**IPL 2022 DATA ANALYSIS USING PYTHON**

**Introduction:**

The dataset that I am using for the task of IPL 2022 analysis is downloaded from Kaggle and the link for the data set is as follows.

**Data set:**

Here is the data set used for IPL 2022 Analysis - <https://www.kaggle.com/datasets/aravindas01/ipl-2022dataset>

**Problem Statement:**

I have consolidated the data for each of the player sold in the IPL 2022 auction and that of the retained players from each franchise.

**The data contains information like** – matches played, runs, wickets, average, strike rate, catches, runouts stumps etc.,

**Based on these data points –** we will try to create the best 11 from these set of players from the current campaign.

**We can summarize our analysis in the following steps:**

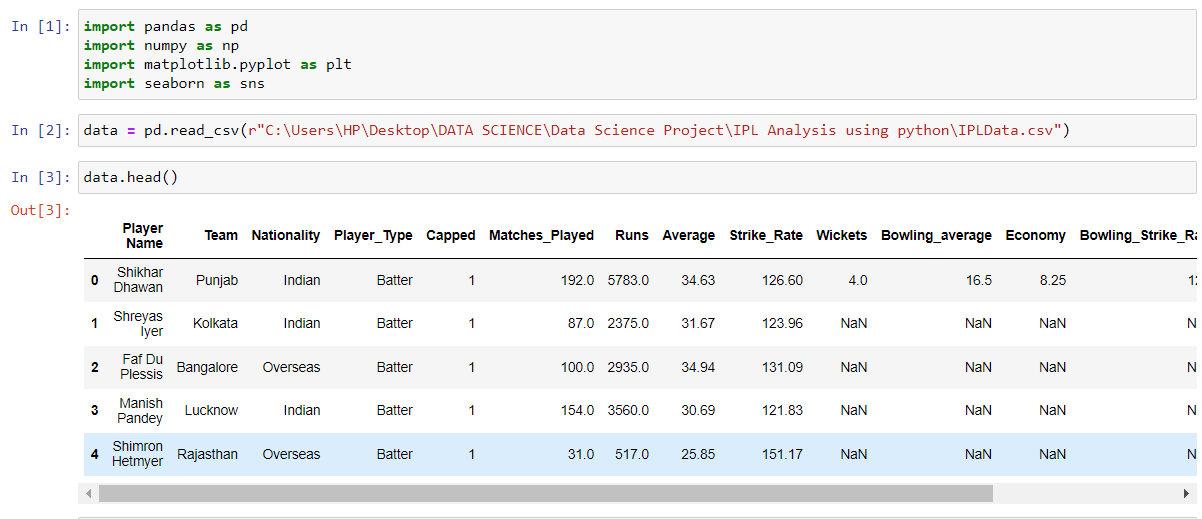
* Extraction and loading the data.
* Cleaning the data and removing the noise.
* Analyze the data on different parameters.
* Visualizing the important statistical findings.
* Making the best team of 11 based on the ICC world cup winning squad formation and last years’ squad formation of the winning team in IPL ie., how many batters, bowlers and all rounders should be included in the team.

**Procedure to perform Data Analysis for IPL 2022:**

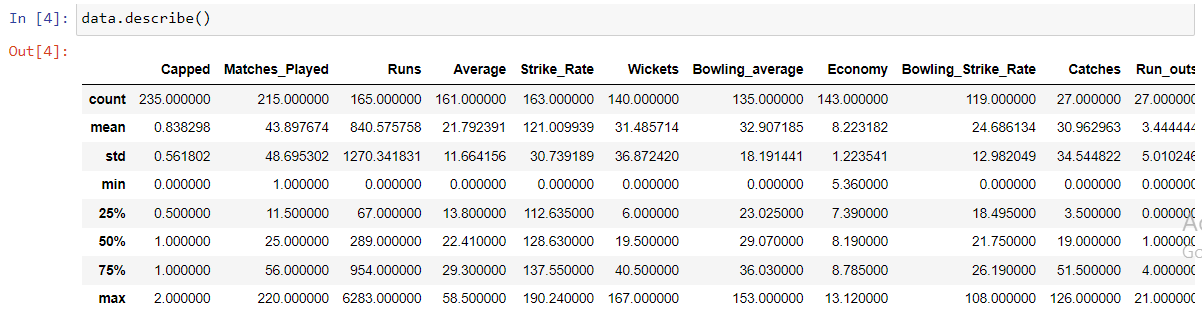
**Step 1**: Import necessary libraries such as Pandas, Numpy, Matplotlib and Seaborn.

**Loading the Data:**

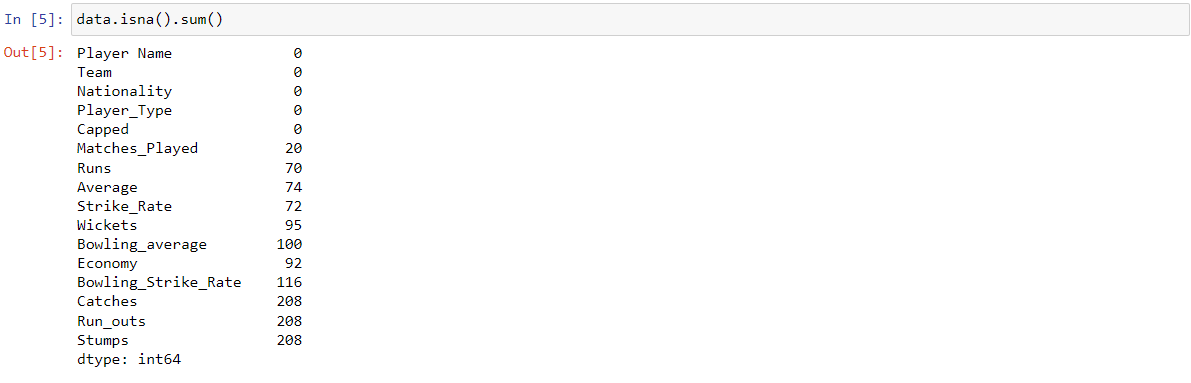
**Step 2:** To import the dataset, we will use the pandas package from the python programming language. The other packages like Matplotlib, seaborn and numpy will be useful for computations and creating visual representations.



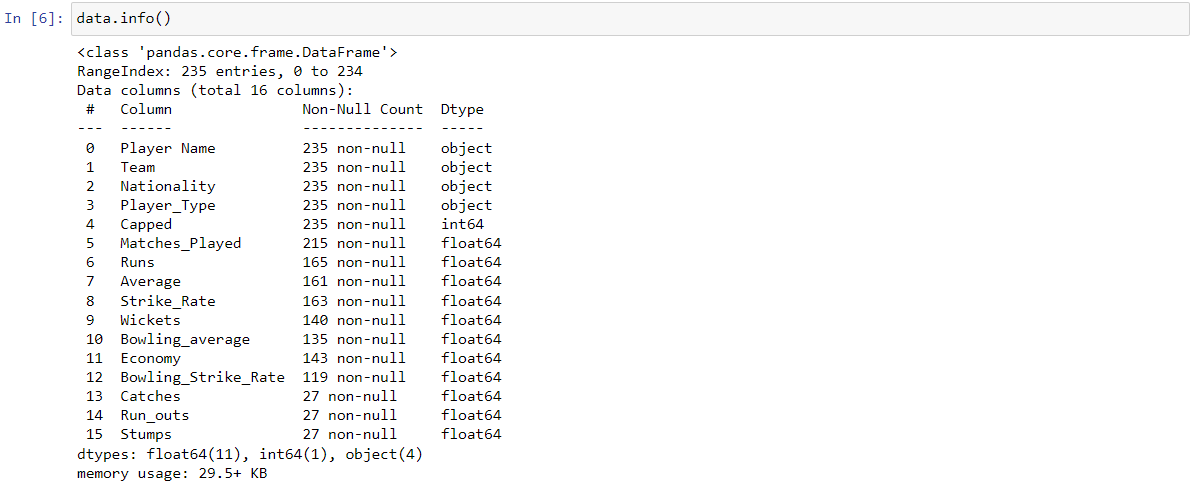
**Step 3:** Describe function will return the necessary information like count, mean, std deviation etc., for the entire dataset.



**Step 4:** We can check the number of null values using the **isna().sum()** method.



**Step 5:** The info method will return the information about the data set like the non null objects and the data type of each of the elements in the data.



**Cleaning the Data:**

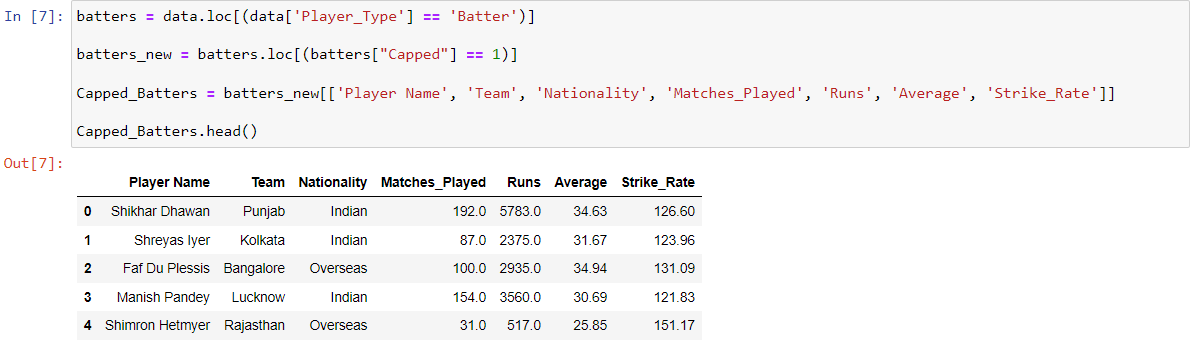
In the above data, the players can be initially segregated based on following parameters,

Capped – Batters, Bowlers, All-rounders, Wicket Keeper

Uncapped – Batters, Bowlers, All-rounders, Wicket Keepers

The uncapped players are playing the IPL for the first time or relevant data isn’t available therefore we have used their domestic T20 performances and T20 international performances.

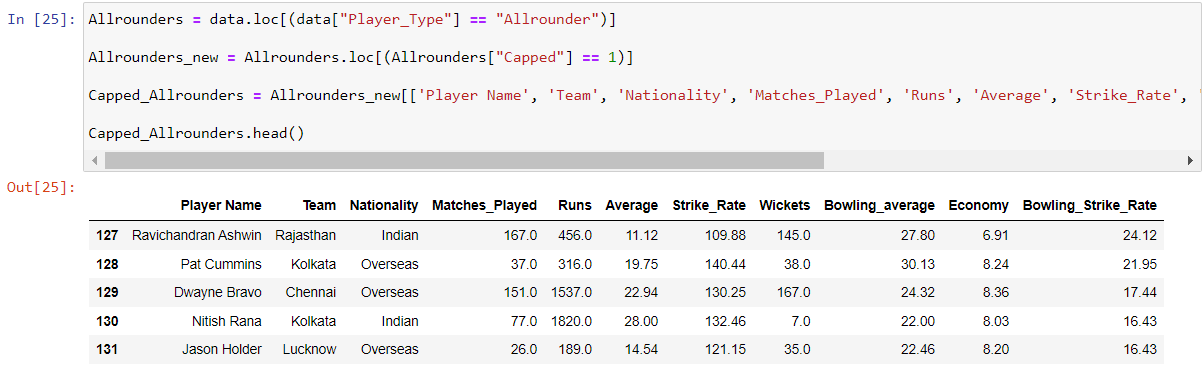
**Step 6:** We have kept the batting parameters to restrict the analysis for batter specific tendencies. The segregation of the batters based on the capped data gives us the batters that have a history of playing IPL matches.



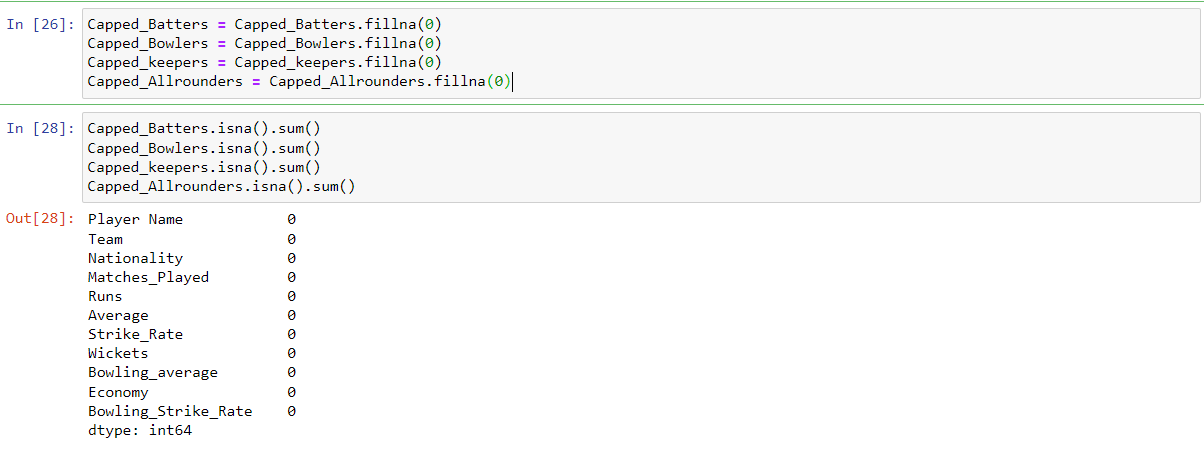
**Step 7:** Since we are segregating data for Capped keepers we are extracting relevant parameters for the keepers, since bowling figures are not needed for the wicket-keepers. The capped keepers’ data consists of keepers that have a history of playing IPL matches.



**Step 8:** Next step is to segregate Capped All rounders data. For that we’ll keep the bowling and batting parameters for the all rounders to analyze all round performance.



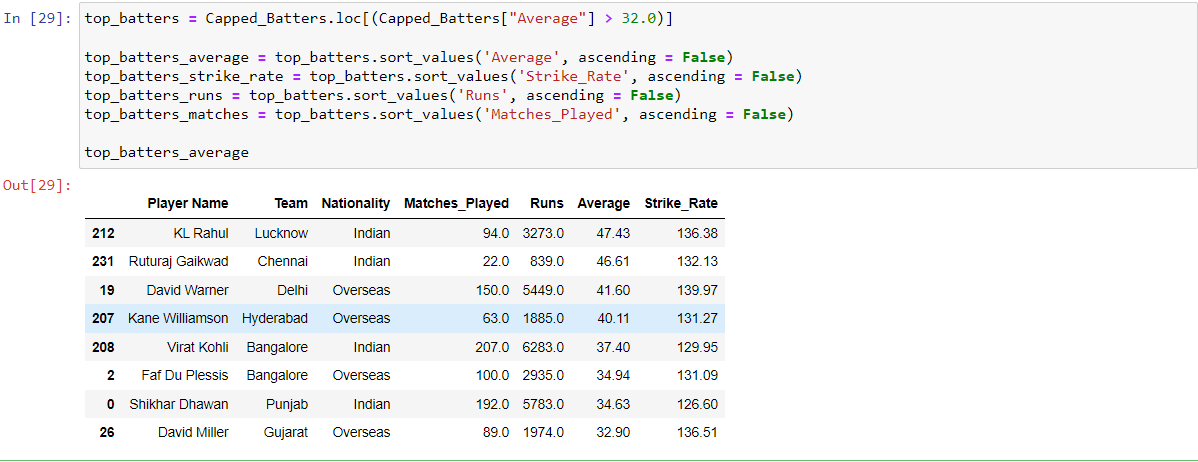
**Step 9:** Cleaning the data by making the null or NaN values 0. Fill the null values to 0 and check for the null values as following.



