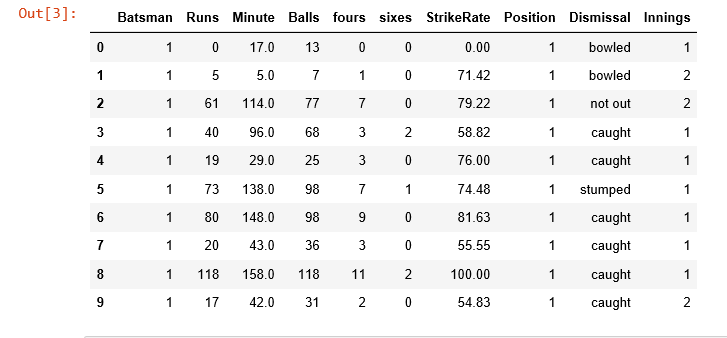
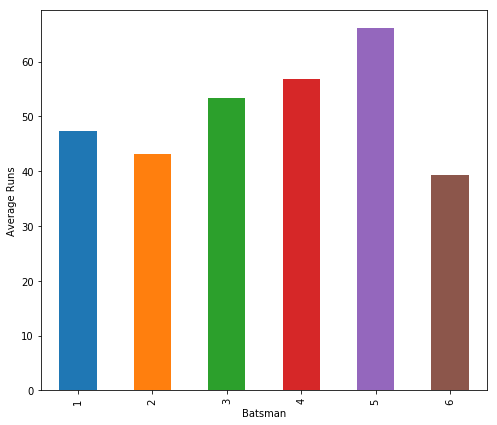
**Introduction:** This report is about data analysis using python notebook and a dataset consist of six batsmen odi career. From this dataset we compare performance of the players, we also given our own observation about the players then we identified players based on their average score and performance. 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,str\_rate,position,dismissal and innings

