International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 2, Mar.-Apr., 2024, pp: 1447-1454 www.ijhssm.org

QUANTITATIVE TECHNIQUES – II
EXCEL PROJECT- STATISTICAL ANALYSIS OF SIXES SCORED IN IPL (CRICKET)

Ananya Singh, Ansh Chheda, Arnav Kalra, Aryan Agrawal
Ashmi Jain

Project Submitted in Partial Fulfilment of the Requirements of
B.B.A. 2023-26

Programme: B.B.A. Semester II
Division: I
Group Number 2

Date of Submission: 09-04-2024 Date of Acceptance: 23-04-2024

I. INTRODUCTION

In the world of cricket, the ability to hit sixes is often celebrated as a mark of a powerful batsman. The frequency and distribution of sixes per innings gives a promising idea of the changing strategies and players' performance during the match. In this data analysis project, we dive into the fascinating world of cricket statistics, focusing specifically on the frequency of sixes across various matches and players.

Our analysis uses a multi-method approach combining techniques such as moving averages, descriptive statistics, least squares statistics and statistical analysis across the supply chain. This technique is a powerful tool for uncovering complex patterns and hidden threads in data, providing insight into the dynamics of a single run. Using this method of analysis, we aim to not only present all the issues across many situations, but also identify underdog performance, main effect and be able to make predictions across the cricket spectrum.

Through careful analysis and interpretation of data, we strive to provide essential information to make the right decisions in the sport and deepen our understanding of the nuances of good cricket. Join us in this special edition as we explore the fascinating world of cricket statistics, uncovering the hidden stories in the numbers and uncovering the events that led to the six-time surge in some time.

Objectives:

- To thoroughly analyze how many sixes are hit in IPL cricket matches throughout a year, by using descriptive statistics, the least squares method, and chain-based index construction
- To understand better how power-hitting works in cricket.
- To learn more about the factors that contribute to hitting sixes and how they change over time.
- To analyze measures such as mean, median, mode, standard deviation, skewness, and kurtosis.
- To formulate a regression equation that models the relationship between a Year and the number of sixes hit in a year.
- To track and analyze the percentage change in the number of sixes hit from one time period to another within the designated years.
- To understand moving average analysis to reveal the underlying trends in the number of sixes hit over the time of 3,4 and 5 years.

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 2, Mar.-Apr., 2024, pp: 1447-1454 www.ijhssm.org

II. DATA EXPLANATION

Player	Runs	BF	SR	4s	6s	Against	Venue	Match Date
Brendon McCullum	158	73	216.43	10	13	RCB	M. Chinnaswamy Stadium	18-Apr-08
Praveen Kumar	18	15	120	1	2	KKR	M. Chinnaswamy Stadium	18-Apr-08
Michael Hussey	116	54	214.81	8	9	PBKS	IS Bindra Stadium	19-Apr-08
Suresh Raina	32	13	246.15	2	3	PBKS	IS Bindra Stadium	19-Apr-08
James Hopes	71	33	215.15	10	3	CSK	IS Bindra Stadium	19-Apr-08
Subramaniam Badrinath	31	14	221.42	3	2	PBKS	IS Bindra Stadium	19-Apr-08
Yuvraj Singh	23	13	176.92	1	2	CSK	IS Bindra Stadium	19-Apr-08
Kumar Sangakkara	54	33	163.63	5	2	CSK	IS Bindra Stadium	19-Apr-08
Ravindra Jadeja	29	23	126.08	2	2	DC	Arun Jaitley Stadium	19-Apr-08
David Hussey	38	43	88.37	1	3	DEC	Eden Gardens	20-Apr-08
Shaun Pollock	28	12	233.33	3	2	RCB	Wankhede Stadium	20-Apr-08
Mark Boucher	39	19	205.26	4	2	MI	Wankhede Stadium	20-Apr-08
Ross Taylor	23	12	191.66	1	2	MI	Wankhede Stadium	20-Apr-08
Adam Gilchrist	23	22	104.54	1	2	KKR	Eden Gardens	20-Apr-08
Andrew Symonds	32	39	82.05	2	2	KKR	Eden Gardens	20-Apr-08
Shane Watson	76	49	155.1	5	5	PBKS	Sawai Mansingh Stadium	21-Apr-08
Yuvraj Singh	57	34	167.64	6	3	RR	Sawai Mansingh Stadium	21-Apr-08
Karan Goel	26	21	123.8	3	2	RR	Sawai Mansingh Stadium	21-Apr-08
Virender Sehwag	94	41	229.26	10	6	DEC	Rajiv Gandhi Intl. Cricket Stadium	22-Apr-08
Rohit Sharma	66	36	183.33	6	4	DC	Rajiv Gandhi Intl. Cricket Stadium	22-Apr-08
Harbhajan Singh	28	14	200	1	3	CSK	Chidambaram	23-Apr-08
Suresh Raina	53	37	143.24	3	3	MI	Chidambaram	23-Apr-08
Matthew Hayden	81	46	176.08	12	2	MI	Chidambaram	23-Apr-08
Andrew Symonds	117	53	220.75	11	7	RR	Rajiv Gandhi Intl. Cricket Stadium	24-Apr-08
Yusuf Pathan	61	28	217.85	4	6	DEC	Rajiv Gandhi Intl. Cricket Stadium	24-Apr-08
Mohammad Kaif	34	16	212.5	2	3	DEC	Rajiv Gandhi Intl. Cricket Stadium	24-Apr-08
Shane Warne	22	9	244.44	2	2	DEC	Rajiv Gandhi Intl. Cricket Stadium	24-Apr-08
Graeme Smith	71	45	157.77	9	2	DEC	Rajiv Gandhi Intl. Cricket Stadium	24-Apr-08
Ross Taylor	44	20	220	6	3	RR	M. Chinnaswamy Stadium	26-Apr-08
Matthew Hayden	70	49	142.85	6	3	KKR	Chidambaram	26-Apr-08

