Dashboards and Visualization(MSIS 2629)

Individual project

IPL(Indian Premier League): Analysis for picking best players

Aman Sihmar

W1266471

Under the guidance of:

Prof. Michael Schermann

Objective:

To help the franchise owners in picking the best players, on the basis of player's past records, with the help of interactive dashboards.

About IPL:

The **Indian Premier League** (**IPL)** is a professional Twenty20 cricket league in India contested during April and May of every year by teams representing Indian cities. The tournament is based on a Twenty20 format where each team plays a maximum of 20 overs. The cricket is notorious for being a 5 day game and this tournament proved to be a game changer. The tournament has been a huge hit among the masses because the games last for about 3 hours, hence also bringing in audiences who were not die hard cricket fans. Every tournament starts with a bidding process where the franchise owners bids for the player which they think will play a crucial role in getting their team win the tournament. Usually, the owners are business tycoons and bids large amount of money for the players. This tool will help them in making the choices based on their needs.

Understanding the game:

Understanding cricket and the stats associated with it is little difficult to interpret for anyone who is not a follower of the game. Following are some terminologies which will be useful in understanding the game and the dashboard.

Runs:
This is the primary thing which is used to keep scores between the team. The team batting first scores the run and the team batting second chases the runs scored by the first team. In the end the team who has scored more runs wins the game. On a typical ball a batsmen can score minimum of 0 run and maximum of 6 runs. The runs are scored by the batsmen.

Ball:
A ball is bowled by the bowler and is played by the batsmen. Batsmen will try to score runs on the ball while bowler will try to stop runs. The best way to stop runs is to get batsmen out. There are 10 ways in which batsmen can get out but mainly they are bowled (ball hitting the stumps) or catch out.

Over:
An over constitutes of 6 legal balls (also called deliveries). A bowler bowls an over before another bowler comes to ball another over. In a Twenty20 match a bowler can bowl a maximum of 4 overs.

Batting average:
The average number of runs scored per innings by a batsman, calculated by dividing the batsman's total runs scored during those innings in question by the number of times the batsman was out.

Batting strike rate:
A percentage equal to the number of runs scored by a batsman divided by the number of balls faced. In other words, runs scored per hundred balls.

Bowling average:
A bowler's bowling average is defined as the total number of runs conceded by the bowler divided by the number of wickets taken by the bowler.

Bowling strike rate:
The average number of deliveries bowled before a bowler takes a wicket.

Economy rate:
The average number of runs given by the bowler per over (per 6 balls).

Note: **A batsman is considered good if his batting average and strike rate are on the higher side whereas a bowler is considered good if his bowling average, bowling strike rate and economy rate is on a lower side.**

Data:

The data for IPL(2008-2016) was available on Kaggle. The first csv file(deliveries) has ball by ball details of every match and the second file(matches) has the other details like who won the toss, won the match and player of the match. There were some extra columns which were not needed for the analysis and hence was removed using pandas(python). Both the files have a field match_id and this field has been used to join the two files in Tableau. This is done so that the player's performance can be measured year on year.

Dashboard1 (Batsman Performance)

This interactive dashboard is prepared keeping in mind the characteristics that are desired by a franchise owner in a Batsman.

Broadly batsmen are of 2 types, aggressive and traditional. The aggressive batsmen score runs quickly i.e with a higher strike rate whereas the traditional batsmen will score runs at a slower pace. A team needs both kinds of batsmen because chances are that the aggressive batsmen will get out quickly (lower batting average) while the traditional will score runs at a moderate pace (lower strike rate) but are less likely to get out.

This dashboard uses bar chart because it is very easy to understand. Text tables are also used for the same reason.

Filters used:

1. **Strike Rate/Batting Average:** This is calculated based on what **parameter** has been selected by the user. If the user want to see the details of 'Aggressive' batsmen then this field will display 'strike rate' because this used so that the franchise owner can shortlist player on the basis of their strike rate. For other selection this field will display the 'Batting Average".

2. **Batsmen Run:** The strike rate of batsmen, who has scored very less runs, can be very high and hence can be very misleading. This is why a filter on Batsmen run has been provided.

3. **Performance year wise:** This is provided because a player might have performed very well in the past but the franchise might be interested in the recent performance/form.

4. **% runs in boundaries:** Towards the ends on innings or chasing a high total, the team needs someone who can score boundaries (4 and 6 runs per ball). This filter will give the franchise a better perspective so that they can bid and get both type of players(High boundary scorer and low boundary scorer).

Parameters: As discussed above, batsmen can be selected based on two parameters namely 'Aggressive' and 'Orthodox'. Changing these parameters will change the calculated field accordingly.