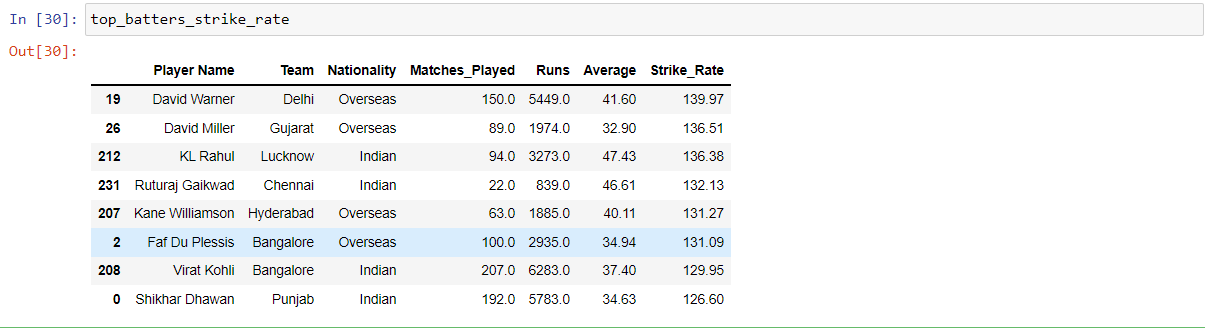
**Initial Analysis:**

Here we will analyze data for batters, bowlers, keepers and all rounders.

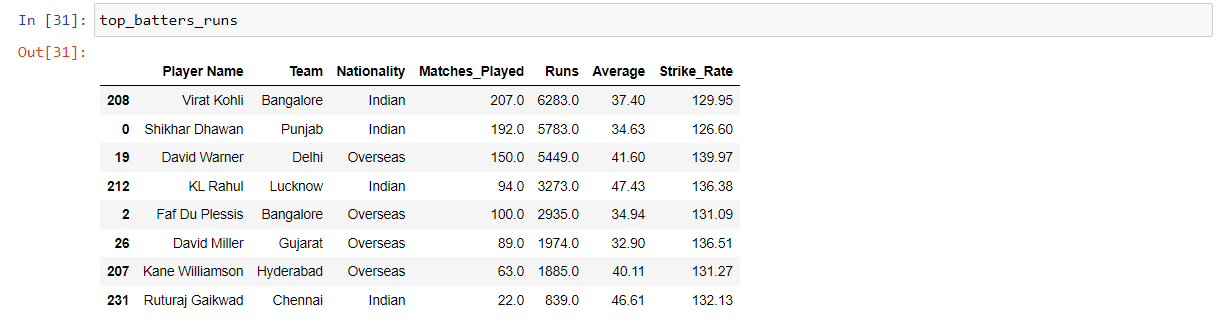
**Step 10:** Here we have narrowed our analysis to batters who have a batting average more than 32.0 and it has been sorted in descending order.



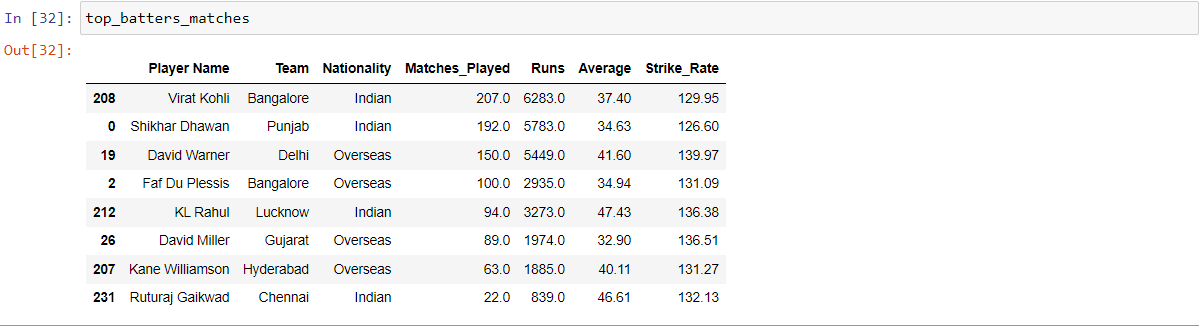
**Step 11:** Lets’ view the batters data in the descending order of strike rate.



**Step 12:** Check batsman runs in descending order.



**Step 13:** Check the batters matches’ plays data in descending order.

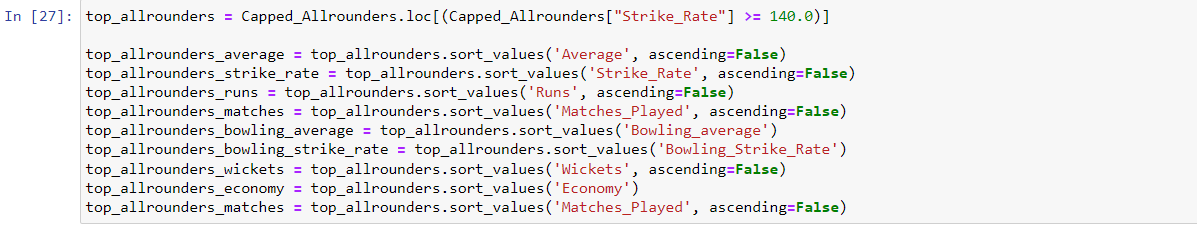


From our analysis if we rank from 0-10, the top three batters that will come while analysis each of the above data are:

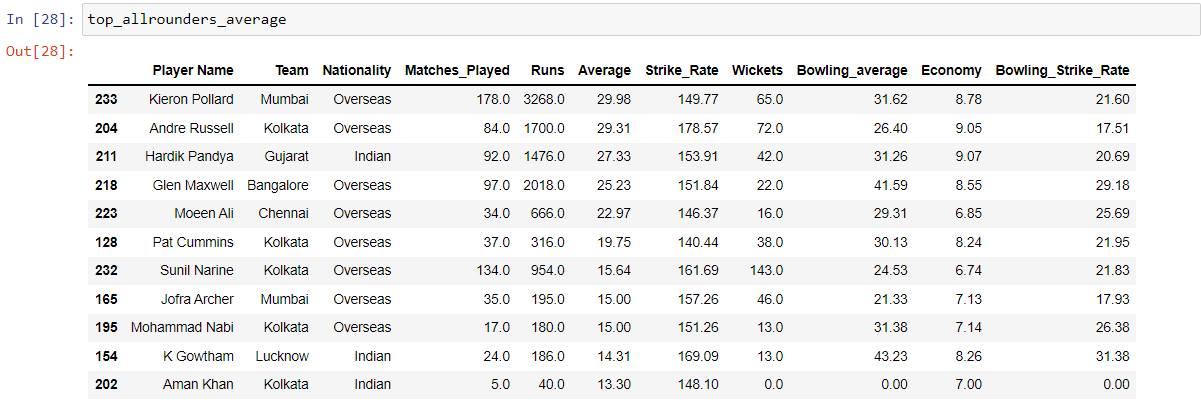
1. David Warner
2. KL Rahul
3. Virat Kohli

**Analyzing all-rounders’ data:**

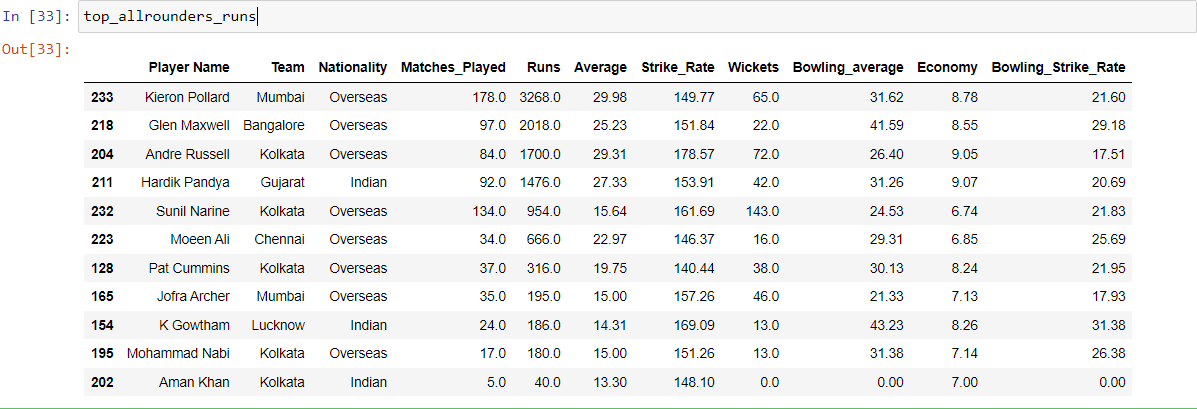
**Step 14:** Analyze all-rounders data by applying the following functions for individual values



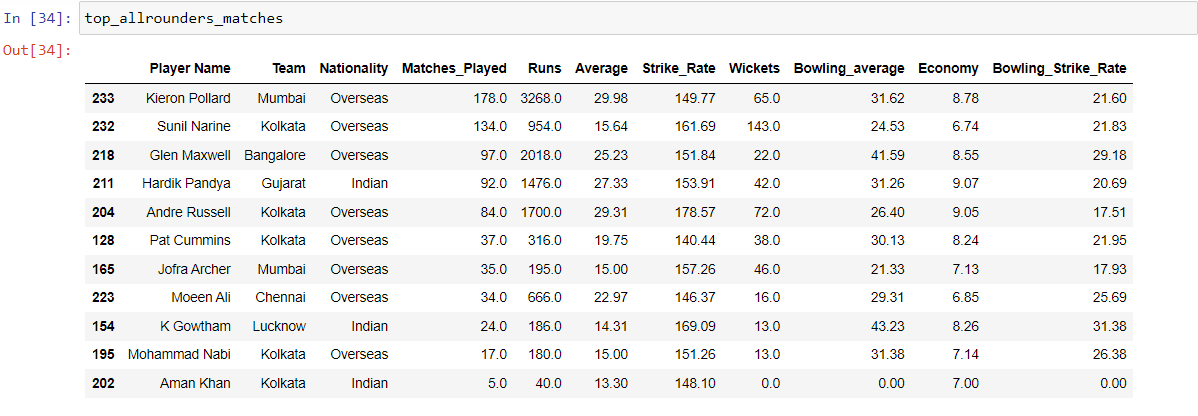
**Step 15:** Here is the average values of top all-rounders



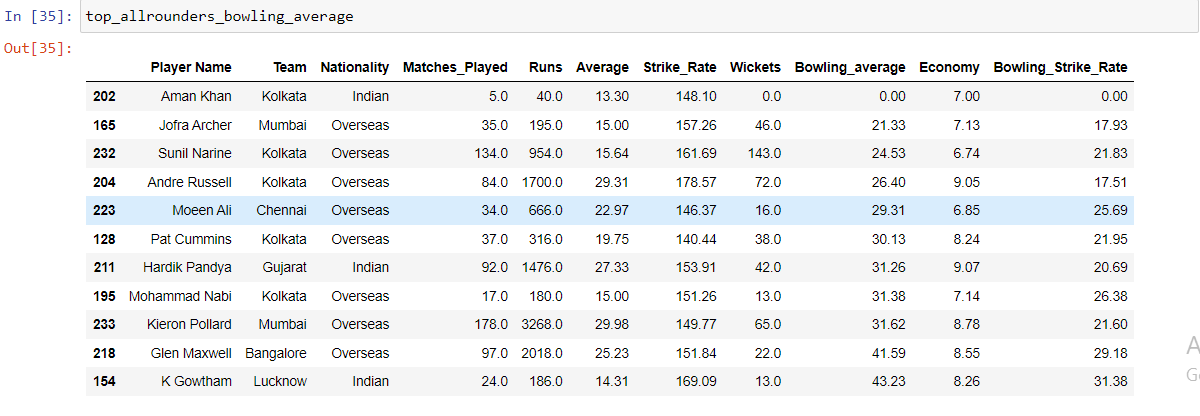
**Step 16:** Here is the average runs value of top all-rounders



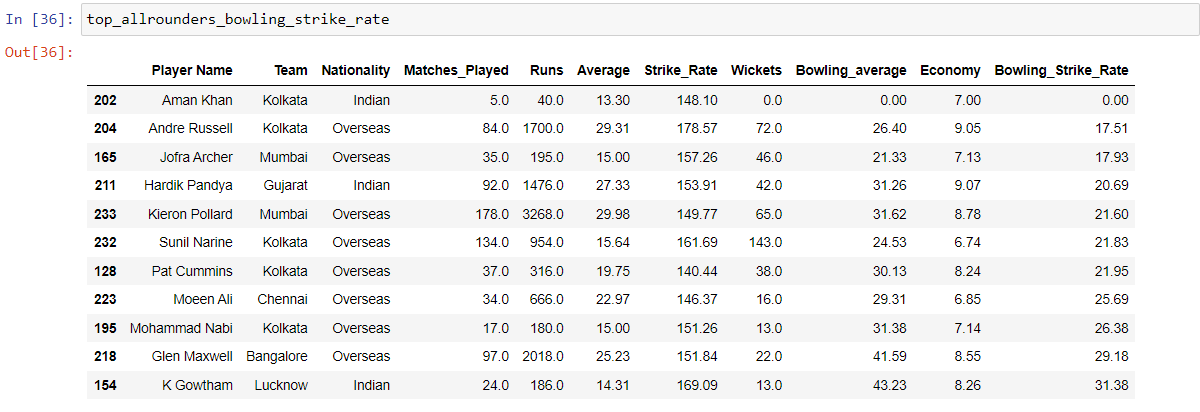
**Step 17:** Here is the matches played by top all-rounders



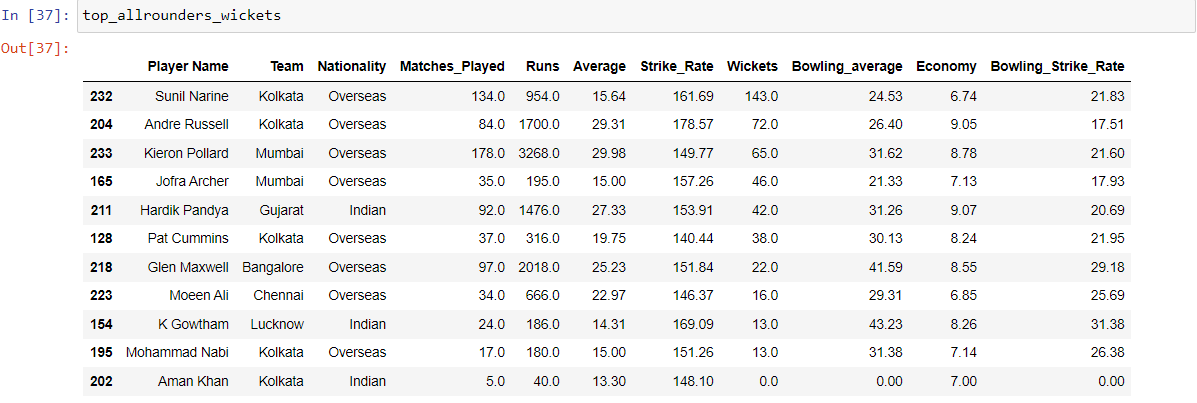
**Step 18:** Here is the bowling average of top all-rounders



