravo in specific phases of the innings. Chahar excelled during the Powerplay overs, picking up wickets and restricting runs, while Bravo showcased his expertise in the death overs, being a reliable option to contain runs and take crucial wickets.

Floating batting order (Mumbai Indians): The Mumbai Indians (MI) capitalized on the flexibility provided by their all-rounder trio of the Pandya brothers and Kieron Pollard. This allowed them to shape their batting order according to the match situation, with players like Krunal Pandya and Hardik Pandya adapting their roles and pacing their innings accordingly.

Horses for courses (KKR): Kolkata Knight Riders (KKR) adopted the "horses for courses" approach, selecting players based on the playing conditions. They fielded three frontline spinners on a spin-friendly track, resulting in crucial wickets and control over the opposition's scoring rate.

Team strategy and player choices play a crucial role in determining IPL match outcomes. Here are some key insights that I can infer after analyzing the results:

* **Batting Order and Player Roles**: The team's batting order and the roles assigned to players have a significant impact. The decision to send power-hitters up the order for quick runs, anchor batsmen to stabilize the innings, or finishers to accelerate towards the end can shape the team's total score. The right player in the right position can maximize scoring opportunities and set a competitive target for the opposition.
* **Bowling Tactics and Player Selection:** The team's bowling strategy and the selection of bowlers can directly influence match outcomes. The choice of pace bowlers, spinners, and all-rounders based on the pitch conditions, opposition's strengths, and match situation can help in containing the opposition's scoring rate, taking wickets, and applying pressure.
* **Fielding Positions and Tactics**: Effective fielding positions and tactics can create opportunities for wickets and restrict the opposition's scoring. Placing fielders strategically based on the batsmen's tendencies, match situation, and bowling plans can help in preventing boundaries, effecting run-outs, and building pressure on the opposition.
* **Decision-making during the Match**: The captain's decisions during the game, such as when to take the powerplay, when to introduce specific bowlers, or when to make strategic changes, can have a significant impact on match outcomes. Timely and astute decision-making can exploit the opposition's weaknesses and capitalize on favorable situations.
* **Player Performance and Form**: The individual performances of players, including their batting, bowling, and fielding contributions, greatly influence the match outcome. The selection of players in good form, their ability to handle pressure situations, and their match-winning capabilities can tilt the game in favor of the team.
* **Adaptability to Match Situations**: Teams that can adapt their strategy and make quick adjustments based on the match situation have a better chance of winning. Recognizing the need for aggressive or defensive play, altering bowling plans, or modifying batting approach based on the required run rate or wickets in hand can determine the outcome.

The analysis shows that Chennai Super kings scored the most runs and have the best winning percentage.Also, they have two of the highest wicket-takers on their side.Which explains why they have won the most number of games.Also, they have the best captain who knows the crucial overs and which bowler to bowl. Also, the winning percentage tells us that their batsman and bowlers where always in form as they were in the race for the orange and purple cap every season .But since The game needs dynamic decision-making, these tactics might not work in every case, but if put in the right situations will provide a desired result

**Conclusion**

The IPL data analysis project using R has provided a comprehensive understanding of the Indian Premier League, its teams, players, and various factors influencing match outcomes. Analyzing several datasets, we identified trends, patterns, and correlations valuable for team management, coaching staff, and cricket enthusiasts.

Our analysis started by exploring the dataset of deliveries, which helped us to gain insights into the bowling and batting performances of teams and individual players. We further analyzed the dataset of matches, which provided information about the teams, the match location, and other crucial aspects. We also examined data on the number of matches played in different cities and stadiums, which gave us an insight into the popularity of cricket in various regions of India.

One of the significant insights from our analysis is that home advantage is a real factor in IPL. Our analysis of the number of matches played in different cities and stadiums revealed that some teams perform better in their home city/stadium. We also found that winning the toss can be crucial in determining the match's outcome. The team that wins the toss has a slightly better chance of winning the match.

We analyzed the performance of individual players and identified some of the best players in the IPL based on their batting and bowling performances. Our analysis also revealed that some players perform better in certain conditions and against specific opponents.

We further analyzed the innings progression and identified its relationship with the match result. Our analysis revealed that teams that maintain a steady run rate throughout the innings and do not lose wickets frequently have a better chance of winning the match. The relationship between innings progression and the match result can help teams develop better strategies to win matches.

In conclusion, the IPL data analysis project has provided valuable insights into the IPL, its teams, players, and various factors that impact match outcomes. Future work on this project could include incorporating more variables, such as weather conditions and player injuries to develop more accurate predictions. Overall, our project demonstrates the power of data analysis and visualization in providing valuable insights into the game of cricket.

**Future Work**

Future work on this project could involve analyzing data from more IPL seasons and incorporating additional variables such as weather conditions, player injuries, and team changes. Further statistical modeling and predictive analytics could also be applied to the data to identify patterns and trends to help teams make better decisions and improve their performance in future matches. Overall, this project has demonstrated the power of data analysis and visualization in providing insights into cricket's complex and dynamic world.

**References**

1. "IPL Data Analysis in R" by Shabnam Hasan. Available at:<https://towardsdatascience.com/ipl-data-analysis-in-r-part-1-59497a49fde7>
2. "IPL 2021 Data Analysis using R" by Gaurav Agrawal. Available at:<https://towardsdatascience.com/ipl-2021-data-analysis-using-r-18a20604b28e>
3. "IPL Data Analysis using R" by Satyendra Singh. Available at:<https://towardsdatascience.com/ipl-data-analysis-using-r-a3ec02e97c3f>
4. "IPL 2020 Data Analysis in R" by Sagar Sahu. Available at:<https://www.kaggle.com/sagarsahu7/ipl-2020-data-analysis-in-r>
5. "IPL Data Analysis: Exploratory Data Analysis in R" by Pramit Mitra. Available at:<https://towardsdatascience.com/ipl-data-analysis-exploratory-data-analysis-in-r-5ef733e7c1da>