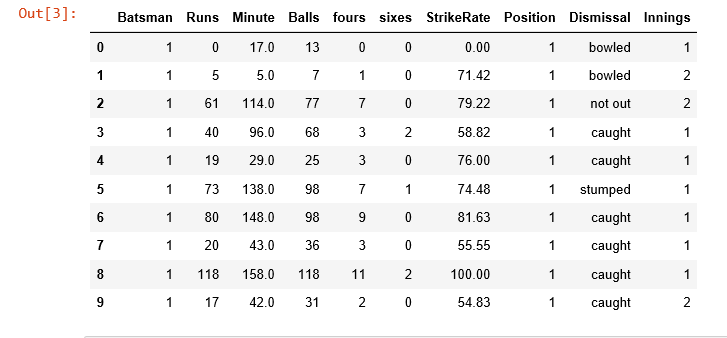
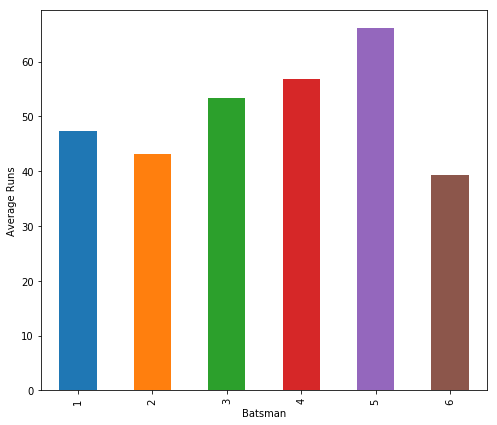
**Introduction:** This report is about data analysis using python notebook and a dataset consist of six batsmen ODI career. From this dataset we compare performance of the players, we also given our own observation about the players then we identified players based on their average score and performance. To implement this task we used anaconda distribution of python version 3 also NUMPY, PANDAS as well as MATPLOTLIB library.   
  
**Objectives:**

Our main job was to observe the data and analyze to find out comparative relation between different batsman and different attributes. We analyzed the data in our own way and got some conclusions on different relation like batsman strike rate etc.  
 **Data Frame :** The dataset we worked on consist total 225 samples including 6 batsmen’s runs, minute, balls, fours, sixes, strike rate, position, dismissal and innings.

**Batting Average:** Batting average of six batsmen****

Batsman 1 47.440000

Batsman 2 43.150000

Batsman 3 53.333333

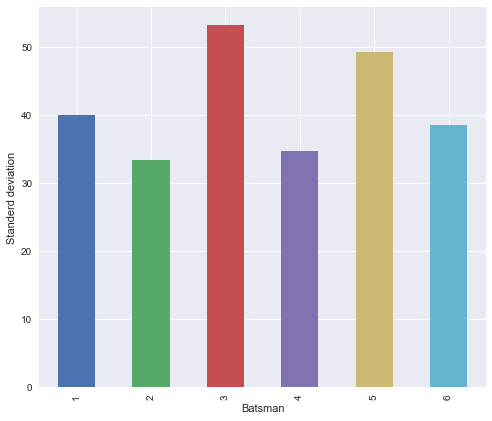
Batsman 4 56.914286

Batsman 5 66.146341

Batsman 6 39.386364

From analysis of this bar plot we can easily conclude that batsman 1 has the highest average and batsman 6 with the lowest.