An extract of our dataset

We obtained this dataset from Kaggle. The dataset we have selected has 2101 rows, giving us ample data to analyze. The tenure or the time period of our dataset extends from 18-04-2008 to 15-10-2021.

The data contains 9 columns. Here is the description of what each column depicts:

1. Player: The player column shows the name of the player who hit the six.
2. Runs: This column depicts the runs made by the player in a match.
3. BF: The BF column shows the number of balls the player faces in that match.
4. SR: SR is the strike rate of the player in the match.SR = Average Runs per ball multiplied by 100.
5. 4s: This column displays the number of 4s that the player hit in that match.
6. 6s: This column displays the number of 6s that the player hit in that match.
7. Against: Against column lists the team's name against which the match was played.
8. Venue: The Venue column contains the stadium where the match was played.
9. Match date: This column displays the date of the match.

Since one of the columns of the given data set displays the number of the sixes hit in a year, the selected data will help us to get the insights of the same. We will get to know the trends in the sixes with respect to years,

III. DATA ANALYSIS AND INTERPRETATION

MOVING AVERAGES

Since our topic is number of sixes scored overtime, finding out the moving averages provided us with insights based on the performances of the players and the trends overtime.

Here is how finding moving averages helped us:

Smoothing trends: Moving Averages make it easier for us to identify long-term trends by smoothing out the fluctuations in the data. In our data it can help us understand the trajectory of sixes scored over the years.

Identifying patterns: By calculating moving averages we can find and identify patterns over the years. It can also help us identify the high and low points in the performances of players and analyze cyclical trends, if any.

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 2, Mar.-Apr., 2024, pp: 1447-1454 www.ijhssm.org

Forecasting: By calculating moving averages, we can make predictions for the future by analyzing historical data. We can thus make informed predictions and decisions for the number of sixes scored in the upcoming seasons or matches.
Comparison: By finding out 3 yearly 4 yearly and 5 yearly moving averages, we can compare contrast and understand the differences between short term and long-term trends. It also helps us understand the evolution of the performances.

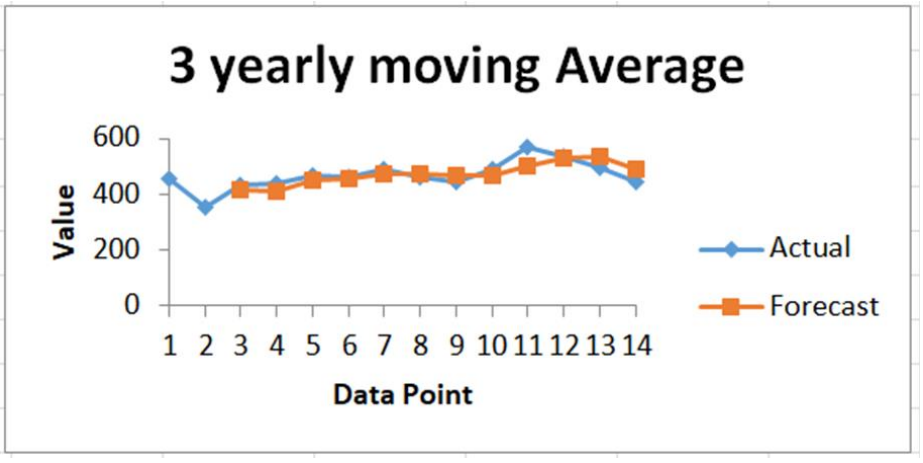
Data interpretation: By being aware of the past trends and historical data, it allows the cricket analysts to paint a picture that is easier for the audience to understand even those that are unfamiliar with the intricacies of cricket statistics.
From our data, the year conceding the highest number of sixes was 2018 with 569 sixes, whereas the one with the lowest was 2009 with 352 sixes.

