IPL Data Visualisation using Tableau

PROJECT REPORT

18IPE415J – FOUNDATION OF ANALYTICS
III Year/ V Semester
Academic Year: 2023 -2024
By

PREETHI V(RA2111003010042)

ADITHYA P MALLYA (RA2111003010059)

Under the guidance of

Dr. A. REVATHI

Assistant Professor

Department of Computational Intelligence

FACULTY OF ENGINEERING AND TECHNOLOGY

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

Kattankulathur, Chengalpattu District

NOVEMBER 2023

COLLEGE OF ENGINEERING & TECHNOLOGY SRM INSTITUTE OF SCIENCE & TECHNOLOGY SRM NAGAR, KATTANKULATHUR- 603203, CHENGALPATTU DISTRICT

(Under Section 3 of UGC Act, 1956)

BONAFIDE CERTIFICATE

This is to certify that 18CSE415J – FOUNDATION OF ANALYTICS project report titled "IPL Data Visualisation using Tableau" is the bonafide work of PREETHI V (RA2111003010042) and ADITHYA P MALLYA (RA2111003010059) who undertook the task of completing the project within the allotted time.

SIGNATURE
Dr. A. Revathi
Course Faculty
Assistant Professor
Department of Computational Intelligence
SRM Institute of Science and Technology
Kattankulathur.

SIGNATURE
Dr. M. Pushpalatha
Head of the Department
Professor
Department of Computing Technologies
SRM Institute of Science and Technology
Kattankulathur.

ABSTRACT

The Indian Premier League (IPL) is a professional cricket league in India. It is the world's most-watched cricket league and the country's most valuable sports property. The IPL was founded in 2008 and is contested by ten teams, representing eight Indian cities and two from abroad.

This paper explores the use of Tableau to visualize IPL data. Tableau is a powerful data visualization tool that can be used to create interactive charts, graphs, and maps. In this paper, we will use Tableau to visualize the following aspects of IPL data:

- Team statistics: We will visualize the performance of each team over the years, including their win-loss records, run rates, and bowling averages.
- Player statistics: We will visualize the performance of individual players, including their batting and bowling averages, strike rates, and field saves.
- Match analysis: We will visualize the outcomes of individual matches, including the scores, wickets taken, and runs scored.

ACKNOWLEDGEMENT

We express our heartfelt thanks to our honorable Vice Chancellor Dr. C. MUTHAMIZHCHELVAN, for being the beacon in all our endeavors.

We would like to express my warmth of gratitude to our Registrar Dr. S. Ponnusamy, for his encouragement.

We express our profound gratitude to our Dean, College of Engineering and Technology, Dr. T. V.Gopal, for bringing out novelty in all executions.

We would like to express my heartfelt thanks to Chairperson, School of Computing Dr. Revathi Venkataraman, for imparting confidence to complete my course project

We are highly thankful to our Course project Faculty Dr.A.Revathi, Assistant Professor, Department of Computational Intelligence, for his/her assistance, timely suggestion and guidance throughout the duration of this course project.

We extend my gratitude to our HoD Dr.M.Pushpalatha, Professor, Department of Computing Technologies and my Departmental colleagues for their Support.

Finally, we thank our parents and friends near and dear ones who directly and indirectly contributed to the successful completion of our project. Above all, I thank the almighty for showering his blessings on me to complete my Course project.

TABLE OF CONTENTS

CHAPTERS	CONTENTS
1.	**INTRODUCTION**
	1.1)MOTIVATION
	1.2)OBJECTIVE
	1.3)PROBLEM STATEMENT
	1.4)SCOPE OF PROJECT
2.	**REQUIREMENTS**
3.	**DATASET DESCRIPTION**
4.	**EXPLORATORY DATA ANALYSIS**
	4.1) DATA VISUALIZATION
5.	**INTERACTIVE DASHBOARD USING TABLEAU**
6.	**CONCLUSION & FUTURE ENHANCEMENT**
7.	**REFERENCES**
APPENDIX-A	**SCREENSHOTS**

CHAPTER 1

INTRODUCTION

Introduction:

The Indian Premier League (IPL) is one of the most popular and exciting cricket tournaments in the world, captivating millions of fans globally with its electrifying matches and star-studded rosters. The IPL's wealth of data offers a treasure trove of insights into the intricacies of the game, the performance of teams and players, and the strategies that determine victory. This project aims to harness the power of Tableau, a powerful data visualization tool, to uncover hidden patterns, trends, and stories within the IPL data, providing a deeper understanding of the tournament's dynamics.

Motivation:

The motivation behind this project stems from the desire to delve into the rich tapestry of IPL data, transforming it into meaningful and impactful visualizations. The project seeks to:

- Uncover hidden patterns and trends: By analyzing the vast amounts of IPL data, we can identify patterns and trends that may otherwise be overlooked, revealing insights into team performance, player contributions, and match dynamics.
- Enhance data comprehension: Through the creation of compelling visualizations, we aim to make the data more accessible and understandable, allowing a wider audience to grasp the intricacies of the IPL.

- Gain a deeper understanding of the game: By analyzing the data from various angles, we can gain a more comprehensive understanding of the factors that influence match outcomes, player performance, and team strategies.

Objective:

The primary objective of this project is to create a series of interactive data visualizations that effectively communicate the insights derived from the IPL data. These visualizations should be:

- Engaging and informative: The visualizations should be visually appealing and easy to understand, captivating the viewer's attention and effectively conveying the data's message.
- Interactive and exploratory: The visualizations should allow users to interact with the data, enabling them to explore different perspectives and uncover new insights.
- Comprehensive and insightful: The visualizations should cover a wide range of aspects of the IPL, providing a holistic understanding of the tournament's dynamics.

Problem Statement:

The Indian Premier League, or IPL, is a T20 cricket league, which was founded in 2008 and is held every year. It sees participation from both national and international players, and eight teams representing eight Indian cities compete in a double round-robin format in the league stages, which are followed by playoffs. Over the years, IPL has become one of the most watched and most attended live sporting events all over the world.

You work as a data analyst at IFP, a nationally recognized news agency, which is based out of New Delhi, and provides news reports and feeds to magazines, newspapers, and TV broadcasters all over the country. The Sports Editor of the agency has approached you to build a Tableau dashboard of IPL statistics over the years since its inception to create an infographic for a newsletter that their team is working on. For this newsletter, in some cases, they will use the visual representations for their infographic, and in a few other cases, they will use important statistics after trying out the different filters and customizations that you have provided for interactivity. Therefore, you are expected to build an interactive dashboard in Tableau for this purpose

Scope of project:

The scope of this project encompasses the following aspects of IPL data visualization:

- Team performance: Analyzing win-loss records, run rates, bowling averages, and other metrics to assess team performance across seasons.
- Player performance: Evaluating batting and bowling averages, strike rates, field saves, and other player-specific statistics to identify top performers and emerging talents.
- Match analysis: Visualizing match outcomes, scores, wickets taken, runs scored, and other match-related data to understand game dynamics and factors influencing victory.
- Exploring trends: Identifying patterns and trends over time, such as changes in batting and bowling techniques, team strategies, and the impact of new rules and regulations.

- Creating interactive dashboards: Developing interactive dashboards that allow users to explore the data in a variety of ways, filtering, drilling down, and uncovering hidden insights.

CHAPTER 2

REQUIREMENTS

- Dataset: A diverse and representative dataset of resumes categorized by professional fields (e.g., IT, Marketing, Engineering, Finance, etc.). The dataset should be prepared in a structured format, allowing for easy integration into the project.
- Tableau Desktop: Tableau Desktop is the main software tool used to create the visualizations. A free trial version is available for download.
- Data preparation tools: Data preparation tools may be required to clean and transform the data before it can be imported into Tableau.

CHAPTER 3

DATASET DESCRIPTION

Datasets provided

• **matches.csv** - It contains match-level information for each, and every match held in IPL from 2008 to 2017.
• **deliveries.csv** - It contains ball-by-ball information for each of the matches.

Each row in deliveries.csv data set contains

- Match related information (batting team vs bowling team)
- Player information (Bowler, Batsman, Non-striker)
- Delivery information (Runs scored, Wickets, Extras, etc.)

Each row in matches.csv data set contains:

- Match specifics (umpires, ground, etc.)
- Results (which team won, win type,

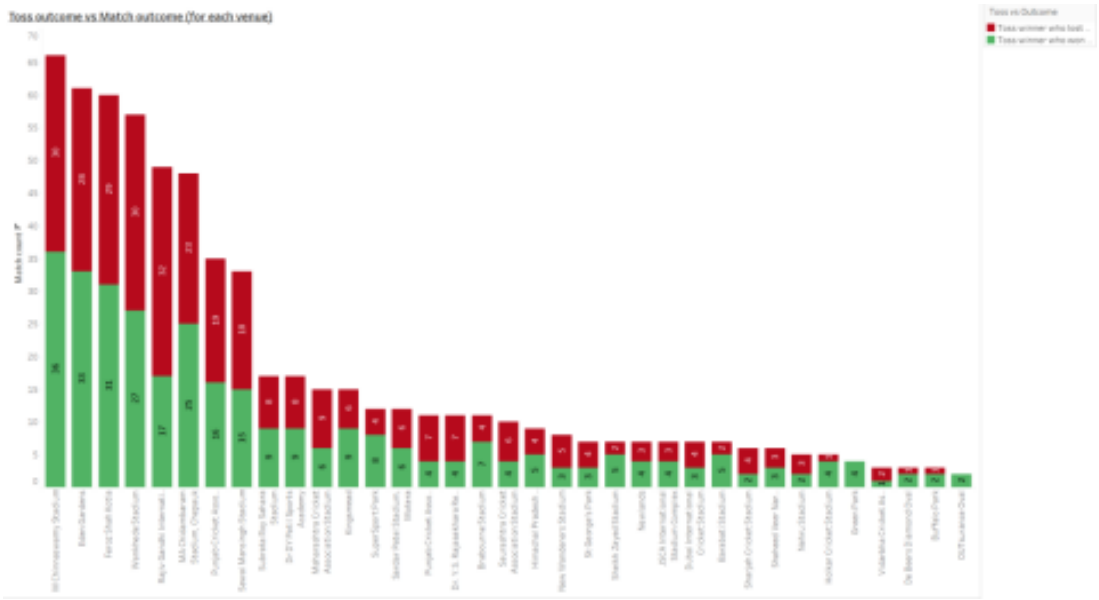
player of the match, etc.)

