DEPARTMENT OF BIG DATA ANALYTICS

ST. JOSEPH'S COLLEGE (AUTONOMOUS)

BENGALURU – 560027

PROJECT REPORT

Submitted by:

MUHAMMED HISHAM T

21BDA43

INTRODUCTION

Indian Premier League is a famous cricket league played in India every year during the period between March and May. It is played in a professional T20 format of the game. It is played by 8 franchises representing 8 different cities across India. The league started in 2008 and it has been a big hit and it is the most attended league in the World. Each team should consist of 18 – 25 players, with a maximum of 8 overseas players.

Statistics have always had a significant role in sports. Sports analytics is on the rise and will continue to play a significant role in how teams operate, pick their players, how they play the game, etc. The runs scored by a batsman, the wickets taken by a bowler, or the matches won by a cricket team – these are all examples of the most important numbers in the game of cricket. The teams and the individual players can dig deep into this data and find areas of improvement. It can also be used to assess an opponent's strengths and weaknesses.

The analysis is done on the performance of each player.

Purpose

To analyse the performance of players as batsmen and bowlers on the basis of strike rate, economy rate, Sum of runs and runs (scored/conceded) per season and the strike rate of selected batsmen against a selected bowler in a season Provided they have played against each other and to represent this information visually.

DATASET DESCRIPTION

The data for the study was collected from kaggle.com. It is a ball-by-ball dataset from every game from 2008 to 2020.

Number of attributes= 22

Number of instances= 193573

Missing values = N/a

All these attributes were included in the dataset:

MATCH ID, INNING, BATTING TEAM, BOWLING TEAM, OVER, BALL, BATSMAN,

NON-STRIKER, BOWLER, BATSMAN RUNS, PLAYER OF THE MATCH, EXTRA RUNS, TOTAL RUNS, PLAYER DISMISSED, DISMISSAL KIND, SEASON, CITY, DATE, TOSS WINNERS, TOSS DECISION, WINNER, VENUE.

Source: https://www.kaggle.com/patrickb1912/ipl-complete-dataset20082020/metadata

METHODOLOGY

Area of study: IPL is a very popular sport event widely well received and known by most people, In cricket at any point of time, players will either be a batsman, bowler or a fielder. The idea behind this dashboard was to create simple visualisations to comprehend how the players performed in both batting and bowling throughout the seasons showing their consistencies or inconsistencies.

Data cleaning: out of the 22 attributes of the dataset we took all these attributes in to consideration:

SEASON, BATTING TEAM, BOWLING TEAM, OVER, BALL, BATSMAN, NON-STRIKER, BOWLER, BATSMAN RUNS, EXTRA RUNS, TOTAL RUNS, PLAYER DISMISSED, DISMISSAL KIND.

Data Analysis:

- Statistical analysis was the important thing before doing any kind of visualization.
- Graphs were plotted on the basis of performance of each player in each season. This helps the user to understand about each season and players.

  We took all these formulas into consideration:

- Strike rate per season= (Total runs per season /total No. of balls faced) * 100
- Economy rate per season=sum of total runs conceded per season/ overs bowled
- Strike rate of batsmen vs bowler per season= runs scored by batsmen against the bowler / balls faced