**Standard Deviation:**

****

Batsman

1 39.987581

2 33.456861

3 53.300209

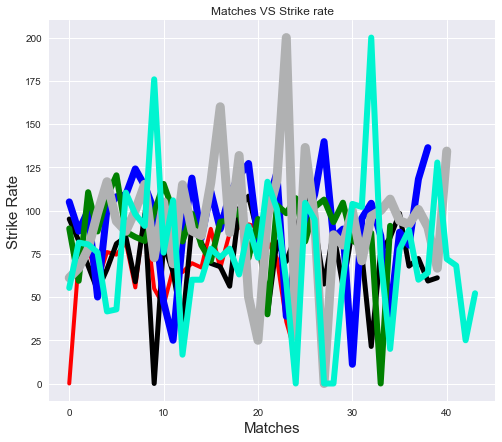
4 34.748145

5 49.296836

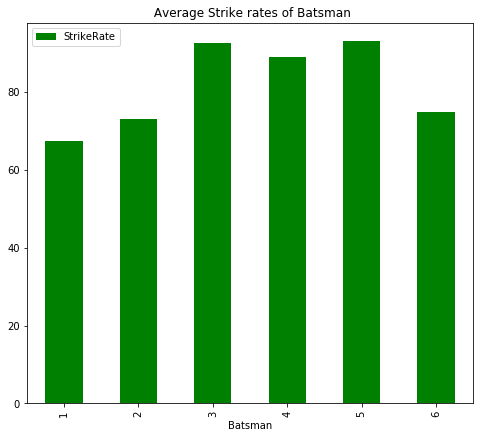
6 38.566895

This graph compares the average runs of six different batsman’s. Batsman1 has a higher average the other batsman’s.

**Career Strike Rate:**

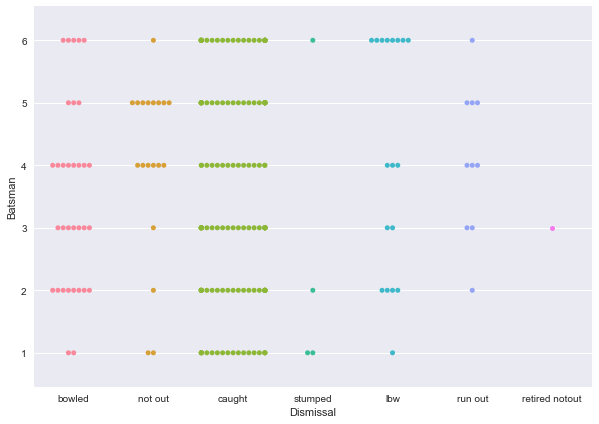
****

This graph shows the players individuals strike rate per match of those six players.

**Average Strike Rate:  
**

That figure is a comparison of six different batsman’s strike rate.

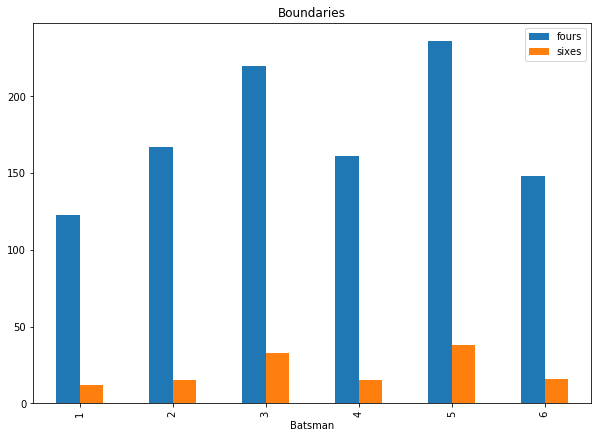
**Dismissal:**

****

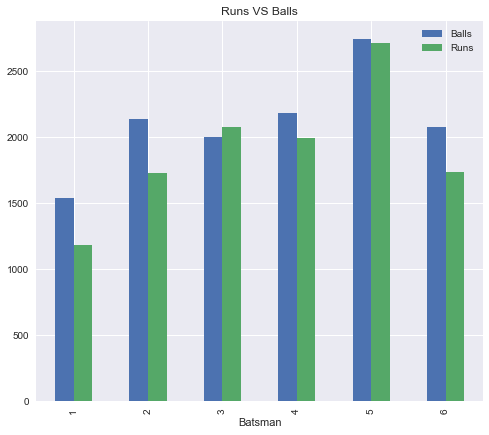
This bee swarm plot shows that batsman 6 loses his wicket more times by **LBW** than the other batsman’s. It also shows that batsman 4 and batsman 5 remains **not out** more times than others batsmen.

.

