  **Cricket**

**Vikash Shrivastava**

**Appendix**

1. [About](#About)
2. [Part 1: Description of Cricket Data Used](#Part1)
3. [Part 2: Data Exploration](#Part2)
4. [Part 3: Extracting Details from a dataset to answer a series of questions](#Part3)
5. [Graphical Representation showing various insights of the Cricket Data](#Graph)

**Link to my Selected Dataset-**

##### <https://www.kaggle.com/saivamshi/cricket-world-cup-2019-players-data>

##### I have attached a Script file (written in R) which I have used for my exploratory data analysis of cricket data.

##### Short Description of my Dataset

I have taken this Dataset from Kaggle.com as per the above link, further processed, cleaned, and used for my analysis. I am using Cricket World Cup 2019 Players Dataset from Kaggle to analyse all Batsman and Bowler’s Statistics playing since 2013 till 2019

Below is each CSV file description which I have used for my analysis: -

*Batsman\_Data.csv* – Contains all Batsman’s performance

*Bowler\_data.csv* – Contains all Bowler’s performance

*Ground\_Averages.csv*- It consists information for all matches played in specific grounds across the world during the period 2013-2019

*ODI\_Match\_Totals.csv*- It has information for all ODI cricket matches played during the period 2013-2019

*Wc\_Players.csv* - It lists all players with their unique ID and country name from which they play.

**[About, what is Cricket](#About)**

Cricket is a game which is played between 2 nations consisting of 11 players on each side where one team either win or lose. It is one of the fast-growing sports all over the world. I believe it is a game of self-confidence, hard word, dedication, and lot of patience where the team holding their nerves till the end takes the match.

As a huge cricket fan, I found Cricket dataset best to work on. Through my exploratory data analysis of Cricket Data, I have tried to explore as much data as possible for a better insight. After reading my data analysis of cricket, I believe people will have a clearer idea about game of cricket

For better data analysis, I have explained my presentation in three parts -part 1, part2, part 3

**[Part 1: Description of Cricket Data used](#Part1)**

I have imported 5 CSV files into my R studio.

* *Batsman\_Data.csv* – Contains all Batsman’s performance
* *Bowler\_data.csv* – Contains all Bowler’s performance
* *Ground\_Averages.csv*- It consists information for all matches played during the period 2013-2019
* *ODI\_Match\_Totals.csv*-It has information for all ODI cricket matches played during the period 2013-2019
* *Wc\_Players.csv* - It lists all players with their unique ID and country name from which they play.

After importing, I have cleaned each file manually in my code (please refer R Script file), removed all NAs if any and replaced it with zero. I am using Batsman\_Data.csv to merge with Wc\_Players.csv based on player ID as primary key. I have used inner join to merge these 2 data frames and produced a new Data frame Batsman\_stats1 where I am doing exploratory analysis to answer a series of questions (Please refer attached R Script file for code review).

**[Part2: Data Exploration](#Part2)**

**I have used 4 Data Frames for Exploratory analysis of Cricket Data.**

*Data Frame 1: Batsman\_stats1*

* **This newly produced data frame is produced after joining Batsman\_Data.csv with Wc\_players.csv.**
* **Batsman\_stats1** lists each player performance (**Player, Player ID, Country, X1,Bat1, Runs, BF, SR, Fours, Sixes, Opposition, Ground, Start Date, Match\_ID , Batsman, out\_or\_not** ) for all ODI matches played since 2013 till 2019.
* These information’s are crucial to get an overall idea of a batsman statistics.
* Looking at the new data frame**- Batsman\_stats1,** we might have got the fields required for further analysis, but it is difficult for any person to get any specific information.

Hence, using this data frame, I have simplified it further to answer a series of questions.

*Data Frame 2: Ground information*

* I have created this data frame from the csv file Ground\_Averages.csv
* I have used Ground\_Averages.csv to extract meaningful information from the dataset. In Part 3, I have listed series of question using the data frame Ground information
* Ground information has data for all ODI matches played in different grounds all over the world with column such as Span, Mat, Won, Tied, NR, Runs, Wkts, Balls, Ave and RPO
* I have tried to explore this data frame by focusing on a particular subset.