Claim 1:

Chris Gayle is a very good player. The franchise should bid aggressively to get him on the team.

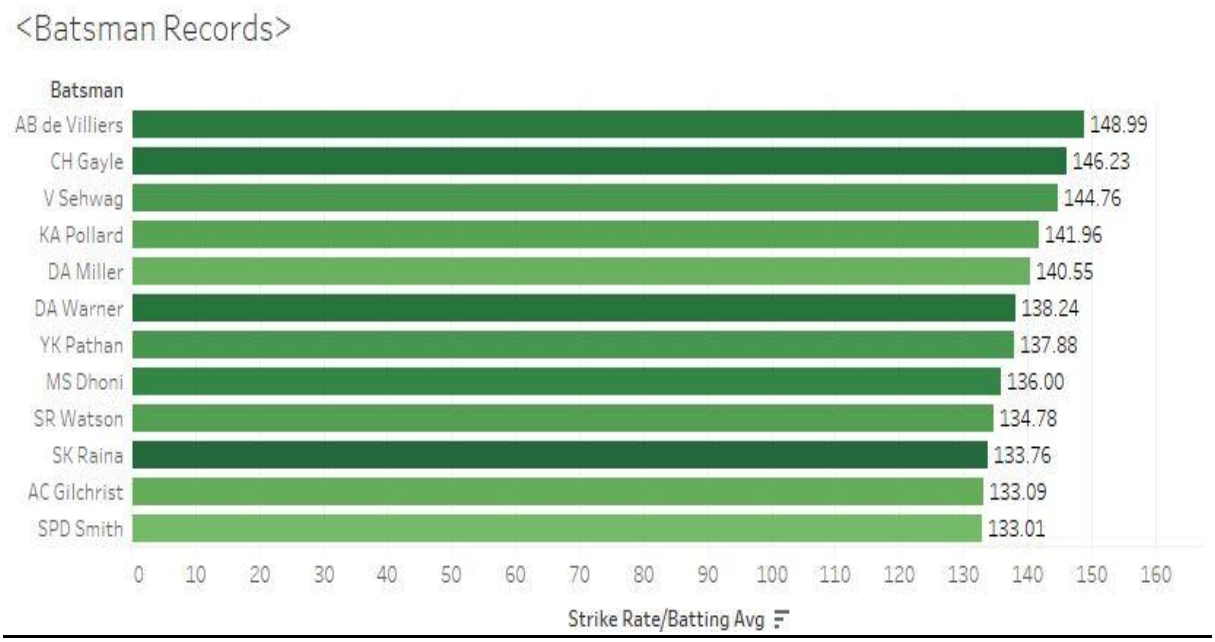
Evidence:

1. **Big Hitter:** Chris Gayle has hit most number of 6's in the IPL history. His tally of 252 6's is unmatched, second on the list is Suresh Raina with 162 sixes. Ability to 6's is a great ability just like hitting home runs in baseball.

Batsman	% Runs in boundries	6s	Batsman Runs
CH Gayle	76	252	3,451
SK Raina	59	161	4,106
YK Pathan	64	143	2,779
AB de Villiers	60	142	3,270
MS Dhoni	55	140	3,270
DA Warner	64	134	3,373
KA Pollard	65	125	1,959
SR Watson	68	121	2,557
V Sehwag	72	106	2,728
AC Gilchrist	73	92	2,069
DA Miller	57	74	1,480
ML Hayden	68	44	1,107
SPD Smith	52	33	1,241

2. **% Runs in Boundaries:** When it comes to % Runs in boundaries, Chris Gayle again leads the chart with 76% of his runs coming in 4's and 6's.

3. **Batting Strike Rate:** is the most desired thing to look for batsmen in Twenty20 game. Chris Gayle's strike rate of 146.23 is second only to AB de Villiers(among batsman who have scored more than 1000 IPL runs)



Rebuttal:

When we take a look at the stats of year 2016, we find that the current form of Chris Gayle is not that good. Even though he is on 2nd place in the list of % of runs scored in boundaries and on 4th number for the number of 6's, he only managed to score only 227 runs which is not a good number.

Sheet 3

Batsman	% Runs in bo..	6s	Batsman Runs	Batting_Avera..
CR Brathwaite	86.7	8.0	83.0	13.8
CH Gayle	85.5	21.0	227.0	22.7
ER Dwivedi	83.3	2.0	24.0	12.0
BCJ Cutting	80.0	6.0	65.0	21.7
MJ McClenaghan	76.5	3.0	34.0	11.3
JA Morkel	75.0	2.0	16.0	16.0
LMP Simmons	75.0	1.0	8.0	8.0
N Rana	73.1	8.0	104.0	34.7

Dashboard 2 (Bowler's Performance)

This interactive dashboard is prepared keeping in mind the characteristics that are desired by a franchise owner in a Bowler.