**Batsman’s Fours and Sixes:**

****

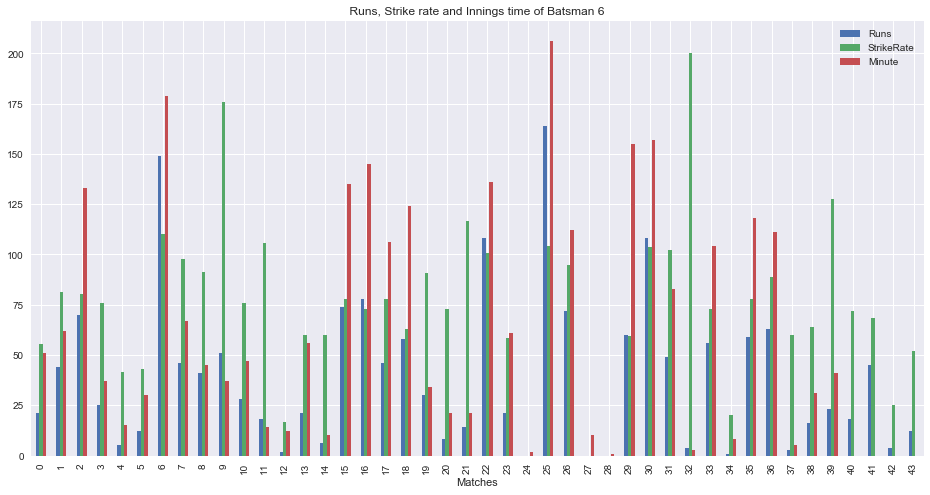
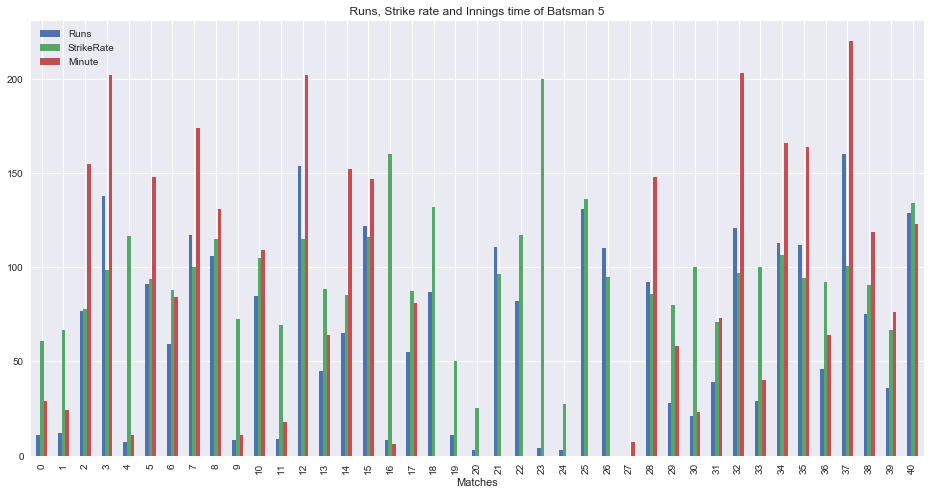
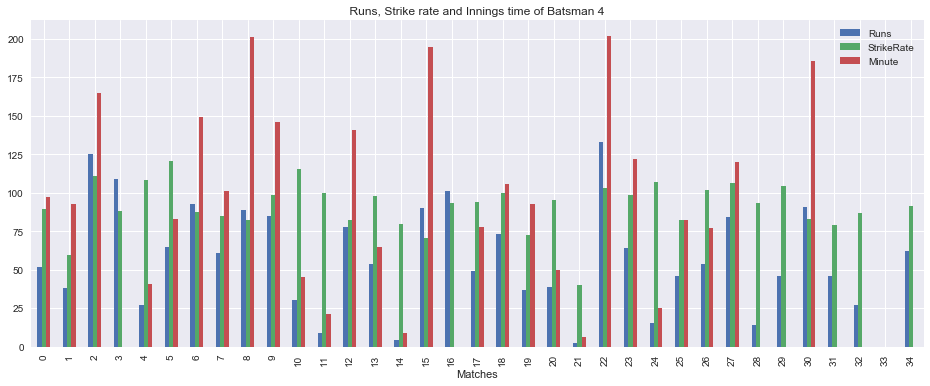
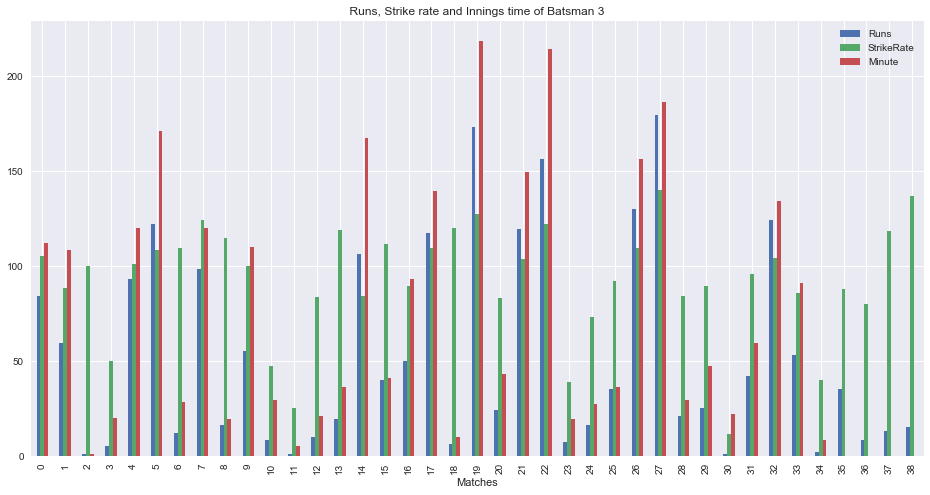
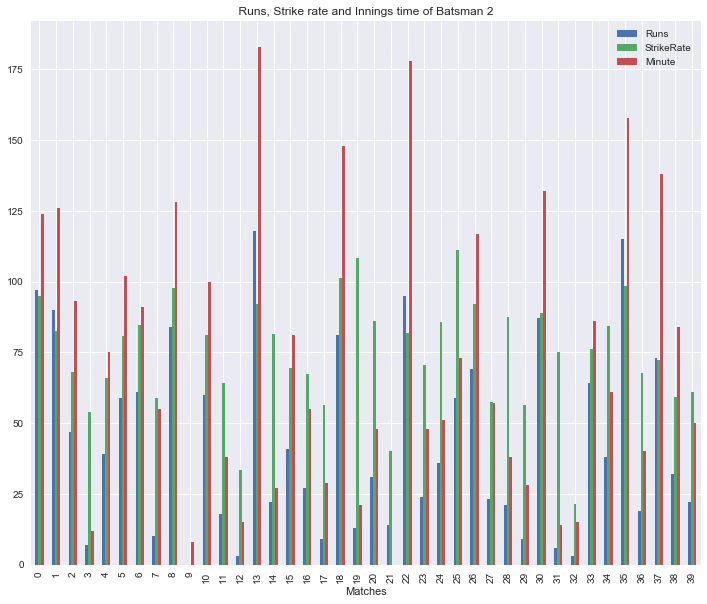
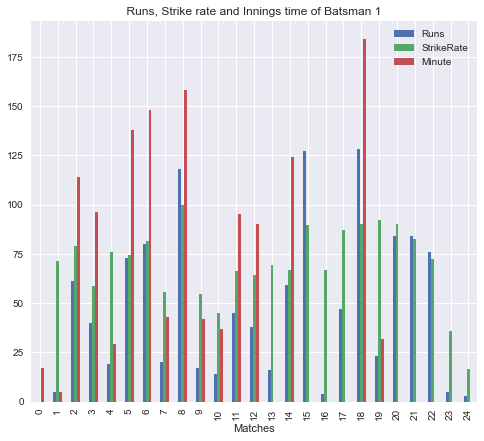
From this scenario it is clear that **batsman 5** and **batsman 3** scored more runs with hitting boundaries then rest of the batsman’s we are considering here.

