Analyzing IPL data

Krishna Solanki

Applied Data Science Capstone

1. **Introduction**

1.1 Background

Indian Premier League (IPL) is the one of the best cricket league around the globe with the millions of fans worldwide. Which team will be champion that depends on the teams each players performance. IPL auction occurs every 3 years and players get bid on their performance during a year in international match. It’s good platform to debut for young players and one who perform best get emerging player of the tournament award and likely to get a chance in international team.

1.2 Problem

Data will help to analyze players performance on which delivery player got out and how he got out. This data will help team management to improve players performance and bring out their weakness and improve them as well. And by analyze this data we can predict that which team will become champion.

1.3 Interest

Each team management needs to analysis of each player before auction. They want to know player performance throughout the year and past seasons. Then they get a idea to put a bid on a player and which player they will need to balance the team.

1. **Data acquisition, cleaning and analyzing**

2.1 Data Sources

You can find [IPL dataset](https://www.kaggle.com/manasgarg/ipl) here captured from Kaggle. This dataset contains IPL cricket match data between 2008 and 2016. This dataset contains 2 files named deliveris.csv and matches.csv

[deliveries.csv](https://www.kaggle.com/manasgarg/ipl?select=deliveries.csv) is the ball-by-ball data of all the IPL matches including data of batting team, batsman, bowler, non-striker, runs scored, etc. [matches.csv](https://www.kaggle.com/manasgarg/ipl?select=matches.csv) contains details related to the match such as location, contesting teams, umpires, results, etc. Geo-coordinates of city will be obtained with the geocoder tool in the notebook.

2.2 Methodology

Firstly, we need to look at data and remove unnecessary data. For that we will drop column “umpire3” since all values are none. And we’ll replace team name with their abbreviation for our convenience.

After that we’ll get latitude and longitude of the city where match has been played. For that we’ll import “Nominatim” from geocoder.

from geopy.geocoders import Nominatim

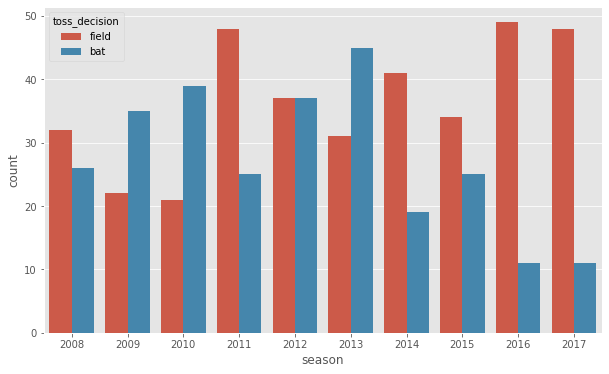
locator = Nominatim(user\_agent='myGeocoder')

Graphical user interface, text, application, email

Description automatically generated

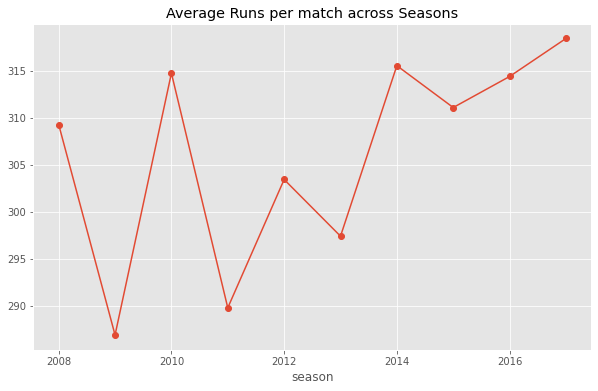
Toss Decision

We will start our analysis with the team that wins the highest number of matches and player that won most number of the man of the match awards. When we come up to toss decision after winning the toss in 2011 and 2014 to 2017 it’s most likely to team opt for a field.



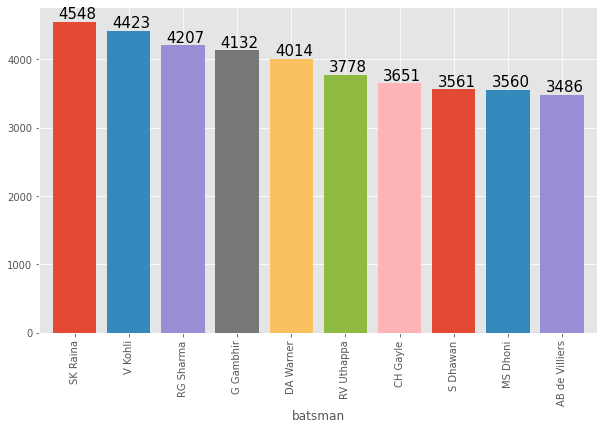
Average runs per match across season

Below graph shows the average runs per match by both teams combined. We can clearly see hiking in runs in 2010 compare to 2009. And it’s highest in 2016.