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

Batsman 1 47.440000

Batsman 2 43.150000

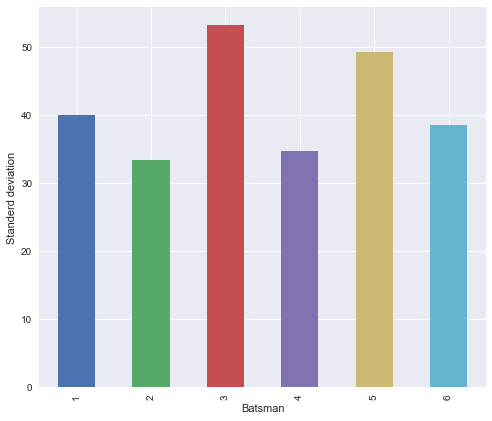
Batsman 3 53.333333

Batsman 4 56.914286

Batsman 5 66.146341

Batsman 6 39.386364

**Standard Deviation :**

****

Batsman

1 39.987581

2 33.456861

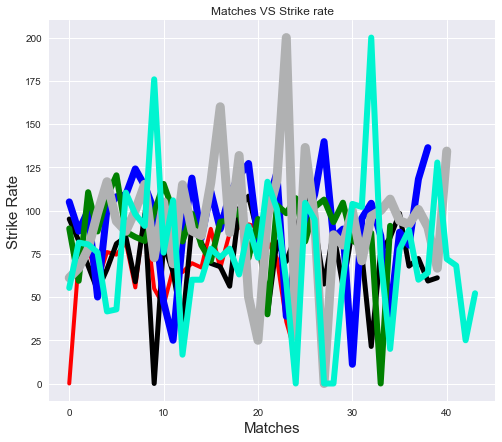
3 53.300209

4 34.748145

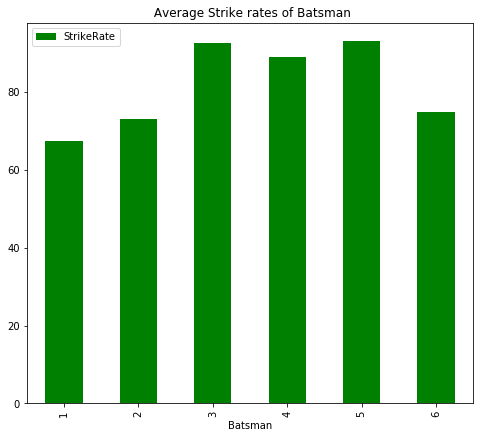
5 49.296836

6 38.566895

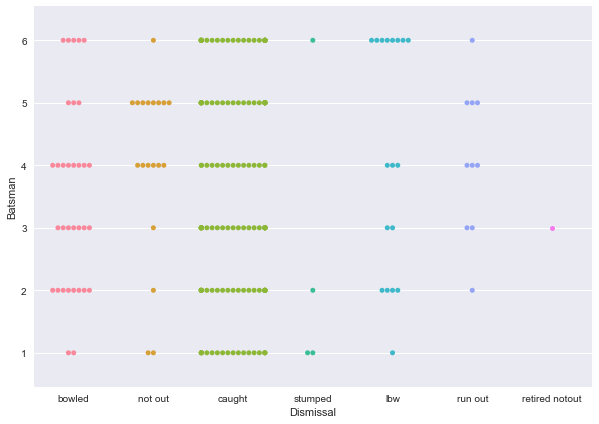
**Career Strike Rate:**

****

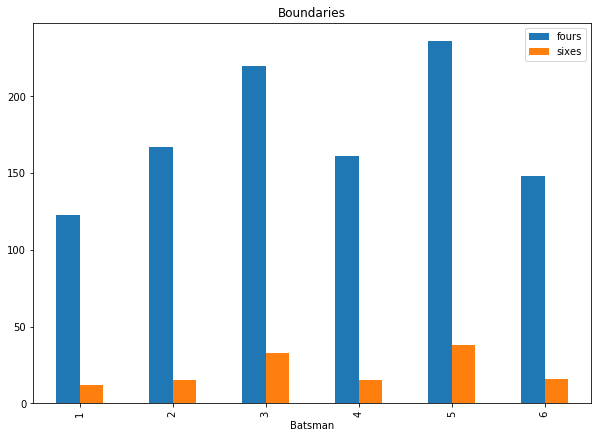
This graph shows the players individuals strike rate per match.

