**PROJECT TITLE**

Predicting the Scores of Players in different Stadiums.

**PROJECT DESCRIPTION AND OBJECTIVES**

Predicting Raina’s and Dhoni’s Scores for the next match in Chennai for the year 2016.

Predicting Raina’s and Dhoni’s Scores for the next match in Delhi for the year 2017.

Thereby, comparing the performance of Raina and Dhoni.

**PROJECT TEAM**

Sushmitha Somashekar: 01FB15ECS319

Tejaswini Krishnamurthy: 01FB15ECS325

Daivik U D: 01FB15ECS085

**INTRODUCTION TO THE IPL AND HOW WE ARE USING DATA ANALYTICS IN IPL PREDICTION**

The Indian Premier League (IPL), created by the Board of Control for Cricket in India (BCCI) and sanctioned by the International Cricket Council (ICC), is a Twenty20 cricket competition. It is held in April and May of every year by teams representing Indian cities. Each team plays all the other teams twice throughout the season- once at the franchise home ground and once away from home. The IPL is the most-attended cricket league in the world and ranks sixth among all sports leagues.

Data Analysis is the process of systematically applying statistical and/or logical techniques to describe and illustrate, condense and recap, and evaluate data.

For our analysis, we are extracting data from the year **2008 to 2016**.

We will be using **Simple Exponential Smoothing** as a Data Analysis Technique, which is used to predict data value for the next time interval.

**TECHNIQUES USED**

* Simple Exponential Smoothing: Used to predict the next time interval’s realized value (ft+1). We will be using the following formula for different values of weight (α) and then analyze, predict suitable scores for both the players based on Error Components.
* Error Components:Mean Error(ME), Mean Square Error(MSE), Mean Absolute Deviation(MAD) and Mean Absolute Percentage Error(MAPE).