CHAPTER 4

EXPLORATORY DATA ANALYSIS

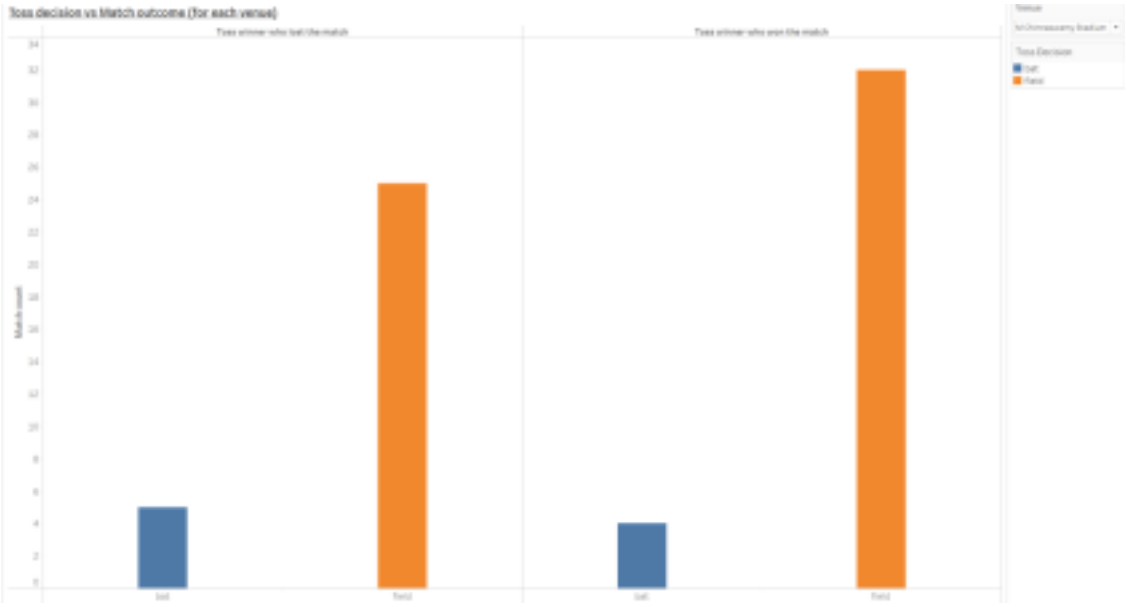
Match Statistics

Toss outcome vs Match outcome (for each Ground/Venue)



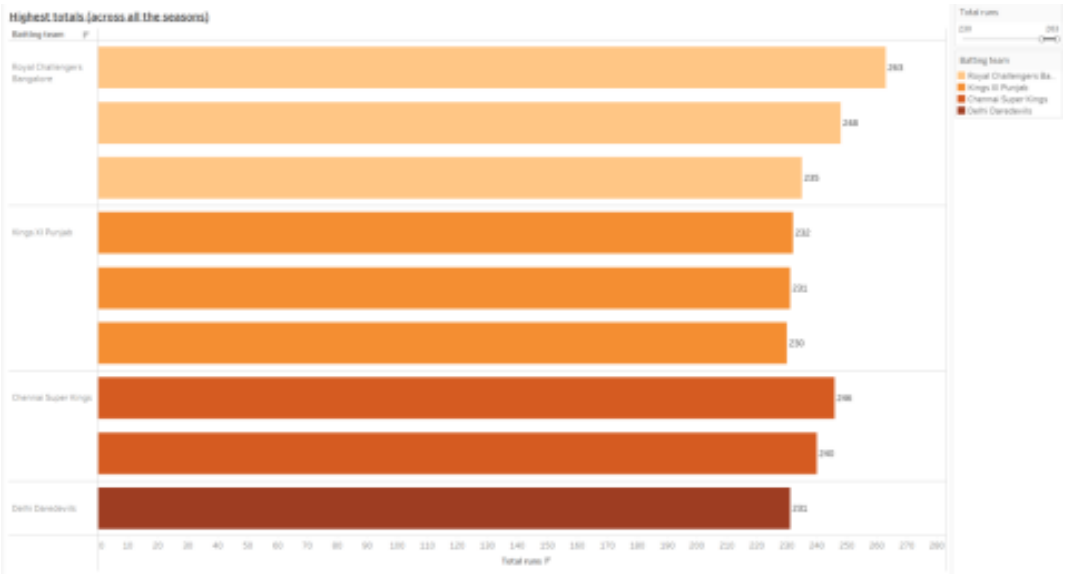
The above graph indicates the number of matches the toss winner won and lost the matches for each venue. Toss winner is derived from the data. The green color indicates the number of matches the toss winner won and red color indicates the number of matches the toss winner lost against each venue. Maximum number of matches were played at M Chinnaswamy, Eden gardens and at Feroz Shah Kotla stadiums. Maximum number of matches were won at M Chinnaswamy stadium and maximum matches were lost at Rajiv Gandhi International stadium.

Toss outcome vs Match outcome (for each Ground/Venue)



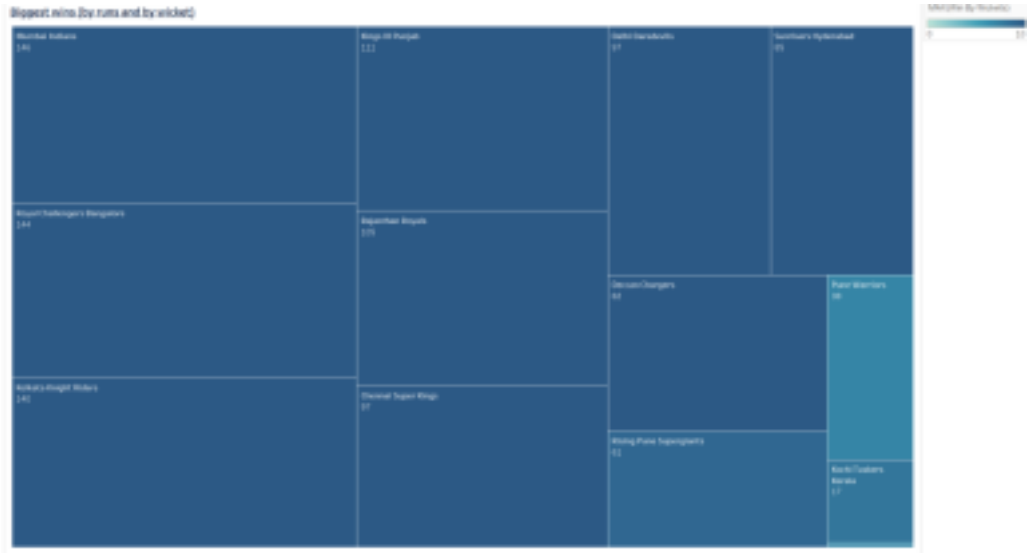
The graph displays the toss winners who chose to bat and field respectively and the corresponding number of matches they lost or won. Toss winner is derived from the data. The color shows the details of the toss decision, and the data is filtered on Venue. Toss winners at M Chinnaswamy stadium choose to field first according to the above graph.

Highest totals (across all the seasons)



The total runs score in each match is calculated and the highest runs are filtered based on ranges of total. Color indicates the different batting teams. RCB has the highest scores of 263 and 248 across all seasons.

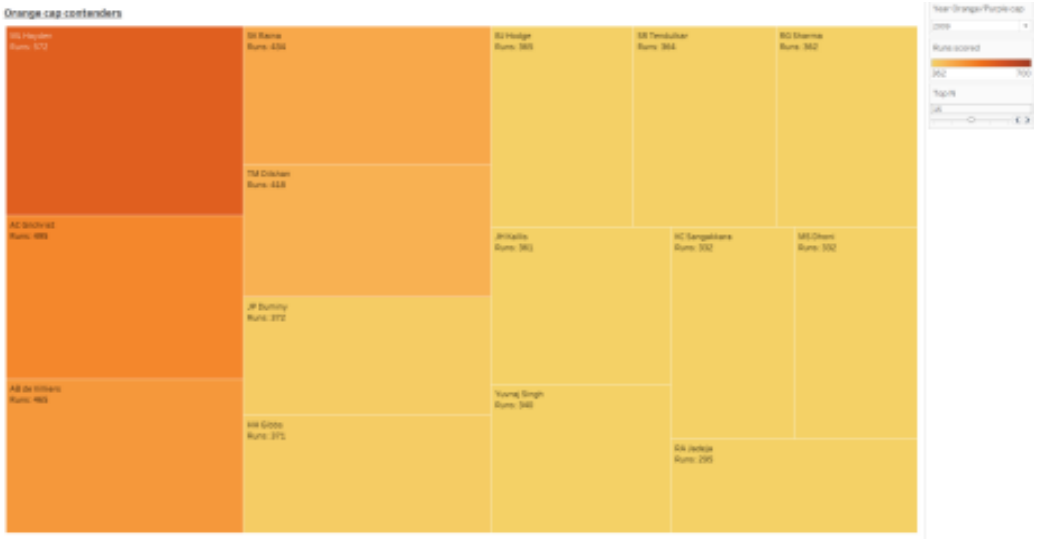
Biggest wins (by runs and by wicket)



In the above graph, maximum Win by Wickets is indicated by color intensity. The size of each tile indicates the maximum Win by Runs. The maximum win by wickets is 10 and the maximum win by runs is by Mumbai Indians.