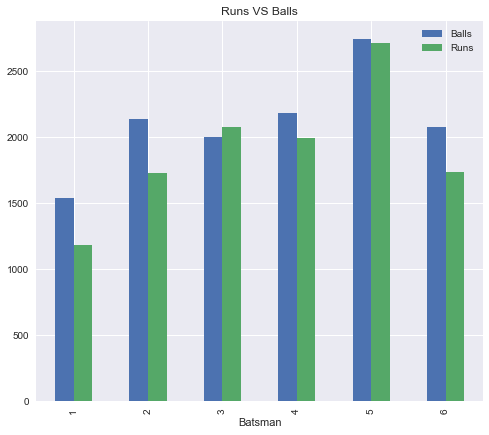
**Average Strike Rate:  
**

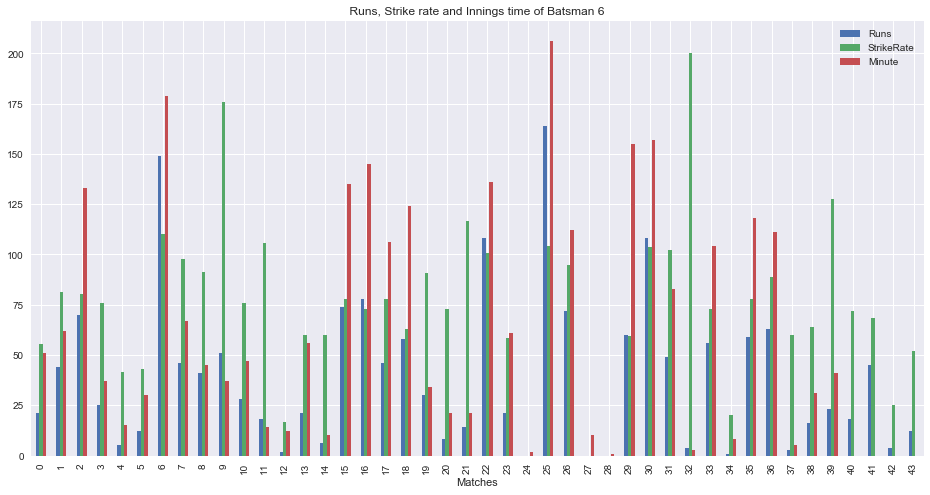
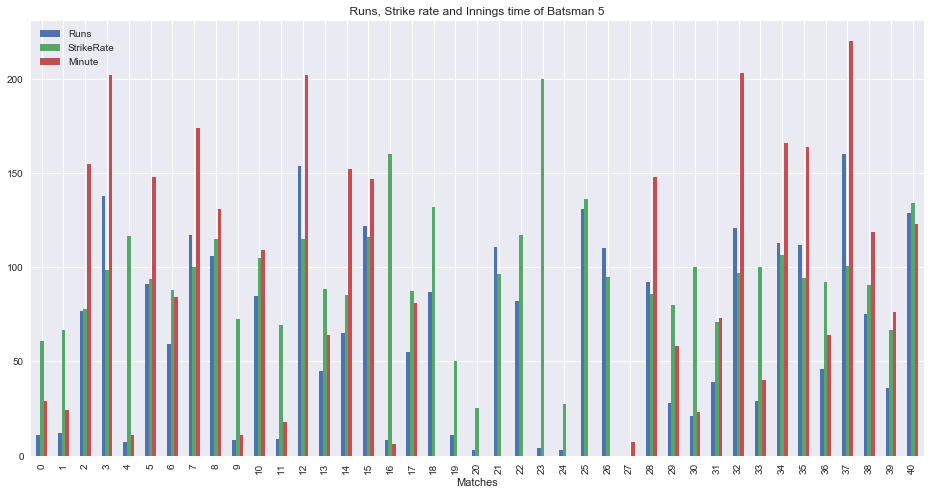