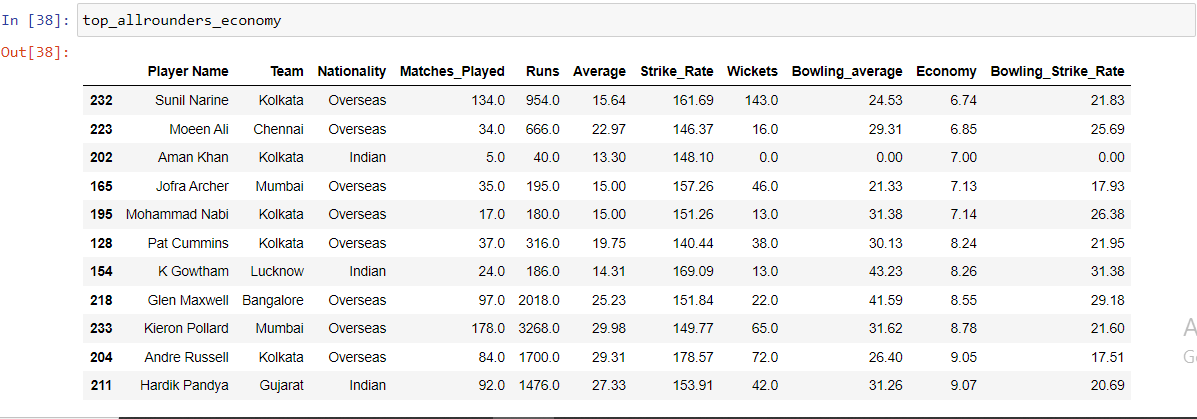
**Step 19:** Here is the bowling strike rate of top all rounders



**Step 20:** Here is the wickets data of top all-rounders



**Step 21:** Here is the economy data of top all rounders

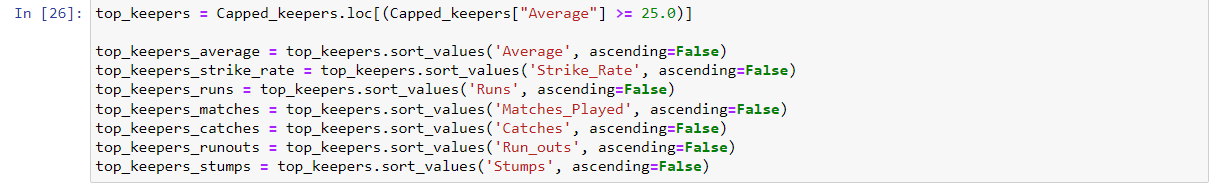


From the above analysis, if we rank down the all rounders from 1-10 on various parameters the top all rounder options are as follows:

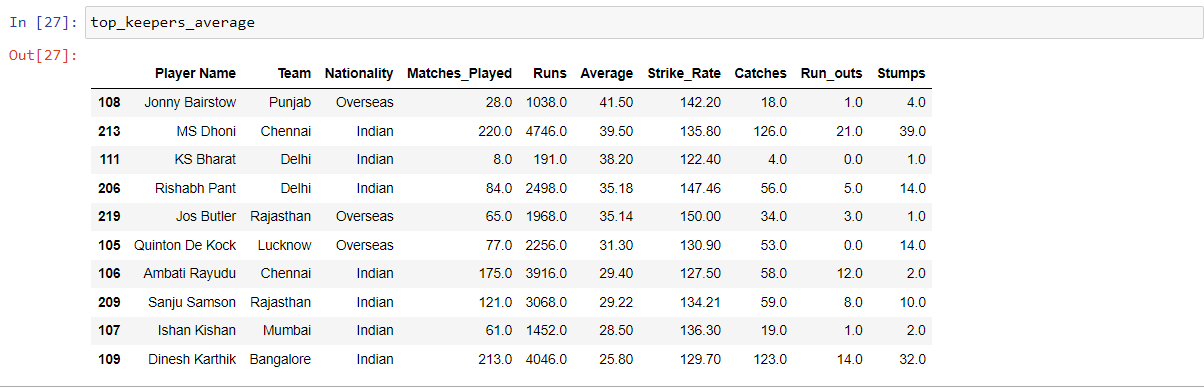
1. Andre Russell
2. Sunil Narine
3. Hardik Pandya
4. Jofra Archer

**Analyzing the keepers’ data:**

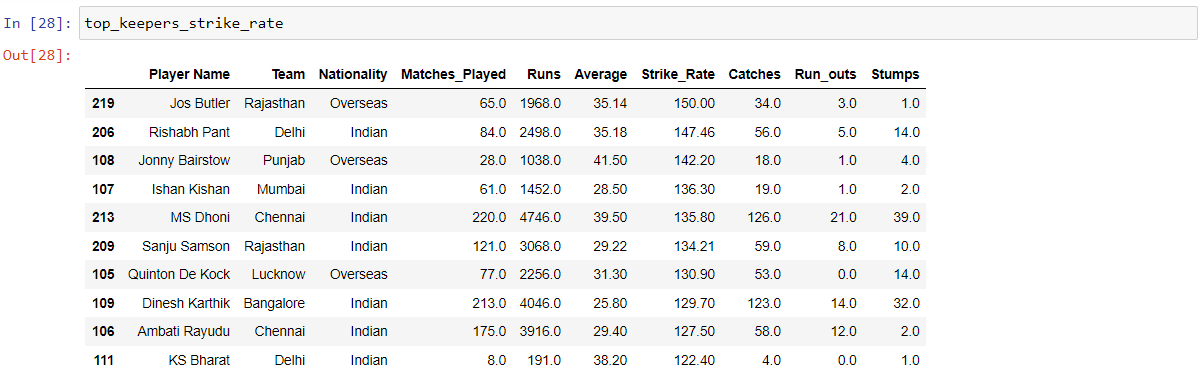
**Step 1:** We have narrowed our analysis down to keepers averaging more than 25.0



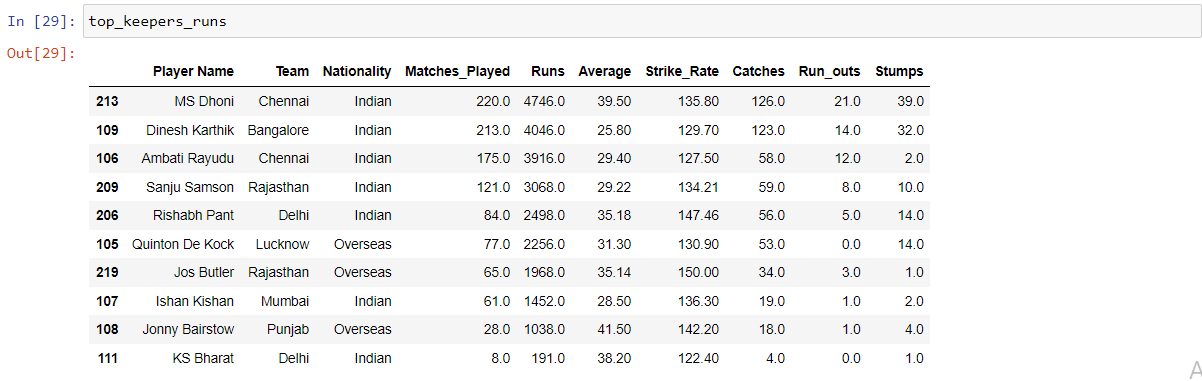
**Step 2:** From the above analysis here is the average values of top keepers



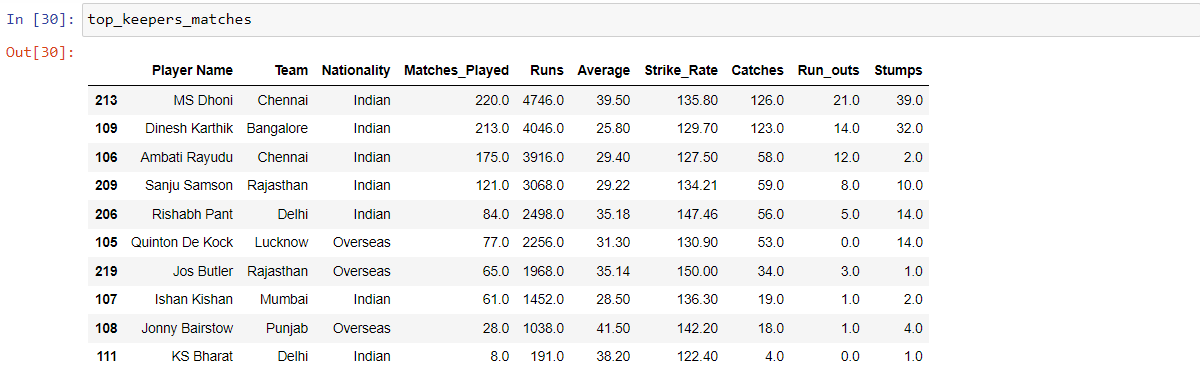
**Step 3:** Here is the strike rate of top keepers



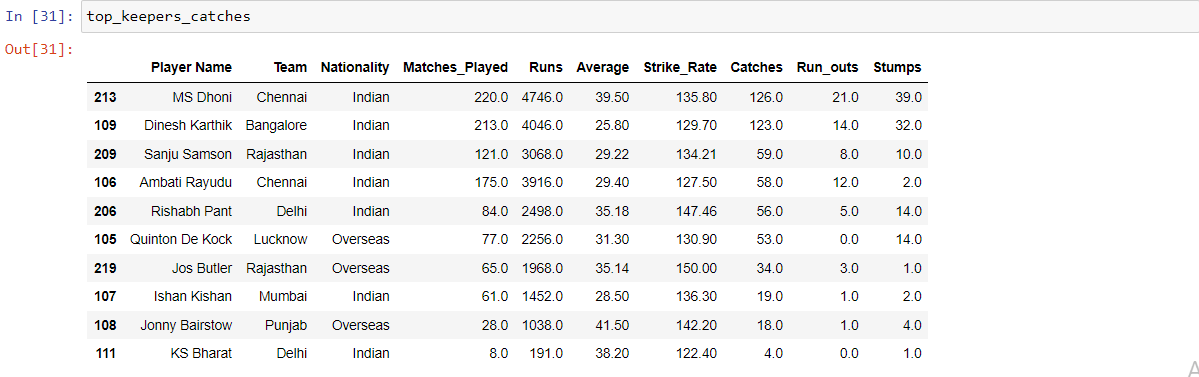
**Step 4:** Here is the runs of top keepers



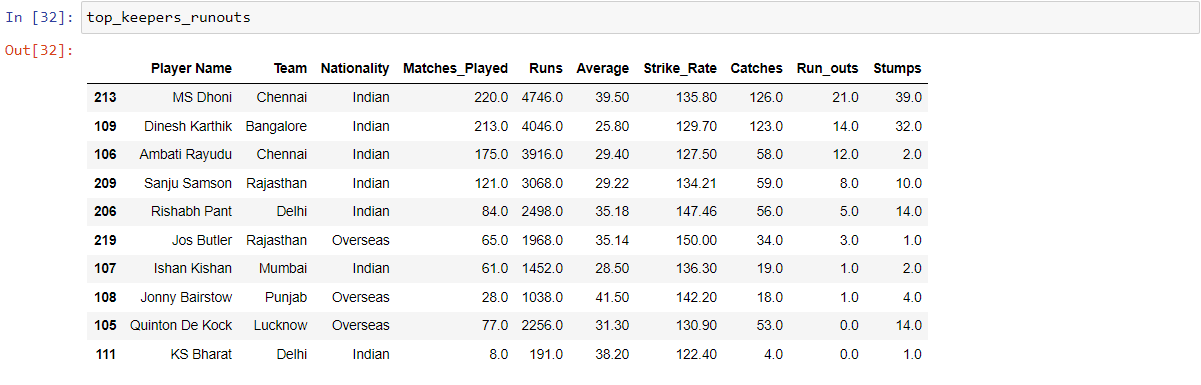
**Step 5:** Here is the number of matches played by the top keepers



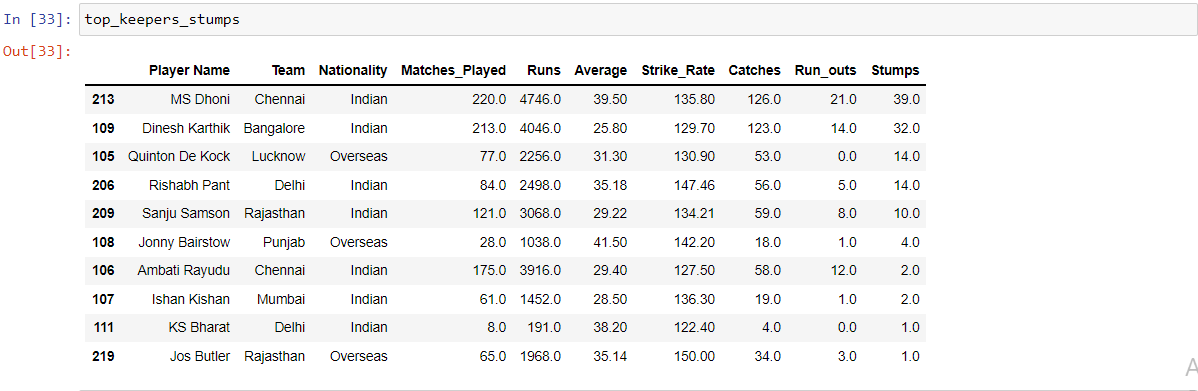
**Step 6:** Here is the catches of top keepers



**Step 7:** Here is the runouts of top keepers



**Step 8:** Here is the stumps values of top keepers



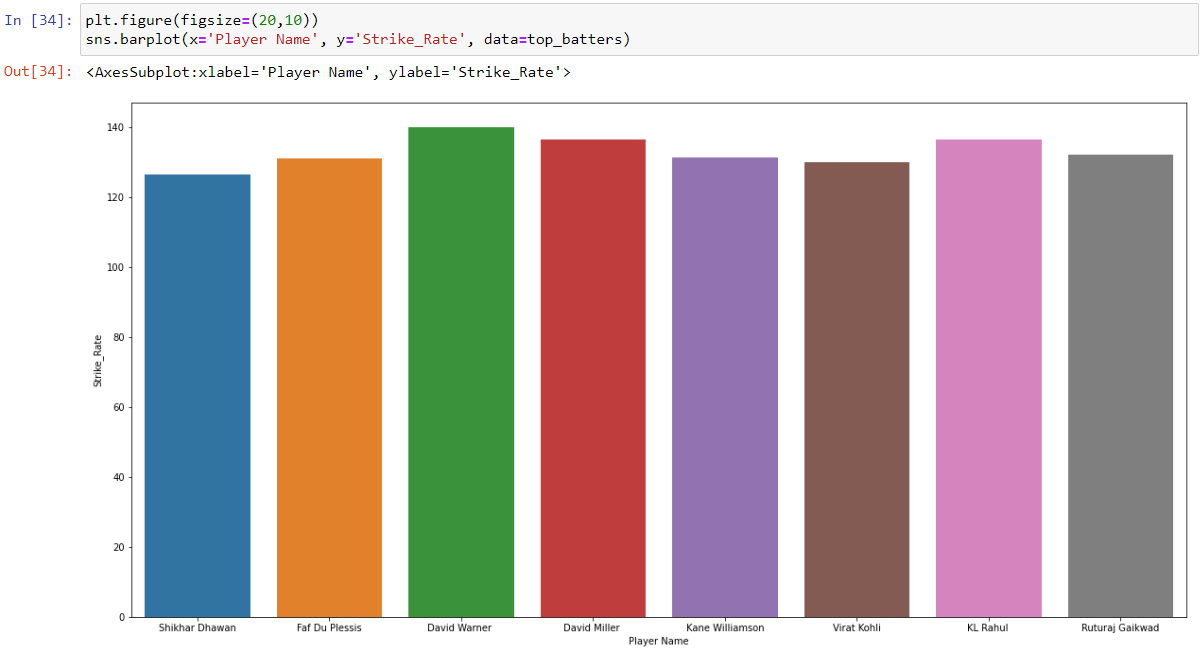
According to the data we extracted we have ranked the keepers in order of 1 – 10 and the top 3 keepers will be as follows:

1. MS Dhoni
2. Dinesh Karthik
3. Rishabh Pant

**Visualization for Enhanced Analysis:**

The initial analysis that has been done on the segregated data. We can create a few visual representations of the same to get better analysis of the data.