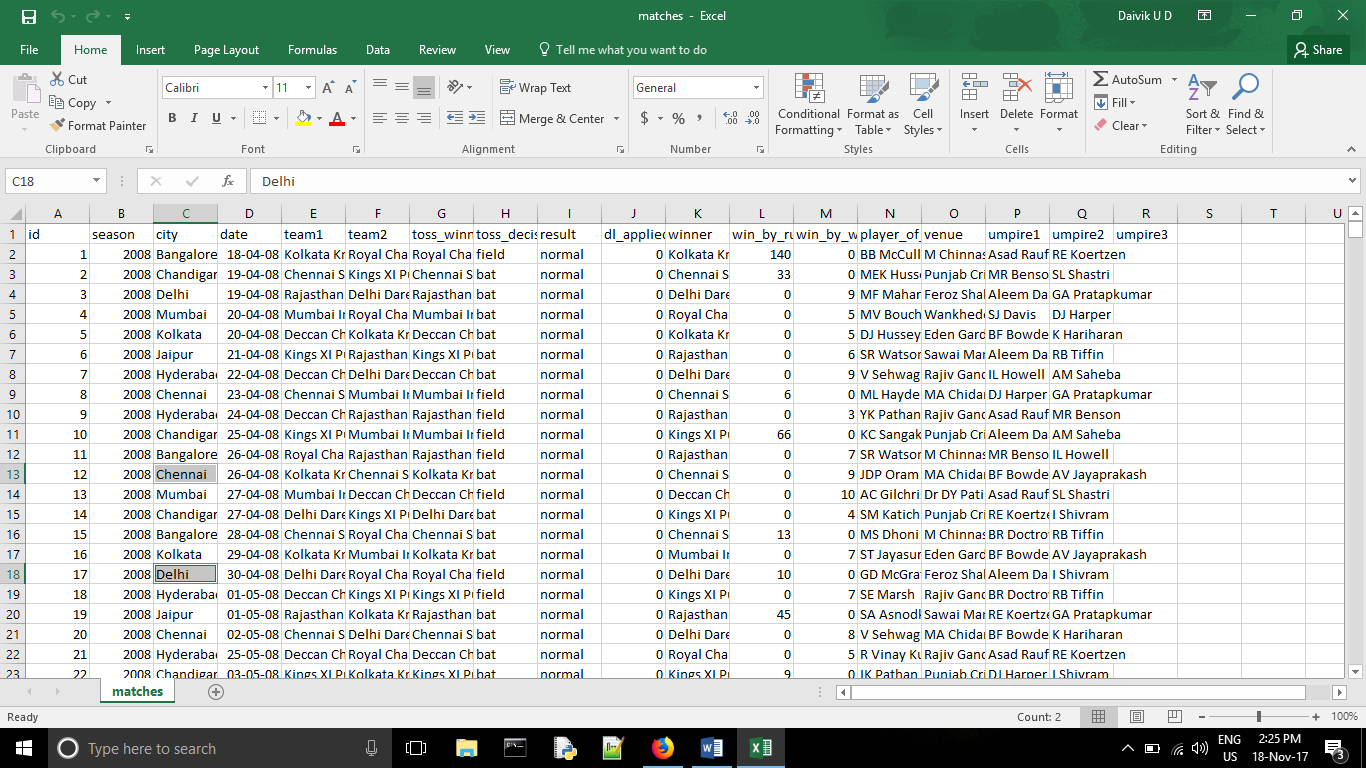
**ft+1 = ft + α (yt-ft)**

* We have also plotted Moving Average curves to prove Higher Order MA Models are smoother than Lower Order MA Models.
* Summary Statistics: Summarizes various parameters and this acts as a strong evidence for our prediction.
* R Programming Language, MS Excel and GRETL are the various ways through which we handled the data and came to our conclusion.

**GETTING THE DATA**

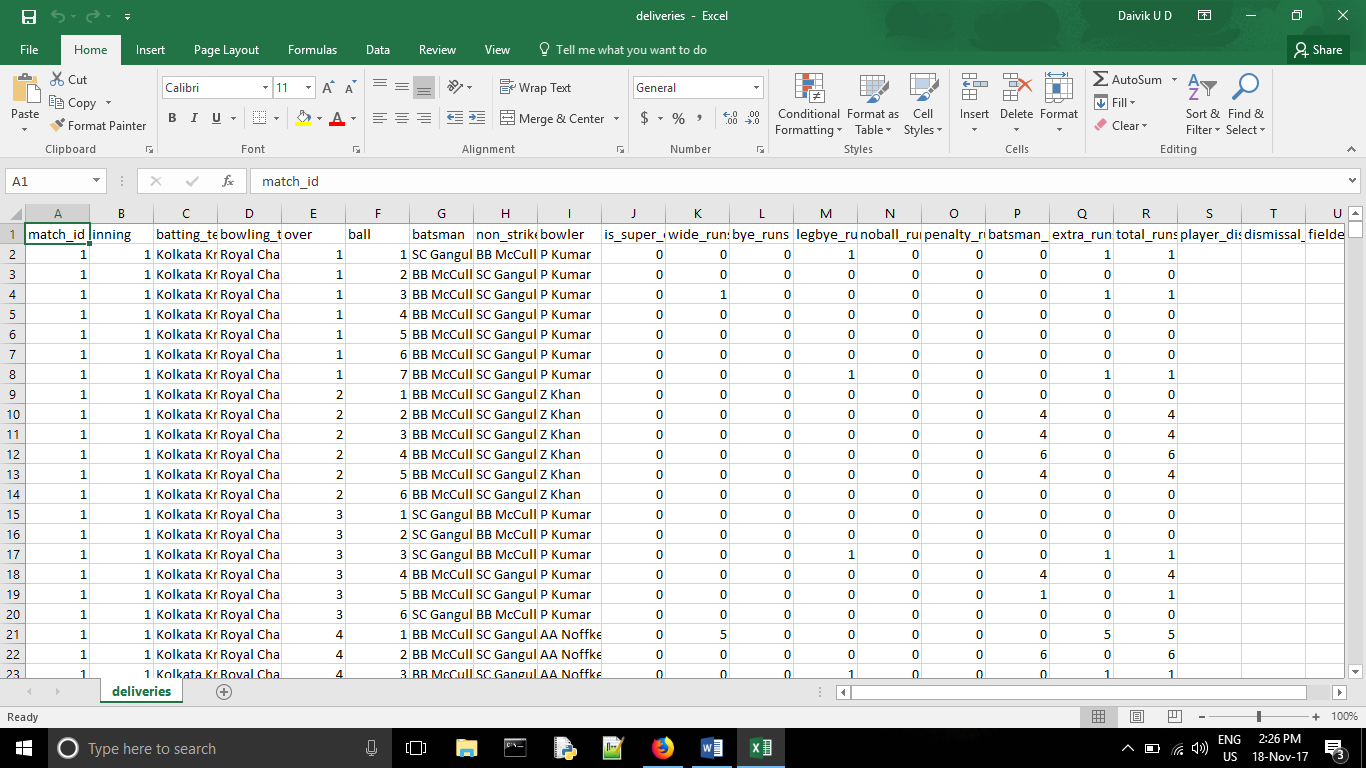
[**https://cricsheet.org/**](https://cricsheet.org/) has a sparse dataset of the IPL matches held from 2008 to 2016. This dataset contains ball to ball information about every match held from 2008 to 2016, as a sparse dataset. On running python scripts **mapper.py** and **combiner.py**, we scraped the data to suit our requirements.