The table for the moving averages:

Row Labels	Sum of 6s
2008	455
2009	352
2010	436
2011	438
2012	470
2013	462
2014	489
2015	462
2016	447
2017	489
2018	569
2019	533
2020	498
2021	442
Grand Total	6542

Year	Sum of 6s	3-Yearly MA	4 yearly MA	5 yearly MA
2008	455	#N/A	#N/A	#N/A
2009	352	#N/A	#N/A	#N/A
2010	436	414.33	#N/A	#N/A
2011	438	408.67	420.25	#N/A
2012	470	448.00	424.00	430.20
2013	462	456.67	451.50	431.60
2014	489	473.67	464.75	459.00
2015	462	471.00	470.75	464.20
2016	447	466.00	465.00	466.00
2017	489	466.00	471.75	469.80
2018	569	501.67	491.75	491.20
2019	533	530.33	509.50	500.00
2020	498	533.33	522.25	507.20
2021	442	491.00	510.50	506.20

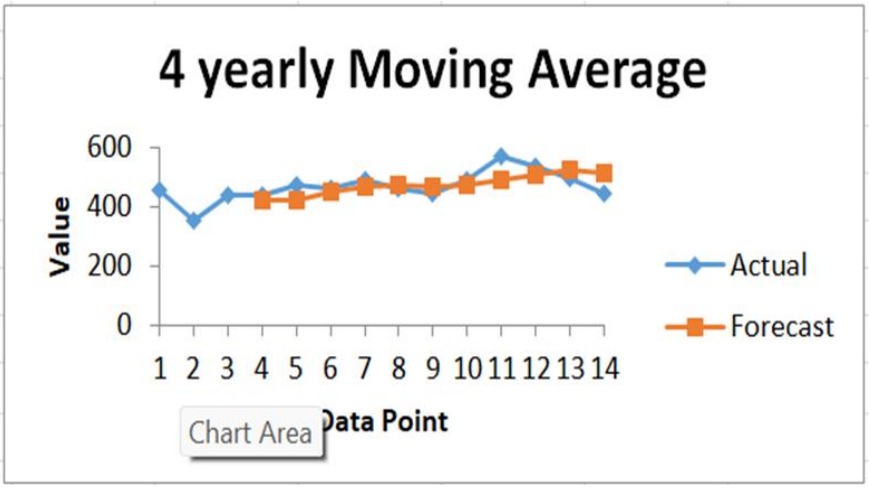
3-YEARLY MOVING AVERAGES
Calculating this involves making groups of three years each, adding the respective observations and dividing them by 3. Typically, the moving average for the first 2 years of the data comes to be N/A.
While there is a minute fall in the moving average of 2011 compared to that of 2010, a constant increase can be seen from 2012 to 2015. There is again a fall from 2015 to 2016 after which it remains constant till 2017. After 2017 till 2020, there is a general increase in the same. A major fall is witnessed from 2020 to 2021.
As seen in the graph below, after a slight fall in 2011, there is a steady increase up to 2015 and then a fall in 2016. It witnesses a steady growth till 2020 and falls in 2021. A fluctuation in trends can be seen in the short term.



International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 2, Mar.-Apr., 2024, pp: 1447-1454 www.ijhssm.org