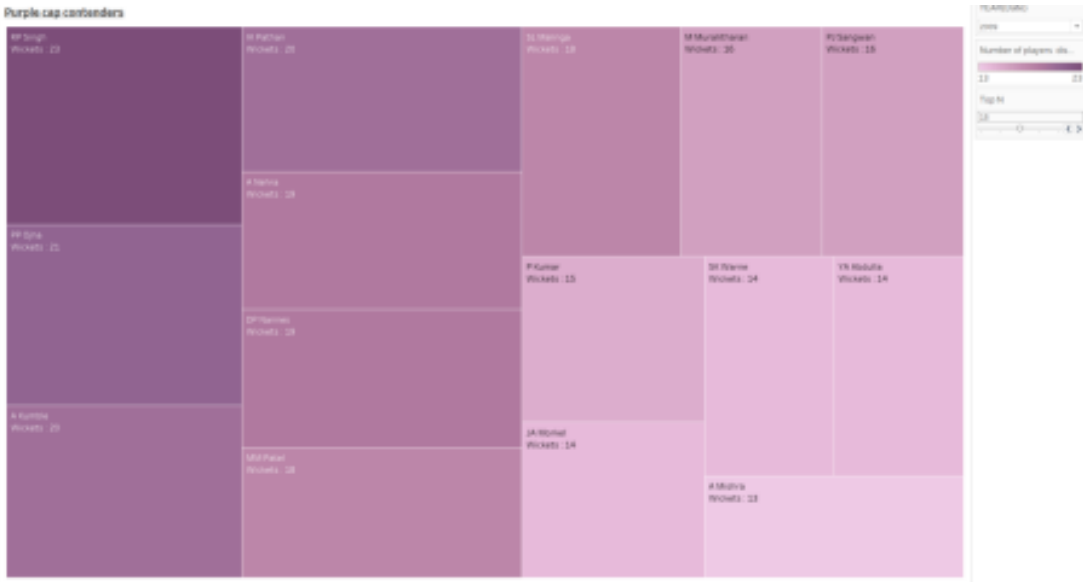
Player Statistics

Orange Cap contenders (The batsmen who have scored the maximum number of runs in a particular season)



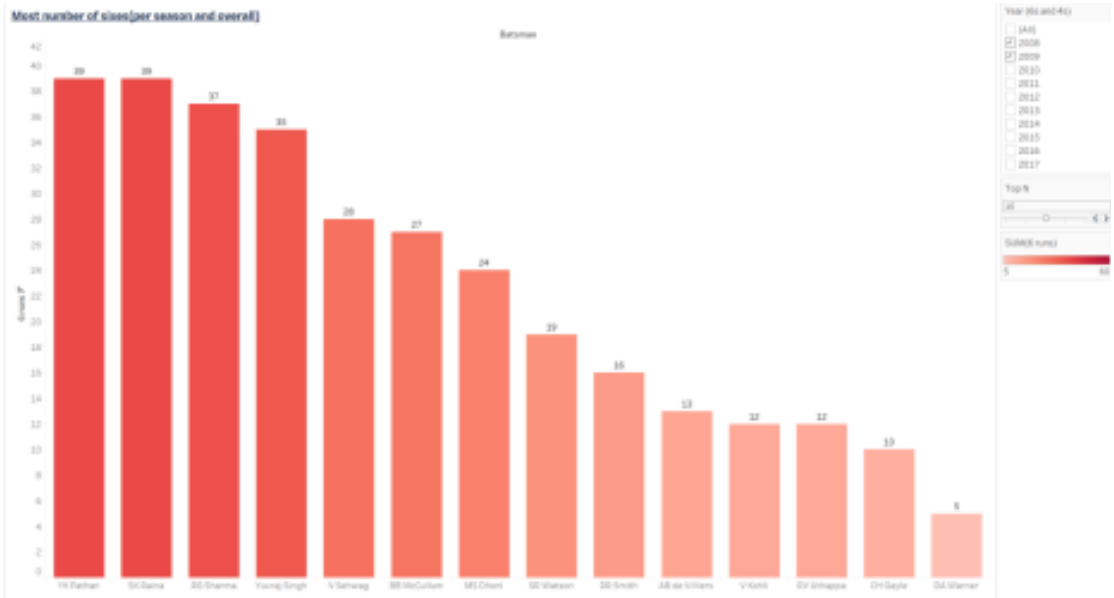
Color and intensity of the graph indicates the total runs scored by a batsman for a corresponding year. The context is filtered using Year and "top N" orange cap contenders are displayed. In 2009, the orange cap contender was ML Hayden who had amassed 572 runs in that season followed by AC Gilchrist who had scored 495 runs in the entire season. Although treemaps are not the best fit for the above visualization, it is visually appealing for purple cap and orange cap contenders.

Purple Cap contenders (The bowlers who have taken the maximum number of wickets in a particular season)



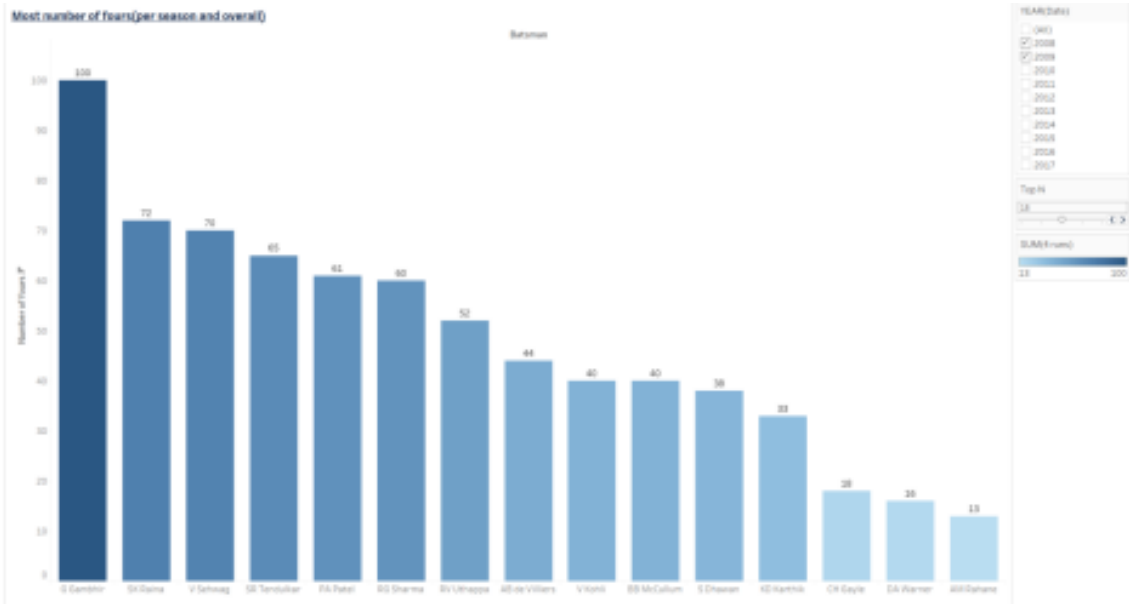
Color and intensity of the graph indicates the total numbers of players dismissed (wickets taken) for a corresponding year. The context is filtered using Year and "top N" purple cap contenders are displayed. In 2009, RP Singh has taken 23 wickets followed by PP Ojha who has taken 21 wickets in that season.

Batsmen who have hit the most number of sixes (per season and overall)



The graph displays the number of sixes that each batsman has hit. The data is filtered on Year and “top N” number of batsmen are displayed. In the graph, a cumulative sum of number of sixes in 2008 and 2009 is displayed. G Gambhir has hit a total of 484 fours across all IPL seasons.

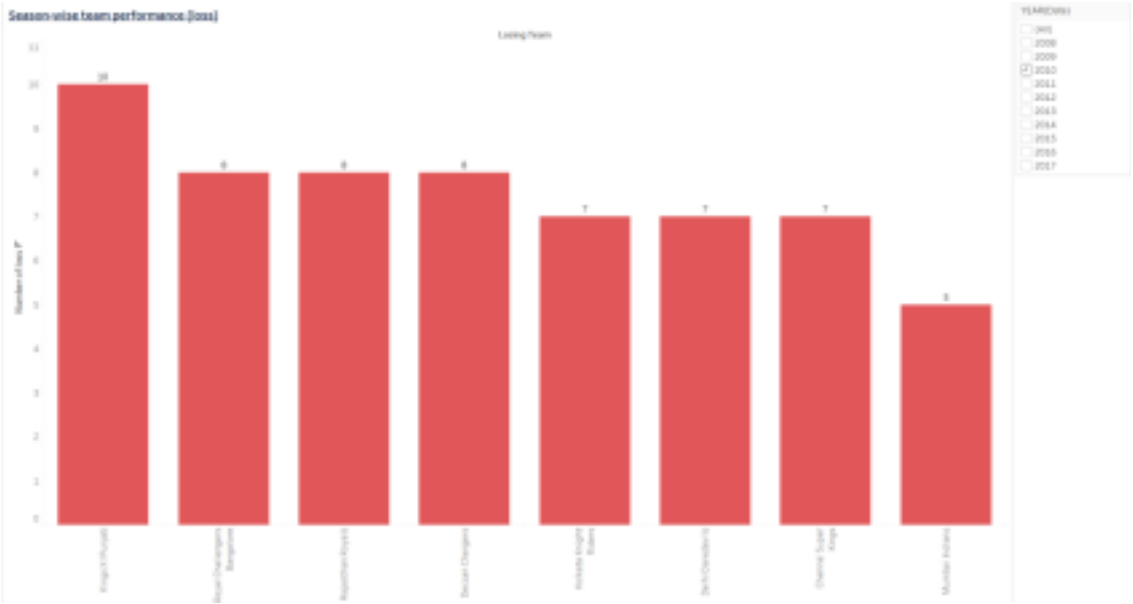
Batsmen who have hit the most number of fours (per season and overall)



The graph displays the number of fours that each batsman has hit. The data is filtered on Year and “top N” number of batsmen are displayed. In the graph, a cumulative sum of the number of fours in 2008 and 2009 is displayed. CH Gayle has hit a total of 266 fours across all IPL seasons.