Matches.csv contains:



1. Teams which played against each other in a particular venue.
2. Result of the match.
3. Other match relevant details.

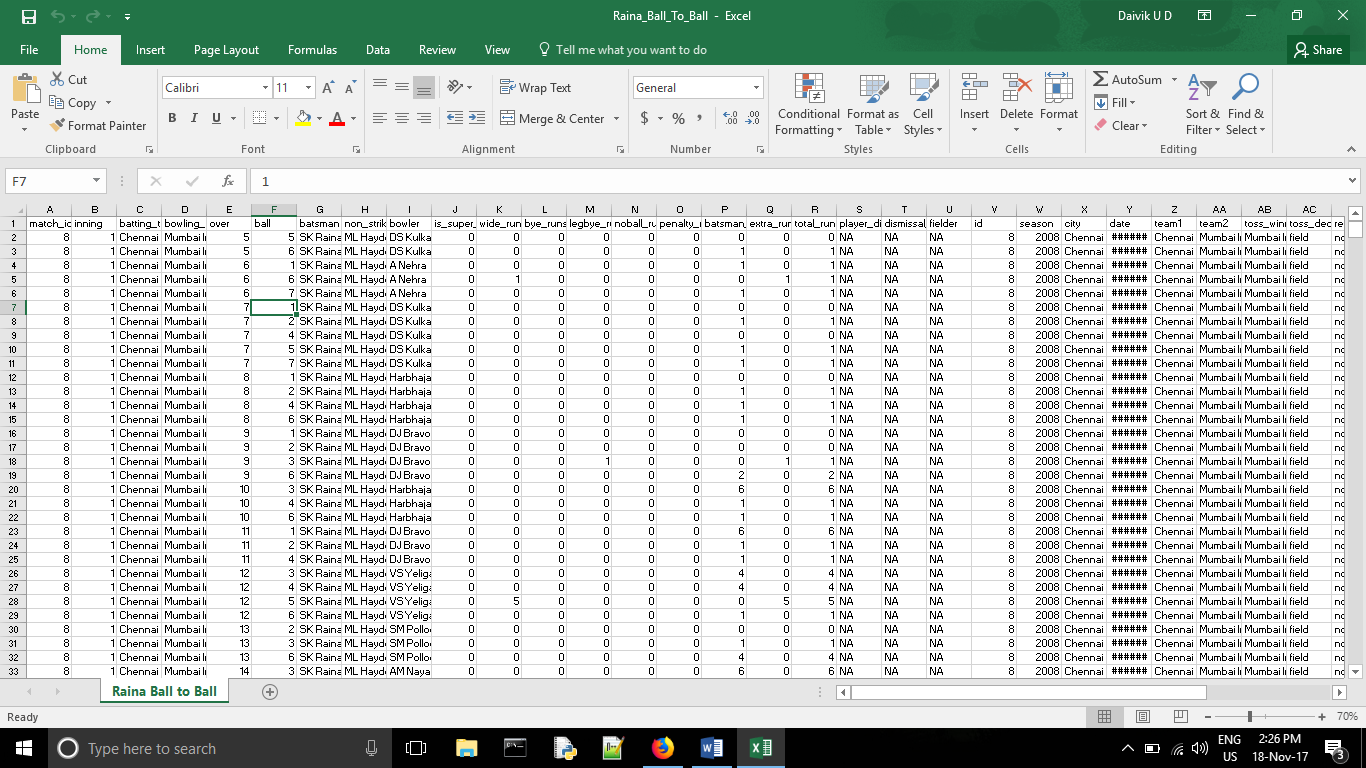
Deliveries.csv contains:



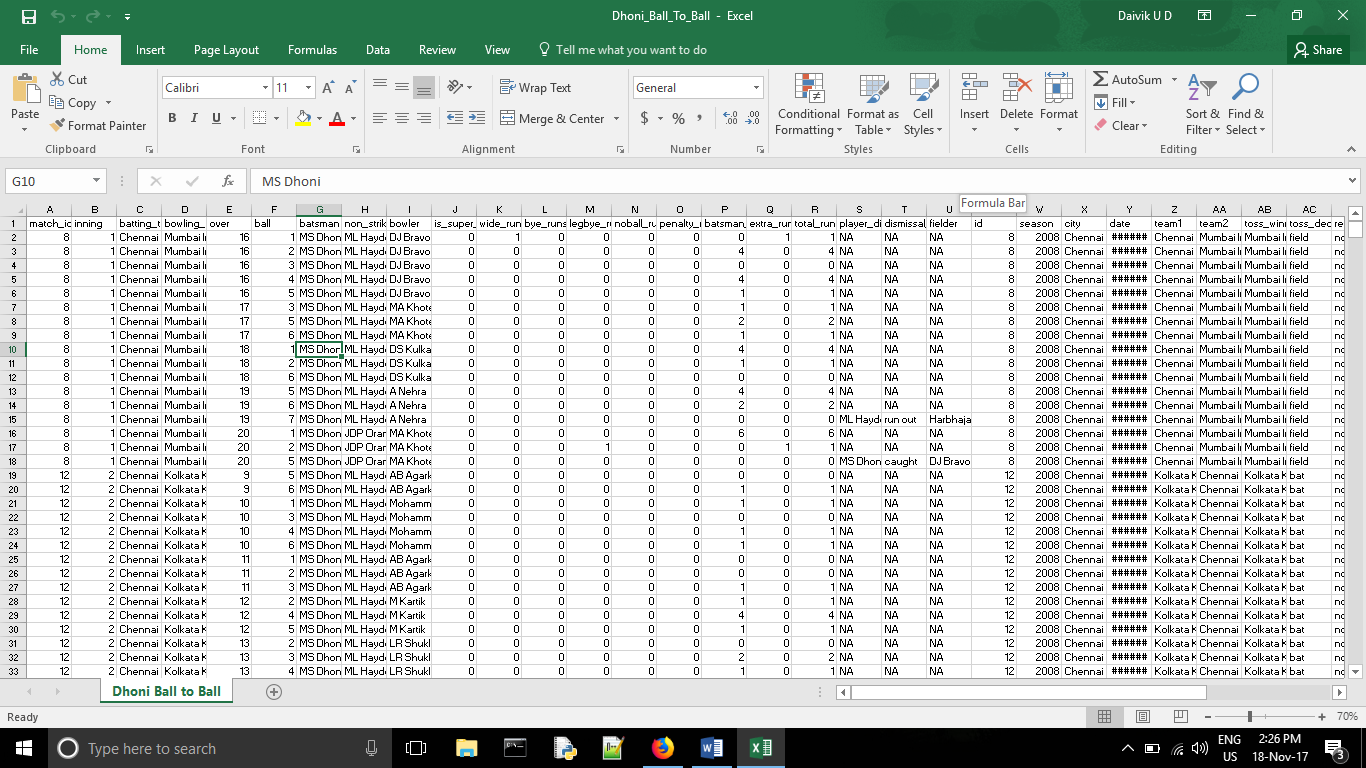
Information for every match(both innings). It contains the ball to ball details of every match. It shows the exact result for every ball played –including the striker, non-striker, bowler and how the runs were scored for that particular ball.

Using R studio, we read deliveries.csv and matches.csv. Then we filter out Raina and Dhoni’s data in MA Chidambaram, Chepauk in Chennai and Feroz Shah Kotla in Delhi.

Raina ball to ball.xls:



Dhoni’s ball to ball.xls:



Eventually, we have calculated the total number of runs scored by Raina and Dhoni in both the stadiums. The following Codes are used to filter out our required data:

(1. Raina\_Chennai\_code.txt) (2. Raina\_Delhi\_code.txt)

(3. Dhoni\_Chennai\_code.txt) (4. Dhoni\_Delhi\_code.txt)

Raina\_Total\_Chennai, Raina\_Total\_Delhi, Dhoni\_Total\_Chennai, Dhoni\_Total\_Delhi are EXCEL files.

This has the total number of runs scored by Raina and Dhoni in a particular match (match id). We observe that from 2008 to 2015, Raina has played 46 matches in Chennai and 8 matches in Delhi. We observe that from 2008 to 2017, Dhoni has played 43 matches in Chennai and 5 matches in Delhi.

**DATA ANALYSIS**

Chennai Super Kings were not a part of the IPL in 2016, Raina and Dhoni were not a part of Chennai Super Kings. Hence, no matches were played at Chennai in 2016. So, we are trying to predict their scores in Chennai for the year 2016 and analyze what their score would have been if they had played. Let X be the first Match ID which they play in Chennai of 2016.

Raina and Dhoni were a part of Gujarat Lions and Rising Pune Supergiants in 2016. In Delhi, for the year 2016, Raina has played 2 matches with match IDs 540 and 576 and Dhoni has played 1 match with match ID 550. Note that there are also previous matches from the year 2008 to 2015 in Delhi for Chennai Super Kings. Let Y be the first Match ID which they play in Delhi of 2017.

**We are predicting X and Y values using SES.**

**SUMMARY STATISTICS**

Kurtosis is a statistical measure that's used to describe the distribution, or [skewness](https://www.investopedia.com/terms/s/skewness.asp), of observed data around the mean, sometimes referred to as the volatility of volatility. Kurtosis is used generally in the statistical field to describes trends in charts. Kurtosis can be present in a chart with fat tails and a low, even distribution, as well as be present in a chart with skinny tails and a distribution concentrated toward the mean.

Coefficient of Variation (CV) is the ratio of the standard deviation to the mean. The **higher the coefficient of variation**, the **greater the level of dispersion around the mean**. It is generally expressed as a percentage.

**Raina’s Summary Statistics**

* Average runs scored by Raina in Chennai = 28.304 and Average runs scored by Raina in Delhi = 28.625.