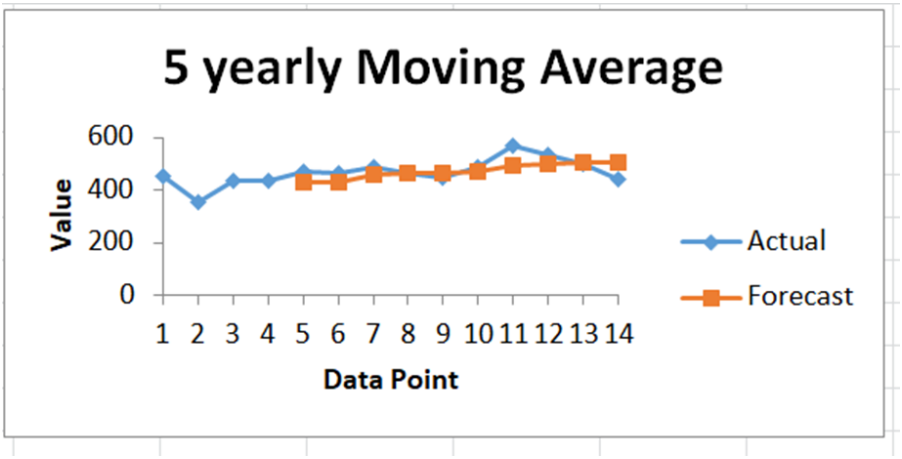
4-YEARLY MOVING AVERAGES

Calculating this involves making groups of four years each, adding the respective observations and dividing them by 4. Typically, the moving average for the first 3 years of the data comes to be N/A. There is a steady increase from 2011 to 2015 after which it falls in 2016. From 2016 to 2020, a progressive increase is witnessed. But in the year 2021 it seems to have fallen.



5-YEARLY MOVING AVERAGES

Calculating this involves making groups of five years each, adding the respective observations and dividing them by 5. Typically, the moving average for the first 4 years of the data comes to be N/A. Talking about the 5 yearly moving averages in our data, there seems to be a constant increase in the same over the years. An overall increase in the long-term trends can be witnessed.



DESCRIPTIVE STATISTICS

Descriptive statistics is a summary of the statistical data derived from our data set. It consists of various statistical functions such as mean, median, mode, standard deviation, skewness etc. The values of mean and standard deviation in the summary statistics were instrumental in finding the coefficient of variation subsequently determining the consistency of our data set.

The table below shows the descriptive statistics of our data.

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 2, Mar.-Apr., 2024, pp: 1447-1454 www.ijhssm.org

Sum of 6s	
Mean	467.29
Standard Error	13.45
Median	462.00
Mode	462.00
Standard Deviation	50.34
Sample Variance	2534.07
Kurtosis	1.99
Skewness	-0.16
Range	217.00
Minimum	352.00
Maximum	569.00
Sum	6542.00
Count	14.00

coefficient of variation
10.77274677

The variation coefficient was found using standard deviation and mean, 10.77274677. When we multiply it by 100, we get 1077.274677. We can thus conclude by saying that the given data is inconsistent. The dependent variable in the data was the number of sixes scored which displayed deviation over the years resulting in the data being inconsistent.

LEAST SQUARE METHOD

Here we use the equation, y = a +bx to project the trendline on a chart and get the estimated values for the total sixes scored each year based on actual totals.

In this equation,

y = the sum of sixes

a = y-intercept

b = coefficient of x, it determines the slope of the trendline

x = year – middle year

Year	Sum of 6s	x = year - middle year	y(estimated) = a + bx
2008	455	-6.5	421
2009	352	-5.5	428
2010	436	-4.5	435
2011	438	-3.5	442
2012	470	-2.5	449
2013	462	-1.5	457
2014	489	-0.5	464
2015	462	0.5	471
2016	447	1.5	478
2017	489	2.5	485
2018	569	3.5	492
2019	533	4.5	499
2020	498	5.5	507
2021	442	6.5	514
2022		7.5	521
2023		8.5	528
2024		9.5	535
2025		10.5	542
2026		11.5	549
2027		12.5	557
2028		13.5	564
2029		14.5	571
2030		15.5	578

Using this technique to analyze the dataset offered the following insights:

Modelling Trends: The dataset of the total sixes scored annually exhibits a trend over time. Whether it's an increasing, decreasing, or stable trend, the least squares method can help identify and model this trend by fitting a line or curve that minimizes the sum of squared differences between the observed data points and the predicted values.

Quantitative Relationship: The least squares method allows for the quantification of the relationship between variables. In this case, it can help quantify how the number of sixes scored in

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 2, Mar.-Apr., 2024, pp: 1447-1454 www.ijhssm.org

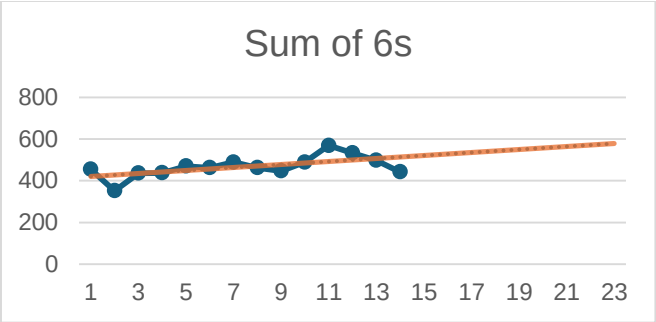
cricket changes over time, providing valuable insights into the dynamics of the game and the factors that influence scoring patterns.