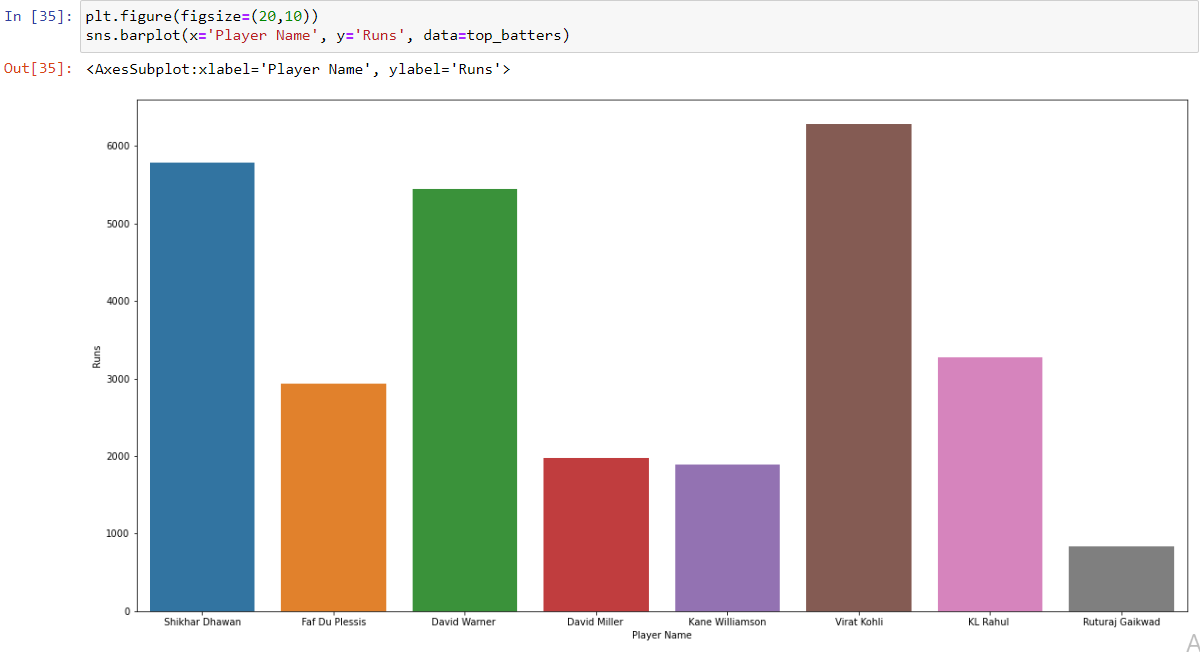
**Step 1:** Visualizing batters data by plotting each of the top batters strike rate. Here the strike rate is visualized for each player by plotting the player name to x-axis and their strike rate to y-axis.



From the above graph we can easily view the top 3 ranks of players strike rate and the players are

1. David Warner
2. David Miller
3. KL Rahul

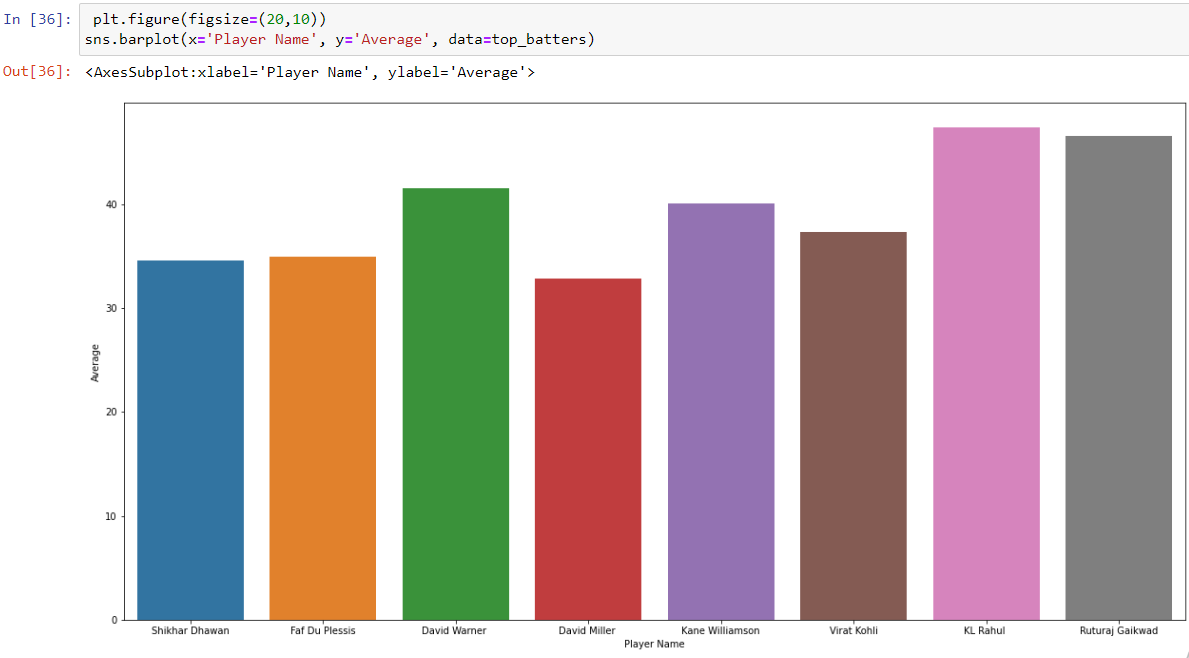
**Step 2:** We can also visualize the data for runs rate. Here the runs rate is visualized by plotting the players name to x-axis and Runs value to y-axis.



From the above graph we can easily view the top 3 ranks of players with highest runs and the players are as follows,

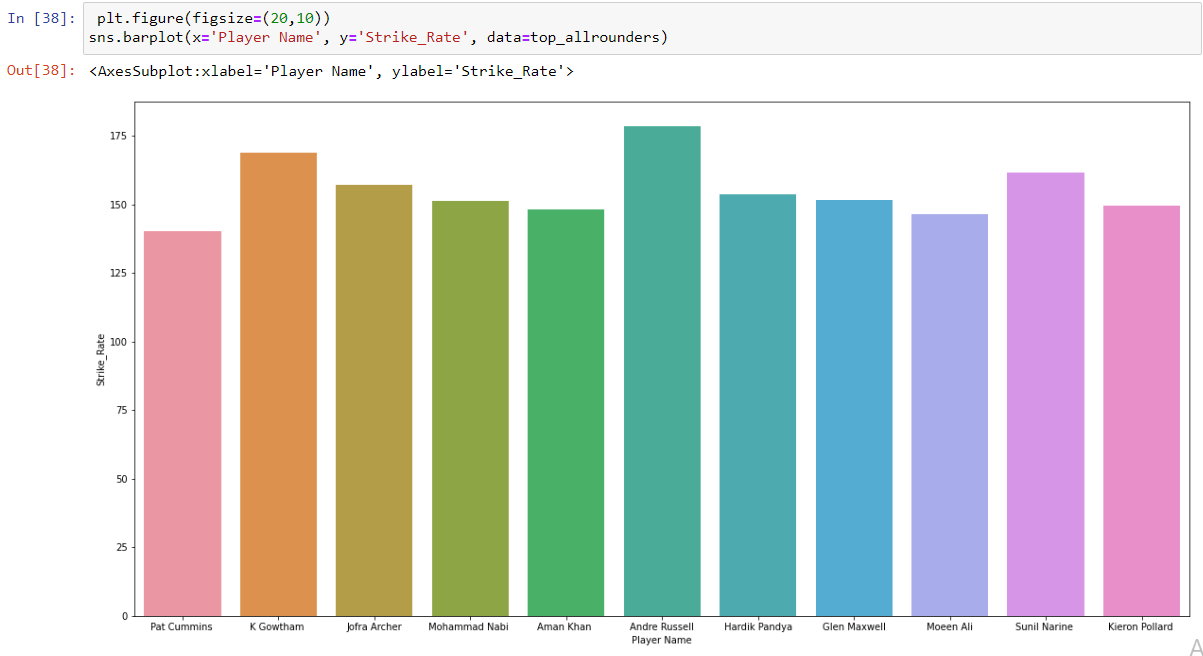
1. Virat kohli
2. Shikar Dhawan
3. David Warner

**Step 3:** We can also view the average values of top batters as following

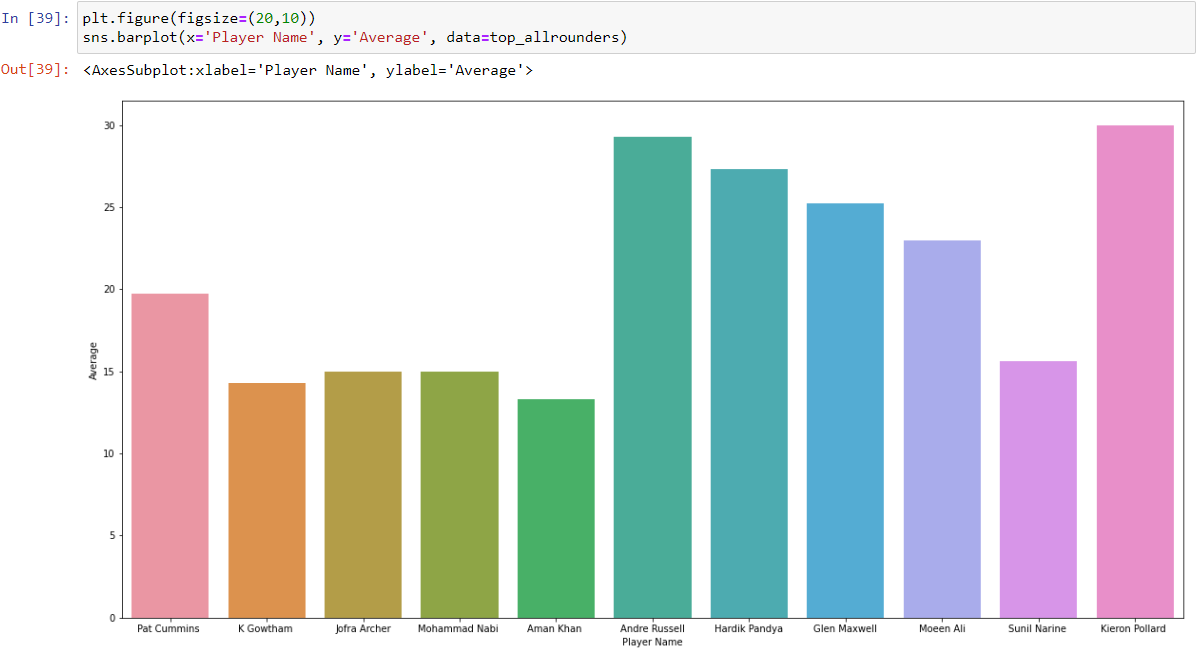


**Visualization of All-rounders data:**

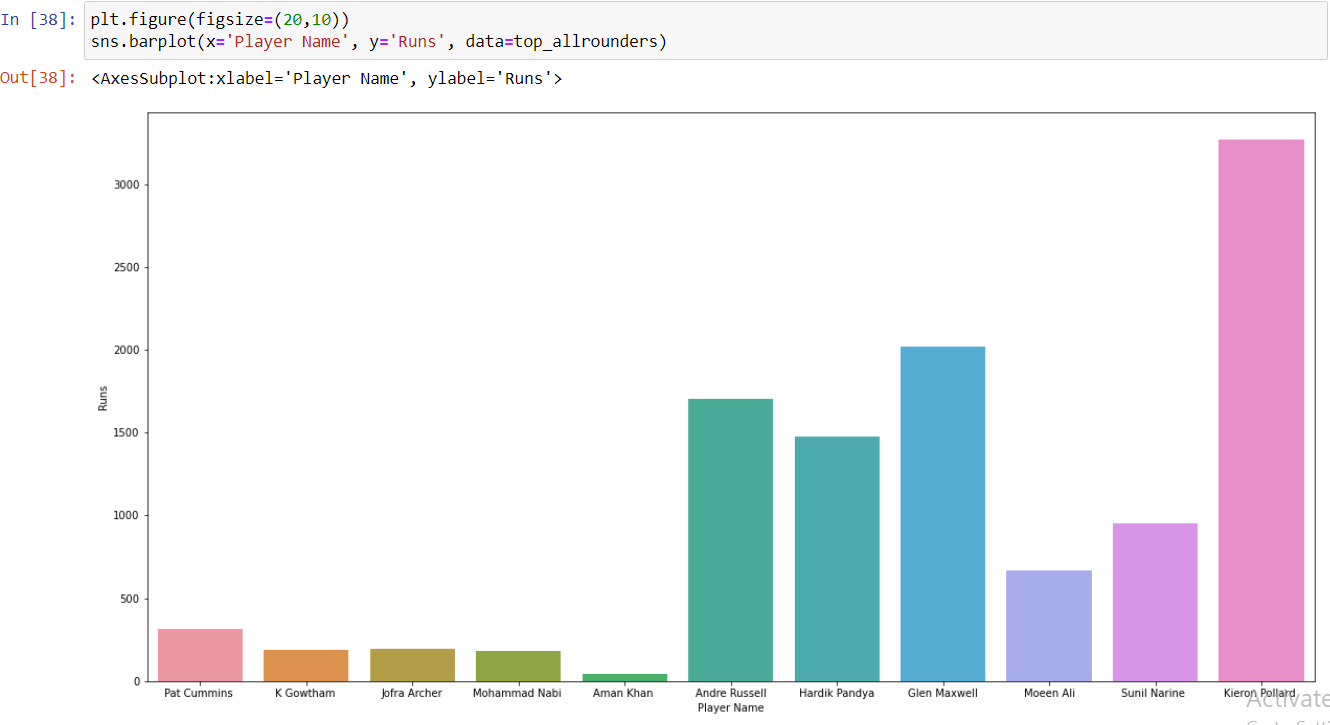
**Step 1:** This plot shows the strike rate of top all rounders



**Step 2:** This plot shows the top all rounders average by plotting players name in x-axis and Average values in y-axis.

