

Data analysis of IPL statistics

AUTHOR

Group 11 : Smriti Triphati, Garima Verma, Rohit Kumar,
Jayant Jha

PUBLISHED

November 16, 2023

Under the guidance of Dr. Dootika Vats

Acknowledgement

We extend our heartfelt gratitude to Professor Dootika Vats for her unwavering support and invaluable guidance throughout the development of this IPL Data Analysis project . Her expertise and mentorship have been instrumental in shaping the project's direction, refining analytical methodologies, and fostering a deeper understanding of cricket statistics. We are sincerely appreciative of her commitment to academic excellence, which has inspired us to explore and analyze the dynamic world of the Indian Premier League comprehensively. Her encouragement and insights have been instrumental, contributing significantly to the success of this project.

1. Introduction

1.1 About Indian Premier League (IPL)

The Indian Premier League (IPL) is a professional Twenty20 cricket league in India. It was founded by the Board of Control for Cricket in India (BCCI) in 2007, with the first season taking place in 2008. The IPL has since become one of the most popular and lucrative cricket leagues globally.

Key features of the IPL include:

- **Franchise-Based Teams:** The IPL consists of franchise-based teams representing different cities or regions across India. These franchises bid for players in an annual auction, creating dynamic and diverse teams.
- **T20 Format:** Matches in the IPL are played in the Twenty20 (T20) format, which is a short and fast-paced version of cricket. Each team faces a limited number of overs, and matches typically last around three hours.
- **International Player Participation:** The IPL attracts top international cricketing talent, with players from various countries participating in the league. This international flavor adds to the competitiveness and global appeal of the tournament.
- **Auctions and Transfers:** Before each season, teams participate in an auction to buy, sell, or trade players. This dynamic player market contributes to the ever-changing team compositions and strategies.
- **Global Viewership:** The IPL has a massive fan following, not just in India but around the world. The league is broadcast in numerous countries, and its fan base includes cricket enthusiasts and casual viewers alike.

- **Economic Impact:** The IPL has a significant economic impact, contributing to the growth of the cricketing industry. It generates revenue through sponsorships, broadcasting rights, ticket sales, and merchandise.

The IPL has played a crucial role in popularizing T20 cricket and has become a prominent event in the international cricket calendar. It has also provided a platform for young and domestic talent to showcase their skills on a global stage. The league typically takes place annually, with the season spanning a couple of months, attracting attention from cricket fans worldwide.

1.2 Objective

The primary objective of this project is to conduct an in-depth analysis of IPL statistics spanning the years 2013 to 2023. This comprehensive analysis aims to provide users with a multifaceted understanding of the Indian Premier League, empowering them to engage in friendly competition with their peers by creating and managing their own IPL fantasy teams. Additionally, the project seeks to enhance user experience by incorporating a feature that allows users to geographically locate the countries of their favorite IPL players.

1.2.1 IPL Statistics Analysis:

- Collect and process detailed statistics from IPL seasons 2013-2023, encompassing player performances, team dynamics, match outcomes, and key trends.
- Utilize data visualization techniques to present compelling insights into team and player dynamics, highlighting patterns, trends, and significant statistical anomalies.

1.2.2 Fantasy Team Creation and Competition:

- Develop a user-friendly interface enabling users to create and manage their own IPL fantasy teams.
- Implement a competitive framework that allows users to compete with friends by assembling and refining their fantasy teams based on real-time IPL statistics.
- Integrate scoring mechanisms that align with actual player performances, providing an immersive and engaging fantasy cricket experience.

1.2.3 Geolocation Feature for Player Countries:

- Implement a feature allowing users to geographically locate the home countries of their favorite IPL players on an interactive map.
- Enhance user engagement by providing additional information about the countries, such as cricketing culture, notable achievements, and team affiliations.

2. Scraping and Cleaning the Data

The data for different purposes has been scraped from various sources.

We have scraped the following data:

1. Year-wise(2013-2023) performance of individual batters in IPL
2. Year-wise(2013-2023) performance of individual bowlers in IPL
3. Final Points Table of each team from 2013-2023

4. Country from where each player belongs

5. Latitude of different countries

Data sets 1, 2, 3 and 4 were scraped from the official site of ESPNCRICINFO.

The 5th data set was manually entered and the data was obtained from Wikipedia.

The raw datasets that we obtained from the respective webpages were not ready-to-use. There were many missing values and there also were some discrepancies in the data. We had to clean the data and also organize the data in a way that is easy to use.

We used the following libraries for doing these tasks:

1. rvest
2. tidyverse
3. dplyr

For example, here is a snippet of code that was used to extract data for batting performances of players in IPL2023:

2.1 Example Data set

This is the data that was obtained after running the above code:

1	1	MEK Hussey (CSK)	17	17	3	733	95	52.35	566	129.50	-	6	-	81	17
2	2	CH Gayle (RCB)	16	16	4	708	175*	59.00	453	156.29	1	4	-	57	51
3	3	V Kohli (RCB)	16	16	2	634	99	45.28	457	138.73	-	6	-	64	22
4	4	SK Raina (CSK)	18	17	4	548	100*	42.15	365	150.13	1	4	3	50	18
5	5	SR Watson (RR)	16	16	2	543	101	38.78	380	142.89	1	2	-	59	22
6	6	RG Sharma (MI)	19	19	5	538	79*	38.42	409	131.54	-	4	-	35	28
7	7	KD Karthik (MI)	19	19	1	510	86	28.33	411	124.08	-	2	2	54	14
8	8	AM Rahane (RR)	18	18	4	488	68*	34.85	458	106.55	-	4	-	42	11
9	9	R Dravid (RR)	18	17	1	471	65	29.43	425	110.82	-	4	-	64	5
10	10	MS Dhoni (CSK)	18	16	5	461	67*	41.90	283	162.89	-	4	-	32	25
11	11	AJ Finch (PWI)	14	14	0	456	67	32.57	336	135.71	-	4	1	54	16
12	12	RV Uthappa (PWI)	16	16	0	434	75	27.12	371	116.98	-	2	1	42	12
13	13	KA Pollard (MI)	18	18	8	420	66*	42.00	281	149.46	-	3	-	27	29
14	14	DA Miller (KXIP)	12	12	5	418	101*	59.71	254	164.56	1	3	-	28	24
15	15	DR Smith (MI)	13	13	0	418	68	32.15	341	122.58	-	4	1	40	19
16	16	DA Warner (DC)	16	16	3	410	77	31.53	323	126.93	-	4	2	41	14
17	17	G Gambhir (KKR)	16	16	0	406	60	25.37	343	118.36	-	4	2	51	5
18	18	AB de Villiers (RCB)	14	14	4	360	64	36.00	219	164.38	-	2	-	34	15
19	19	YK Pathan (KKR)	16	16	5	332	72	30.18	240	138.33	-	1	2	30	14
20	20	DPMD Jayawardene (DC)	15	15	0	331	66	22.06	313	105.75	-	2	-	39	6
21	21	M Vijay (CSK)	15	15	1	312	55	22.28	286	109.09	-	2	-	27	8
22	22	S Dhawan (SRH)	10	10	2	311	73*	38.87	253	122.92	-	3	1	37	5
23	23	JH Kallis (KKR)	16	16	1	311	41	20.73	322	96.58	-	-	2	35	3
24	24	EJG Morgan (KKR)	14	13	2	307	51	27.90	240	127.91	-	1	-	28	13
25	25	SE Marsh (KXIP)	8	8	0	300	77	37.50	249	120.48	-	3	-	39	4

2.2 Description of our Data

2.2.1 In the batters data set for each year, the following data points are included:

1. Player name: Name of the Player.
2. Matches: Number of matches played by the player.
3. Innings: Number of Innings played by the player.
4. Not outs: Number of times the player has remained not-out at the end of the innings.
5. Runs: Total Runs scored by the player over the whole season.
6. Highest Score: The highest runs scored by a player in an innings over the whole.
7. Average: Batting average of the player. It is the total number of runs scored by a batter divided by the number of times they have lost their wicket.
8. Bowls Faced: Total balls faced by the player in a season.
9. Strike Rate: The Strike rate at which the player bats. It is number of Runs scored by a batsmen divided by the number of balls faced by the faced multiplied by 100.
10. 100s/50s/0s: Number of centuries/half-centuries and ducks played by the player in a season.
11. 4s/6s: Number of boundaries hit by the player in a season.