Forecasting: Once a trend has been identified and modelled using the least squares method, it can be used for forecasting future values. By extrapolating the fitted line or curve, analysts can make predictions about the number of sixes that might be scored in future years, helping teams, coaches, and analysts prepare for upcoming matches and tournaments.

Evaluating Performance: It also allows for the evaluation of model performance. Analysts can assess how well the fitted line or curve captures the variation in the observed data and whether there are any systematic deviations or patterns left unaccounted for. This information can guide further refinement of the model and improve its predictive accuracy.

Identifying Outliers and Anomalies: Through this method, analysts can identify outliers and anomalies in the dataset—data points that deviate significantly from the overall trend. Understanding such unusual events can provide insights into exceptional performances, rule changes, or other factors affecting the scoring dynamics in cricket.

	Coefficients
Intercept	467.285714
x = year - middle year	7.14285714



Given above is the Y intercept of the trendline, which is 467.285714 and the slope of the trendline which is7.14285714.

As we can observe, there is a very slight increase in the sum of sixes over the years which could be due to various qualitative factors such as improvement in batters or a gradual decrease of efficient bowlers.

CHAIN-BASED INDEX NUMBERS

Using a chain-based index to analyze this dataset offered the following insights:

Year-on-Year comparison: In cricket, the number of sixes scored annually might fluctuate due to various factors such as rule changes, player performance, and match conditions. Using a chain-based index, we can normalize these fluctuations, providing a clearer picture of the actual trend in the summation of sixes scored from one year to the next.

Elimination of Base Years: Traditional index numbers often use a fixed base year, which can introduce bias if the chosen year has atypical characteristics or abnormal circumstances. Chain-based indices, however, do not rely on a fixed base year. Instead, they link consecutive years, making them more robust against base year selection bias.

Accommodation of Structural Changes: Cricket, like any other sport, undergoes structural changes over time, such as modifications in rules, formats, or the introduction of new tournaments. A chain-based index can better accommodate such structural changes by capturing the relative growth rates between consecutive periods, thus providing a more accurate reflection of the underlying trends.

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 2, Mar.-Apr., 2024, pp: 1447-1454 www.ijhssm.org

Year	Sum of 6s	Link Relative	Chain Base Index Numbers
2008	455	100	100
2009	352	77.36	77.36
2010	436	123.86	95.82
2011	438	100.46	96.26
2012	470	107.31	103.30
2013	462	98.30	101.54
2014	489	105.84	107.47
2015	462	94.48	101.54
2016	447	96.75	98.24
2017	489	109.40	107.47
2018	569	116.36	125.05
2019	533	93.67	117.14
2020	498	93.43	109.45
2021	442	88.76	97.14

As evident, the year 2018 witnessed the highest relative increase as compared to the first year and the year 2009 saw the lowest cumulative sixes scored.

LIMITATIONS

While the process helped us to get a detailed analysis of the data, we did face limitations.
The following are the limitations faced by us:

- Qualitative factors like the quality of the pitch, the bowler and the weather conditions were not taken into consideration during the calculations.
- Restricted access to the data may have hindered our ability to make accurate calculations.
- Due to our topic being sixes scored in a year we could not select a definitive base year and had to resort to chain based index numbers.
- The presence of years where the play was affected due to covid 19 can be considered as abnormal years and therefore clouded the forecasts.
- Future predictions might not be entirely accurate due to the reasons stated above.

REFERENCES

[1]. The data set was obtained from Kaggle.
[2]. The statistical concepts and their significance was referenced from the ppts provided by the professor.
[3]. The graphs and tables were obtained from the excel file mailed prior.
[4]. Plagiarism certificate obtained from turnitin.

CONCLUSION

By conducting our survey and analyzing the data, we aimed to achieve the following objectives:

- To show number of sixes scored by batsmen in an IPL match
- The data shows highest sixes in an inning of a match from the first season of IPL back in 2008
- As a T20 format game, sixes are an integral part of it. Boundaries help you score runs quicker
- Chris Gayle scored the highest number of sixes in an inning (17). He also is the player with the highest individual score in a match against Pune warriors (175).
- Whereas AB de Villiers has the record for the greatest number of fours in a match (19) against MI in Wankhede.

- We can see that in the year 2008 the lowest number of sixes scored in an IPL edition (352). Whereas in 2018 highest number of sixes in an edition was recorded (569)

- A total of 6542 sixes have been hit in the history of IPL

- The excel sheet has a table made for least square which predicts the upcoming number of sixes scored in the next few editions of IPL

- We were able to analyze and have a better understanding of the data. It helped us forecast, understand trends, both short term and long term and find out the consistency of the data.