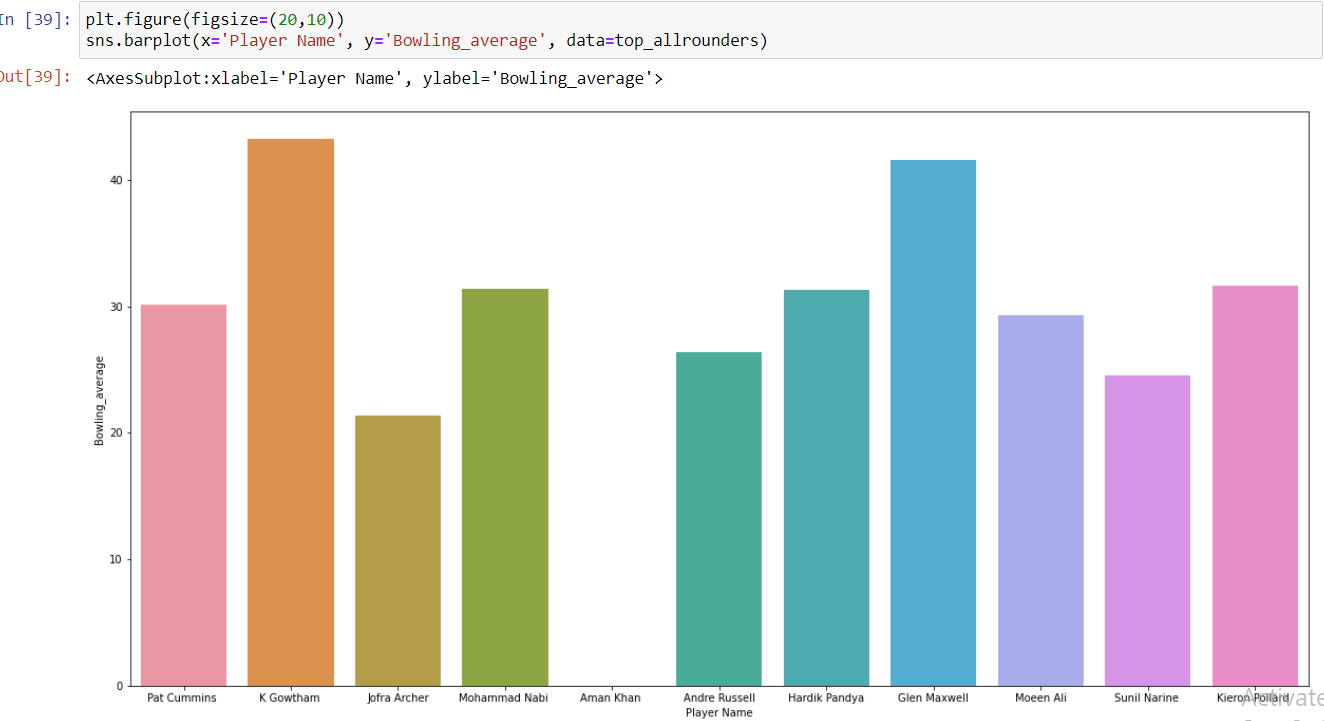


**Step 3:** This plot shows the top all rounders runs value.

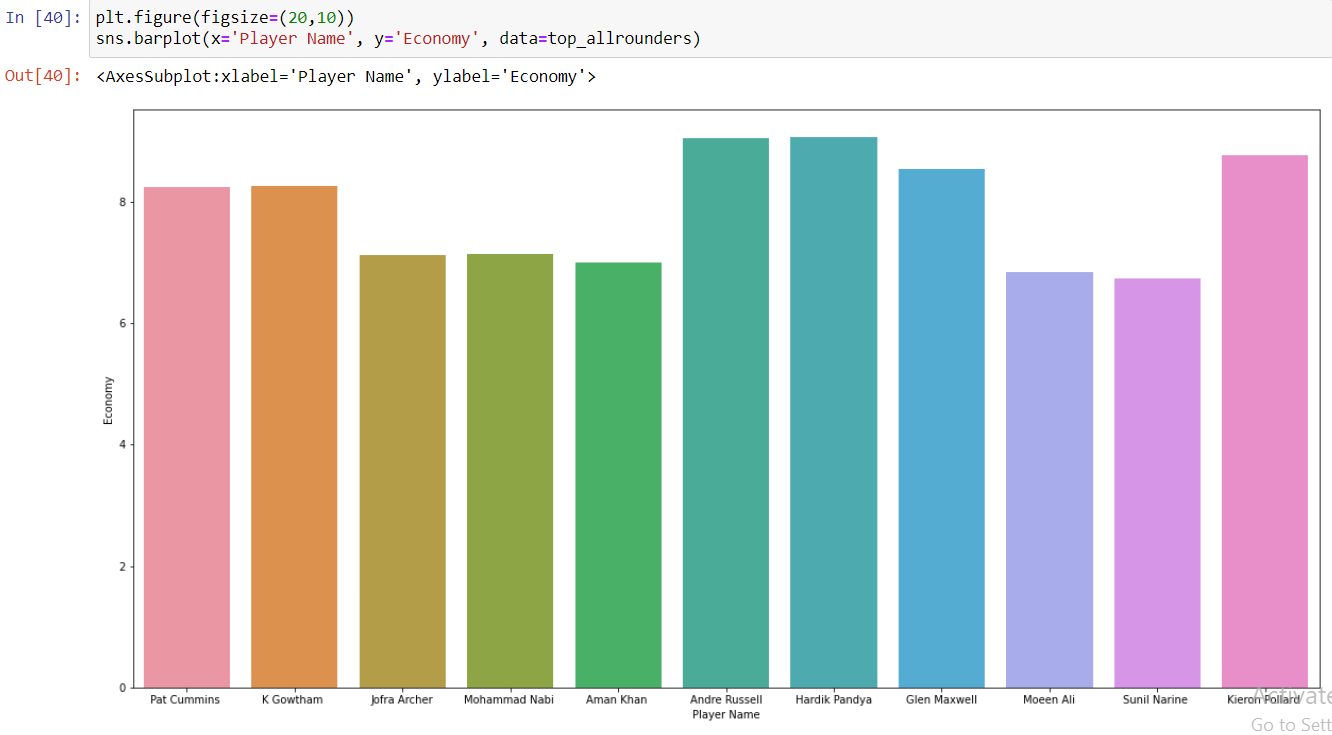


From this chart we can view that Kieron Pollard possess more runs since he has played more matches.

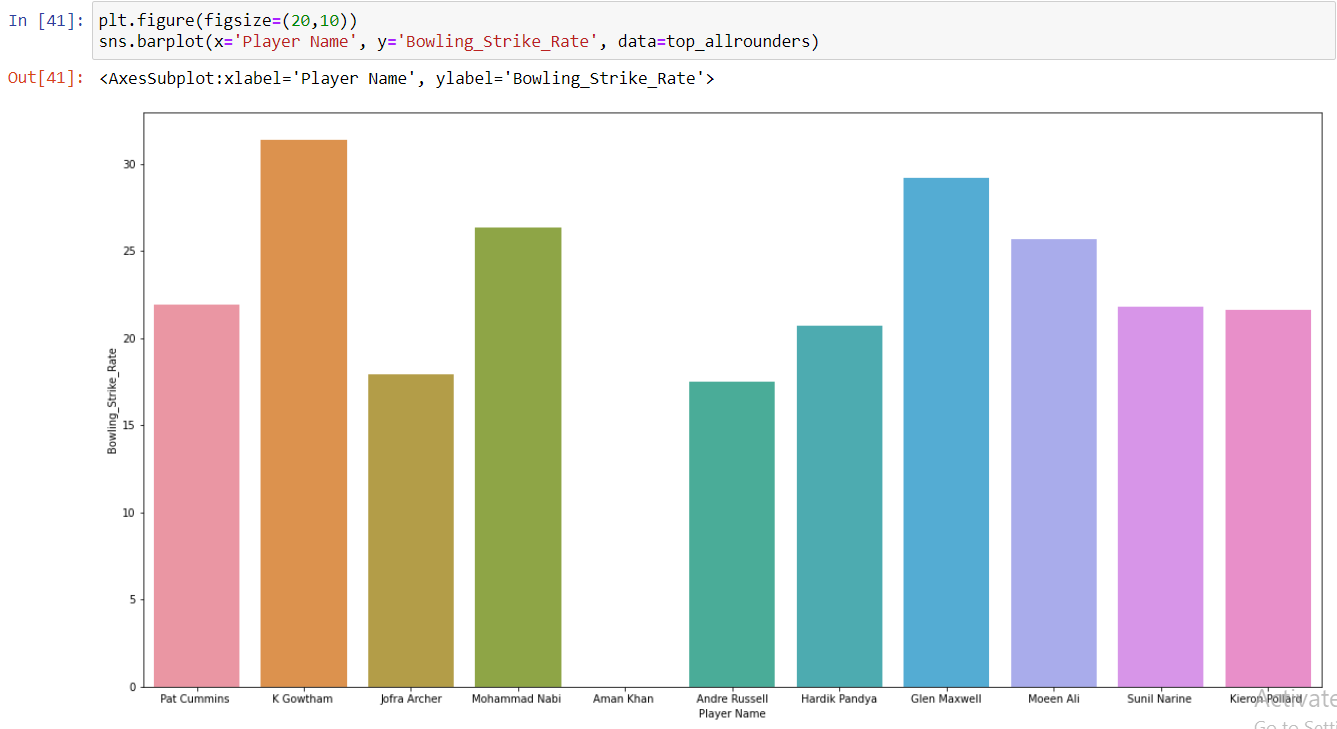
**Step 4:** Here we can view the Bowling average of the top all rounders



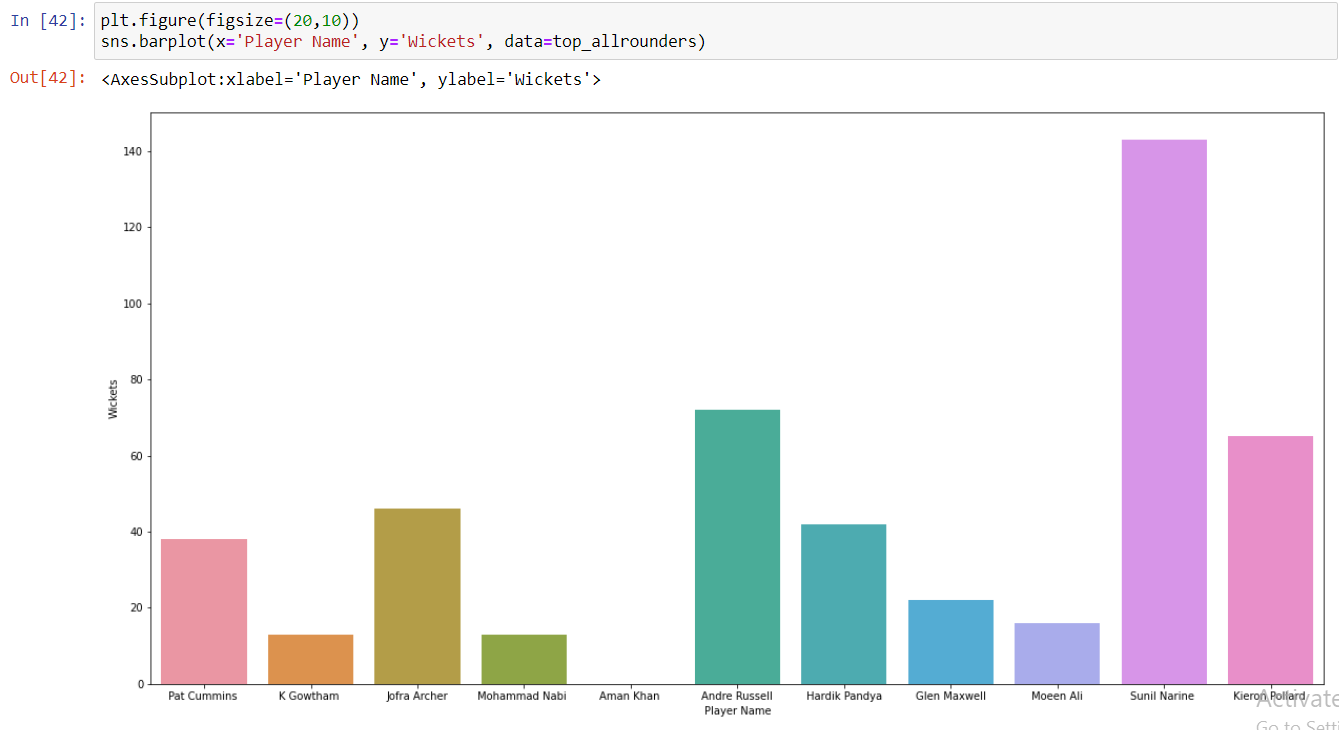
**Step 5:** This plot shows the economy of the top all rounders.