2.2.2 Meanwhile in the bowlers data set for each year, the following data points are included:

1. Player: The name of the player.
2. Matches: The number of matches played by the player.
3. Innings: The number of innings in which the player bowled.
4. Balls/Overs: The number of balls balled by the player and the number of equivalent overs.
5. Maidens Earned: The number of Maidens(Zero-run) overs bowled by the bowler.
6. Runs: The number of runs conceded by the bowler over the whole season.
7. Wickets: The number of wickets taken by the bowler in the whole season.
8. Best Bowling Innings: The best bowling figures achieved by a player in a match.
9. Average: The average runs conceded in an innings by a player. It is calculated by dividing the numbers of runs they have conceded by the number of wickets they have taken.
10. Economy Rate: The rate at which the player concedes runs. It is the average number of runs conceded per over.
11. Strike rate: The Strike rate at which the Bowler bowled. It is the the average number of balls a bowler takes to pick up a wicket.
12. 4s/6s: The number of boundaries hit by the bowler.

2.2.3 In the Final Points Data set the following points are included:

1. Team Name: The name of the Team.
2. Matches: The number of matches played by the team in a season.
3. Won: The number of matches won by the team in the season.
4. Lost: The number of Matches Lost by the team in the season.
5. Tied: The number of matches for a team that ended up in a tie.
6. Number of Matches Abandoned: The number of Matches that were abandoned due to external factors like rain/technical issues/etc.
7. Points: The total number of points earned by the team.

8. Net Run Rate: The Net Run rate of a team. It signifies how much better a team was compared to other team in a match and the sum of all NRRs of all teams is zero.

2.2.4 The CoordCountBat23 Data Set in addition to all data points of Batting Data Set contains the following Data set:

1. Country: The country which the batter belongs to.
2. Lat/Long: The Latitude and Longitude of the birthplace of the batter within his birthplace.

The CoordCountBowl23 dataset has the same data points as CoordCountBat23 but for bowlers.

3. Visualising the Dataset

3.1 Libraries we used in this process

- ggplot2
- rvest
- tidyverse
- packcircles
- plotly

3.2 Shiny app

The Shiny App made by our group had the following features:

3.2.1. Visualization Tab:

- To present IPL statistics through interactive and insightful visualizations.

3.2.2. Auction Tab:

- To simulate the IPL auction experience for users creating their fantasy teams.

3.2.3. Maps Tab:

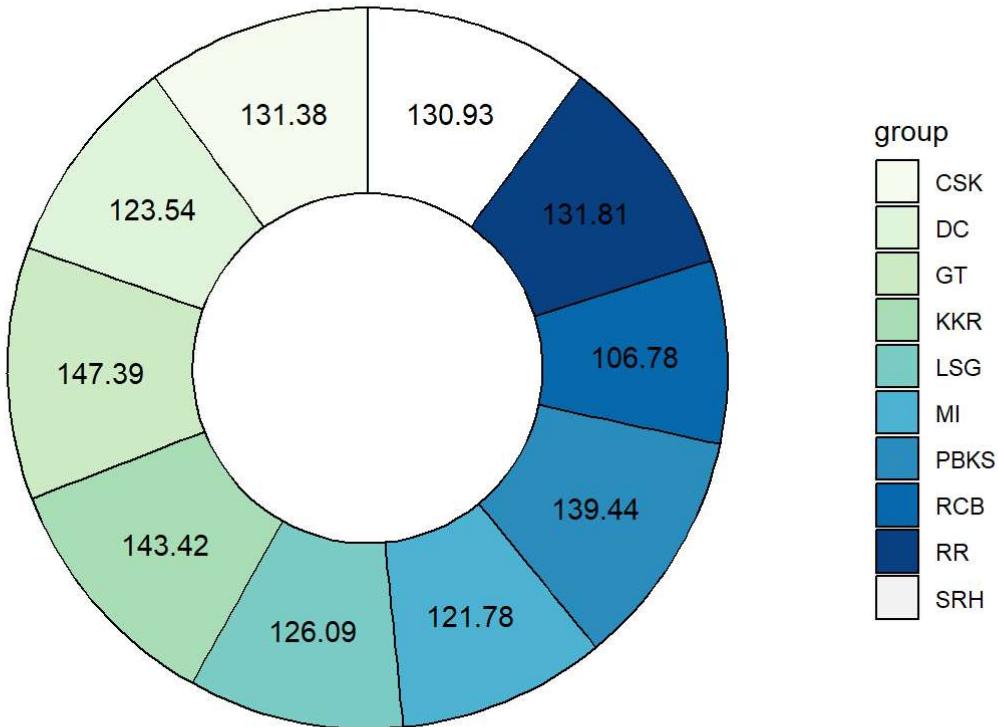
- To geographically locate the home countries of users' favorite IPL players.

3.2.4. Summary Tab:

- To provide a concise overview of key insights and trends.

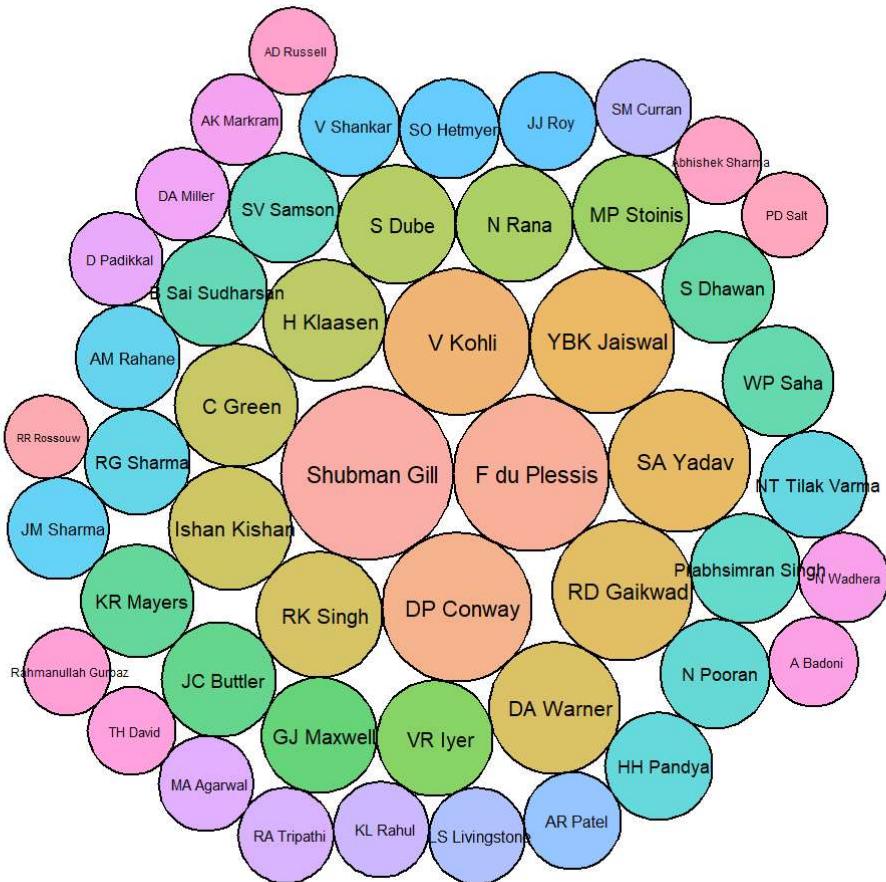
3.3 Generating visualisations for “2023” using our Shiny App

3.3.1 A Donut Chart for Average Batting Strike Rate for each TEAM



- **Observation:** The donut chart highlights a notable variation in the average strike rates among different teams, with some teams exhibiting higher rates and others showing lower rates.
- **Implication:** This variation suggests that teams with higher average strike rates may possess a more aggressive batting approach, potentially aiming for quick runs. The implication is that team management could strategically leverage this insight to tailor their batting line-up and overall game strategy. For teams with lower average strike rates, there might be an opportunity to explore adjustments that balance stability and aggression, enhancing the overall effectiveness of their batting performance.