The average runs scored by Raina in Chennai is very similar to the average runs scored by him in Delhi. By this, we can conclude that his performance does not depend on whether he has played in his franchise home ground(Chennai) or not (Delhi).

* The lowest score of Raina in Chennai is 0 and the highest score of Raina in Chennai is 100. The lowest score of Raina in Delhi is 1 and the highest score of Raina in Delhi is 82.

Since he has played more matches in Chennai, the chances of him scoring a duck or a century is more because the more number of innings a player bats for, the more likely he would play well. But, it is not necessary that he would score well always. His form might not be good, so he could get out for a duck as well.

* Standard Deviation for Raina in Chennai = 21.934. Difference from Mean = 6.37.
* Standard Deviation for Raina in Delhi = 29.213. Difference from Mean = -0.588.

In the upcoming match, Raina would probably play better in Delhi than Chennai.

* (CV) % in Chennai = 77.49%
* (CV) % in Delhi = 102.05 %

Raina is more consistent in Chennai because his CV % value is lesser.

**Dhoni’s Summary Statistics**

* Average runs scored by Dhoni in Chennai = 26.605 and Average runs scored by Dhoni in Delhi = 25.4. Dhoni’s highest score in Chennai is 67 and in Delhi is 44. Dhoni’s Mode is 30 in Chennai and NO MODE in Delhi.

Firstly, the average runs scored by Dhoni in Chennai is very slightly higher to the average runs scored by him in Delhi.

Secondly, his highest score is more in Chennai.

Thirdly, his mode in Chennai is 30 and he might score somewhere around 30 the next time he plays there.

So, he prefers batting in Chennai and may score more in the next match there.

* Standard Deviation for Raina in Chennai = 17.947. Difference from Mean = 8.658.
* Standard Deviation for Raina in Delhi = 14.082. Difference from Mean = 11.318.

In the upcoming match, Dhoni would play better in Chennai.

* (CV) % in Chennai = 67.45%
* (CV) % in Delhi = 55.44%

Dhoni is more consistent in Delhi because his CV % value is lesser. But, this is contradicting to our previous analogy. He has played only 5 matches in Delhi. So, we cannot draw an exact conclusion from CV%. It is better to conclude from more number of data points and taking average as a parameter. Hence, Dhoni is more consistent in Chennai itself looking at his scores.

**SIMPLE EXPONENTIAL SMOOTHING**

The criteria we are using is Mean Square Error (MSE). The least MSE is always the best option and even though Mean Error is close to 0 in some cases, we are dropping it from our analysis.