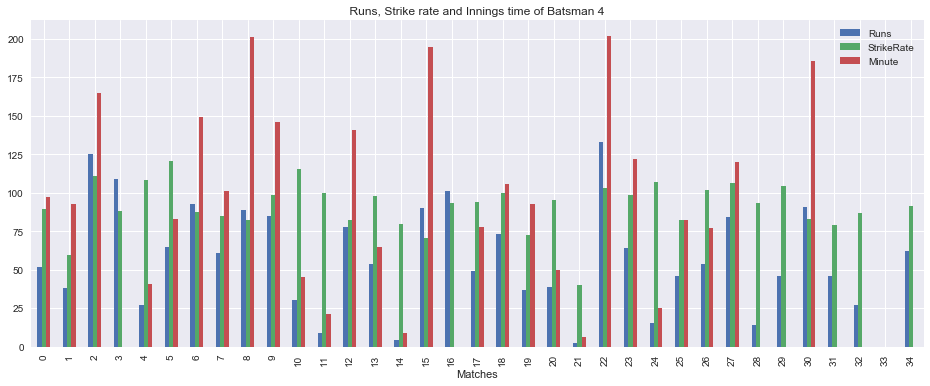
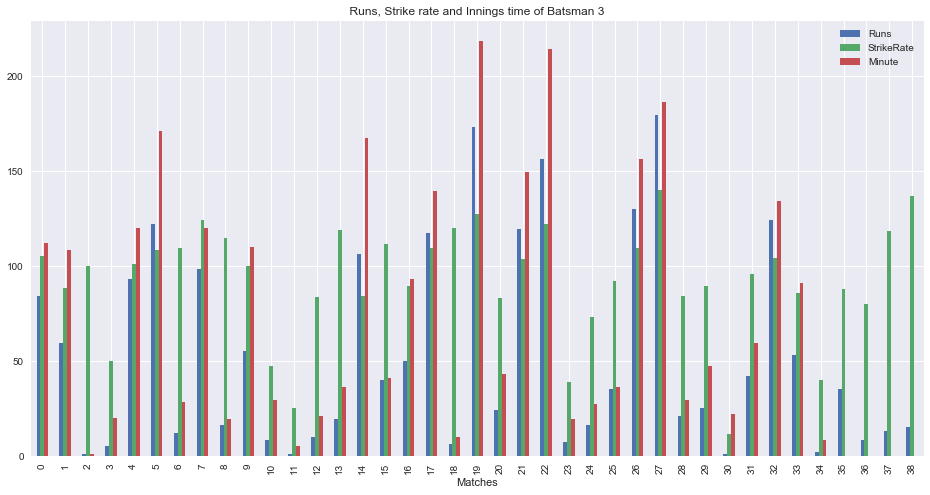
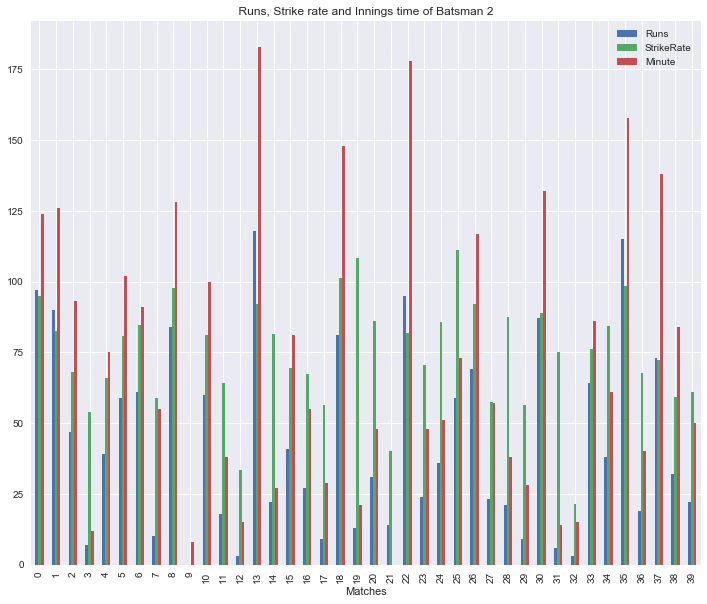
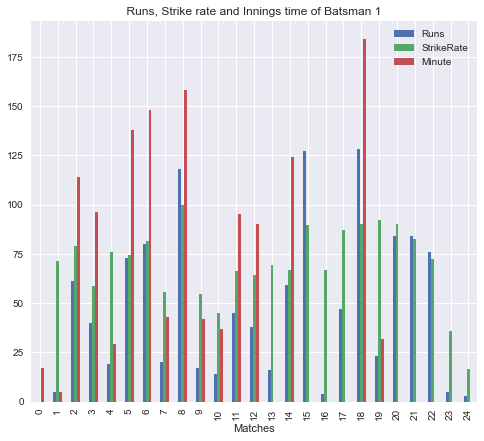
**Dismissal:**