**Step 6:** Here we can view the bowling strike rate of top all-rounders

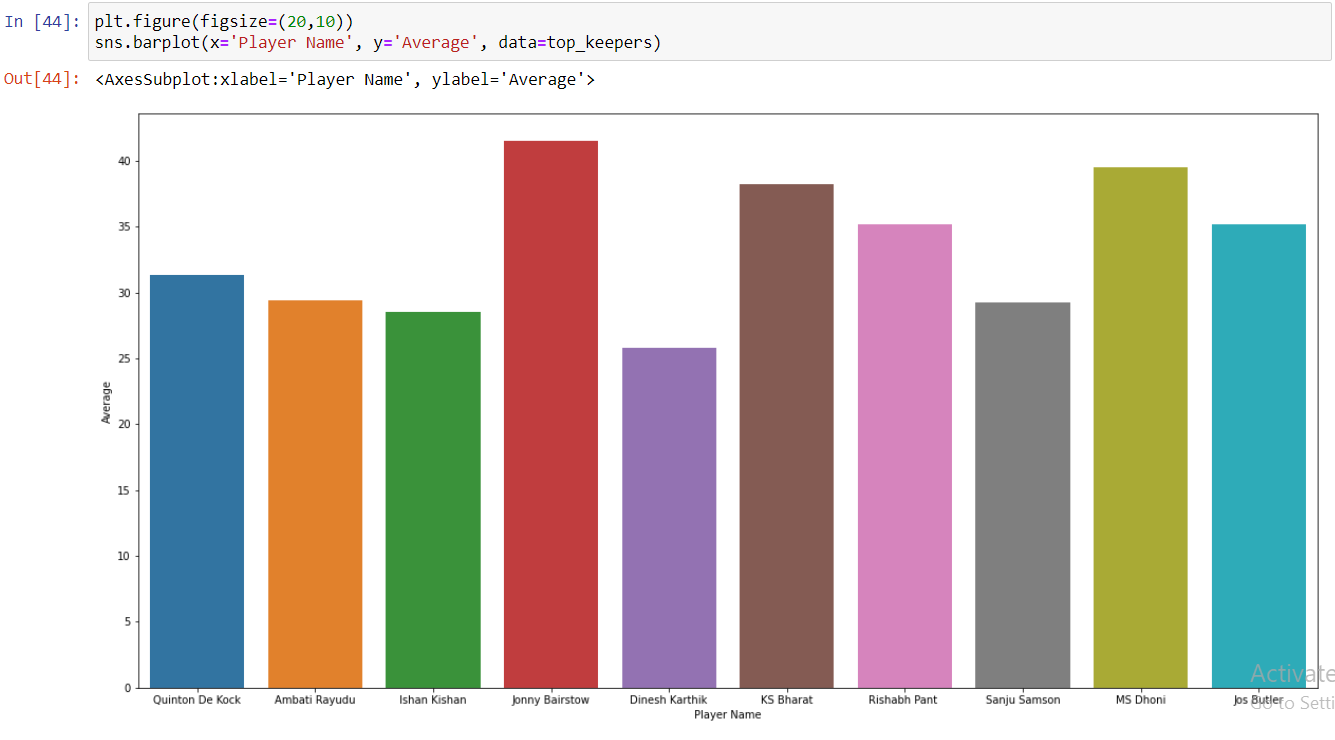


**Step 7:** This plot shows the wickets taken by top all rounders.

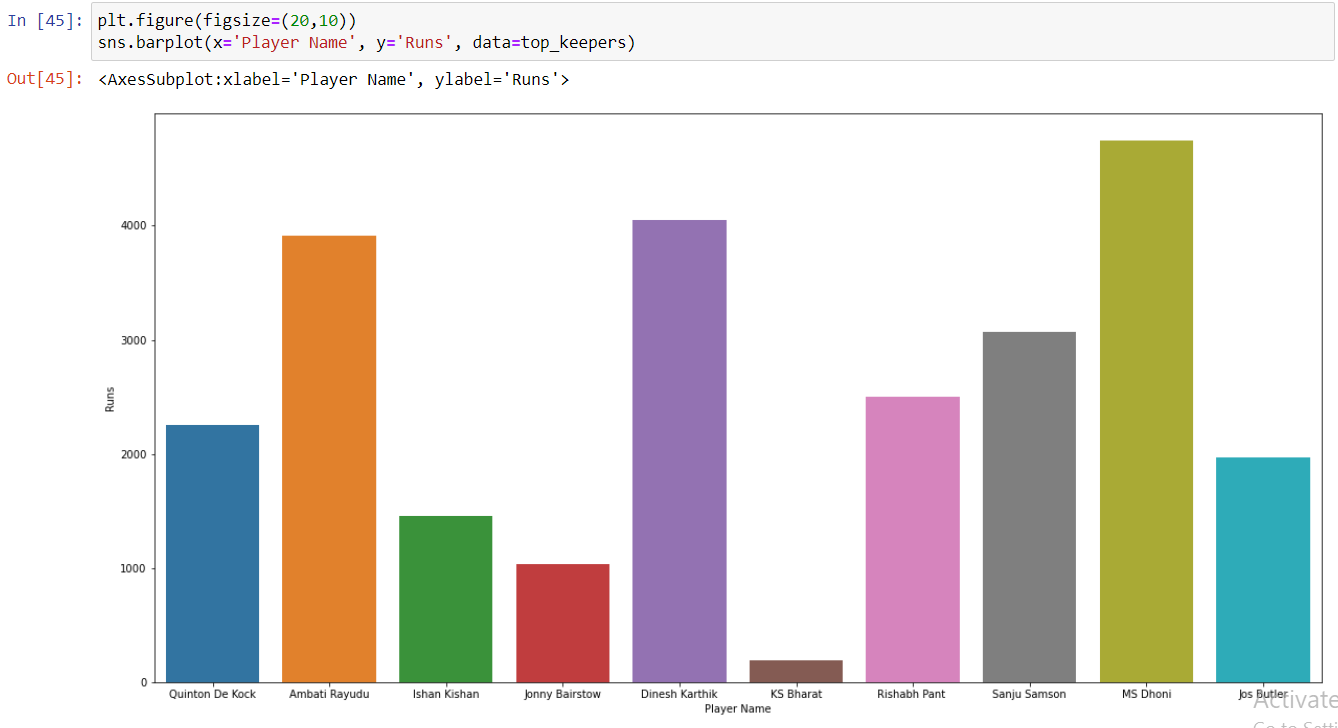


**Visualization of the keepers’ data:**

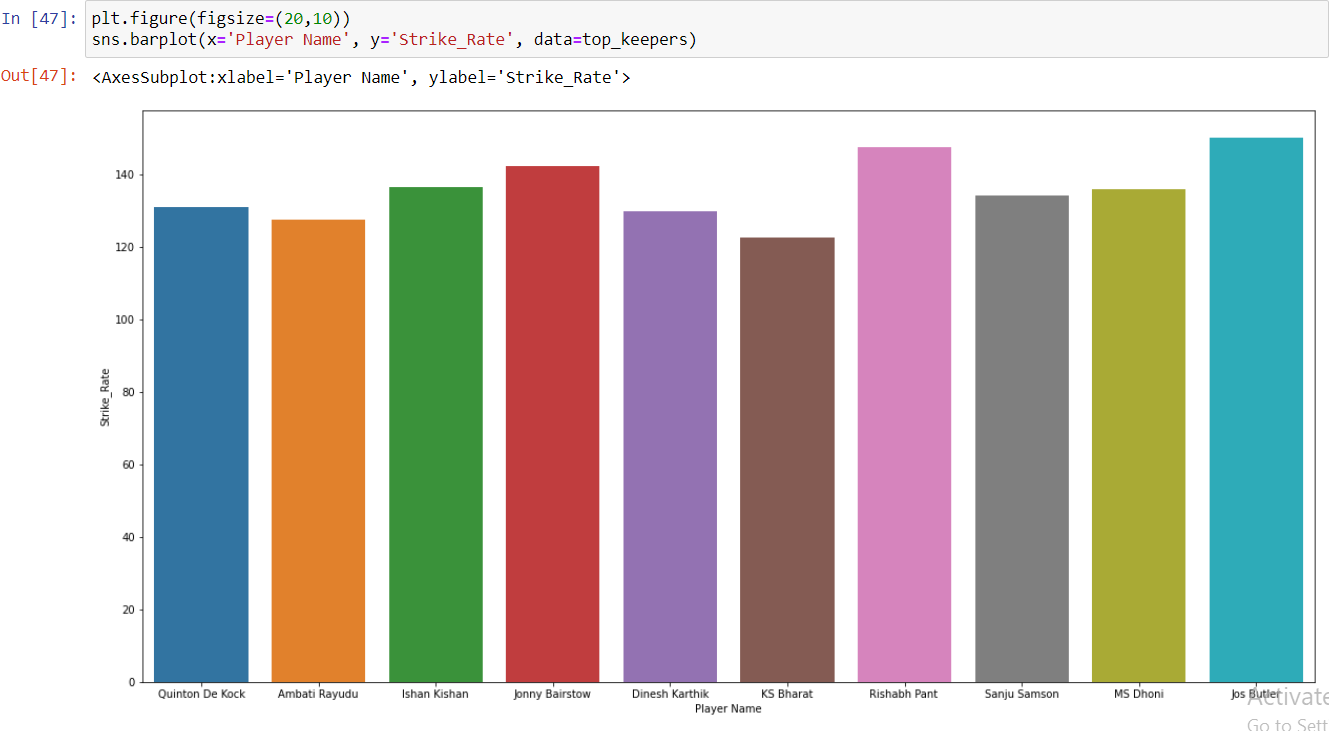
**Step 1:** First visualize the average of top keepers from the following graph



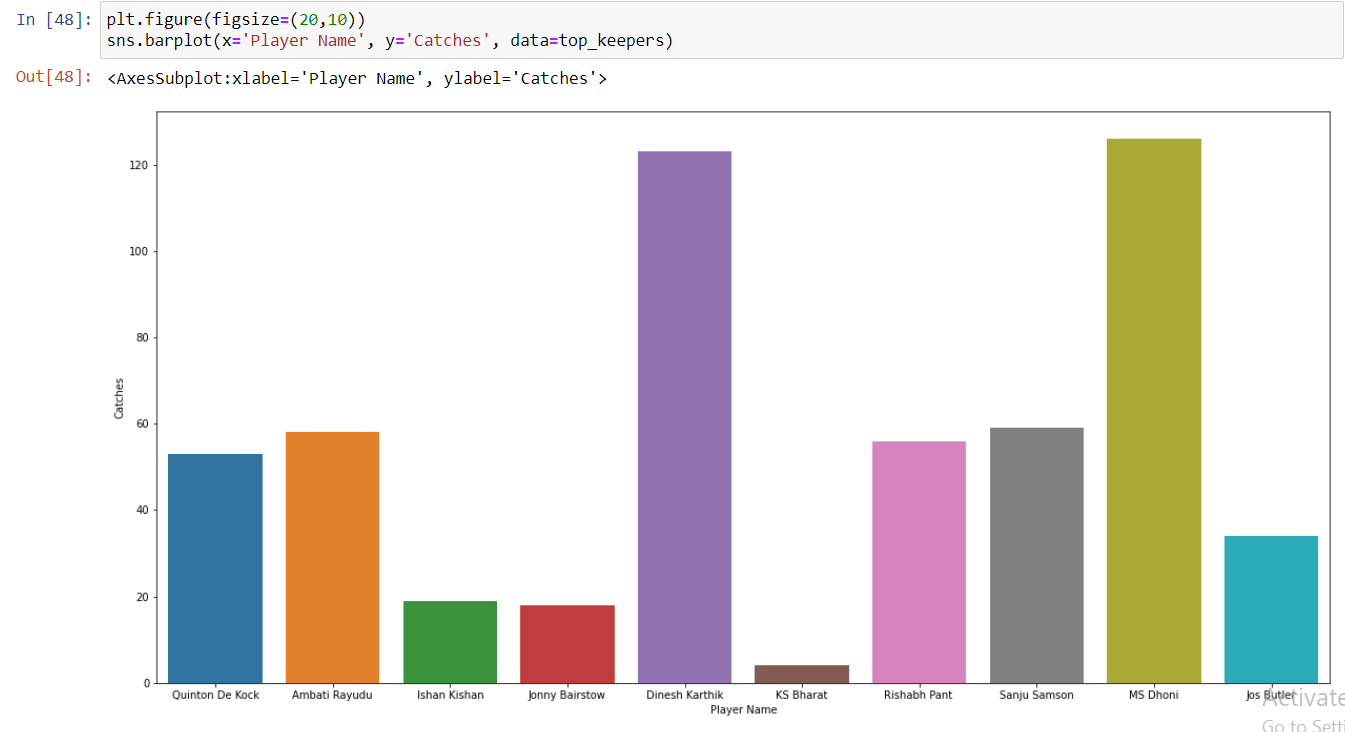
**Step 2:** Lets’ view the runs by top keepers



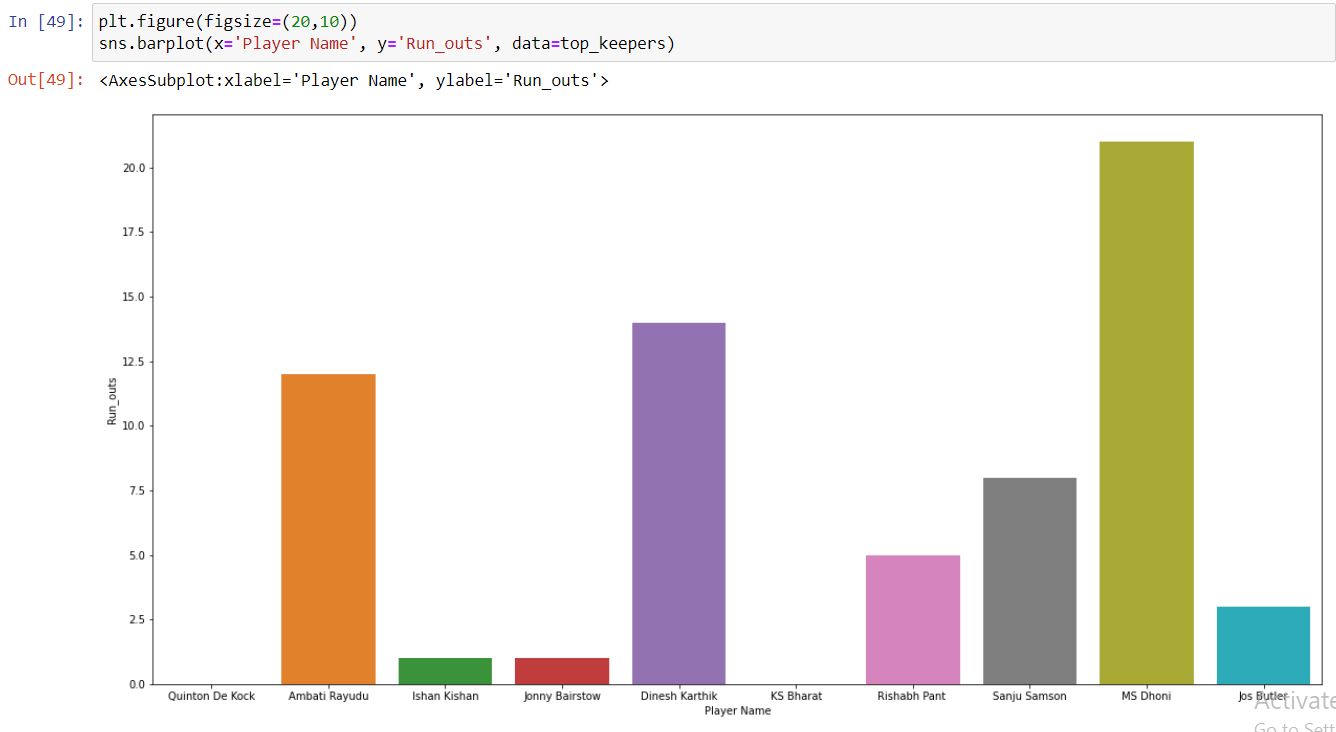
**Step 3:** Here we can view the strike rate of top keepers