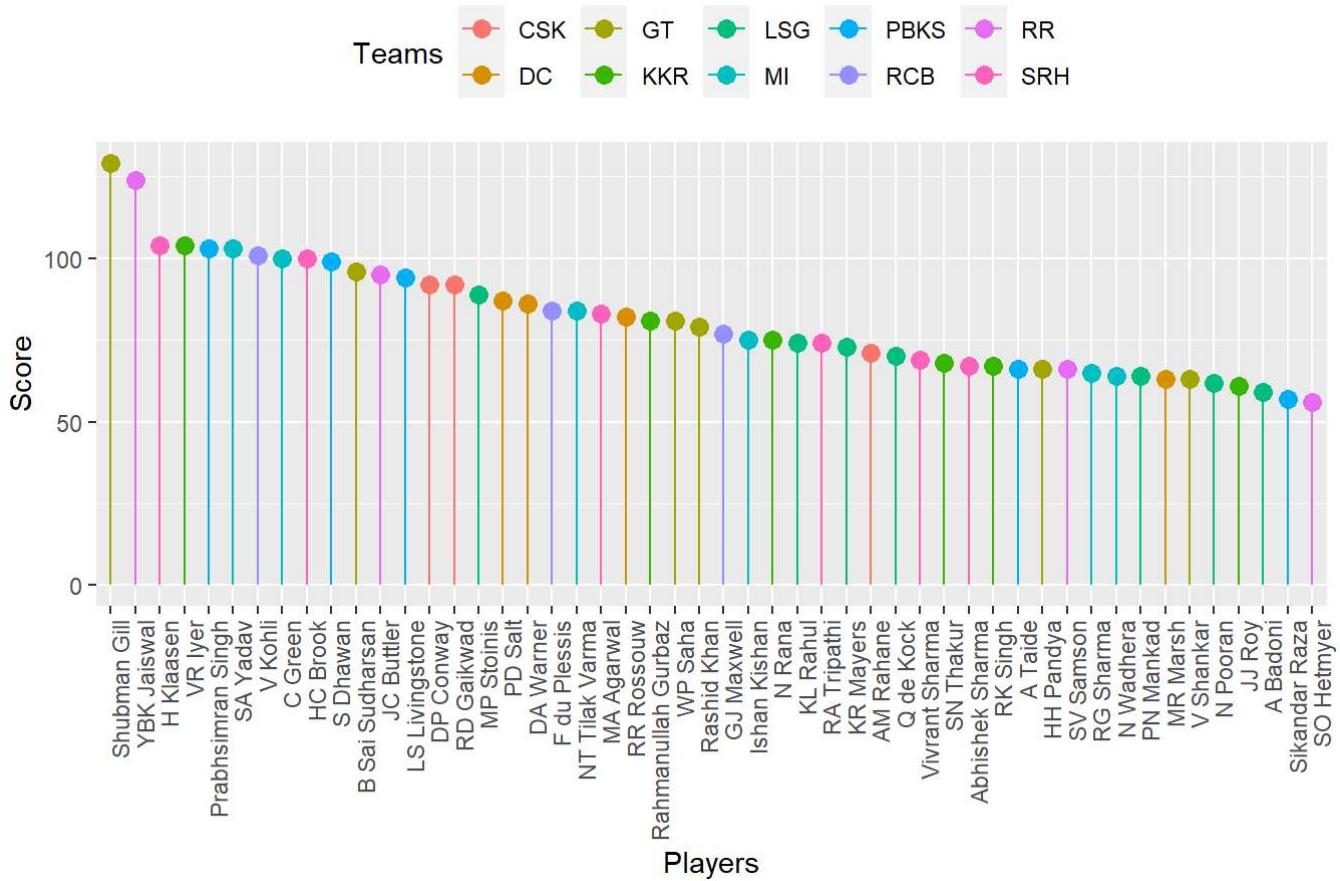
3.3.2 A Circular packing showing Impact Batsmen based on most Runs



- **Observation:** This packing shows Impact Batsmen based on most runs, larger bubble sizes signify players with higher run accumulations, visually emphasizing their significant impact in scoring.
- **Implication:** Team strategists can prioritize and build their innings around these impactful batsmen, leveraging their consistent run contributions for enhanced overall team performance.

3.3.3 A Point and Segment Plot for Batsmen VS Highest Score

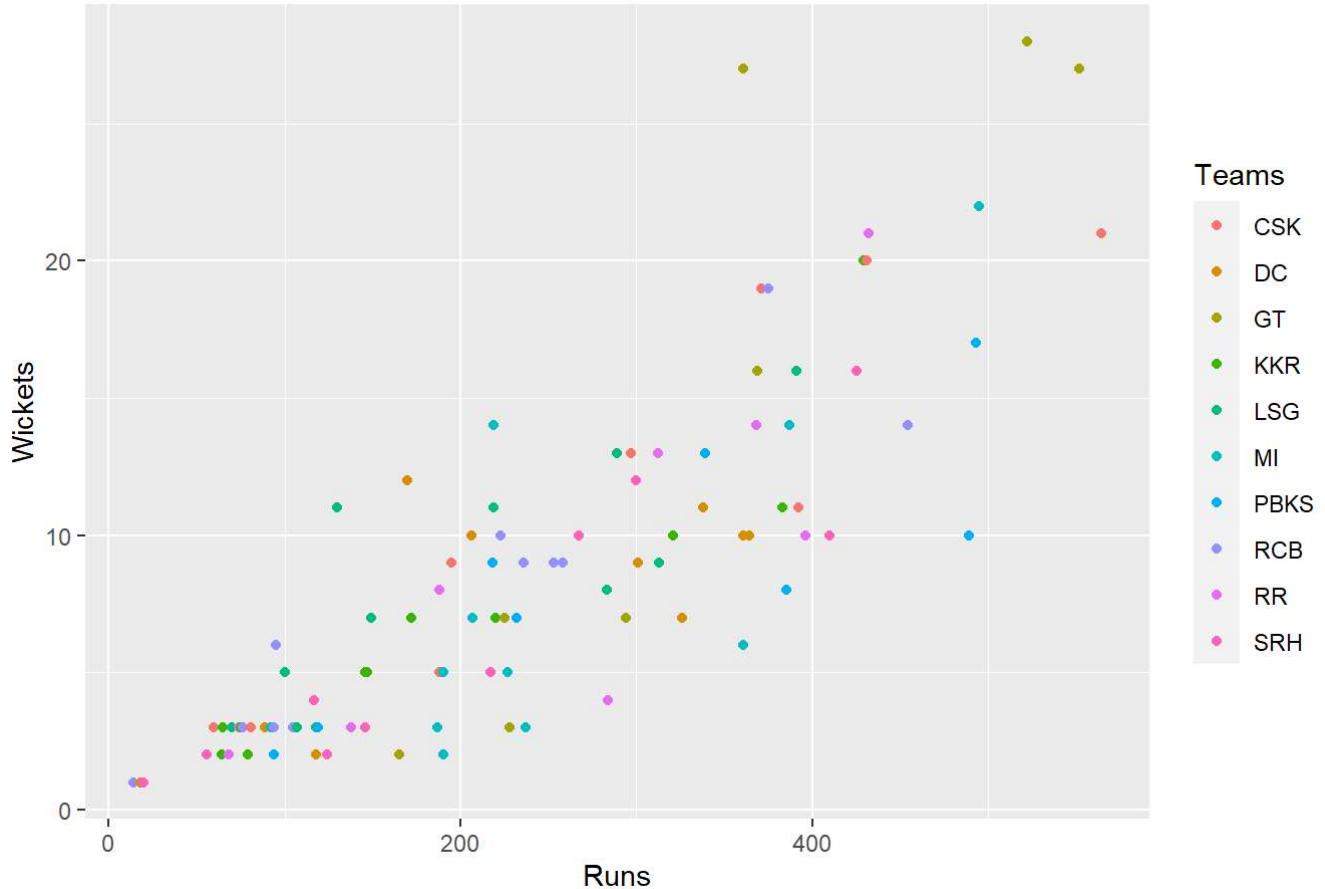
Highest Score



- Observation:** In this Point and Segment Plot, as we move from left to right, batsmen's best scores decrease. Different team colors reveal diverse team performances, showcasing individual player contributions.

