OPEN SOURCE SOFTWARE LAB (15B17CI575)

PROJECT TITLE:

CRICKET PLAYER PERFORMANCE ANALYSIS

Submitted by:

VREETI AGGARWAL

VAISHALI RANJAN

SIDDHANT SINGH

Under the Guidance of:

DR. ANUBHUTI RODA MOHINDRA

DR. CHETNA GUPTA

ANURADHA GUPTA

JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY, NOIDA (U.P.)

INTRODUCTION

Analyzing a player's performance is one of the use cases of Data Science in sports analytics. It would be a great data science project if we analyze the batting performance of a player over the years. The data can be analyzed and provide beautiful insights. We can see total runs scored, average of batsman during the same period, trends of runs scored, number of matches at different batting positions, most runs against teams and many more.

TOOLS AND TECHNOLOGIES USED

Tools and technologies used:

- Language:
  - Python
- Integrated Development Environment:
  - Spyder IDE
- Algorithms used:
  - Pandas
  - Numpy
  - Plotly Python Library

DATASET

Below is the complete information about all the columns in the dataset:

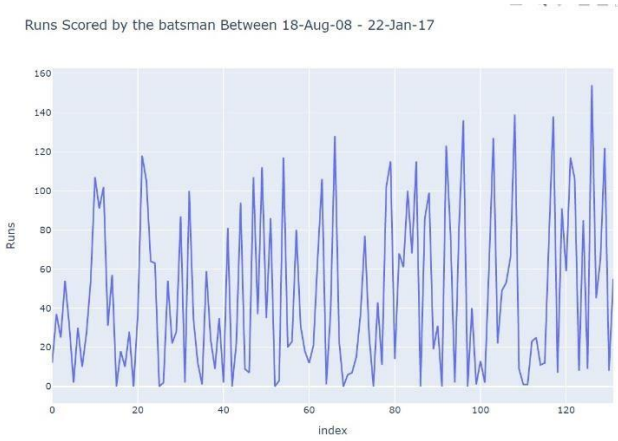
1. Runs: Runs in the match
2. BF: Balls faced in the match
3. 4s: number of 4s in a match
4. 6s: number of 6s in a match
5. SR: Strike Rate in the match
6. Pos: Batting Position in the match
7. Dismissal: How Virat Kohli got out in the match
8. Inns: 1st and 2nd innings
9. Opposition: Who was the opponent of India
10. Ground: Venue of the match
11. Start Date: Date of the match

```
In [46]: runfile('C:/Users/91730/.spyder-py3/pythonProject/SportsAnalysisP
91730/.spyder-py3/pythonProject')
   Runs  BF  4s  6s  ...  Inns   Opposition         Ground  Start Date
0    12  22   1   0  ...     1  v Sri Lanka       Dambulla   18-Aug-08
1    37  67   6   0  ...     2  v Sri Lanka       Dambulla   20-Aug-08
2    25  38   4   0  ...     1  v Sri Lanka  Colombo (RPS)   24-Aug-08
3    54  66   7   0  ...     1  v Sri Lanka  Colombo (RPS)   27-Aug-08
4    31  46   3   1  ...     2  v Sri Lanka  Colombo (RPS)   29-Aug-08
```

Fig. CSV File

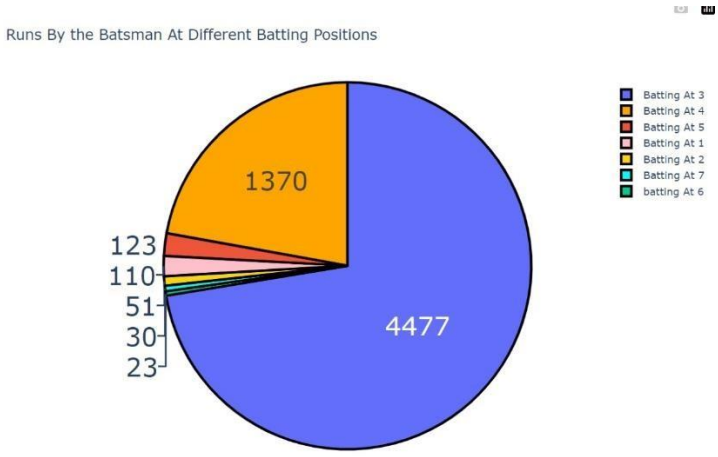
RESULTS AND ANALYSIS

So, as an example we choose Virat Kohli. The trend of runs scored by Virat Kohli in his career from 18 August 2008 to 22 January 2017.