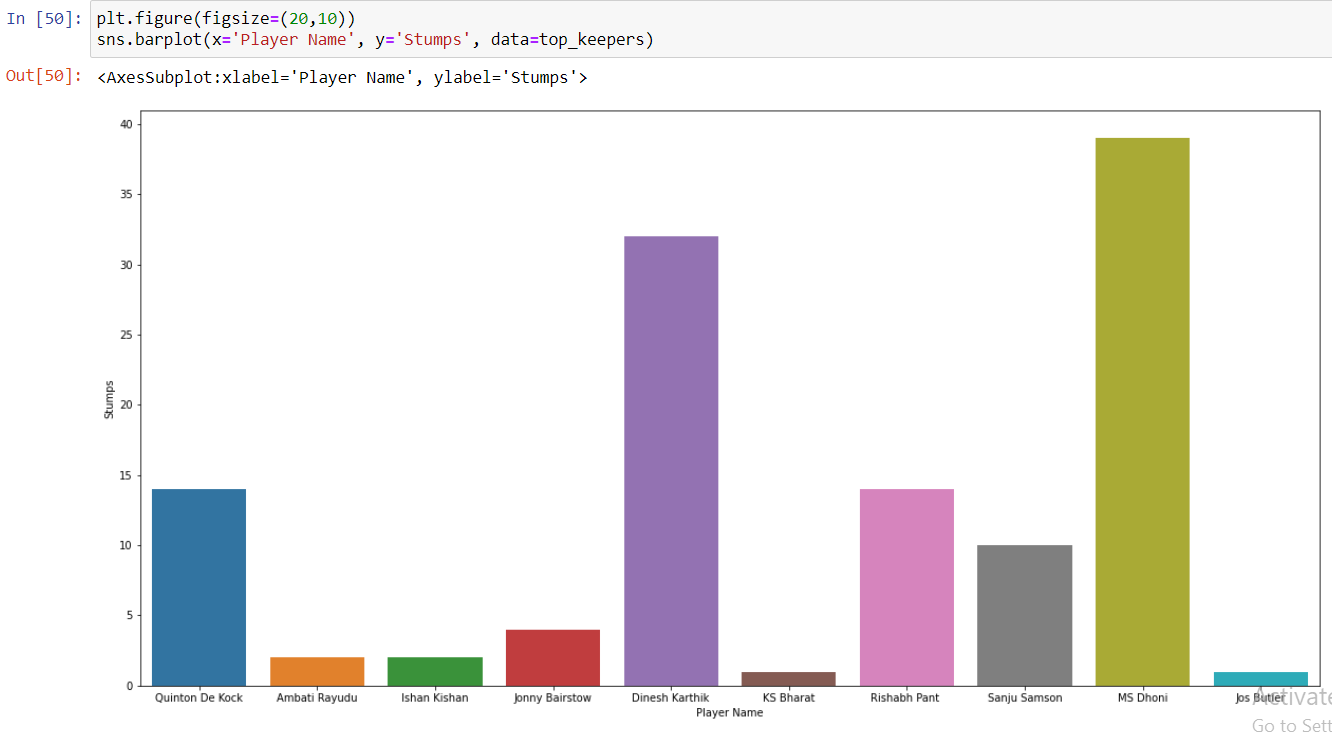
**Step 4:** Then we can view the catches taken by top keepers



**Step 5:** This plot shows the run outs done by top keepers

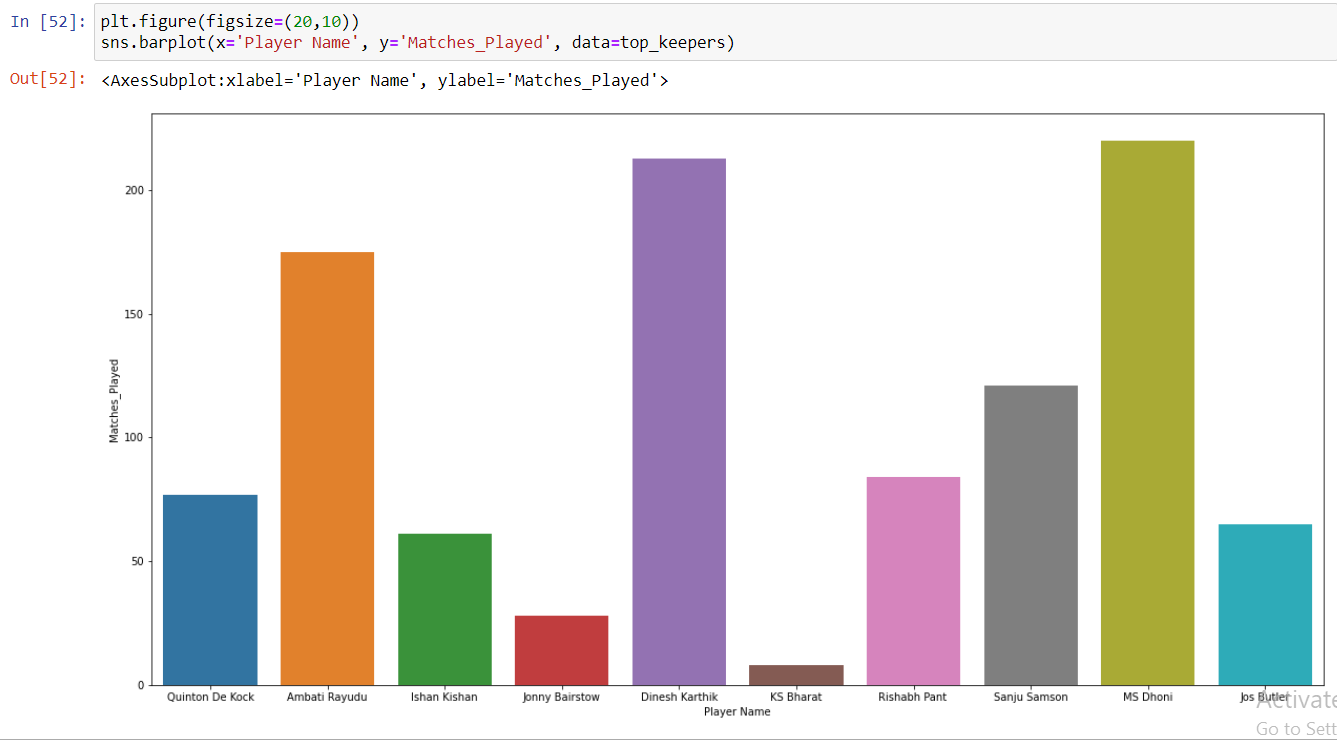


**Step 6:** This plot shows the stumps done by top keepers



From the above analysis, the parameters stumps, runouts and catches Dinesh Karthik ranks number 2.

**Step 7:** This plot shows the details of the matches played by top keepers.



From the graph we can get the top 3 players,

1. MS Dhoni
2. Dinesh Karthik
3. Ambati Rayadu

From the above visualizations we have finalized 3 batsmen, 4 bowlers, 4 all rounders and 3 wicket keepers.

**Forming our best 11 for the Campaign based on the above analysis:**

We will consider the number of players from each category that the t20 world cup winning and last years’ IPL winning team played in their Final matches.

The Australia squad consisted of – 3 Batters, 3 All rounders, 4 Bowlers with 1 spin option and 1 wicket keeper.

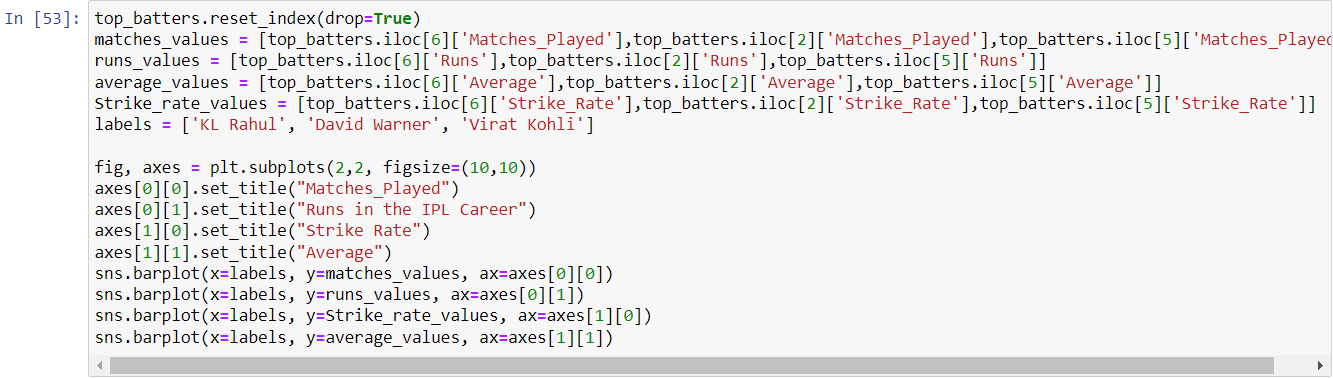
The Chennai squad consisted of – 4 Batters, 3 All rounders, 3 Bowlers and 1 wicket keeper.

For our analysis we will consider the ratio of players in the best 11 as follows:

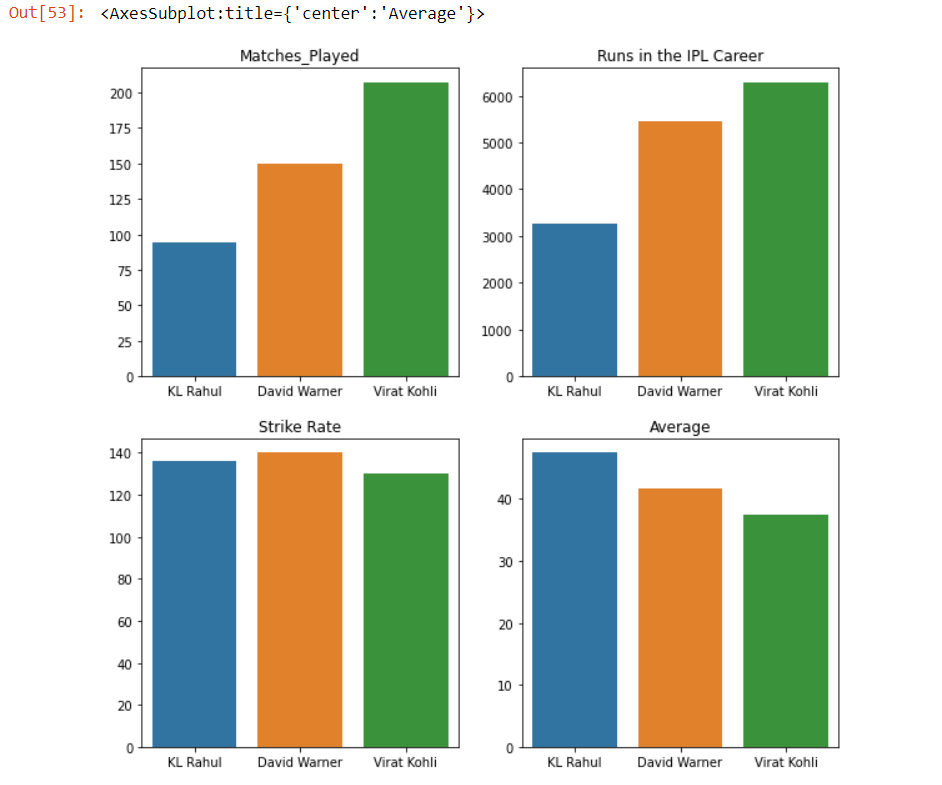
* 3 Batters
* 3 All rounders
* 4 Bowlers with 2 spin options
* 1 wicket keeper

**Representation of all the batsmen with their parameters we have analyzed:**

The following graph shows the performance of 3 specific players namely KL Rahul, David Warner and Virat Kohli based on their analysis parameters like Matches played, Runs in the IPL career, Strike Rate and Average and the screenshot of the chart are as follows,

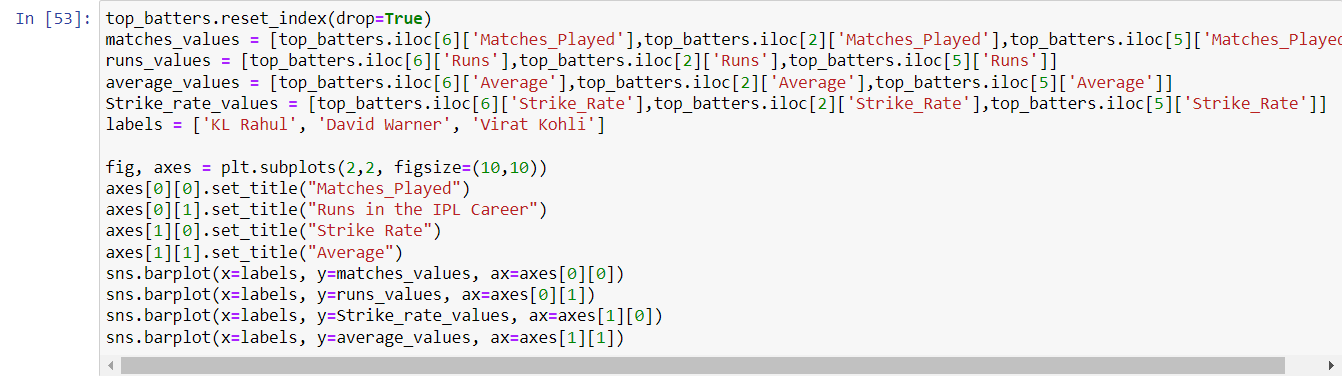


The following Bar chart shows the visual representation of the above analysis.

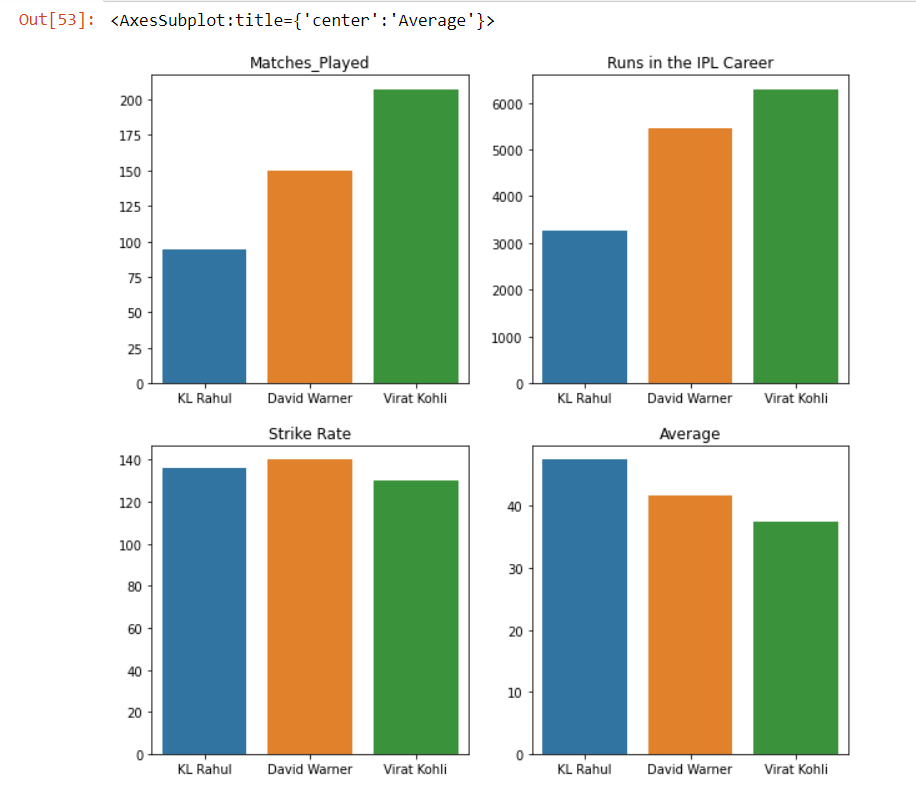


From the above chart we can clearly view the Matches played, Runs in the IPL career, Strike rate and Average values of the players KL Rahul, David Warner and Virat Kohli individually.