3.3.4 A Scatter Plot for Runs conceded Vs Wickets taken by a Bowler

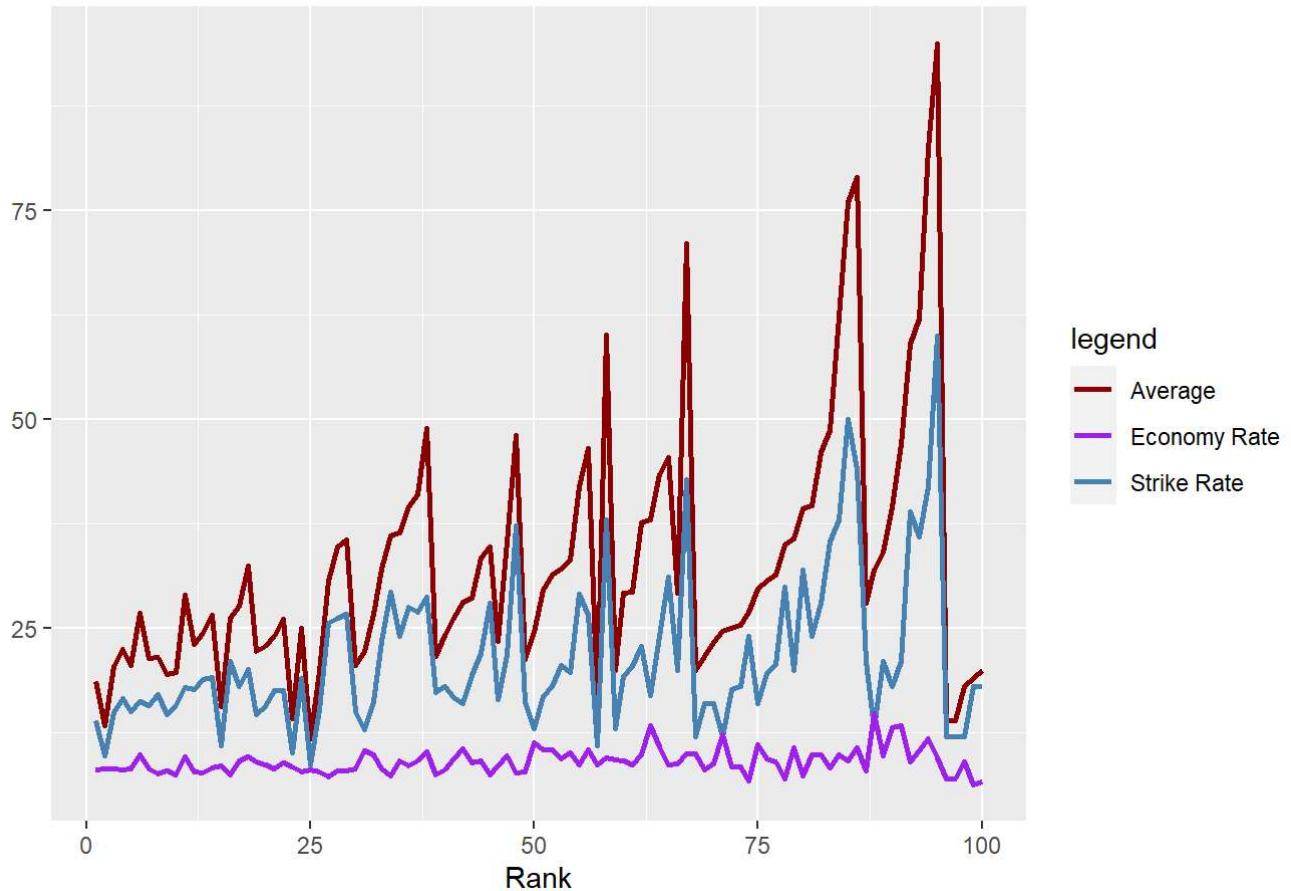
Runs Conceded vs Wickets Taken



- **Observation:** In the Scatter Plot for Runs Conceded vs. Wickets Taken by a bowler, a correlation is evident. As a bowler takes more wickets, there is an expected increase in runs conceded, indicating a relationship but not an exact correlation.
- **Implication:** While the correlation suggests a trade-off between wickets and runs conceded, it emphasizes the complexity of bowling performance. Teams should consider a nuanced approach, valuing wicket-taking ability alongside economical bowling to optimize overall team success.