Bowler's performance is measured by his ability to take wickets and to bowl economically. Bowling economically means giving less runs to the opposition batsmen.

Filters Used:

1. **Bowling Strike Rate:** This is a very important characteristic, the franchise owner can select the bowler whose strike rate is on lower side.

2. **Dot Balls :** Dot balls are like gold. Dot ball is a ball where the batsmen are not able to score runs. It's like strike in baseball.

3. **Wicket:** A wicket is the best dot ball. This is the most sought out stat in a bowler.

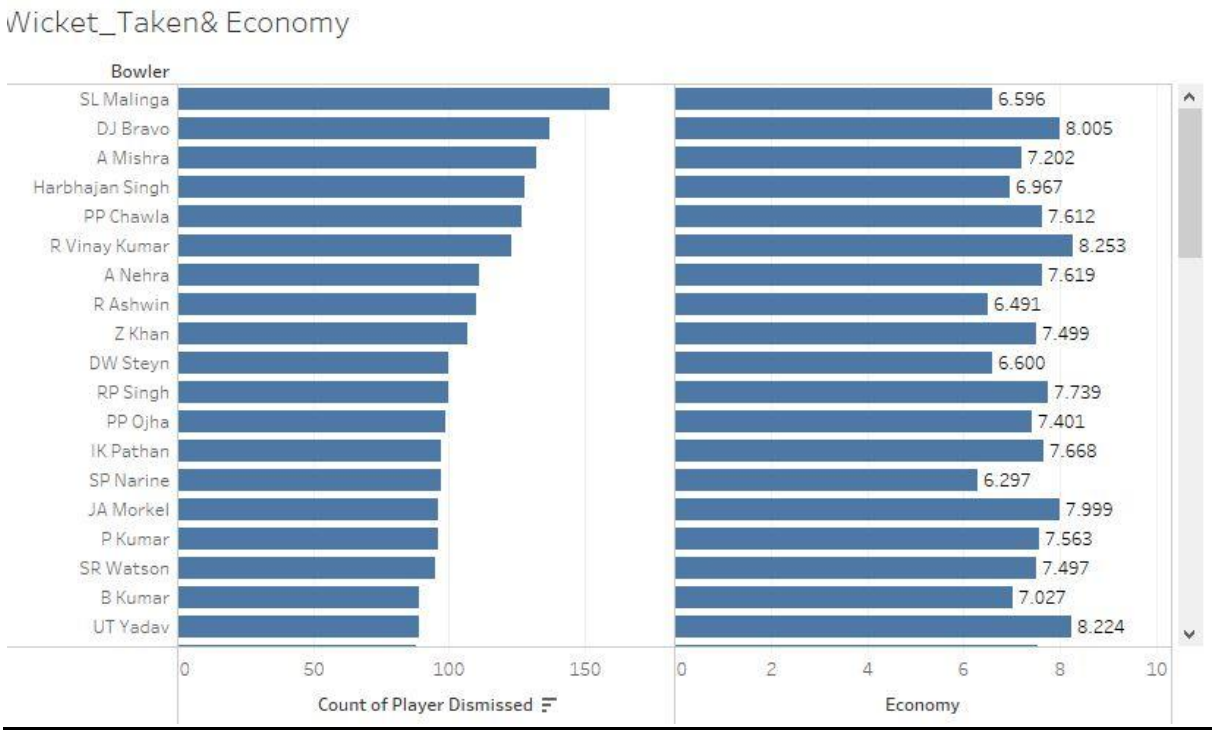
4. **Economy Rate:** Even if the bowler is not among wickets, he can support the other bowler by not giving away too many runs hence increasing the pressure on the opposition.

Claim 2:

Lasith Malinga is the best bowler and franchise should bid to get him in the team.