UI AND CODING

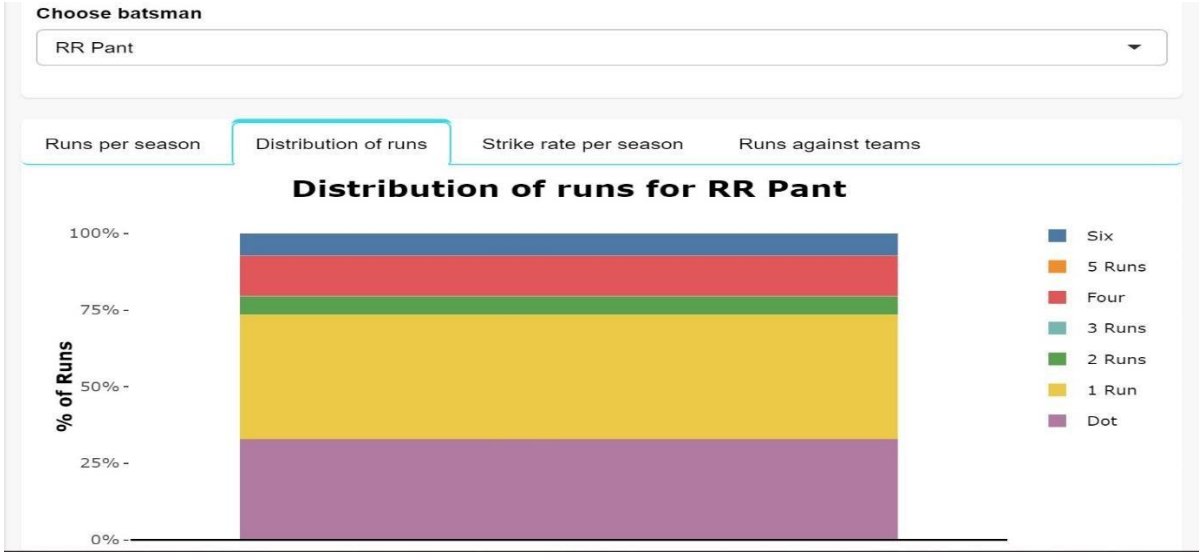
The dashboard was coded on R using Shiny which is an R package that makes it easy to build interactive web apps straight from R.

packages used: tidyverse, dplyr, vroom, ggthemes, plotly, scales, shiny, shinydashboard, dashboardthemes, shinycssloaders, waiter, sever.

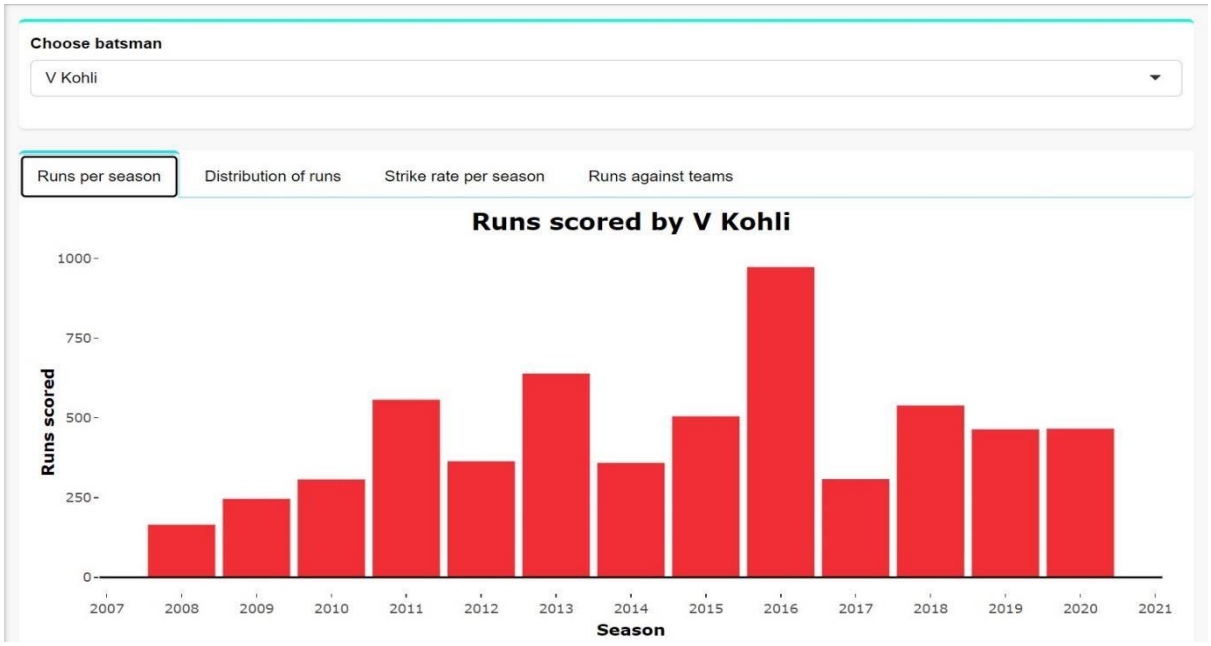
IPL Dashboard:

The about part of the dashboard gives a brief description about IPL followed by a visual representation of teams and their logos.