3.3.5 A Line Plot for comparison between AVERAGE, STRIKE RATE and ECONOMY RATE of Bowlers.

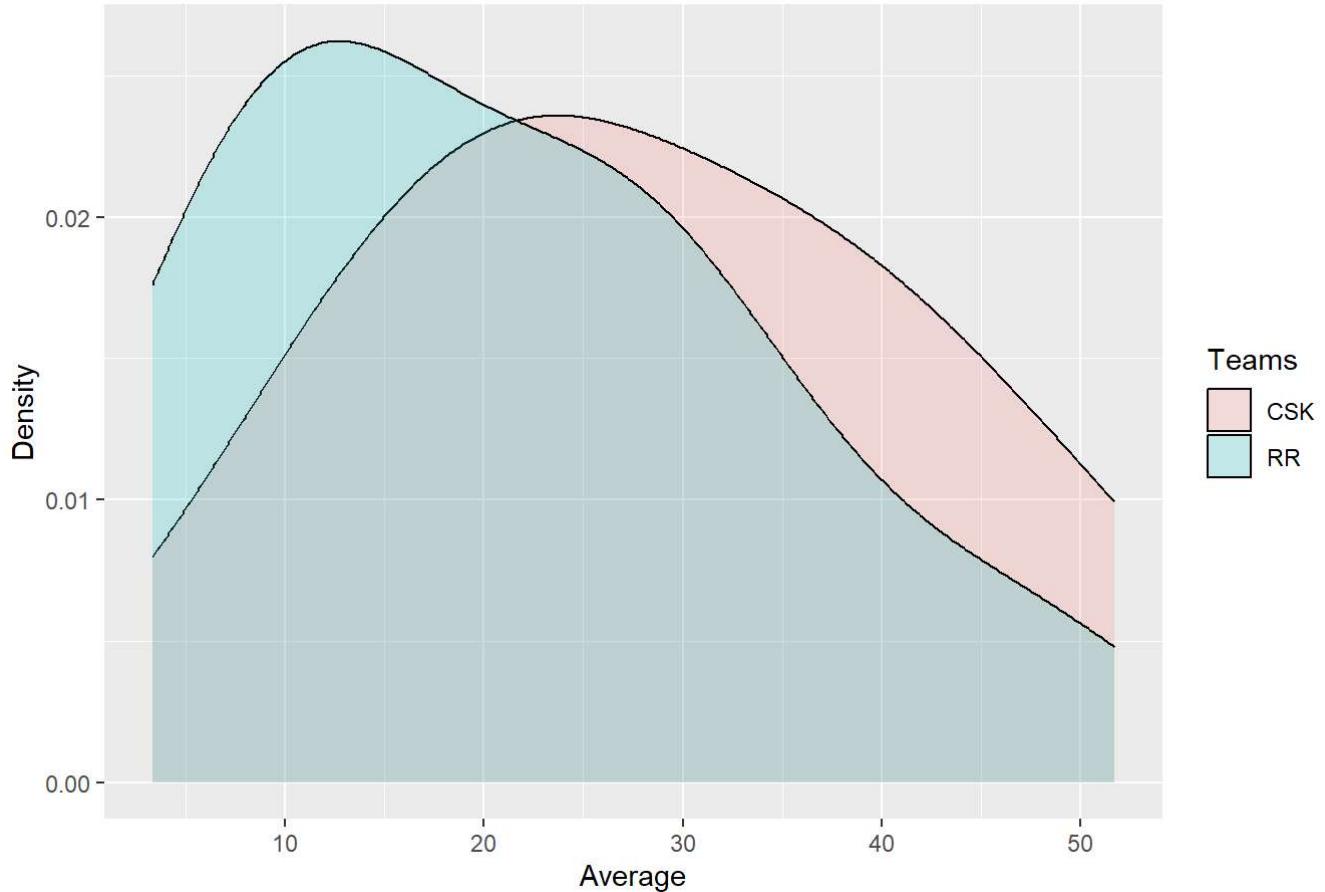
Rank Wise Bowling Comparison



- **Observation:** There is a positive correlation between a bowler's rank in terms of wickets taken, their strike rate, and average.
- **Implication:** Higher-ranked bowlers not only take more wickets but also achieve them with better efficiency (lower strike rate) and effectiveness (higher average).

3.3.6 A Density plot for comparing Average of Batsmen among different TEAMS

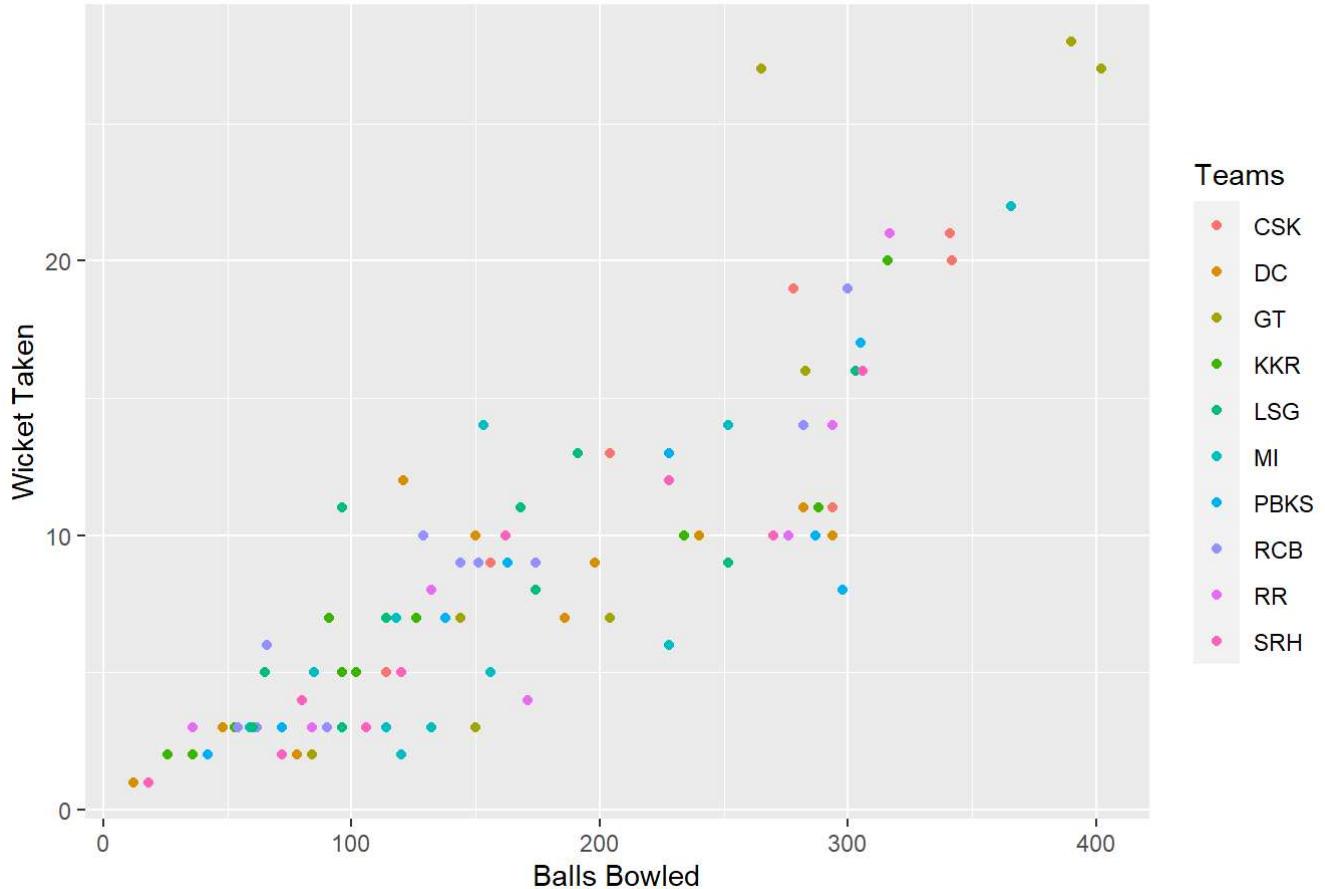
Density plot



- **Observation:** Teams with a large number of players boasting good averages have a higher probability of performing well.
- **Implication:** The team's overall batting performance is strongly influenced by the presence of players with consistently good averages.