**Batsman’s Runs vs Balls:  
**

From this bar plot we can conclude that batsman 3 is more **aggressive** than the other batsman’s. It also shows that batsman 5 is more consistent with his runs compare to the balls he played.

**Batsman’s Run/Innings/Time:**

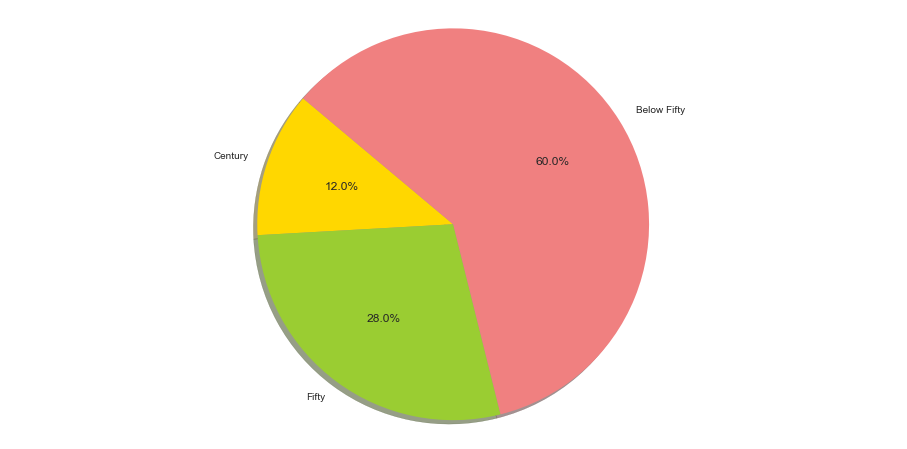
These bar plots shows the comparison of runs of each batsman’s with their strike rate and the length (in minute) of their innings.

****

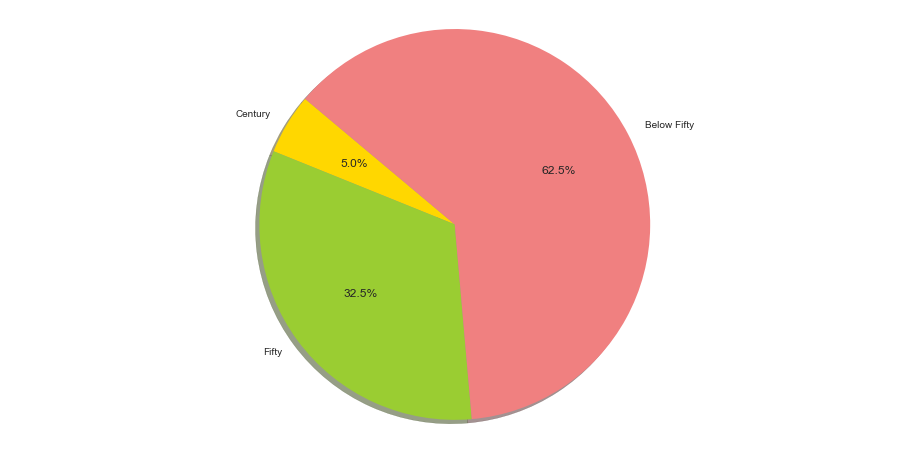
**Batsman’s runs comparison with centuries and half-centuries:**