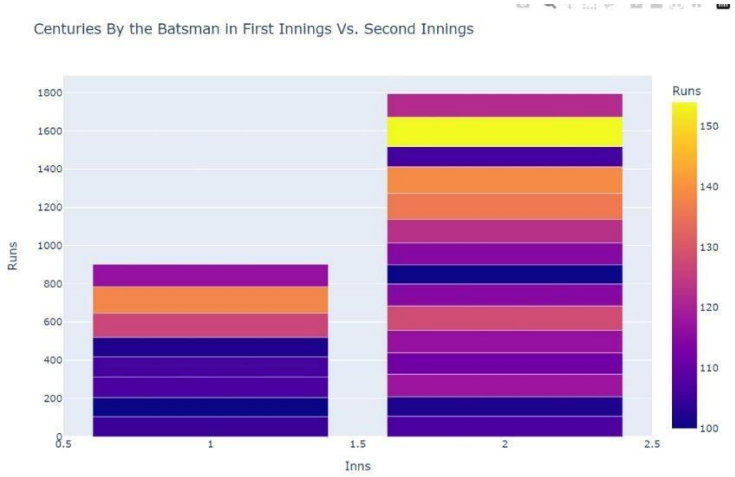
In so many innings played by Virat Kohli, he scored over 100 or close to it. That is a good sign of consistency.

All the batting positions played by Virat Kohli.



More than 72% of the total runs scored by Virat Kohli are while batting at 3rd position. So we can say batting at 3rd position is perfect for Virat Kohli.

The number of centuries scored by Virat Kohli while batting in the first innings and second innings.

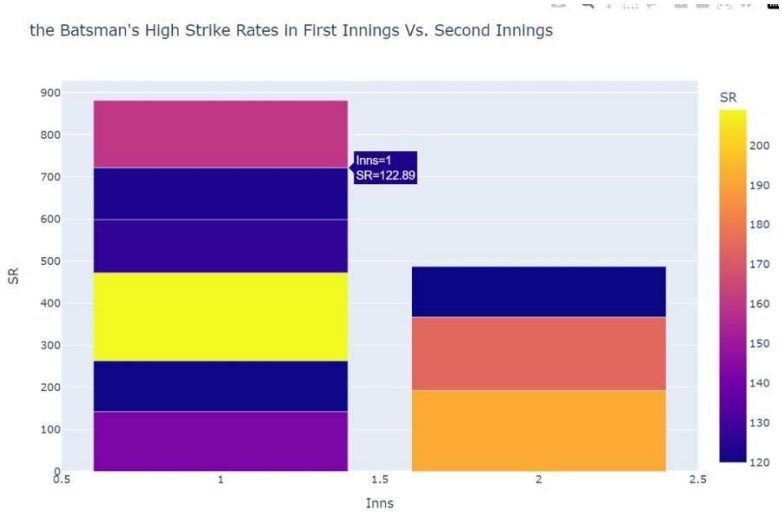


So, most of the centuries are scored while batting in the second innings. By this, we can say that Virat Kohli likes chasing scores.

Now, we check against which team Virat Kohli scored most of his runs.