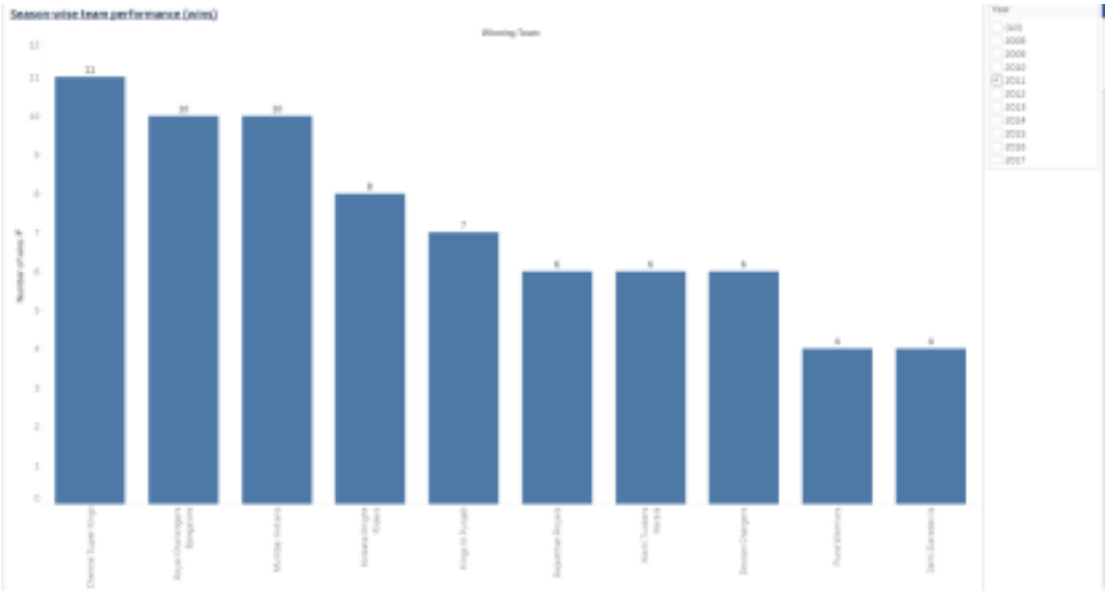
Team Statistics

Season-wise team performance (losses)



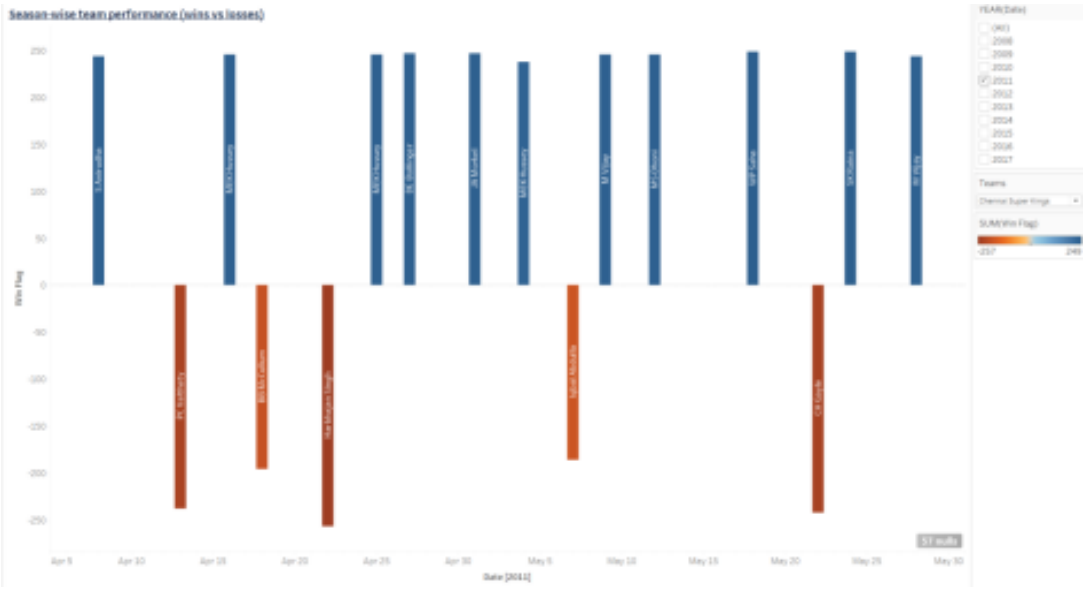
The above graph displays the total number of matches lost by each team participating in a particular season. The data is filtered on date. In 2011, Pune
Warriors and Delhi Daredevils had lost 9 matches. Delhi Daredevils had lost 83 matches in total across all seasons and were the team with who lost the highest number of matches.

Season-wise team performance (wins)



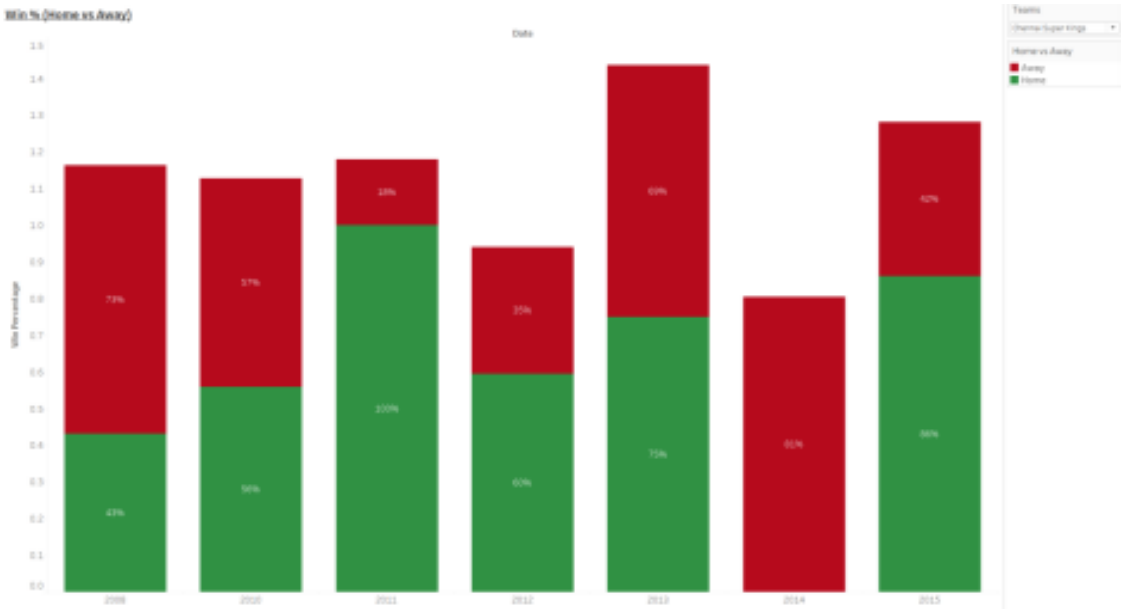
The above graph displays the total number of matches won by each team participating in a particular season. The data is filtered on date. In 2011, CSK had won 11 matches. Mumbai Indians had won 92 matches in total across all seasons and were the team with who won the highest number of matches.

Season-wise team performance (wins vs losses)



The above graph shows the Performance of Team CSK for the Year 2011. The blue bar indicates the matches where CSK won, and red bar indicates the matches where CSK lost. Player of the match is indicated by the label on each bar. In 2011, CSK played 16 matches and it won 11 matches that year. No data is available for CSK and RR in 2016 and 2017 as they had been suspended for two years.

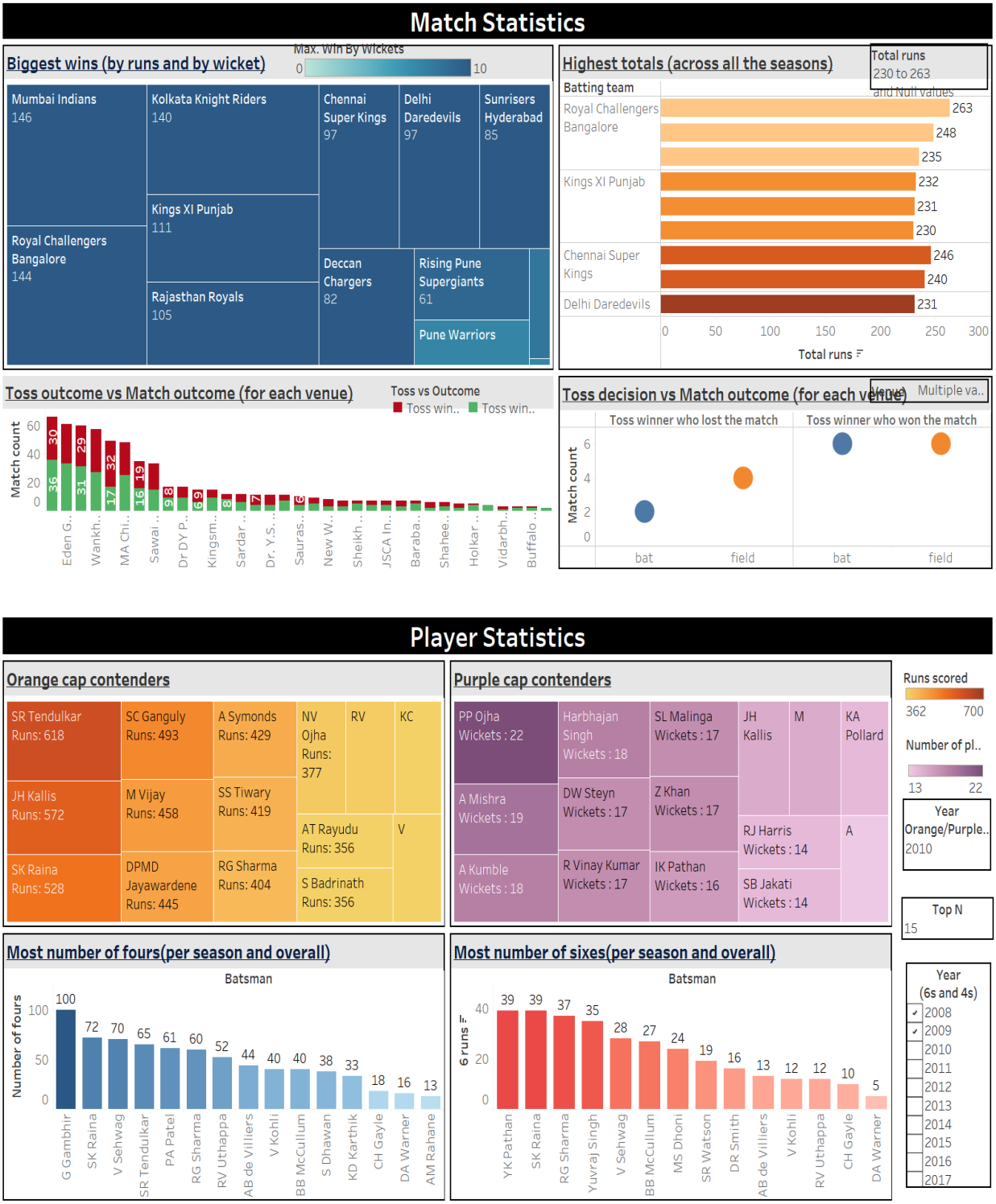
Win %age (home vs away)



The above graph displays the percentage of matches that CSK has won and lost in Home and Away games. In 2011, CSK won all the games played at Chennai.