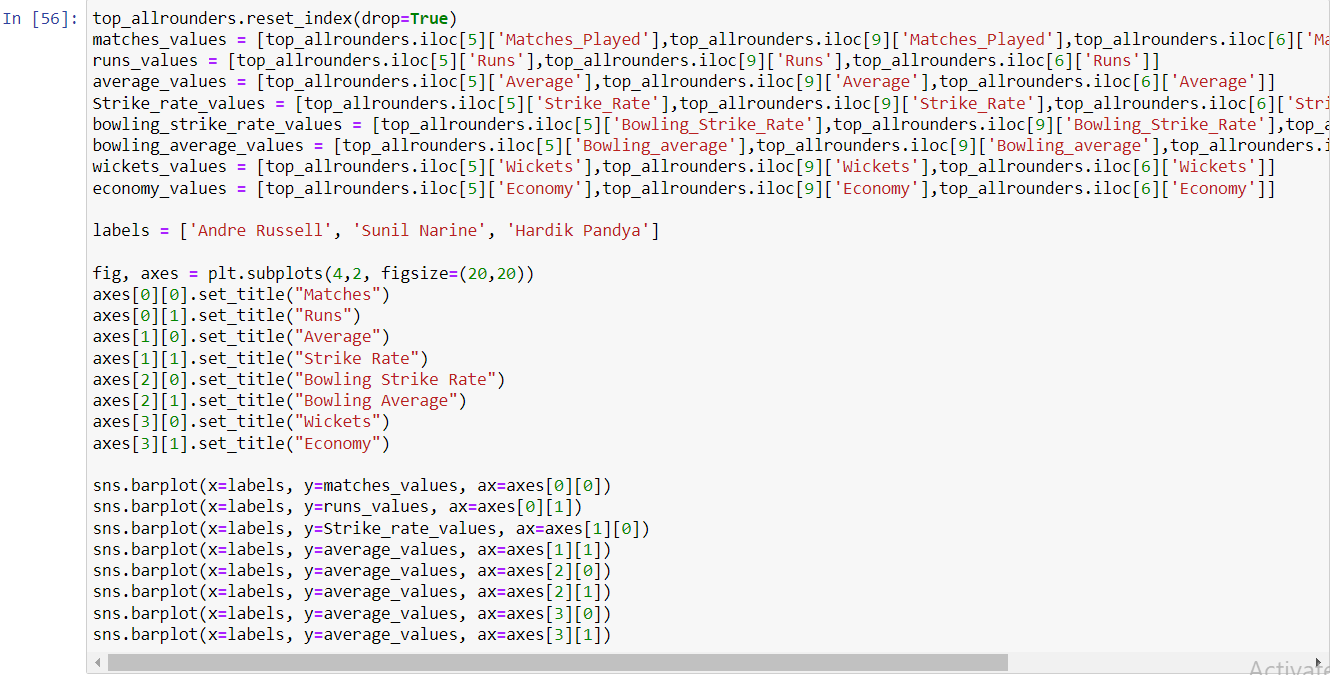
**Finalizing all rounders:**

Here we are storing the values of each top batter in a separate data frame and displaying using the bar plot. 

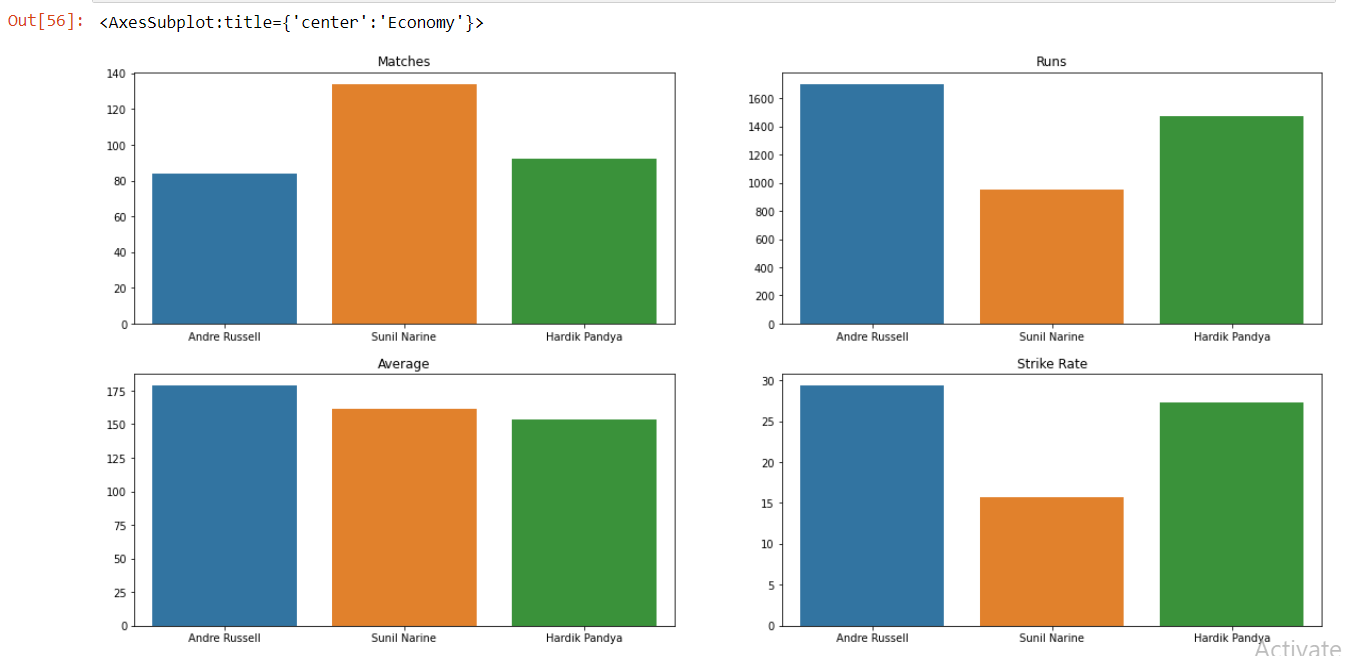
The outputs of values are plotted in the following graph. The following graph shows the details of top batters KL Rahul, David Warner, Virat Kohli matches played, runs, average, strike rate.

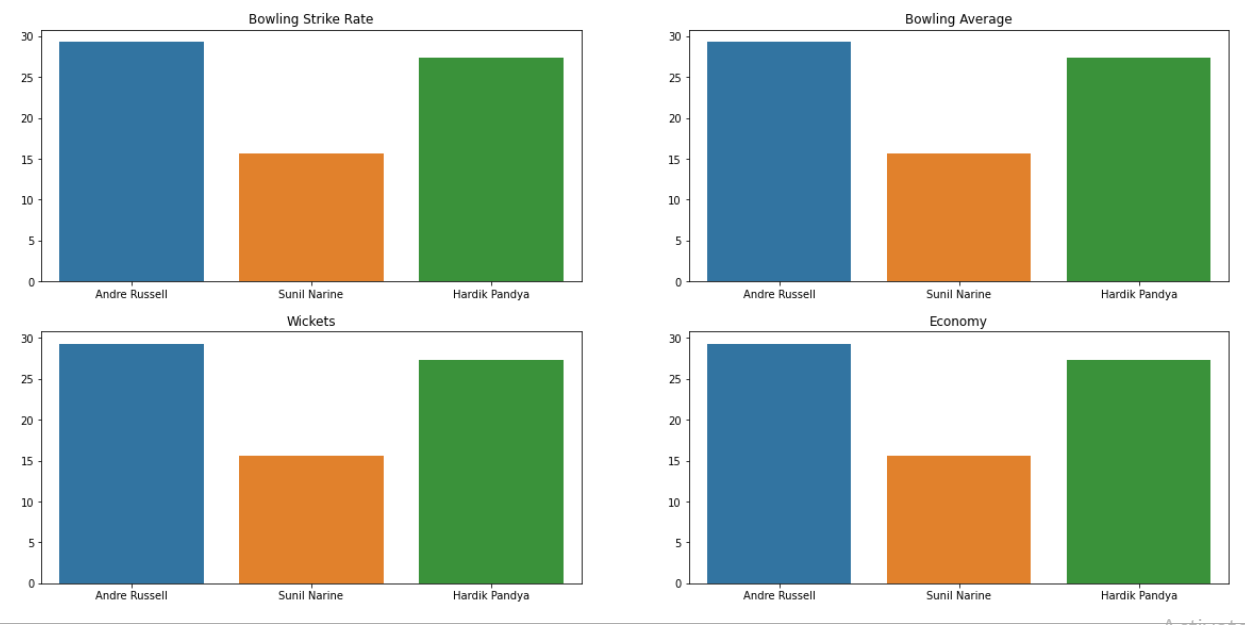


Here we are storing the values of each top all rounders in a separate data frame and displaying using the bar plot.



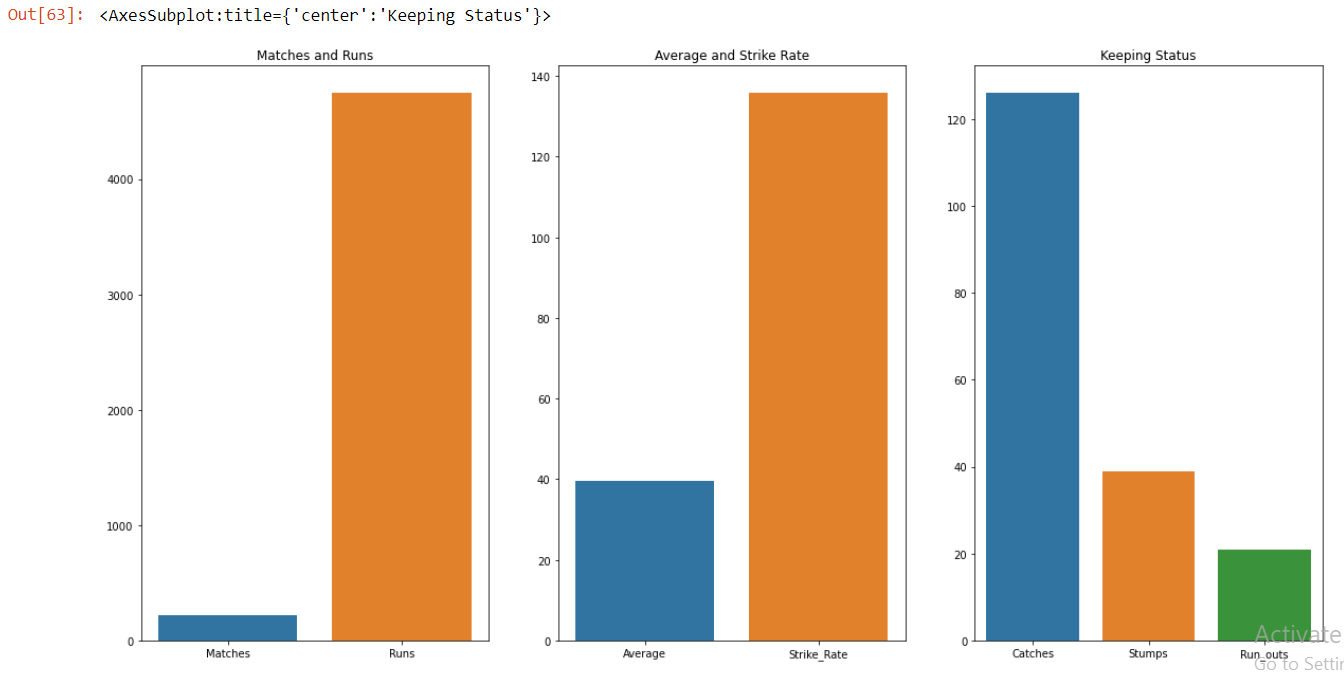
The outputs of values are plotted in the following graph. The following graph shows the details of top all rounders Andre Russell, Sunil Narine, Hardik Pandya and their data on matches played, runs, average, strike rate, Bowling strike rate, Bowling average, wickets, Economy.



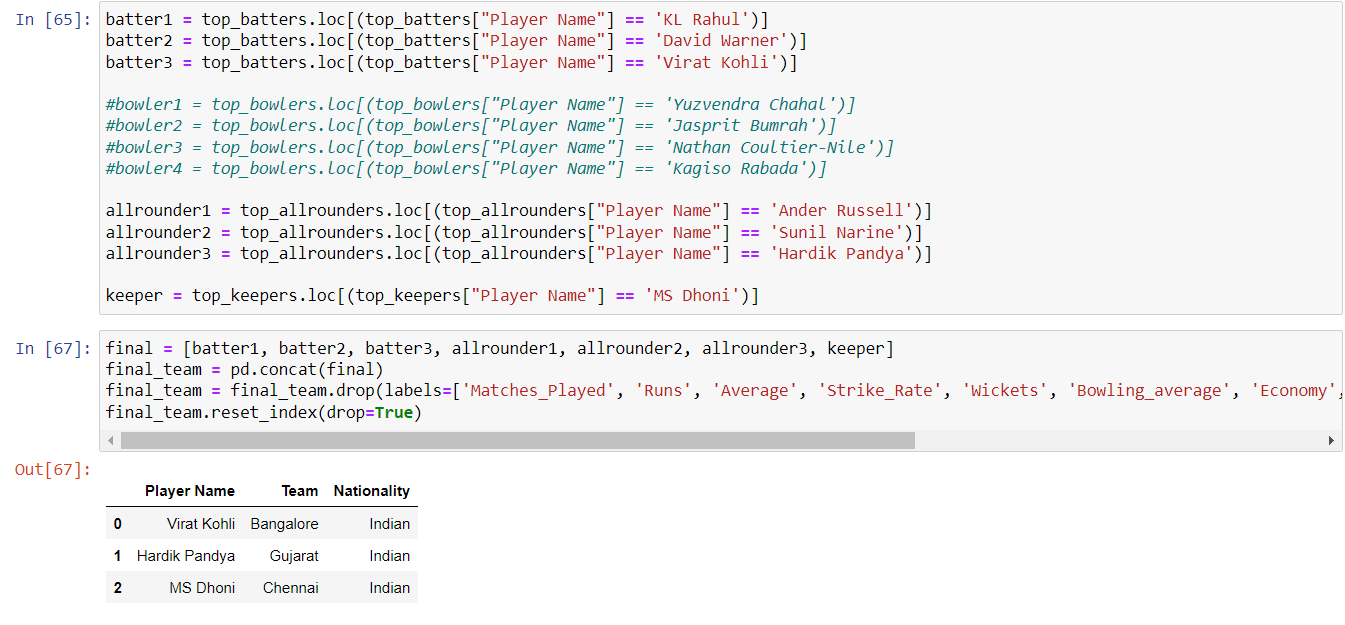


Here we can view the matches, average values and keeping values in the following graph.





From the final analysis we can segregate top batters, top bowlers, top all rounders and top keepers and they are as follows,



Hence from the above analysis we have concluded the final result of top 3 players as follows,

1. Virat Kohli
2. Hardik Pandya
3. MS Dhoni

**Conclusion:**

Here are the final results of the above analysis process of top 10 players.

