CSE303 (Section 1)

[Spring 2022]

Lab Assignment Submission Report

Assignment Title: LAB 04

Submitted by:

Rizvee Hassan Prito

2019-3-60-041

1. Screenshots

Datasets:

Three different distributions correspond to the runs scored in each innings to the IPL by an opening batter, middle-order batter and a lower-order batter. Here, opening batter is Faf Du Plessis, middle-order batter is MS Dhoni, lower-order batter is Ravi Ashwin.

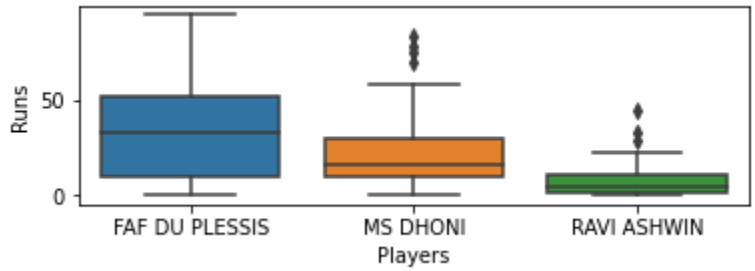
```
In [2]: runfile('D:/Courses/CSE 303/LAB 4/2019-3-60-6
wdir='D:/Courses/CSE 303/LAB 4')

    FAF DU PLESSIS  MS DHONI  RAVI ASHWIN
0                5        16            8
1               88        50            0
2               86         0            2
3                1        18           11
4               76        12            5
5               10        18            6
6               25         0            6
7               41        14            0
8               43         1            2
9               31        11            3
10               0        17            8
11              36         2            7
12              33         3            0
13              95        11            0
14              50         1            3
15              56        14           11
16              50        18           11
17               0        12            2
18              31        18            9
19              43        19            1
20              41         1            9
21              25        16           11
22              10        28            0
23              76         3           14
24               1        21           10
25              86        10           12
26              58        11            8
27              72        47            9
28              43        15           23
29              22         8            1
30              87        32            2
31              17        75            1
32               8        12            0
33               0        37           29
34              58        58            0
35              10        16            1
36               1        84           23
37              25        44            0
38              48        10            1
39              54        37           33
40              43         9           14
41               7         2            6
42              24         5            4
43              45        25            0
44               5        79            0
45               1         5           45
46              39        25            0
47              26        70            0
48              10        26            3
49              67        51           17
50              14        43           16
```

Output:

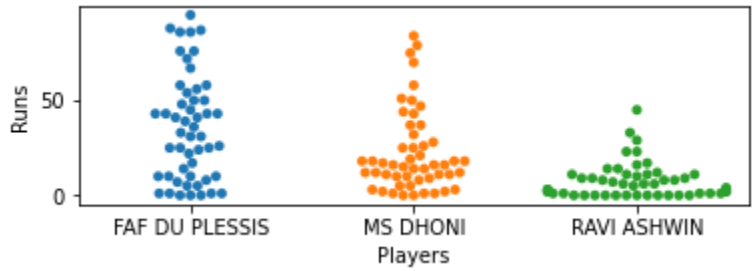
```
                sum_sq     df         F        PR(>F)
Players    20256.169935    2.0  22.90326  2.088374e-09
Residual   66331.725490  150.0       NaN           NaN
```

Here, P-value is 2.088374e-09 which is lower than the 0.05. So, we can reject the null hypothesis and the means of my three datasets are significantly different. Therefore, these three datasets have statistically significant variation.

Boxplot and Swarmplot:



From the boxplot we can also see that means of the three datasets are significantly different.



From the swarmplot we can see the significant variation of those three datasets.

2. Learning outcomes:

From this lab, I have learned how to impellent one-way Anova testing in Python. I have used ols() function of statsmodels package which built the model and fit it . Then I used anova_lm() function for doing the one-way Anova testing. So, by using Python, we can easily find if the means of our datasets are significantly different or not.