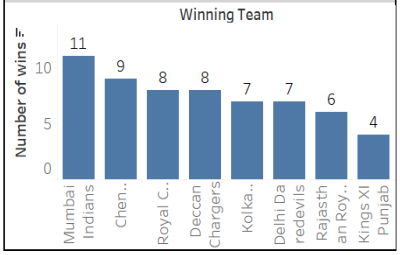
CHAPTER 5

INTERACTIVE DASHBOARD USING TABLEAU

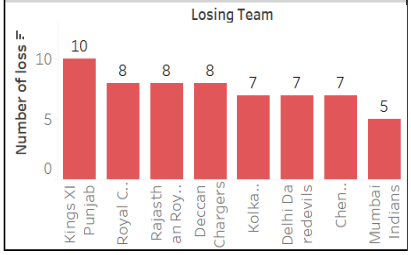


Team Statistics

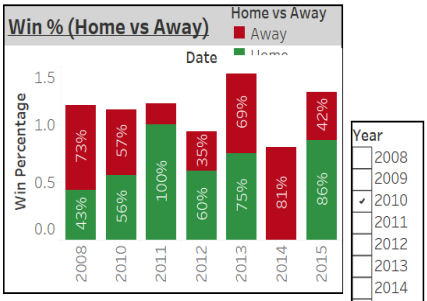
Season-wise team performance (wins)



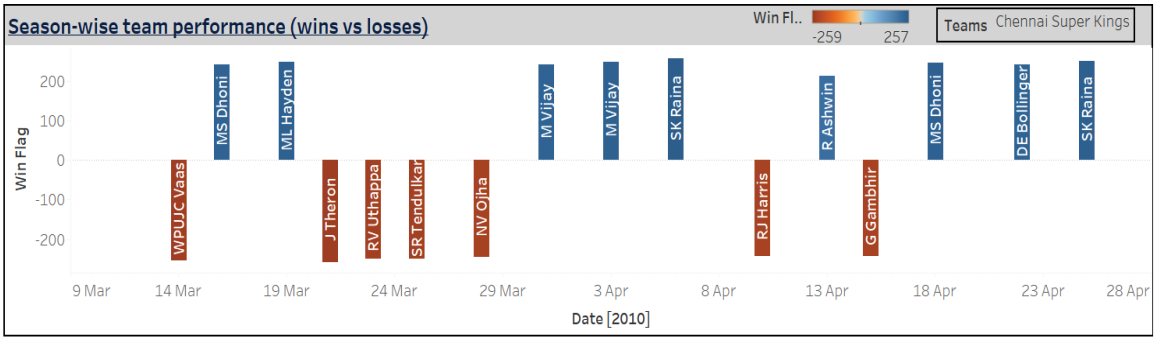
Season-wise team performance (loss)



Win % (Home vs Away)



Season-wise team performance (wins vs losses)



CHAPTER 4

CONCLUSION

RESULTS:

• The highest number of matches were played in Mumbai, Kolkata, Bangalore, and Delhi.

• Maximum number of matches were played at M Chinnaswamy stadium. Toss winners who chose to field have won the matches the most.

• Maximum number of matches were won at M Chinnaswamy stadium and maximum matches were lost at Rajiv Gandhi International stadium.

• RCB holds the highest score in the IPL history of 263 & 248 in the year 2013 and 2016.

• The maximum win by wickets is 10 and the maximum win by runs is by Mumbai Indians. Overall, 10 teams have won the matches by 10 wickets.

• V Kohli holds the record for highest runs in a single season of IPL. • SL Malinga holds the record for taking highest number of wickets in a single season of IPL.

• Gautam Gambhir has hit maximum number of fours across all IPL seasons. • Chris Gayle has hit most no of sixes across all IPL seasons. • The highest number of matches won in a season is 13 and this was achieved by RR in 2008 and MI in 2013.

• The highest number of matches lost in a season is 13 and DD performed the worst in 2013.

• Mumbai Indians and Chennai Super Kings are the most successful teams in the IPL history followed by Kolkata Night Riders.

• CSK is the most successful team in IPL history. Despite a two-year suspension, they are among the top 2 teams who have won the maximum number of matches.

FUTURE ENHANCEMENT:

Enhancing Data Sources and Scope

- Expand data sources: Incorporate data from additional sources, such as player profiles, social media engagement, and fan sentiment, to provide a more comprehensive understanding of the IPL ecosystem.
- Incorporate real-time data: Integrate real-time data feeds to provide live visualizations of match statistics, player performance, and fan reactions, enabling real-time analysis and insights.
- Explore international cricket data: Expand the scope of analysis to include data from international cricket matches, allowing for comparisons of player performance and team strategies across different leagues and formats.

Advanced Visualization Techniques

- Implement interactive maps: Create interactive maps to visualize team locations, player hometowns, and fan demographics, enhancing geographical insights.
- Utilize advanced chart types: Explore advanced chart types, such as heatmaps, treemaps, and Sankey diagrams, to represent complex data relationships and trends effectively.
- Incorporate motion graphics: Integrate motion graphics into visualizations to highlight key moments, trends, and patterns, making the data more engaging and impactful.

Machine Learning Integration

- Apply machine learning algorithms: Employ machine learning algorithms to predict match outcomes, identify player potential, and optimize team strategies, providing data-driven insights for decision-making.
- Automate data analysis: Automate repetitive data analysis tasks using machine learning tools, freeing up time for analysts to focus on more complex and strategic insights.
- Uncover hidden insights: Utilize machine learning techniques to uncover hidden patterns and trends within the data that may not be readily apparent through traditional analysis methods.

User-Centric Enhancements

- Personalize visualizations: Implement user preference settings to customize visualizations based on individual interests, such as favorite teams or players, enhancing user engagement.
- Enable data exploration tools: Provide interactive filters, drill-down options, and data exploration tools to empower users to explore the data independently and discover their own insights.
- Create interactive dashboards: Develop interactive dashboards that allow users to easily switch between different visualizations, track key metrics, and gain a holistic understanding of the data.

CHAPTER 7

REFERENCES

1) https://ninza7.medium.com/ipl-data-analysis-2008-2020-using-python-c031d3e1ae0c

2) https://www.kaggle.com/code/nirajvermafcb/data-visualisation-for-ipl-datasets-1

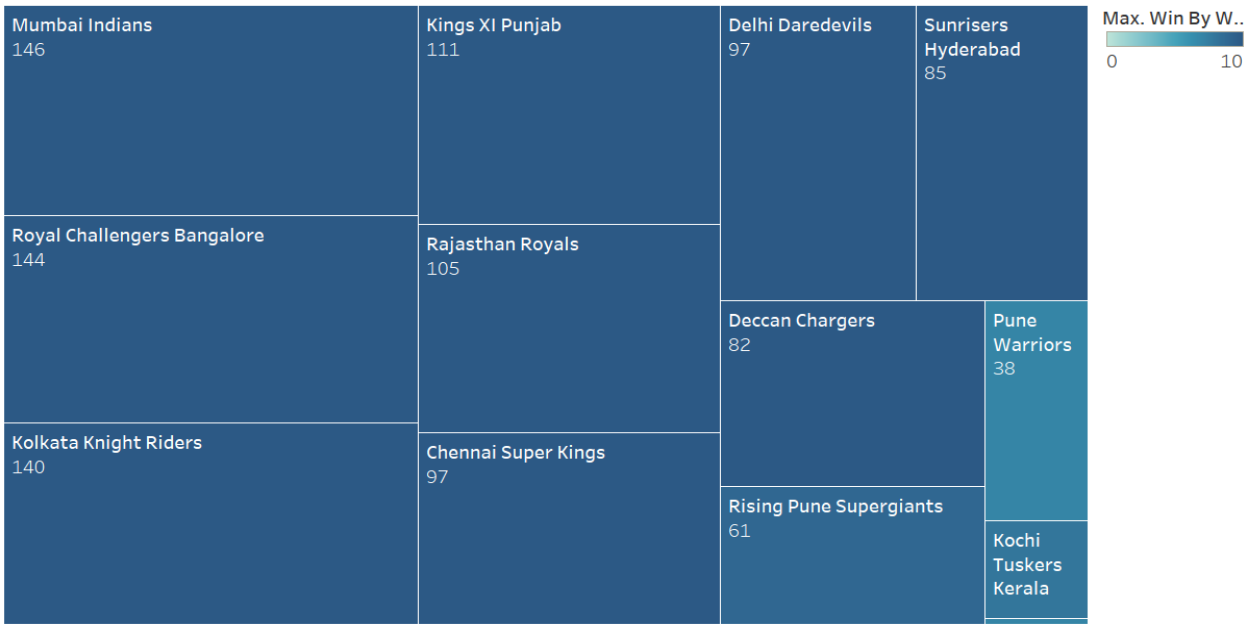
3) https://machinelearningknowledge.ai/ipl-data-analysis-and-visualization-project-using-python/

4) https://www.analyticsvidhya.com/blog/2022/05/ipl-team-win-prediction-project-using-machine-learning/

APPENDIX- A (TABLEAU SCREENSHOTS)

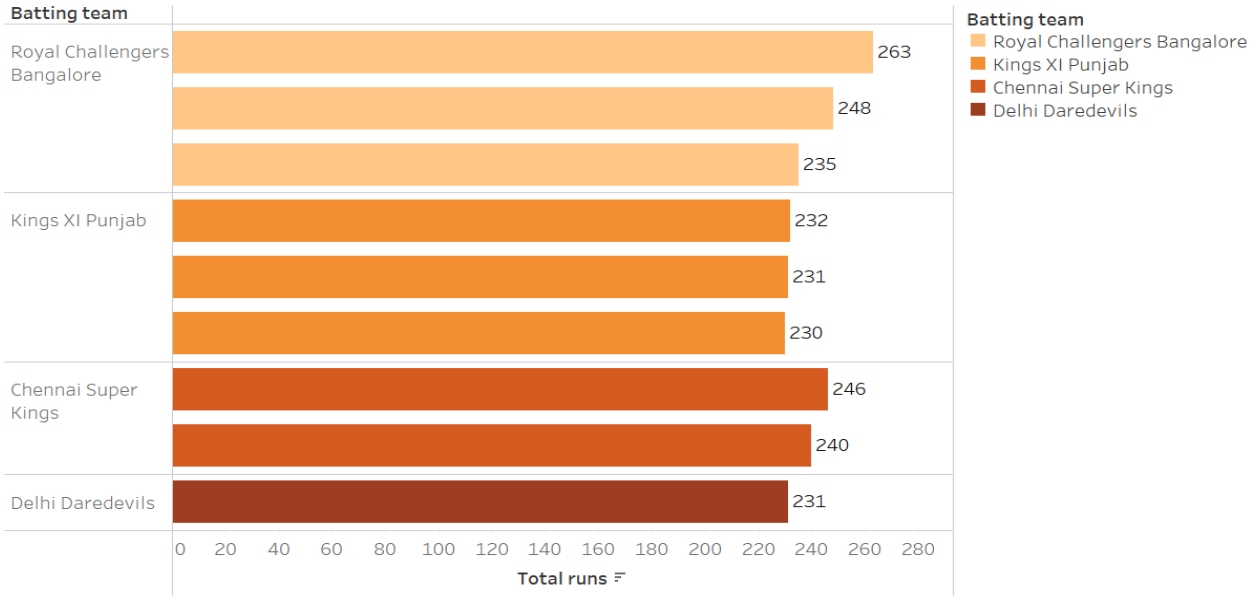
Match Statistics (Worksheets):