**Predicting Raina’s score for the year 2016 in Chennai.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **match\_id** | **Score** | **α = 0.1** | **α = 0.3** | **α = 0.5** | **α = 0.7** | **α = 0.9** |
| 8 | 53 | #N/A | #N/A | #N/A | #N/A | #N/A |
| 20 | 3 | 53 | 53 | 53 | 53 | 53 |
| 26 | 32 | 48 | 38 | 28 | 18 | 8 |
| 32 | 26 | 46.4 | 36.2 | 30 | 27.8 | 29.6 |
| 48 | 21 | 44.36 | 33.14 | 28 | 26.54 | 26.36 |
| 51 | 45 | 42.024 | 29.498 | 24.5 | 22.662 | 21.536 |
| 120 | 6 | 42.3216 | 34.1486 | 34.75 | 38.2986 | 42.6536 |
| 131 | 23 | 38.68944 | 25.70402 | 20.375 | 15.68958 | 9.66536 |
| 143 | 40 | 37.1205 | 24.89281 | 21.6875 | 20.80687 | 21.66654 |
| 147 | 13 | 37.40845 | 29.42497 | 30.84375 | 34.24206 | 38.16665 |
| 152 | 23 | 34.9676 | 24.49748 | 21.92188 | 19.37262 | 15.51667 |
| 163 | 78 | 33.77084 | 24.04824 | 22.46094 | 21.91179 | 22.25167 |
| 165 | 15 | 38.19376 | 40.23376 | 50.23047 | 61.17354 | 72.42517 |
| 176 | 33 | 35.87438 | 32.66364 | 32.61523 | 28.85206 | 20.74252 |
| 189 | 29 | 35.58694 | 32.76454 | 32.80762 | 31.75562 | 31.77425 |
| 203 | 2 | 34.92825 | 31.63518 | 30.90381 | 29.82669 | 29.27743 |
| 213 | 59 | 31.63542 | 22.74463 | 16.4519 | 10.34801 | 4.727743 |
| 217 | 61 | 34.37188 | 33.62124 | 37.72595 | 44.4044 | 53.57277 |
| 230 | 14 | 37.03469 | 41.83487 | 49.36298 | 56.02132 | 60.25728 |
| 238 | 19 | 34.73122 | 33.48441 | 31.68149 | 26.6064 | 18.62573 |
| 248 | 8 | 33.1581 | 29.13908 | 25.34074 | 21.28192 | 18.96257 |
| 249 | 36 | 30.64229 | 22.79736 | 16.67037 | 11.98458 | 9.096257 |
| 261 | 23 | 31.17806 | 26.75815 | 26.33519 | 28.79537 | 33.30963 |
| 273 | 0 | 30.36026 | 25.63071 | 24.66759 | 24.73861 | 24.03096 |
| 275 | 26 | 27.32423 | 17.94149 | 12.3338 | 7.421584 | 2.403096 |
| 284 | 15 | 27.19181 | 20.35905 | 19.1669 | 20.42648 | 23.64031 |
| 288 | 44 | 25.97263 | 18.75133 | 17.08345 | 16.62794 | 15.86403 |
| 293 | 32 | 27.77536 | 26.32593 | 30.54172 | 35.78838 | 41.1864 |
| 305 | 28 | 28.19783 | 28.02815 | 31.27086 | 33.13651 | 32.91864 |
| 321 | 27 | 28.17804 | 28.01971 | 29.63543 | 29.54095 | 28.49186 |
| 322 | 73 | 28.06024 | 27.71379 | 28.31772 | 27.76229 | 27.14919 |
| 327 | 10 | 32.55422 | 41.29966 | 50.65886 | 59.42869 | 68.41492 |
| 337 | 30 | 30.29879 | 31.90976 | 30.32943 | 24.82861 | 15.84149 |
| 340 | 8 | 30.26892 | 31.33683 | 30.16471 | 28.44858 | 28.58415 |
| 351 | 51 | 28.04202 | 24.33578 | 19.08236 | 14.13457 | 10.05841 |
| 355 | 16 | 30.33782 | 32.33505 | 35.04118 | 39.94037 | 46.90584 |
| 359 | 44 | 28.90404 | 27.43453 | 25.52059 | 23.18211 | 19.09058 |
| 366 | 100 | 30.41364 | 32.40417 | 34.76029 | 37.75463 | 41.50906 |
| 387 | 7 | 37.37227 | 52.68292 | 67.38015 | 81.32639 | 94.15091 |
| 460 | 4 | 34.33504 | 38.97804 | 37.19007 | 29.29792 | 15.71509 |
| 462 | 14 | 31.30154 | 28.48463 | 20.59504 | 11.58938 | 5.171509 |
| 483 | 29 | 29.57139 | 24.13924 | 17.29752 | 13.27681 | 13.11715 |
| 488 | 17 | 29.51425 | 25.59747 | 23.14876 | 24.28304 | 27.41172 |
| 495 | 52 | 28.26282 | 23.01823 | 20.07438 | 19.18491 | 18.04117 |
| 500 | 10 | 30.63654 | 31.71276 | 36.03719 | 42.15547 | 48.60412 |
| 504 | 3 | 28.57289 | 25.19893 | 23.01859 | 19.64664 | 13.86041 |
| **X for 2016** | **26.0156** | **26.0156** |  | **15.78644** |  | **5.557289** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Alpha Value** | **Mean Error** | **Mean Square Error** | **Mean Absolute Deviation** | **Mean Absolute Percentage Error** |
| 0.1 | -5.996533823 | 595.9928661 | 19.68561437 | 197.1435686 |
| 0.5 | -1.777364558 | 730.1056272 | 21.0475601 | 198.6483665 |
| 0.9 | -1.20775207 | 1010.425135 | 25.1345547 | 210.3656573 |

The least MSE is **595.992** for **α = 0.1**

Mean of Raina in Chennai = **28.304**

Therefore, his score has to be very close to the mean and from the least MSE criteria,

**X for 2016 = 26.0156**

**Predicting Raina’s score for the year 2017 in Delhi.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **match\_id** | **Score** | **α = 0.1** | **α = 0.3** | **α = 0.5** | **α = 0.7** | **α = 0.9** |
| 28 | 1 | #N/A | #N/A | #N/A | #N/A | #N/A |
| 126 | 49 | 1 | 1 | 1 | 1 | 1 |
| 259 | 17 | 5.8 | 15.4 | 25 | 34.6 | 44.2 |
| 345 | 30 | 6.92 | 15.88 | 21 | 22.28 | 19.72 |
| 395 | 82 | 9.228 | 20.116 | 25.5 | 27.684 | 28.972 |
| 424 | 47 | 16.5052 | 38.6812 | 53.75 | 65.7052 | 76.6972 |
| 540 | 2 | 19.55468 | 41.17684 | 50.375 | 52.61156 | 49.96972 |
| 576 | 1 | 17.79921 | 29.42379 | 26.1875 | 17.18347 | 6.796972 |
| **Y for 2017** | **9.3996** | **16.11929** |  | **9.399606** |  | **2.679921** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Alpha Value** | **Mean Error** | **Mean Square Error** | **Mean Absolute Deviation** | **Mean Absolute Percentage Error** |
| 0.1 | 21.59898686 | 1396.886218 | 31.41438457 | 421.7227092 |
| 0.5 | 3.598214286 | 1237.337612 | 28.83035714 | 742.2545926 |
| 0.9 | 0.092015429 | 1311.158542 | 31.71027029 | 485.466125 |

The least MSE is **1237.337** for **α = 0.5**

Mean of Raina in Delhi = **28.625**

Even though his predicted score is not close to the mean, we are compromising on the mean factor and emphasizing more on the least MSE factor.