3.3.7 A Scatter Plot for Balls Bowled Vs Wickets taken by a Bowler

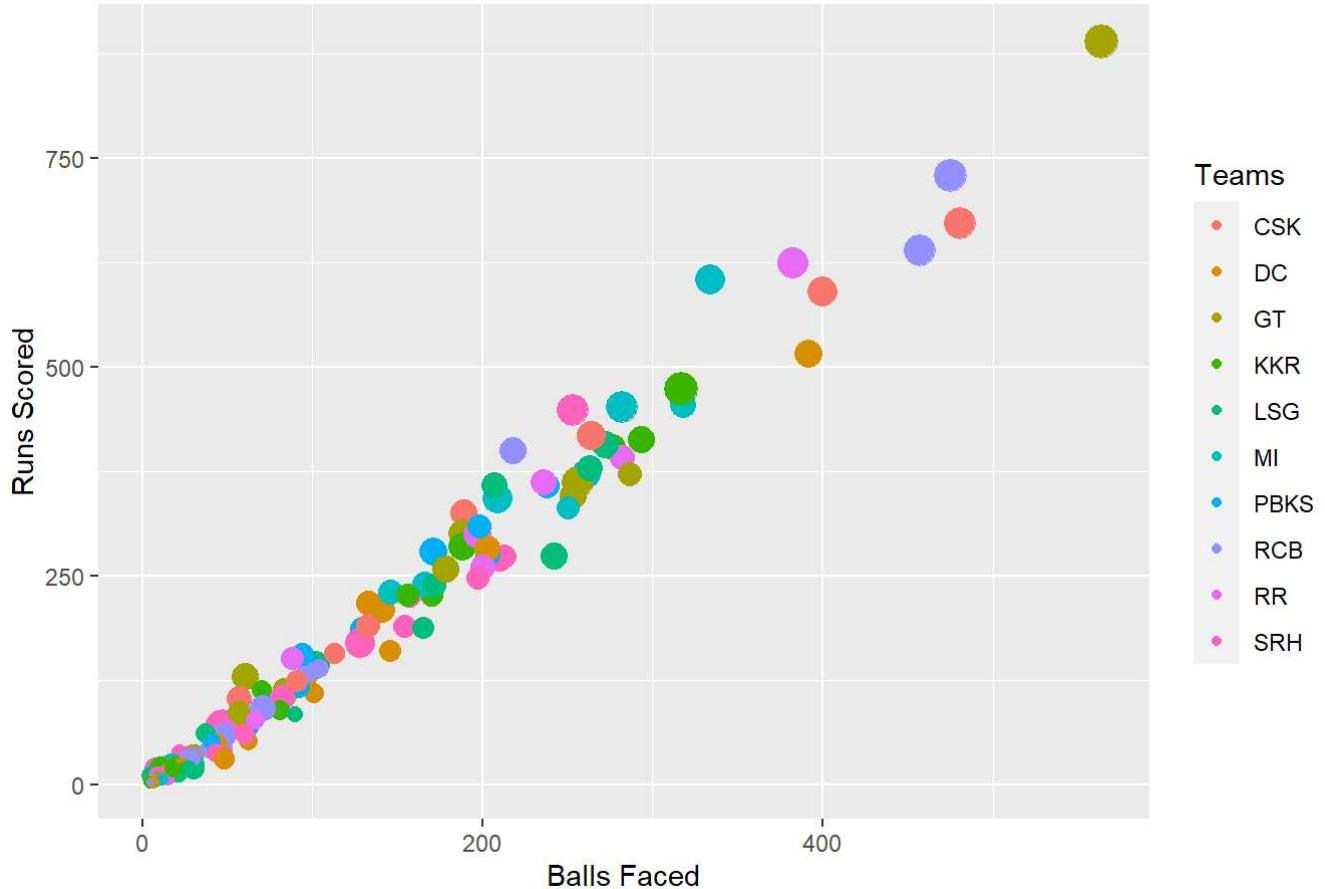
Balls Bowled vs Wickets Taken



- **Observation:** A moderate correlation is observed between the number of balls bowled by a bowler and the wickets taken.
- **Implication:** Bowlers who bowl more balls tend to have a higher probability of taking wickets, highlighting the importance of consistency and persistence in bowling.

3.3.8 A Scatter Plot for Balls Faced Vs Runs Scored by a Batsmen

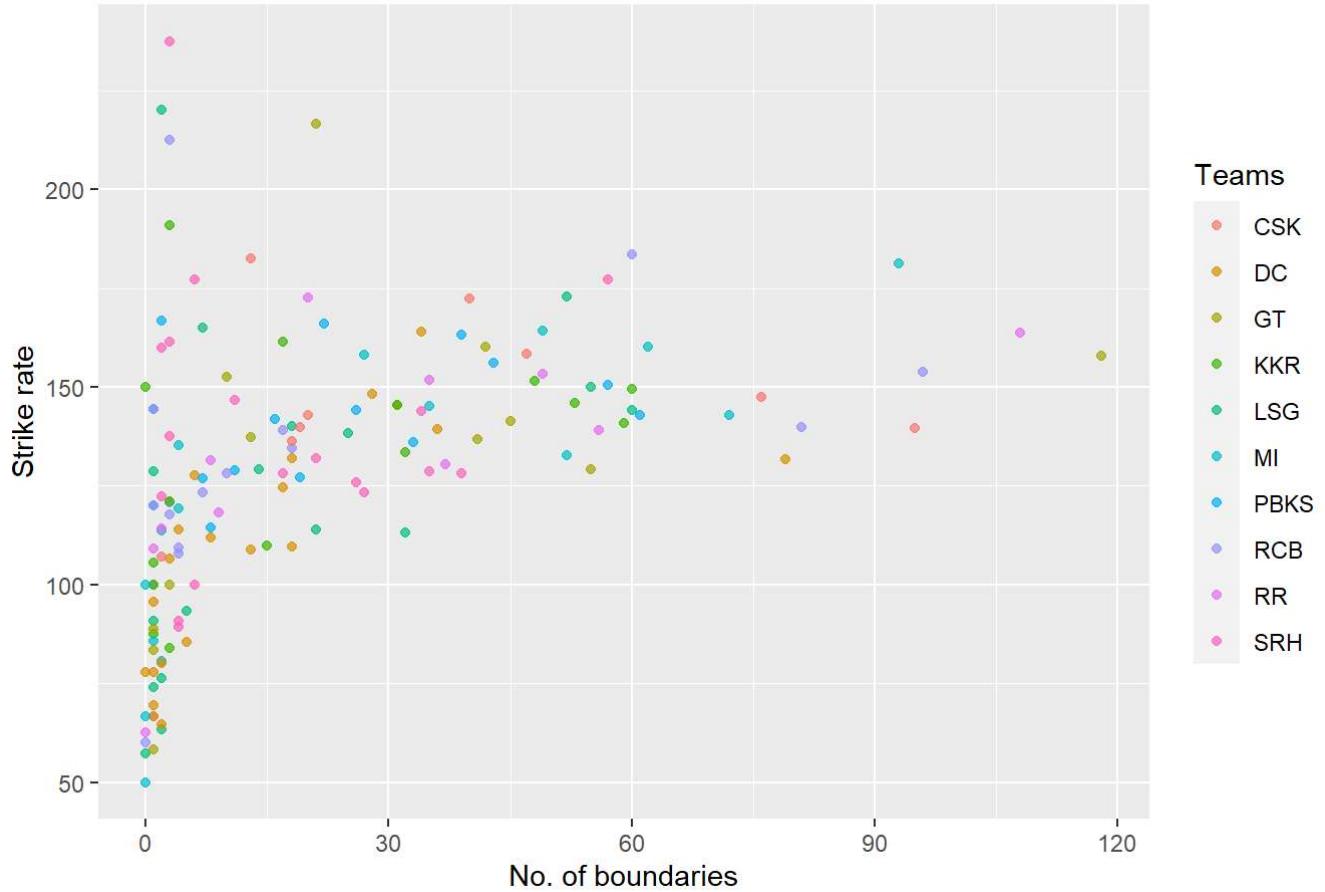
Balls faced vs Runs Scored



- **Observation:** A strong correlation exists between the total number of balls faced and the total runs scored.
- **Implication:** Batsmen who face more balls tend to contribute significantly to the team's total runs.