Biggest wins (by runs and by wicket)



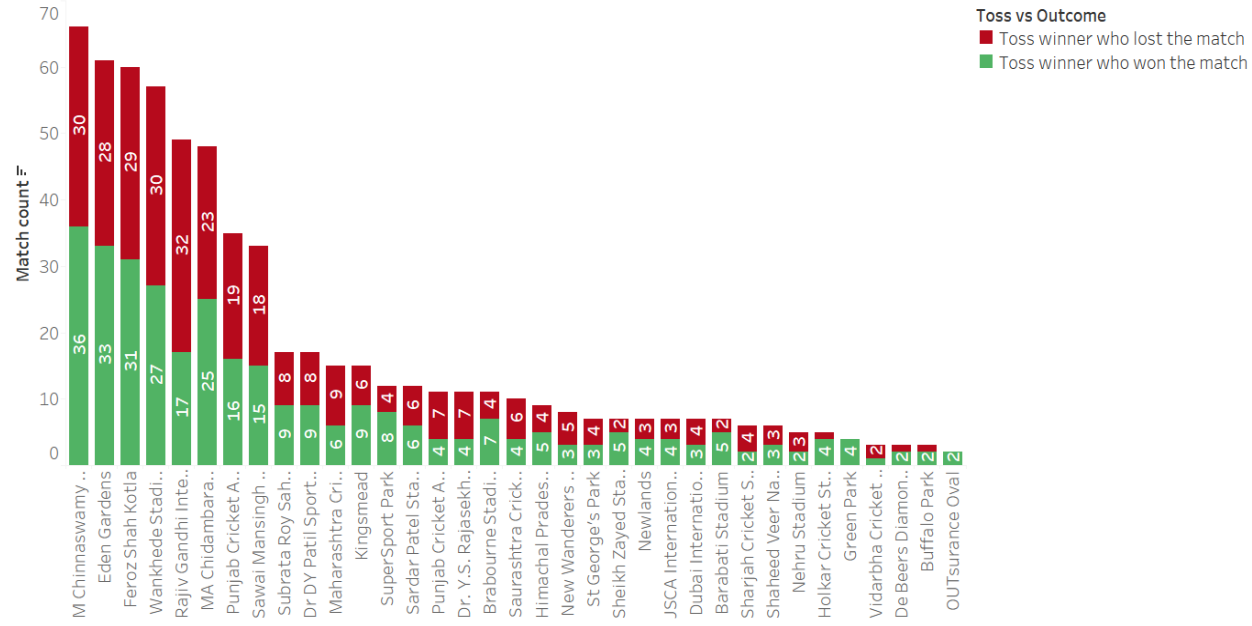
Winning Team and maximum of Win By Runs. Color shows maximum of Win By Wickets. Size shows maximum of Win By Runs. The marks are labeled by Winning Team and maximum of Win By Runs. The data is filtered on Action (Toss vs Outcome,Toss Decision), which keeps 4 members.

Highest totals (across all the seasons)



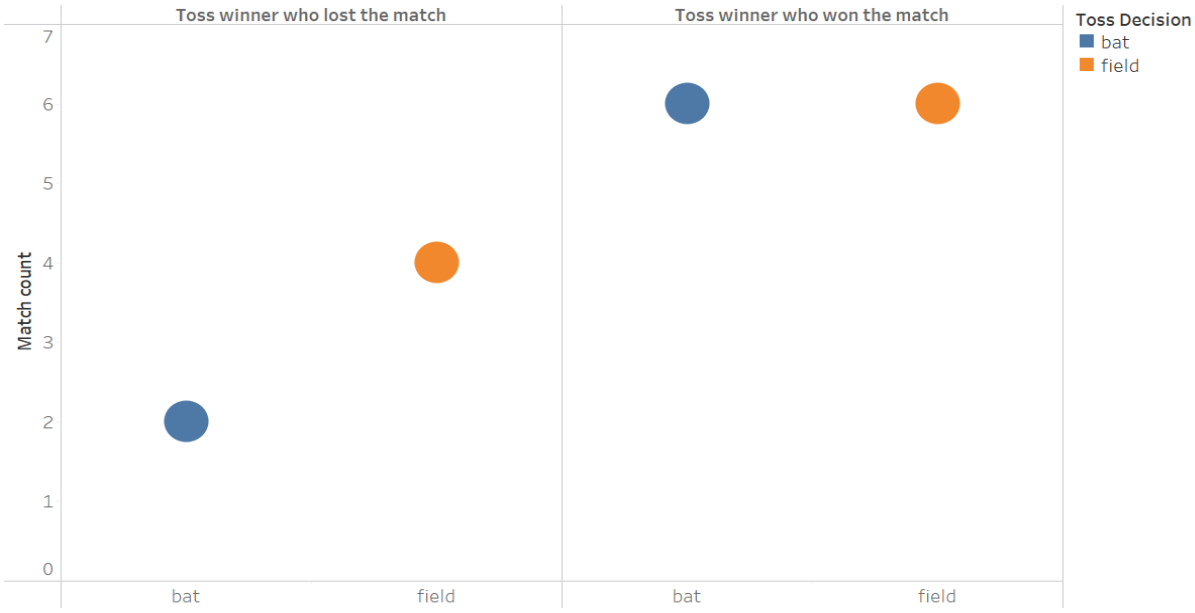
Sum of total_runs for each match_id broken down by Batting team. Color shows details about Batting team. The marks are labeled by sum of total_runs. Details are shown for Date Year. The data is filtered on Action (Winner), Action (Toss vs Outcome,Toss Decision) and Action (Winning Team). The Action (Winner) filter keeps 14 members. The Action (Toss vs Outcome,Toss Decision) filter keeps 4 members. The Action (Winning Team) filter keeps 14 members. The view is filtered on sum of total_runs, which ranges from 230 to 263 and keeps Null values.

Toss outcome vs Match outcome (for each venue)



Distinct count of Id for each Venue. Color shows details about Toss vs Outcome. The marks are labeled by distinct count of match_id. The data is filtered on Action (Winner), Action (Winning Team) and Action (Toss vs Outcome,Toss Decision). The Action (Winner) filter keeps 14 members. The Action (Winning Team) filter keeps 14 members. The Action (Toss vs Outcome,Toss Decision) filter keeps 4 members.

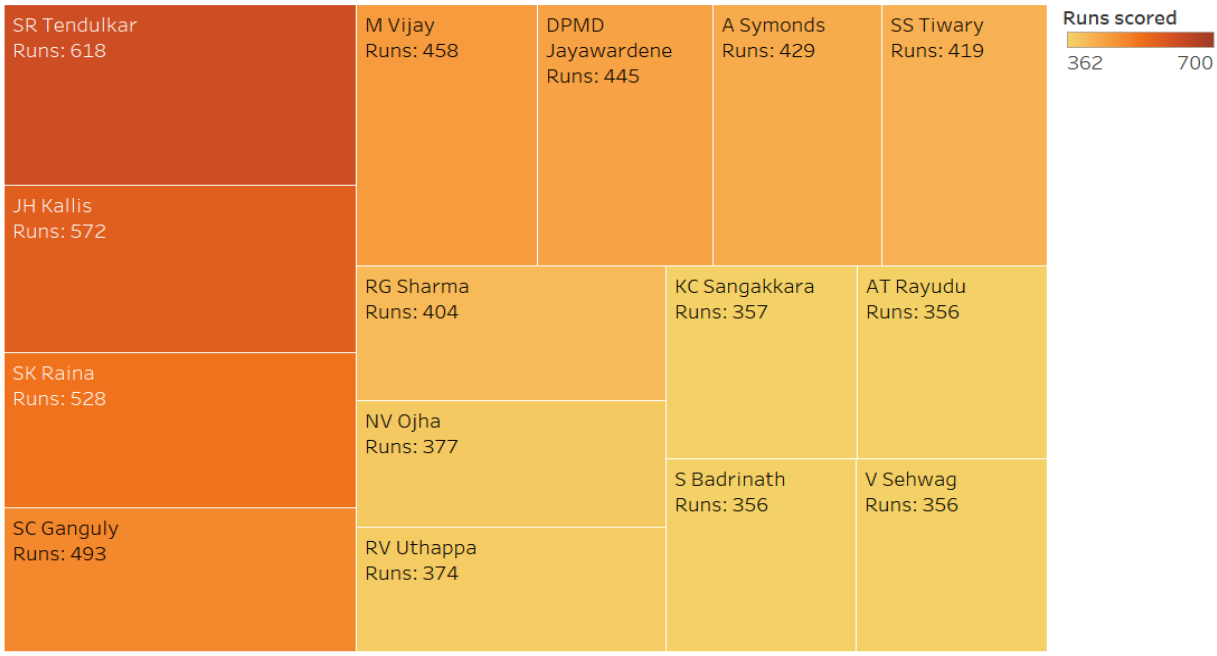
Toss decision vs Match outcome (for each venue)



Distinct count of Id for each Toss Decision broken down by Toss vs Outcome. Color shows details about Toss Decision. The data is filtered on Venue, Action (Winner) and Action (Winning Team). The Venue filter keeps Barabati Stadium and Brabourne Stadium. The Action (Winner) filter keeps 14 members. The Action (Winning Team) filter keeps 14 members.