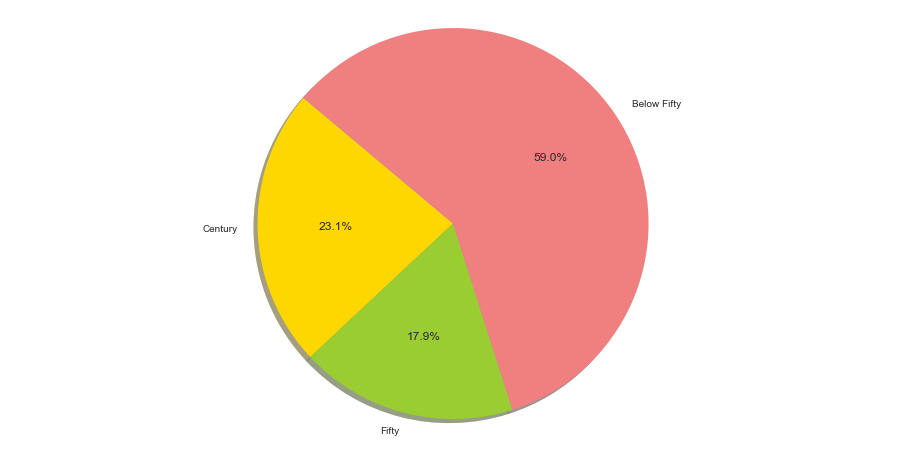
Those pie chart shows the run comparison of those batsman’s with centuries and half centuries