We used dropdown menus with the names of each player to be selected so that plots of their performances may be seen.



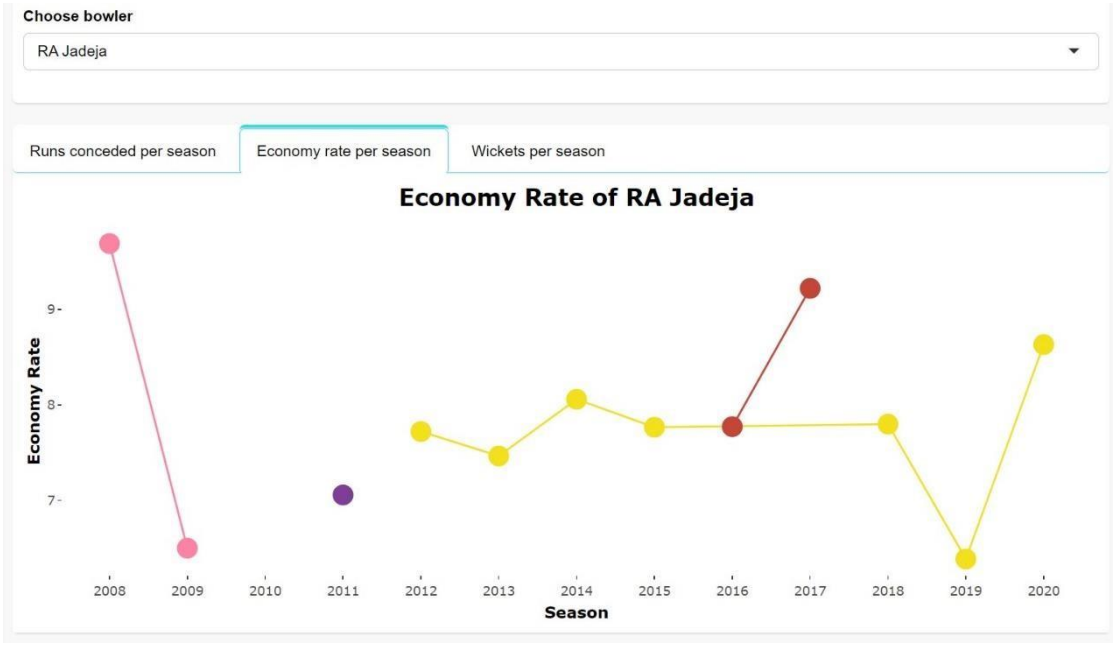
We also colour coded each team such that each plot of players also includes the colour of the team played for or against.



Types of charts used:

- **Bar chart:** A bar chart or bar graph is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. A vertical bar chart is sometimes called a column chart.

- **100% stacked bar chart:** Similar to a stacked bar chart, in a 100% stacked bar chart each series bar represents the percentage of the overall category to which it belongs.
- **Line chart:** A line graph is a graph that is used to display change over time as a series of data points connected by straight line segments on two axes.



RESULT

The aim of the project was to give a visual representation of the performance of players in terms of batting and bowling from 2008 – 2020. We have made this dashboard accordingly that, for a newbie to IPL, they can have an overview of how players have performed in the previous years of the IPL.

It visually represents the aspects of total number of runs scored per season by batsmen, economy rate of bowler, Strike rate of batsmen vs bowler per season, etc. plots were plotted for the easy understanding.

CONCLUSION

IPL's ability to grow and sustain in long term itself is evidence of how much people love this sport. But for a newbie, studying all this i.e., from 2008 – 2020

is not a simple. The IPL dashboard visualizes the performances of each player from 2008 - 2020. It helps us to easily understand the overall overview of IPL.