*Data Frame 3: Bowler*

* I have created this data frame from Bowler\_data.csv
* Bowler data frame contains information for every bowler along with other meaningful data
* Using bar plot, I have answered a question for the Bowler Data Frame ( see page 6 for graph presentation and R script for code review)

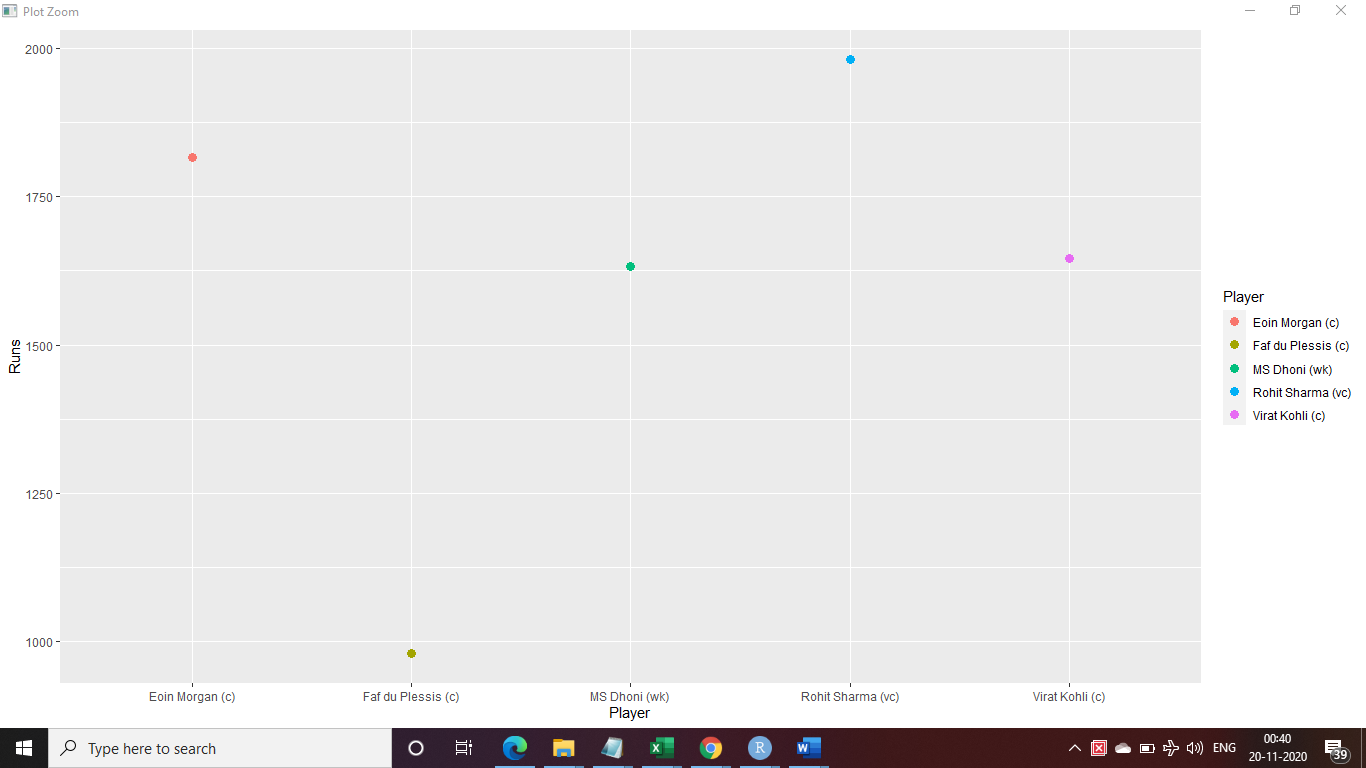
*Data Frame 4: Total\_Matches\_Played\_by\_all\_Teams*

* This data frame is created from "ODI\_Match\_Totals.csv"
* It contains Statistics of all ODI matches Played during the period 2013-2019

**[Part 3: Extracting details from a dataset to answer a series of questions](#Part3)**

I have prepared the below list of question to be pulled from data frame**,** Batsman\_stats1, Ground information, Bowler, Total\_Matches\_Played\_by\_all\_Teams to develop overall understanding of data