**Y for 2017 = 9.3996**

**Predicting Dhoni’s score for the year 2016 in Chennai.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **match\_id** | **Score** | **α = 0.1** | **α = 0.3** | **α = 0.5** | **α = 0.7** | **α = 0.9** |
| 8 | 30 | #N/A | #N/A | #N/A | #N/A | #N/A |
| 12 | 43 | 30 | 30 | 30 | 30 | 30 |
| 20 | 32 | 31.3 | 33.9 | 36.5 | 39.1 | 41.7 |
| 26 | 23 | 31.37 | 33.33 | 34.25 | 34.13 | 32.97 |
| 32 | 60 | 30.533 | 30.231 | 28.625 | 26.339 | 23.997 |
| 48 | 4 | 33.4797 | 39.1617 | 44.3125 | 49.9017 | 56.3997 |
| 51 | 12 | 30.53173 | 28.61319 | 24.15625 | 17.77051 | 9.23997 |
| 120 | 42 | 28.67856 | 23.62923 | 18.07813 | 13.73115 | 11.724 |
| 143 | 14 | 30.0107 | 29.14046 | 30.03906 | 33.51935 | 38.9724 |
| 147 | 0 | 28.40963 | 24.59832 | 22.01953 | 19.8558 | 16.49724 |
| 152 | 31 | 25.56867 | 17.21883 | 11.00977 | 5.956741 | 1.649724 |
| 165 | 0 | 26.1118 | 21.35318 | 21.00488 | 23.48702 | 28.06497 |
| 176 | 29 | 23.50062 | 14.94723 | 10.50244 | 7.046107 | 2.806497 |
| 189 | 22 | 24.05056 | 19.16306 | 19.75122 | 22.41383 | 26.38065 |
| 203 | 26 | 23.8455 | 20.01414 | 20.87561 | 22.12415 | 22.43806 |
| 213 | 21 | 24.06095 | 21.8099 | 23.43781 | 24.83724 | 25.64381 |
| 230 | 63 | 23.75486 | 21.56693 | 22.2189 | 22.15117 | 21.46438 |
| 238 | 9 | 27.67937 | 33.99685 | 42.60945 | 50.74535 | 58.84644 |
| 248 | 22 | 25.81143 | 26.4978 | 25.80473 | 21.52361 | 13.98464 |
| 249 | 4 | 25.43029 | 25.14846 | 23.90236 | 21.85708 | 21.19846 |
| 261 | 40 | 23.28726 | 18.80392 | 13.95118 | 9.357125 | 5.719846 |
| 273 | 28 | 24.95854 | 25.16274 | 26.97559 | 30.80714 | 36.57198 |
| 275 | 15 | 25.26268 | 26.01392 | 27.4878 | 28.84214 | 28.8572 |
| 284 | 1 | 24.23641 | 22.70974 | 21.2439 | 19.15264 | 16.38572 |
| 288 | 34 | 21.91277 | 16.19682 | 11.12195 | 6.445793 | 2.538572 |
| 293 | 34 | 23.1215 | 21.53777 | 22.56097 | 25.73374 | 30.85386 |
| 321 | 23 | 24.20935 | 25.27644 | 28.28049 | 31.52012 | 33.68539 |
| 322 | 14 | 24.08841 | 24.59351 | 25.64024 | 25.55604 | 24.06854 |
| 327 | 51 | 23.07957 | 21.41546 | 19.82012 | 17.46681 | 15.00685 |
| 337 | 33 | 25.87161 | 30.29082 | 35.41006 | 40.94004 | 47.40069 |
| 340 | 10 | 26.58445 | 31.10357 | 34.20503 | 35.38201 | 34.44007 |
| 351 | 21 | 24.92601 | 24.7725 | 22.10252 | 17.6146 | 12.44401 |
| 355 | 67 | 24.53341 | 23.64075 | 21.55126 | 19.98438 | 20.1444 |
| 359 | 18 | 28.78007 | 36.64853 | 44.27563 | 52.89531 | 62.31444 |
| 366 | 2 | 27.70206 | 31.05397 | 31.13781 | 28.46859 | 22.43144 |
| 387 | 58 | 25.13185 | 22.33778 | 16.56891 | 9.940578 | 4.043144 |
| 460 | 30 | 28.41867 | 33.03644 | 37.28445 | 43.58217 | 52.60431 |
| 462 | 53 | 28.5768 | 32.12551 | 33.64223 | 34.07465 | 32.26043 |
| 483 | 41 | 31.01912 | 38.38786 | 43.32111 | 47.3224 | 50.92604 |
| 488 | 3 | 32.01721 | 39.1715 | 42.16056 | 42.89672 | 41.9926 |
| 495 | 29 | 29.11549 | 28.32005 | 22.58028 | 14.96902 | 6.89926 |
| 500 | 39 | 29.10394 | 28.52404 | 25.79014 | 24.7907 | 26.78993 |
| 504 | 13 | 30.09355 | 31.66682 | 32.39507 | 34.73721 | 37.77899 |
| **X for 2016** | **28.3841** | **28.38419** |  | **21.54677** |  | **14.70935** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Alpha Value** | **Mean Error** | **Mean Square Error** | **Mean Absolute Deviation** | **Mean Absolute Percentage Error** |
| 0.1 | -0.384716515 | 349.9406727 | 15.04203827 | 180.064647 |
| 0.5 | -0.347736438 | 483.7899405 | 18.12756063 | 203.9838189 |
| 0.9 | -0.384182559 | 703.5380037 | 22.28872651 | 204.9602862 |