3.3.9 A Scatter Plot for No. of Boundaries Vs Strike Rate of a Batsmen

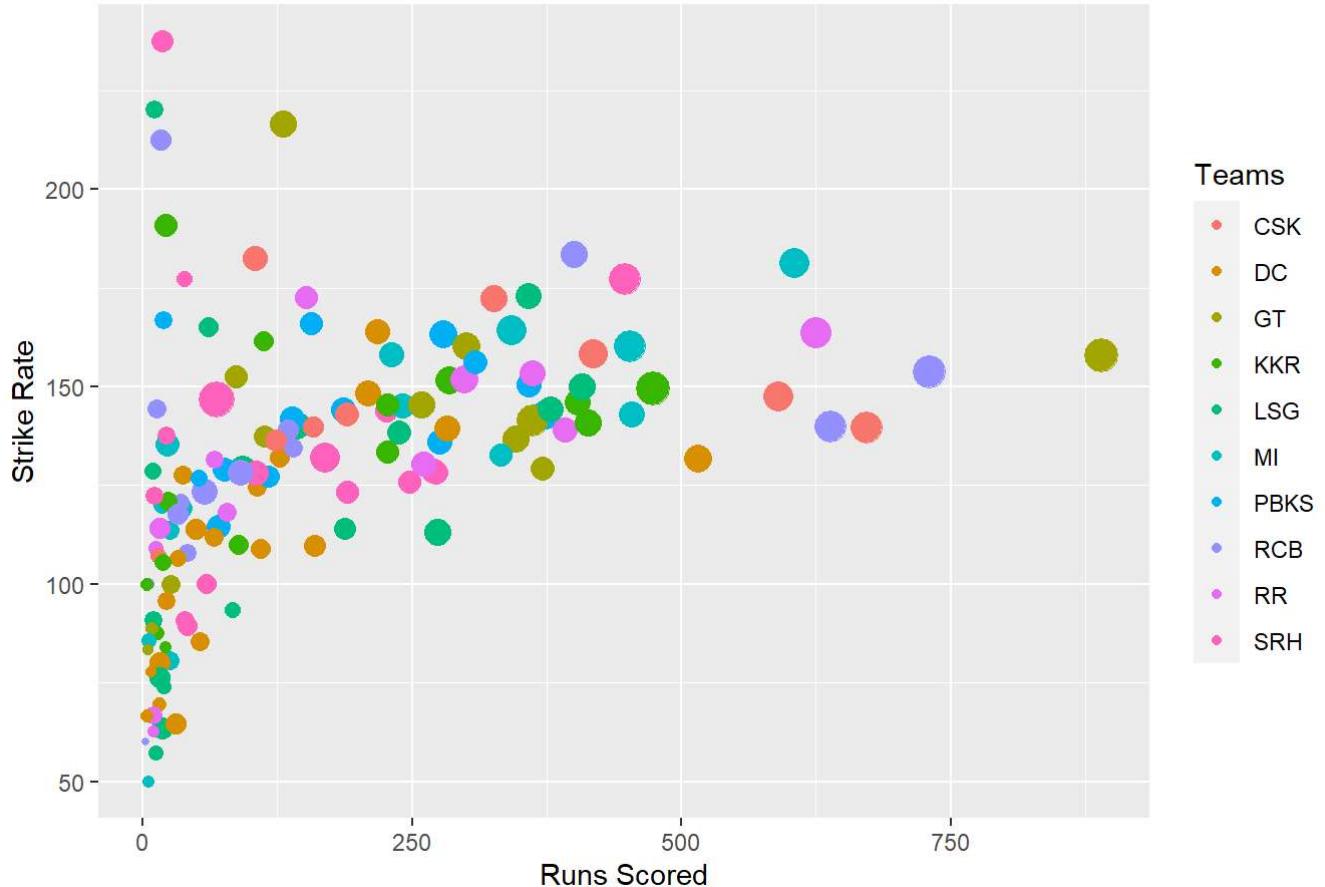
No. of boundaries vs Strike rates



- **Observation:** Players with more boundaries tend to have a better strike rate when they have faced a significant number of balls.
- **Implication:** Aggressive and boundary-hitting players, when given sufficient opportunities, contribute to an enhanced team strike rate.

3.3.10 A Scatter Plot for Runs Scored Vs Strike Rate of a Batsmen

Runs Scored Vs Strike Rate



Observation:

- Batsmen with fewer runs initially exhibit diverse strike rates, ranging from very high to very low.
- As runs increase, there is a noticeable stabilization in strike rates.

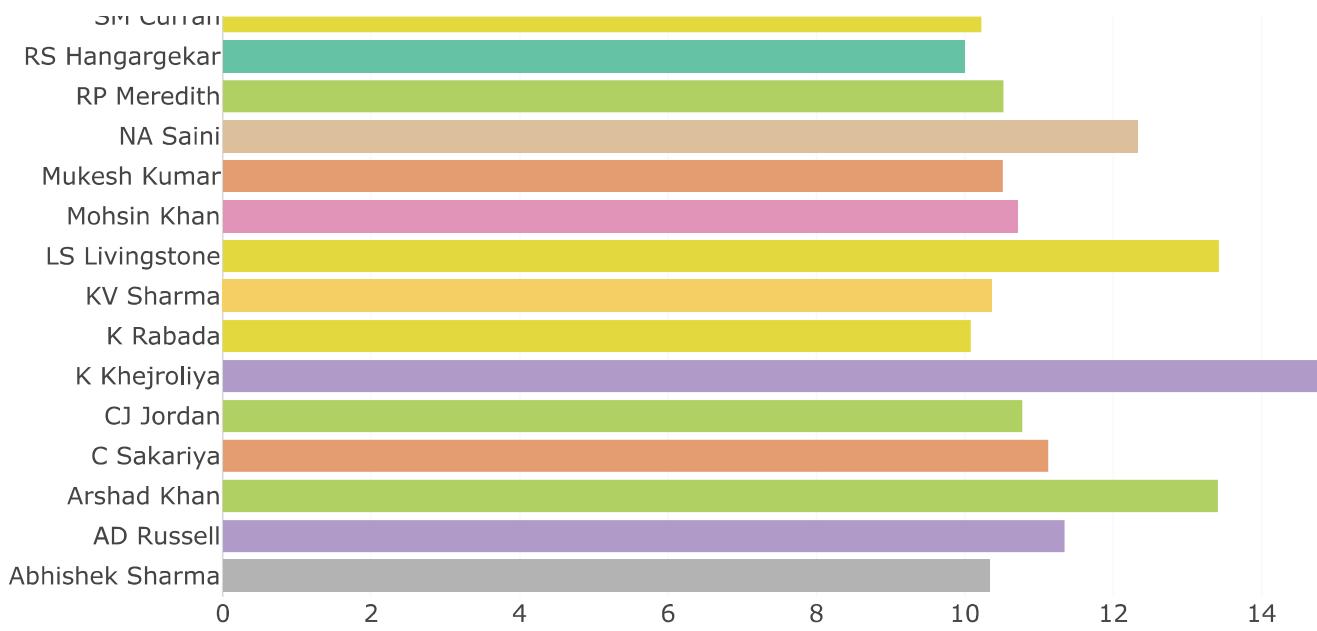
Implications:

- Early Innings Dynamics: Lower run accumulations showcase varied playing styles with diverse strike rates.
- Established Batsmen: Higher run-scorers display a more uniform and predictable approach as indicated by stabilized strike rates.
- Experience Matters: The trend suggests that experienced batsmen develop a consistent batting strategy.

This pattern reveals the evolving dynamics of a batsman's performance and provides insights for strategic decisions related to player roles and team composition.

3.3.11 A Bar Plot for Comparison Between Economy Rates of Top Bowlers





- **Observation:** Teams with lower economy rates among their bowlers tend to have better overall team performance.
- **Implication:** Bowling efficiency, as indicated by a lower economy rate, is a key factor contributing to the team's success.

3.4 Conclusion

- **Summary:** These observations collectively reveal several key relationships within the dataset, emphasizing the interconnected nature of individual player performances and their impact on overall team success.
- **Recommendation:** Consider leveraging these insights for strategic decision-making, player selection, and team composition to enhance performance in future seasons.

By presenting your observations in a structured and clear manner, you convey the significance of the relationships observed in your project data. This approach helps readers understand the implications and potential strategies that can be derived from the analyzed patterns.

4. Summary Tab

A "Summary Tab" is a feature within our App that provides users with a convenient way to access summarized information about bowlers or batsmen for a specific year. The tab is designed to enhance user experience and make it easy for IPL enthusiasts to retrieve relevant data. Here's how it functions :

1. Year Selection -

- The user is presented with a drop-down to select the desired year from 2013-2023. This allows them to focus on the specific time period they are interested in.
- The summary tab is dynamic and adjusts its content based on the selected year.

2. Batsman/Bowler Selection -

- The User can select if he/she wants to see information about a Batsman or a Bowler.

3. Search Bar -

- A search bar is provided to allow users to type in the names of the bowler or batsman they want information about. This feature caters to users who are interested in individual player statistics.

4. Information Display -

- Once the user selects a year and enters the name of the player, the summary tab displays relevant statistical information for the specified player in the chosen year.

Running Code

Below is the "UI Code" that is used to structure the summary tab -

Below is the "Server Code" that outputs the summary tab section -

Upon running the above code, you'll see this on your window -

5. Maps

A "Maps Tab" using Leaflet library displays the nationality of batsmen and bowlers. This is visually engaging and an informative feature for users interested in IPL. The tab consists of -

1. Batsmen/Bowlers -

- Here, you can choose if you want to see the nationalities of 2023 IPL Purple hats or Orange hats!