****

**Batsmen Fours/Sixes:**

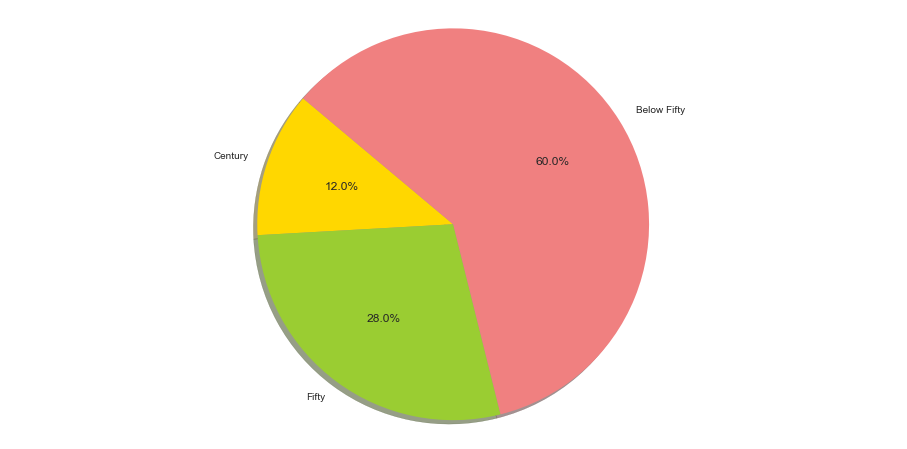
****

**Batsmen Runs vs Balls:  
**

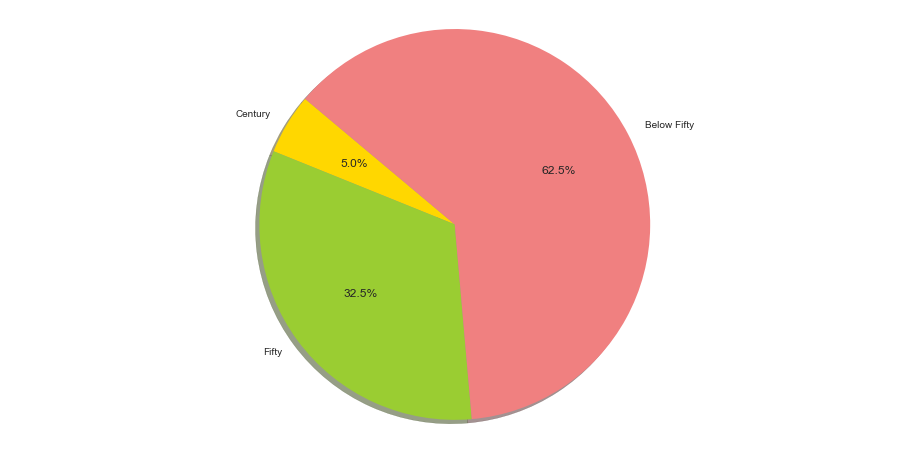
**Batsmen Run/Innings/Time :  
**

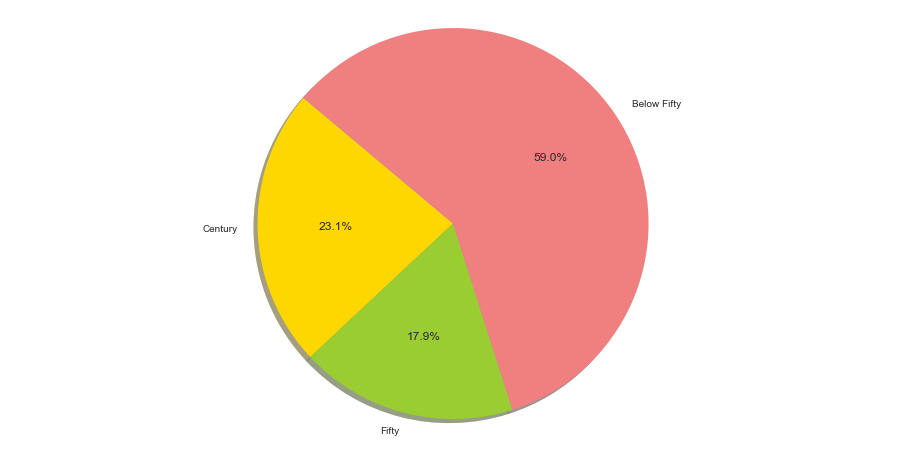
**Batsmen Hundreds/Fifty/Below fifty:** Pie Chart