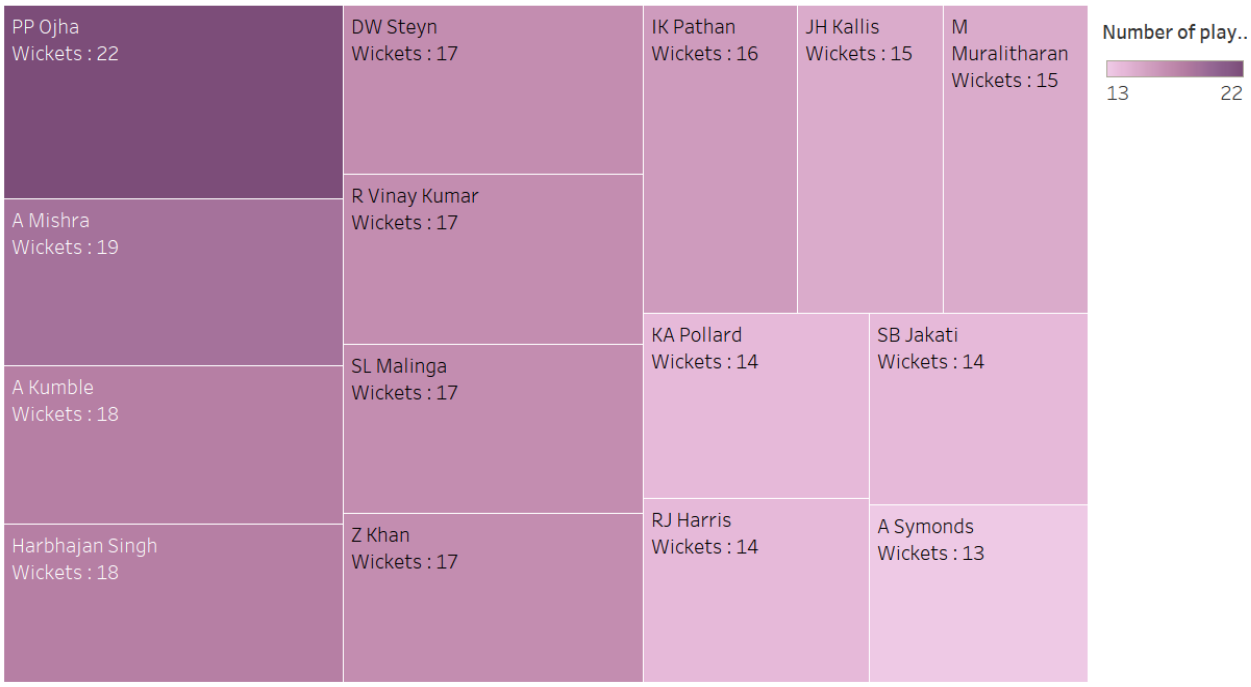
Player statistics(worksheet):

Orange cap contenders



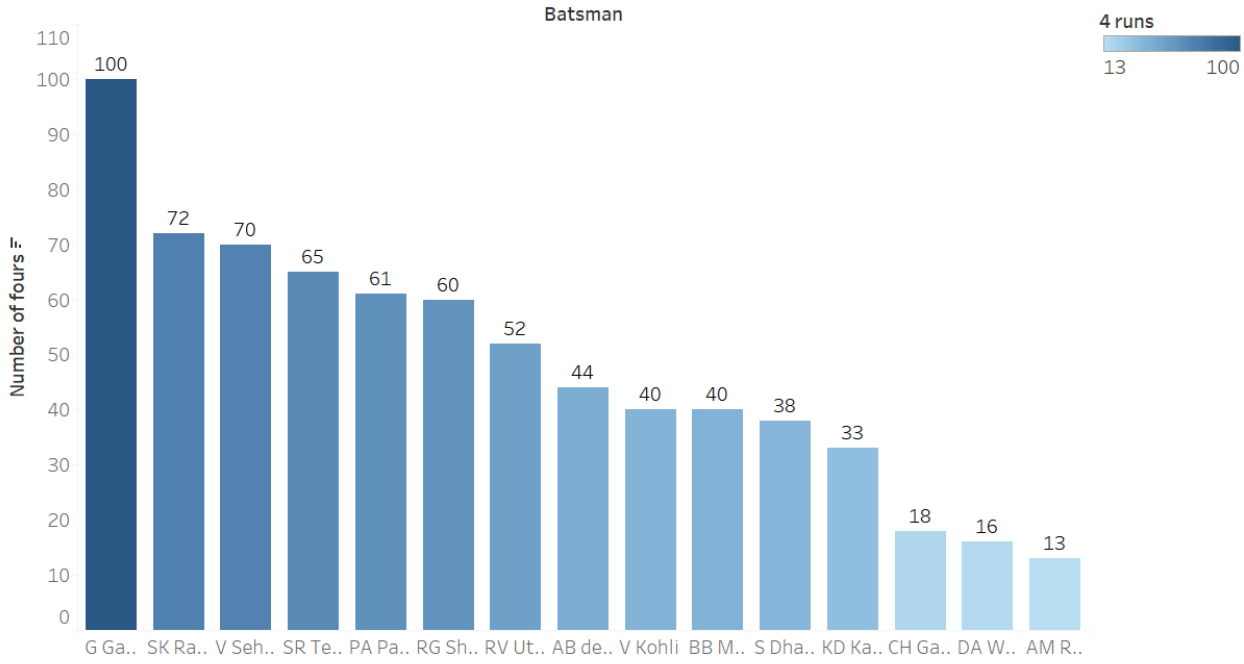
Batsman and sum of batsman_runs. Color shows sum of batsman_runs. Size shows sum of batsman_runs. The marks are labeled by Batsman and sum of batsman_runs. The context is filtered on Date Year, which keeps 2010. The view is filtered on Batsman, which keeps 15 of 461 members.

Purple cap contenders



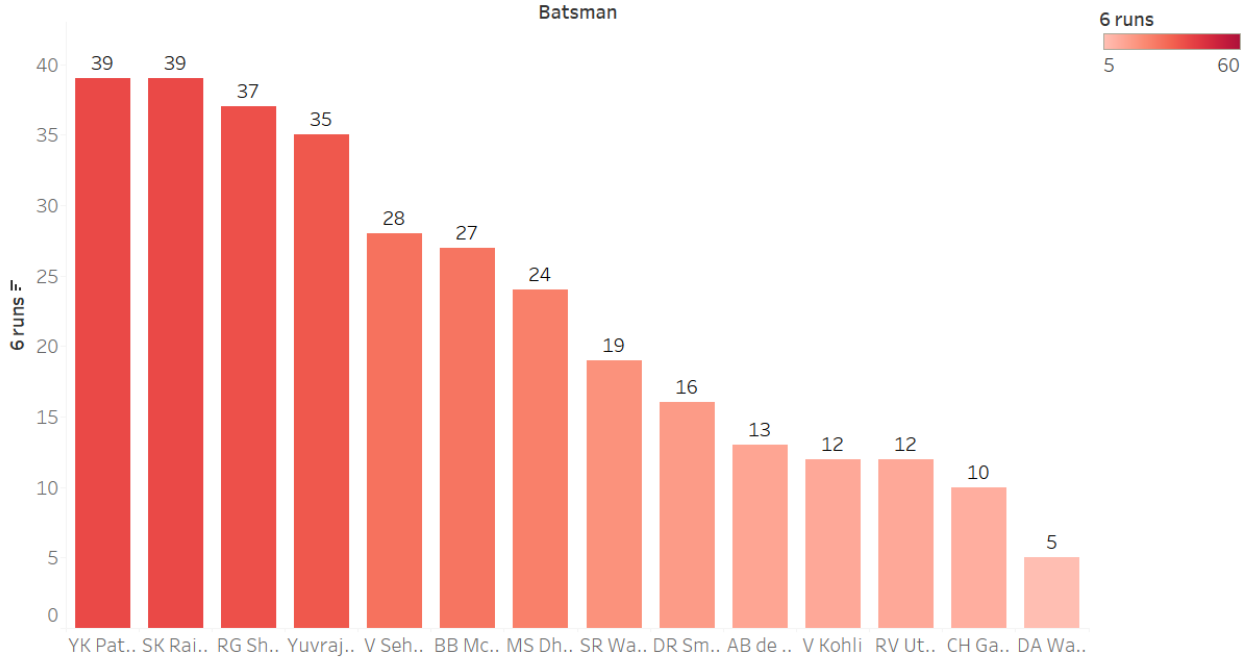
Bowler and distinct count of player_dismissed. Color shows distinct count of player_dismissed. Size shows distinct count of player_dismissed. The marks are labeled by bowler and distinct count of player_dismissed. The context is filtered on Date Year, which keeps 2010. The view is filtered on bowler, which keeps 15 of 356 members.

Most number of fours(per season and overall)



Sum of 4 runs for each Batsman. Color shows sum of 4 runs. The marks are labeled by sum of 4 runs. The data is filtered on Date Year, which keeps 2008 and 2009. The view is filtered on Batsman, which keeps 15 of 461 members.

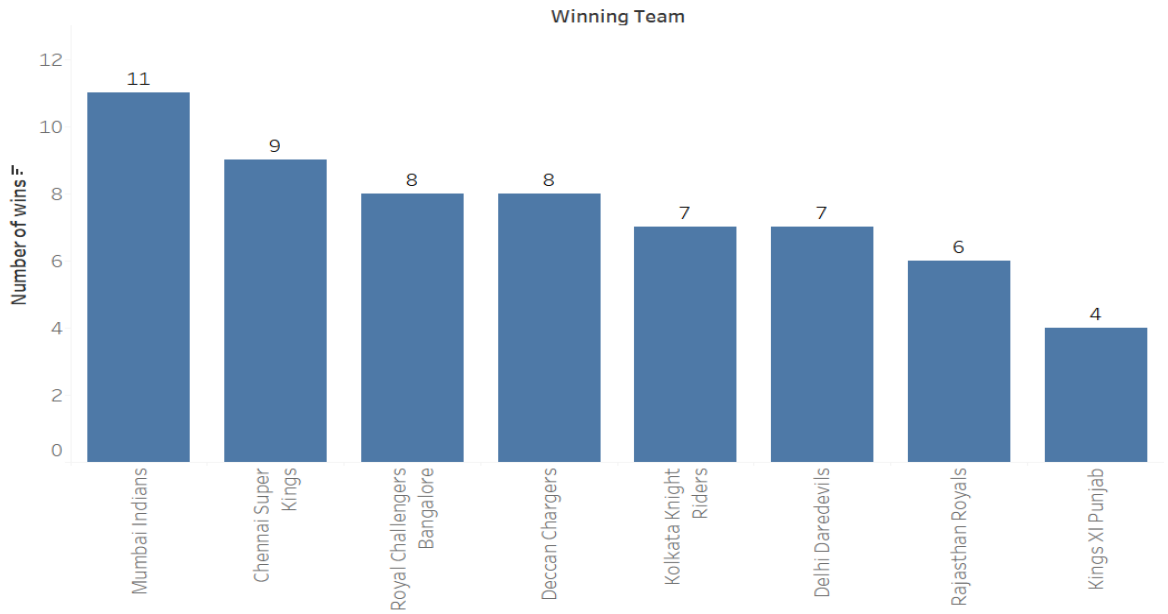
Most number of sixes(per season and overall)



Sum of 6 runs for each Batsman. Color shows sum of 6 runs. The marks are labeled by sum of 6 runs. The data is filtered on Date Year, which keeps 2008 and 2009. The view is filtered on Batsman, which keeps 15 of 461 members.