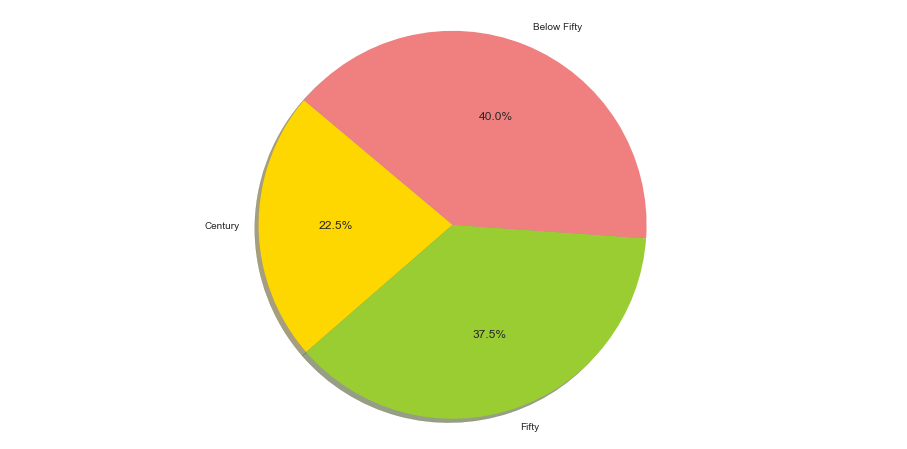
**Batsman 1:**

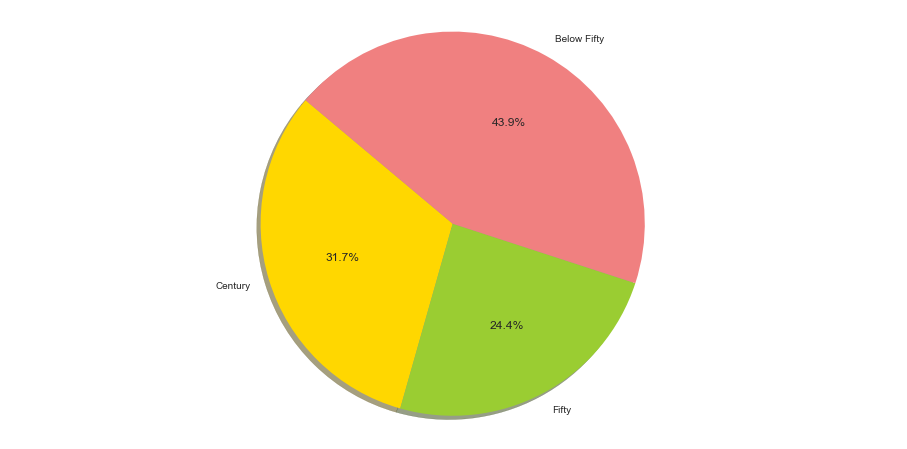
****

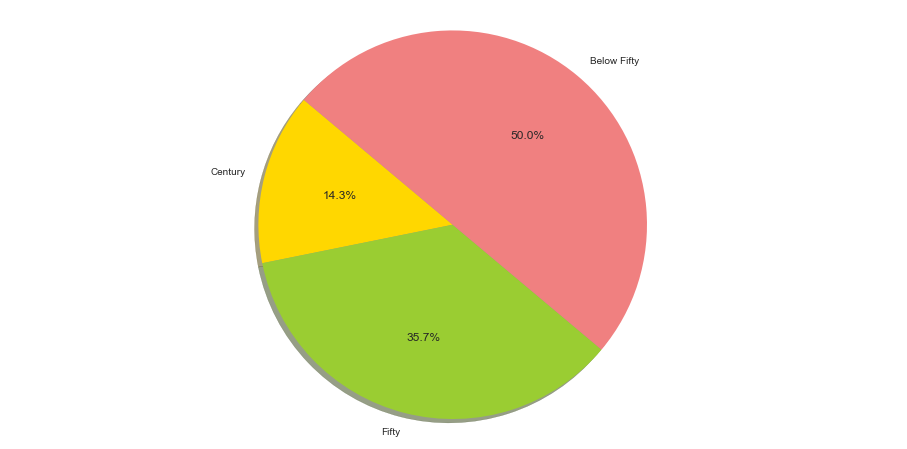
**Batsman 2:**

****

**Batsman 3:**

**Batsman 4:**

**Batsman 5:**

**Batsman 6:**

**Prediction:**

By analyzing our sample data we found the population average of runs of the batsman which is their career average. We used both normal distribution and t distribution to predict career average. From the conﬁdence interval found with 95% conﬁdence level we can predict the range of the average that 95% of the time will fall in the conﬁdence interval.

|  |  |  |
| --- | --- | --- |
| **Batsmen** | **Confidence Interval** | **Comment** |
| 1 | [63.12,31.76] | 95% of the time Batsman1 average will be between 63.12 to 31.76 |
| **2** | [53.65,32.65] | 95% of the time Batsman2 average will be between 63.65 to 30.94 |
| **3** | [70.28,36.38] | 95% of the time Batsman3 average will be between 70.28 to 36.38 |

**Table: Prediction of Batsman Career Average with Z distribution**

|  |  |  |
| --- | --- | --- |
| **Batsmen** | **Confidence Interval** | **Comment** |
| 1 | [**63.65**,30.94] | 95% of the time Batsman1 average will be between 63.65 to 30.94 |
| 2 | [54.00,32.31] | 95% of the time Batsman2 average will be between 54.00 to 32.31 |
| 3 | [70.85,35.82] | 95% of the time Batsman3 average will be between 70.85 to 35.82 |

**Table: Prediction of Batsman Career Average with T distribution**

**Player Identifications:**

**Player 1:** First batsman of this dataset is **Tamim Iqbal**. According to the stats his average runs and strike rate closely matches with recent performance by him. Also his last three ODI top scores (118,127,128) suggests that there’s a possibility **Tamim Iqbal** to be the batsman 1 on dataset. Also aggressively coming down the track and getting stumped frequently another of his nature reflected in the dataset.

**Player 2:** Second batsman is **Kane Williamson**. He is known for aggressive batting and higher strike rate.

**Player 3:** The third player of this dataset is **David warner**. Due to the statistics higher average batting, strike rate and aggressive batting.

**Player 4:** Forth batsman of this dataset is **Joe Root**. Due to the statistics higher average batting and strike rete

**Player 5:** Our fifth player of the dataset is **Virat Kohli**. Due to the statistics higher average batting, strike rate and maximum time no dismissals.

**Player 6**:The last player of this dataset is **Steven Smith** .He has a tendency of losing wicket by LBW.



**Report on data analysis using python notebook and a dataset of six ODI Batsman’s.**

Department of Computer Science and Engineering

**Course Code:** CSE 464

**Course Title:** Advance Database System

**Submitted by:**

Hossain Abedy

ID: 2014-1-60-080

Jafrul Hossain

ID: 2013-3-68-037

Nazmul Hasasn

ID: 2014-2-60-063

**Submitted to:**

Dr. Mohammad Rezwanul Huq

Assistant Professor

Department of CSE

East West University

**Date:** 27th march 2018