The least MSE is **349.9406** for **α = 0.1**

Mean of Dhoni in Chennai = **26.605**

Therefore, his score has to be very close to the mean and from the least MSE criteria,

**X for 2016 = 28.3841**

**Predicting Dhoni’s score for the year 2017 in Delhi.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **match\_id** | **Score** | **α = 0.1** | **α = 0.3** | **α = 0.5** | **α = 0.7** | **α = 0.9** |
| 28 | 33 | #N/A | #N/A | #N/A | #N/A | #N/A |
| 259 | 11 | 33 | 33 | 33 | 33 | 33 |
| 345 | 44 | 30.8 | 26.4 | 22 | 17.6 | 13.2 |
| 424 | 12 | 32.12 | 31.68 | 33 | 36.08 | 40.92 |
| 550 | 27 | 30.108 | 25.776 | 22.5 | 19.224 | 14.892 |
| **Y for 2017** | **29.7972** | **29.7972** |  | **28.554** |  | **27.3108** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Alpha Value** | **Mean Error** | **Mean Square Error** | **Mean Absolute Deviation** | **Mean Absolute Percentage Error** |
| 0.1 | -8.007 | 268.178516 | 14.607 | 102.2944444 |
| 0.5 | -4.125 | 357.3125 | 17.375 | 110.4166667 |
| 0.9 | -2.003 | 603.902516 | 23.457 | 138.9611111 |

The least MSE is **268.178** for **α = 0.1**

Mean of Dhoni in Delhi = **25.4**

Therefore, his score has to be very close to the mean and from the least MSE criteria,

**Y for 2017 = 29.7972**

**CONCLUSION**

**SES Conclusion**

For Raina and Dhoni in Chennai, we observe the SES Graphs.

When α = 0.1, the predicted values are just an average of the actual values. The curve is more flatter and horizontal to the X axis.

When α = 0.9, the predicted values mimic the actual values and hence the jump from one match to the next match is more. It is less probable to predict the right score for the next match.

Therefore, we can conclude that averaging it out is our best bet and our predicted values for both of them in Chennai is for α = 0.1.

**Kurtosis Conclusion**

Kurtosis value of Raina in Chennai = 1.2474

Kurtosis value of Dhoni in Chennai = -0.4933

Kurtosis value of Raina in Delhi = -0.7385

Kurtosis value of Dhoni in Delhi = -1.4243

A more positive Kurtosis value indicates that the player is more consistent and his scores do not deviate from the mean. To compare 2 negative Kurtosis values, more negative the Kurtosis value, more inconsistent the player is.

Hence, Raina plays better than Dhoni in Chennai and Delhi.

**Bar Graphs Conclusion**

We observe the following Bar Graphs: the number of matches played on X axis and the runs scored on Y axis.

Raina has scored only one duck in Chennai compared to Dhoni’s 2 ducks. So, Raina is consistent.

**General Conclusion**

Dhoni usually bats after Raina. So, Dhoni should always have to score more runs as soon as he comes in to bat. So, that is why Dhoni might not be as consistent as Raina. Sometimes, only Raina would have got to bat and the match would have been completed even before Dhoni comes in to bat. But, according to the data, Raina is better player than Dhoni in Chennai and Delhi.

**Prediction Values**

**Raina: X for Chennai 2016 = 26.0156**

**Raina: Y for Delhi 2017 = 9.3996**

**Dhoni: X for Chennai 2016 = 28.3841**

**Dhoni: Y for Delhi 2017 = 29.7972**