2. Country Selection -

- You can filter player nationality visualization country-wise. You can choose from among 10 countries.
- The names of the players will be displayed on the Markers, so that it would be easier for the user to locate the player.

3. Map Display -

- The Map of the selected country with the number of players in that country will be displayed on the top.

The data set to build the Map was -

Rank	Player	Matches	Innings	Not Outs	Runs	Highest Score	Average
1	Shubman Gill (GT)	17	17	2	890	129	59.33
2	F du Plessis (RCB)	14	14	1	730	84	56.15
3	DP Conway (CSK)	16	15	2	672	92*	51.69
4	V Kohli (RCB)	14	14	2	639	101*	53.25
5	YBK Jaiswal (RR)	14	14	1	625	124	48.07
6	SA Yadav (MI)	16	16	2	605	103*	43.21
7	RD Gaikwad (CSK)	16	15	1	590	92	42.14

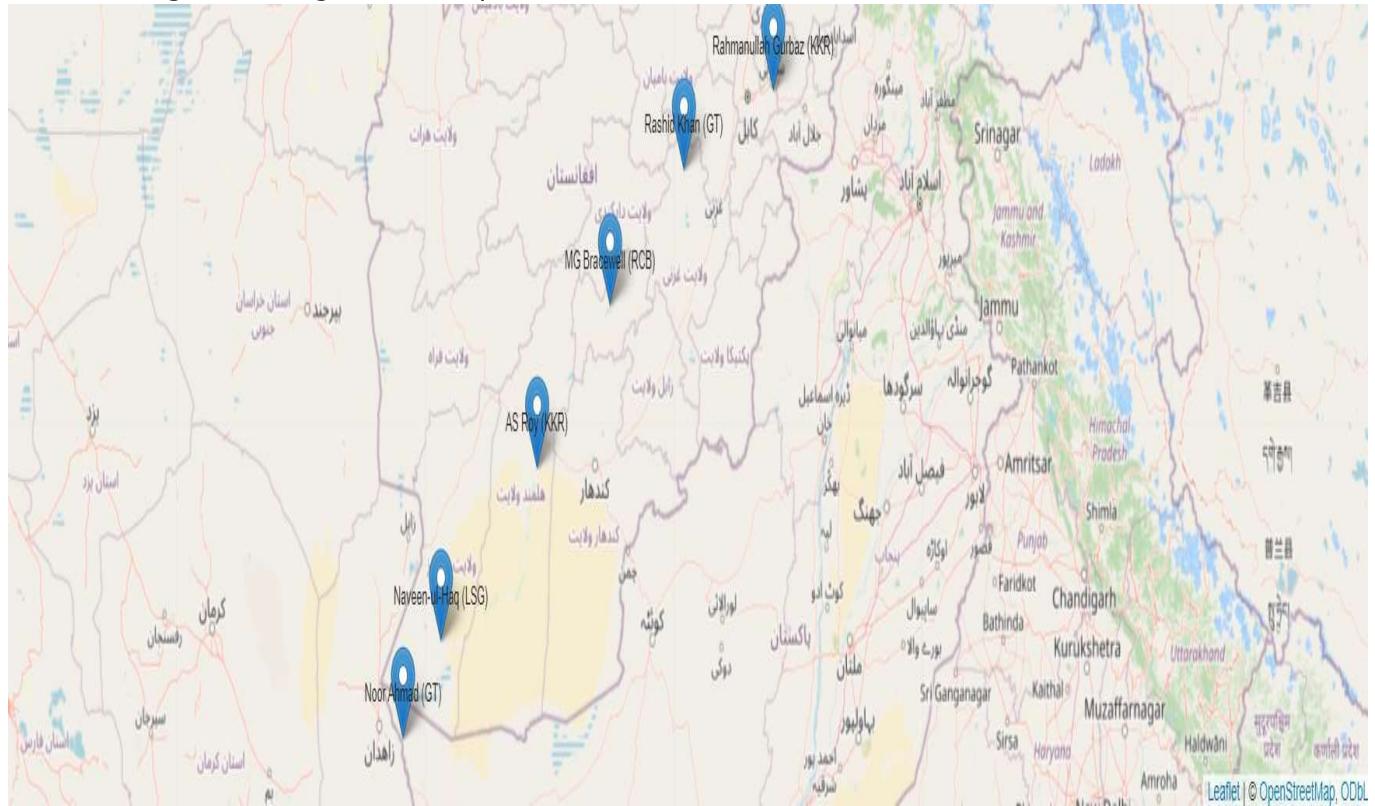
8	8	DA Warner (DC)	14	14	0	516	86	36.85
9	9	RK Singh (KKR)	14	14	6	474	67*	59.25
10	10	Ishan Kishan (MI)	16	15	0	454	75	30.26
		Balls Faced	Strike Rate	100s	50s	0s	4s	6s
1		564	157.80	3	4	1	85	33
2		475	153.68	-	8	-	60	36
3		481	139.70	-	6	1	77	18
4		457	139.82	2	6	1	65	16
5		382	163.61	1	5	1	82	26
6		334	181.13	1	5	1	65	28
7		400	147.50	-	4	-	46	30
8		392	131.63	-	6	2	69	10
9		317	149.52	-	4	-	31	29
10		318	142.76	-	3	-	54	18

Running Code

Libraries Used - maps and leaflet

The code below was used to make the Map -

The following code will give this Output -



6. Auction

The Auction tab is a special tab in our app that allows user to simulate an auction. There will be two teams and the players will be allowed to name their own team.

Eleven random batters and eleven random Bowlers will be randomly chosen among the players who played the 2023 edition of IPL. Then the 22 players will be displayed to the two players.

Each Team will have a starting purse of 50 lakhs. Before the start of Auction for each player, the teams will have to enter how much they want to bid for the player. The strike rate and average of the

players will also be displayed along with the name of the players so that the teams can make an informed decision.

Once the teams have entered the prices, you have to press the 'Start Auction' button. The player will be sold to the team which has a higher bid.

A team cannot buy more than 11 players. After both the teams have bought their eleven players, the probabilities of their teams winning will be displayed.

Calculation of Probability:

For each team, we calculate their 'batting score' and their 'bowling score'.

The Batting strike rate is calculated by first normalizing the strike rates of all the batsmen between 0 and 1. For the Bowling Strike rate we do the same but we subtract the final strike rates by 1 to account for the fact that for a bowler the lower the strike rate, the better.

Then for each team we take average of their bowling scores and batting scores. Then we sum the batting scores and the bowling scores to get the "Team Score".

The final Probability for a Team, say Team 1, will be : $(\text{Team Score 1}) / (\text{Team Score 1} + \text{Team Score 2})$

Below is a picture of the auction tab before the start of the game-

The screenshot shows a user interface for an auction game. At the top, there is a navigation bar with tabs: Mth208Project, Visualization, Auction (which is highlighted), Maps, and Summary. The main area is divided into three panels:

- Hello!**:
 - Enter your Team Name:
 - Enter the money you want to put in
Only spend when you can buy:
 - Remaining Money For Spend More Please:
 - LABEL TEXT's Players:
- Data of the Player**:
 - Players to choose from:
 - \$Name: Player
61 KD Karthik
 - \$Rank: Rank
61 61
 - \$Strike Rate: Strike.Rate
61 134.61
 - \$Average: Average
61 11.66
 - \$Type: Type
61 BATSMAN
- Hello!**:
 - Enter your Team Name:
 - Enter the money you want to put in
Only spend when you can buy:
 - Remaining Money For Spend More Please:
 - LABEL TEXT's Players:

7. Conclusion

In this report we have done a comprehensive analysis of the data of Indian Premier League and drew numerous conclusions from it. We also made a fun game where player could test their knowledge about IPL and their auction skills.

8. References

For Most of the data: <https://www.espncricinfo.com/>