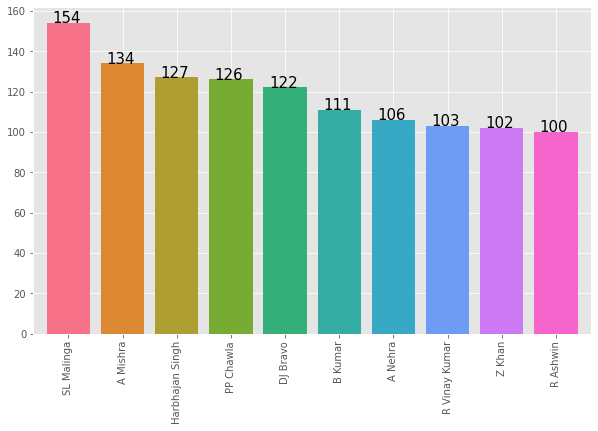
Top batsman

When we see the top 10 batsman of IPL Suresh Raina is on top followed by Virat Kohli.



Top bowler

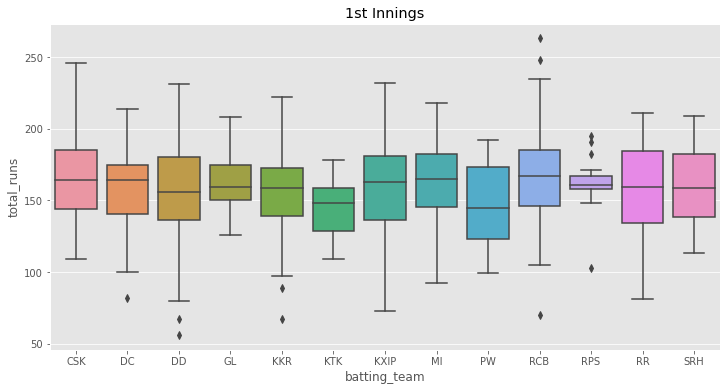
If we go for top 10 IPL bowler then SL Malinga on first number with 154 wickets. Thanks to his unpredictable bowling action. Remaining bowlers in the graph doesn’t have much wicket margin.

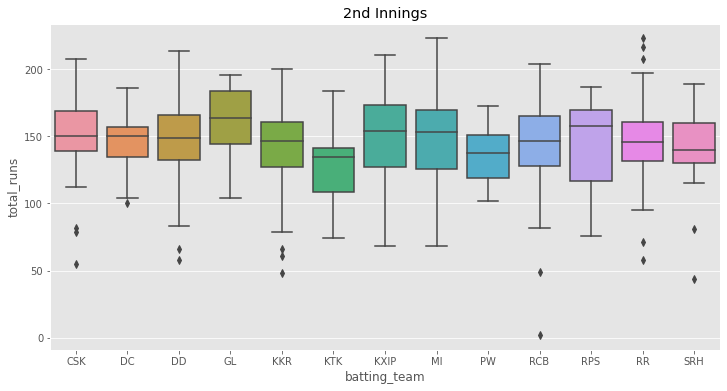


Score distributions for teams by innings

According to 1st boxplot CSK is the best scorer when it comes to batting first and as per 2nd boxplot it shows MI is looking good when comes to 2nd batting.

In 2nd boxplot we see a point near 0 which may seem to be outlier. But it is so because the match was disrupted.





Match venue

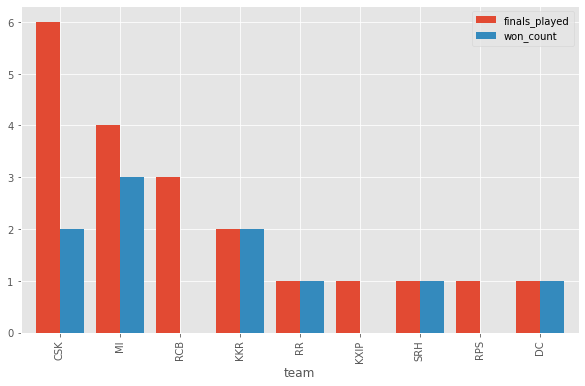
When we come up to venues where match has been played it shows in below map in blue and orange circle. We plotted the map using folium and using “Stamen Terrain” tiles. When you click on the circle it shows the name of stadium where match has been played.

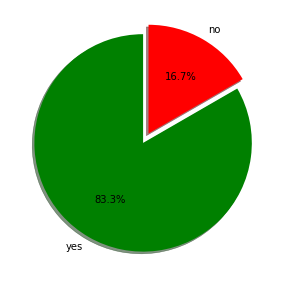
Map

Description automatically generated

1. **Results**

The main thing is champion. We’ll see which team is played final and won cup and how toss is important in finals.





Above bar graph says that CSK played most number of finals and MI becomes most of the times champion. And pie chart shows how important toss winning in finals. Pie chart shows 83.3% chance to win final if you win the toss.

1. **Recommendation**

By above analysis we can say that after winning the toss what to choose bat or field and by that we can predict which team will win and predict champion in finals. During auction this analysis could be helpful for team management by seeing most successful batsman and bowler. And captain get an idea to choose a bat or bowl first by seeing their performance they are good at chase or defend.

1. **Conclusion**

To sum up they should add players experience so during auction team management get an idea to bid on player but here I created graph for most successful batsman and bowler by that owner get to know performance of players and get an idea to put bid on player and buy them during auction.