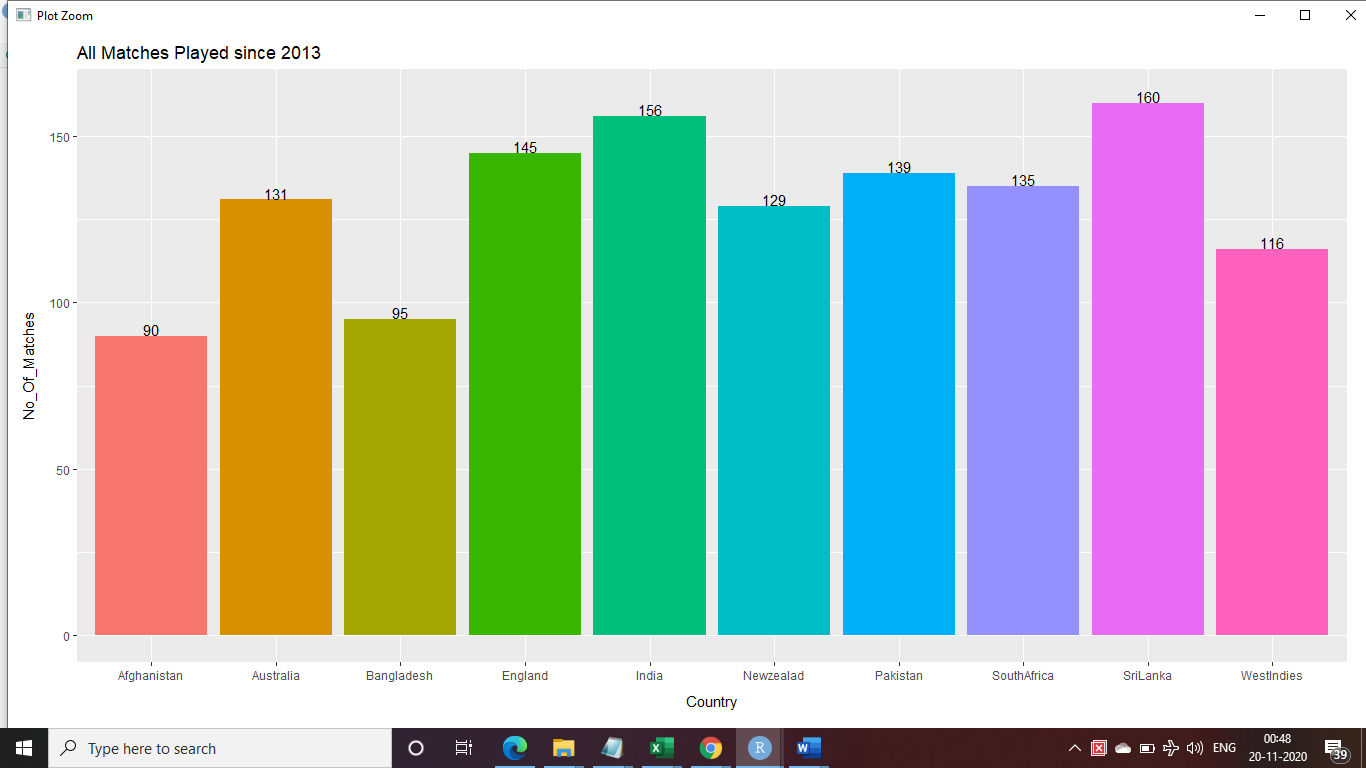
* Top 10 Player with maximum runs scored in ODI Matches since 2013
* Display Top 10 batsman with Maximum number of runs scored only in the year 2019
* Display Top 5 batsman with maximum number of sixes hit since 2013
* Display Top 5 batsman with maximum number of Fours hit since 2013
* Display Top 7 Batsman with best Average since 2013
* Display Statistics of all matches played against Australia
* Display top 5 Batsman with maximum runs scored against Australia since 2013
* Using bar plot, Display Top 5 batsman with maximum runs scored against Australia since 2013 (see page 4 for graph presentation and R script for code review)
* Using bar plot, Display Top 5 batsman with highest Fours scored (see page 4 for graph presentation and R script for code review)
* Using bar plot, Display Top 5 batsman with highest number of Sixes scored (see page 5 for graph presentation and R script for code review)
* Display all the Match Details that have been played only in South African Grounds
* Display only Stadium and Ground Name from all the matches played in South Africa since 2013
* Plot a bar chart displaying all matches played by each team since 2013 (see page 5 for graph presentation and R script for code review)
* Bar Chart showing Top 5 bowlers who took maximum wickets since 2013 (see page 6 for graph presentation and R script for code review)

[U](#Graph)[sing bar plot, Display Top 5 batsman with maximum runs scored against Australia since 2013](#Graph)

Using bar plot, Display Top 5 batsman with Maximum Fours

Using bar plot, Display Top 5 batsman with highest number of Sixes scored

Display all matches played by each team since 2013 through Bar plot



Using Bar plot, Display Top 5 bowlers who took maximum wickets since 2013