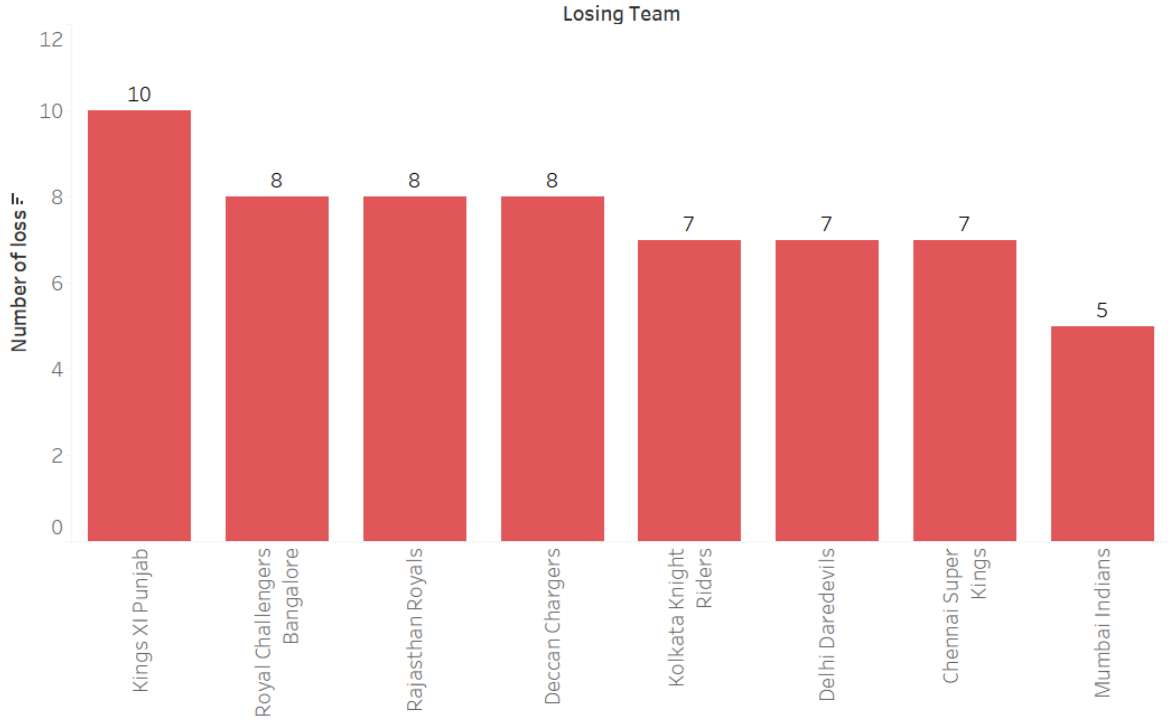
Team Statistics (Worksheet):

Season-wise team performance (wins)



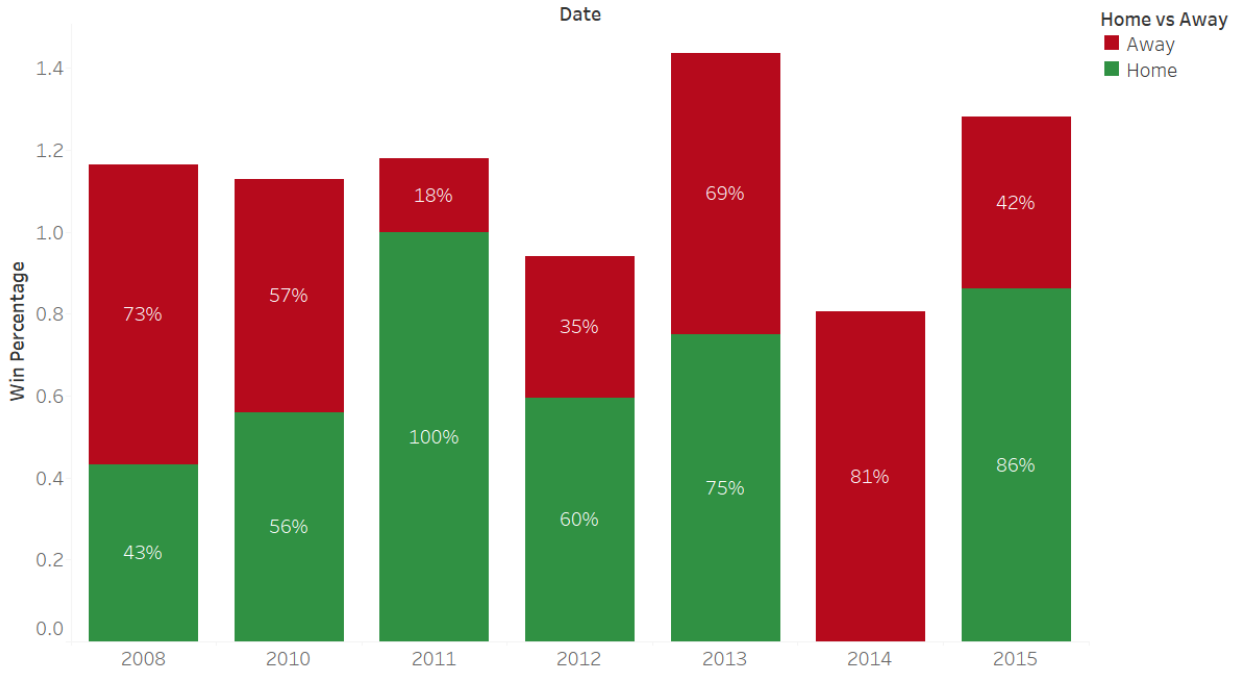
Distinct count of Id for each Winning Team. The marks are labeled by distinct count of match_id. The data is filtered on Date Year, which keeps 2010. The view is filtered on Winning Team, which excludes Null.

Season-wise team performance (loss)



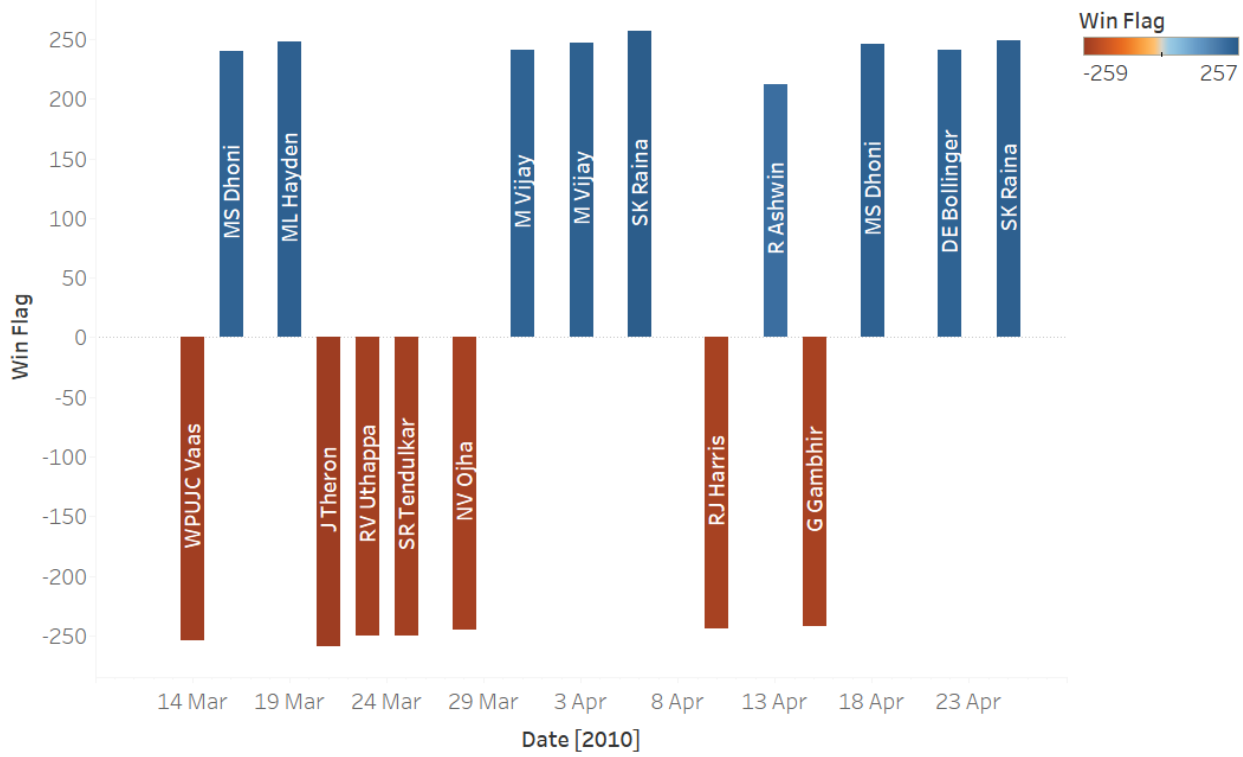
Distinct count of Id for each Losing Team. The marks are labeled by distinct count of match_id. The data is filtered on Date Year, which keeps 2010. The view is filtered on Losing Team, which excludes Null.

Win % (Home vs Away)



Win Percentage for each Date Year. Color shows details about Home vs Away. The marks are labeled by Win Percentage. The view is filtered on Home vs Away, which keeps Away and Home.

Season-wise team performance (wins vs losses)



The plot of sum of Win Flag for Date. Color shows sum of Win Flag. The marks are labeled by Player Of Match. Details are shown for match_id. The data is filtered on Date Year, which keeps 2010.