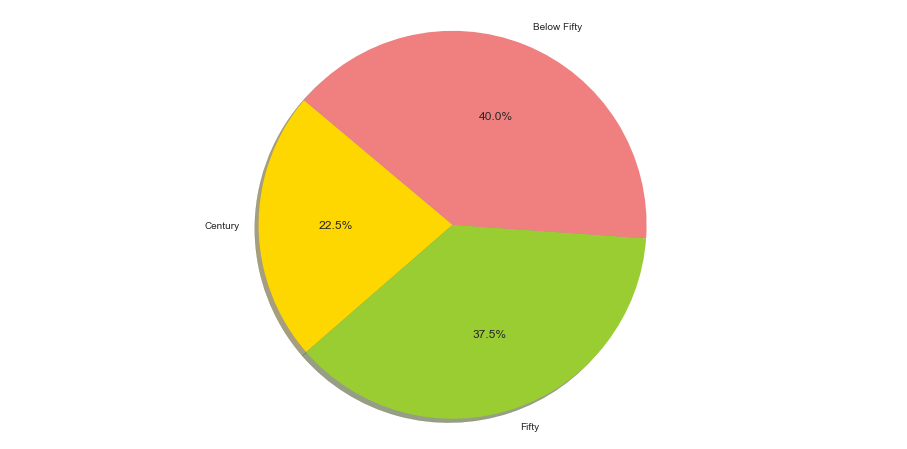
**Batsman 1:**

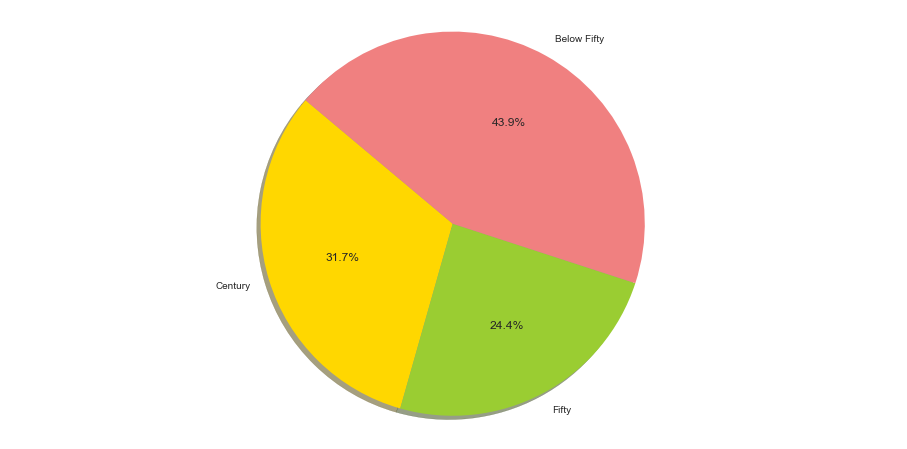
****

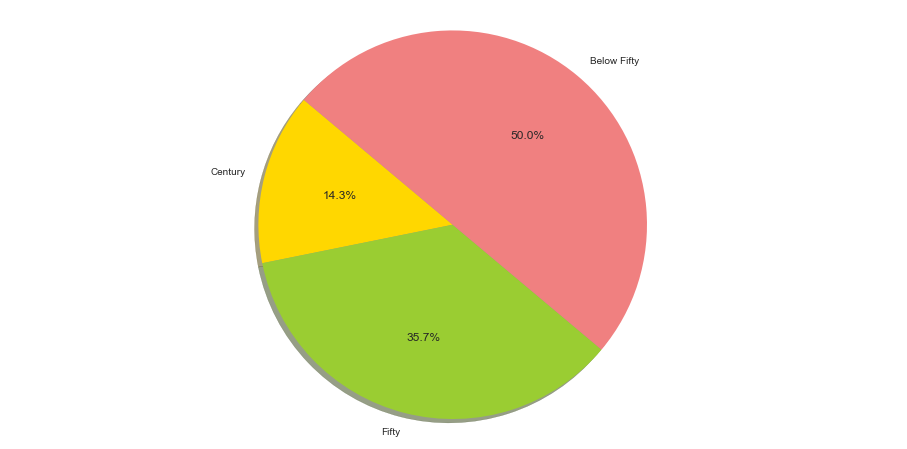
**Batsman 2:**

****

**Batsman 3:**

**Batsman 4:**

**Batsman 5:**

**Batsman 6:**

**Batsman 5:**

**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 tons (118,127,128) suggests that there’s a possibility Tamim Iqbal to be the batsma 1 on dataset.Aslo 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 stats higher average batting+strike rate.

Player 4:Forth batsman of this dataset is Joe Root. Due to the stats higher average batting+strike  
 Player 5: Our fifth player of the dataset is Virat Kohly.Due to the stats higher average batting+strike rate and maximum time no dismissals.

Player 6:The last player of this dataset is Steven Smith.He has a high average but he has a tendency of losing wicket by lbw.