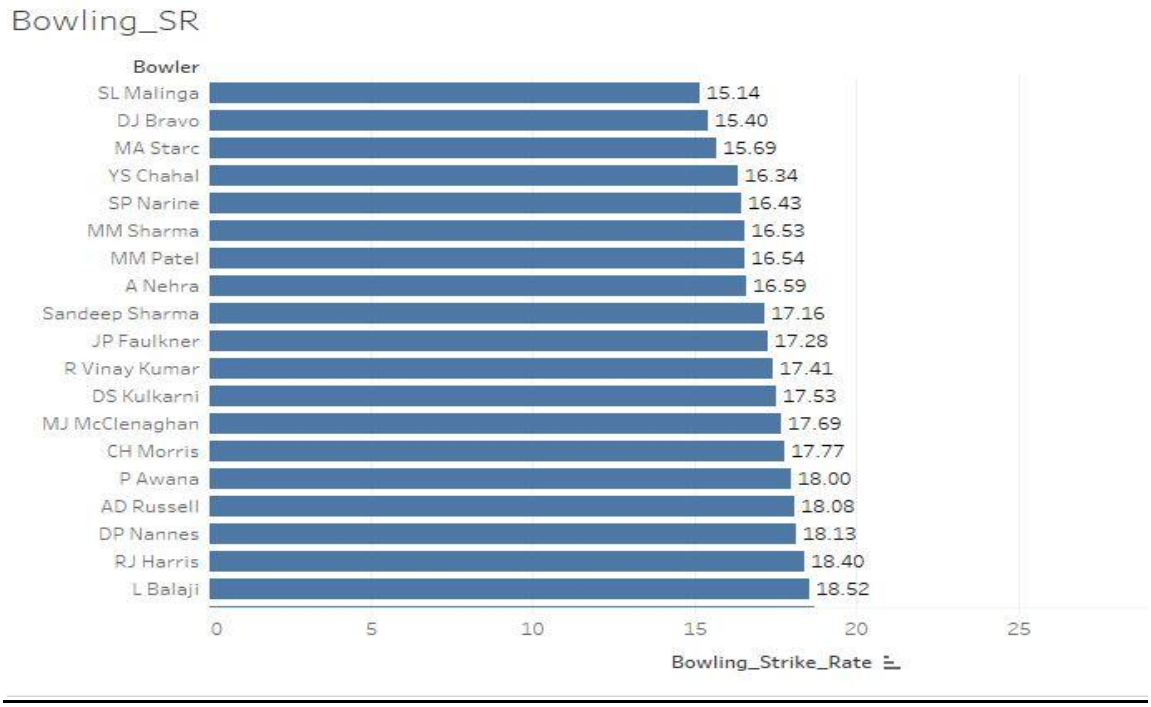
Evidence:

1. **Wickets:** He has the maximum number of wickets in the history of IPL tournament.



2. **Economy Rate:** The economy rate of 6.56 runs/over is among the top three in IPL history.

3. **Best Bowling Strike Rate:** Lasith Malinga has the best strike rate among the bowlers who has bowled more than 100 overs. The strike rate of 15.2 means that he takes a on every 16^{th} ball.



4. **Dot Balls:** Malinga has bowled 970 dot balls, 3rd highest of all time.

Rebuttal

The franchise usually wants the players who are big crowd pullers so that audience come to watch them and they can get a good value of money. If this is what franchise want , they can also bid for R. Ashwin. Since IPL is played in India, bidding for an Indian player will pull bigger crowd.

What could have been done better to improve the analysis(Critique)

As I have said earlier, cricket is such a vast game with so many stats and figure. The analysis of a players performance cannot be measured only by numbers, it is also depends on how he handled pressure and rose to the occasion. There are some other factors that could tell us a little more about the players.

1. **Player contribution in matches in which team won vs in matches in which team lost:** This analysis would help in finding out the match winners.

2. **Inning wise Player's Analysis:** This would tell us how a player performed in 1st innings vs second innings.

3. **Bowling performance in death overs:** Death overs are the last 4-5 overs in the innings. The batsmen try to hit more runs in death overs hence bowlers go for maximum runs. A death over analysis would bring out the best death bowler which every team wants.

4. **Analysis for All rounder's:** All rounder's are those players who can bat and bowl. They are a huge asset to any team. An analysis for all rounder's would be very useful for the user.

5. **Analysis for Biggest crowd puller:** In the end, things comes back to money, an analysis where the franchise can find the biggest crowd puller. These crowd puller will help in generating revenue for the franchise.

Summary:

The dashboards designed are interactive and helpful and its primary purpose is to help the franchise owner in selecting the best teams. Even though, I have provided some suggestions from my side regarding the selection of players, the user can select the players based on the criteria which suit them best. The dashboards have many filters which will be helpful in doing drill down analysis. There are so many factors that plays an important role in selection of a player(other than performance), these factors are not considered in the scope of this analysis.

References:

https://en.wikipedia.org/wiki/Indian_Premier_League
Data Source: https://www.kaggle.com/manasgarg/ipl

Tableau Public Link
https://public.tableau.com/profile/aman1220#!/vizhome/Project_207/Dashboard2

Github Link: https://github.com/AmanSihmar/Dash_Project

Deliveries.csv and matches.csv are the original files downloaded from kaggle.