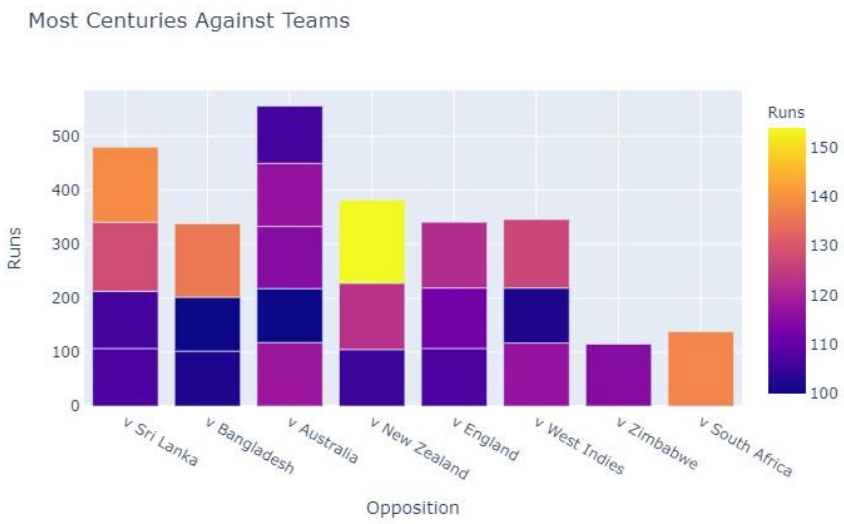
According to the above figure, Virat Kohli likes batting against Sri Lanka, Australia, New Zealand, West Indies, and England. But he scored most of his runs while batting against Sri Lanka.

Virat Kohli plays with high strike rates in the first innings or second innings.



So according to the above figure, Virat Kohli likes playing more aggressively in the first innings compared to the second innings.

Against which team Virat Kohli scored most of his centuries.

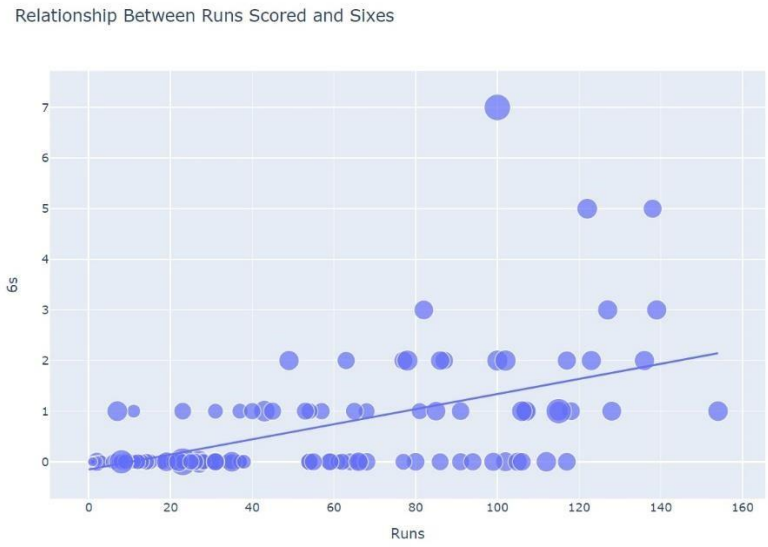


So, most of the centuries scored by Virat Kohli were against Australia.

Now, lets analyze relationship between runs scored by Virat Kohli and fours played by him in each innings.

There is a linear relationship. It means that Virat Kohli likes playing fours.

Relationship with the sixes:



There is no strong linear relationship here. It means Virat Kohli likes playing fours more than sixes.

Console Output:

```
In [46]: runfile('C:/Users/91730/.spyder-py3/pythonProject/SportsAnalysisProject.py', wdir='C:/Users/
91730/.spyder-py3/pythonProject')
   Runs  BF  4s  6s  ...  Inns   Opposition         Ground  Start Date
0    12  22   1   0  ...     1  v Sri Lanka       Dambulla   18-Aug-08
1    37  67   6   0  ...     2  v Sri Lanka       Dambulla   20-Aug-08
2    25  38   4   0  ...     1  v Sri Lanka  Colombo (RPS)   24-Aug-08
3    54  66   7   0  ...     1  v Sri Lanka  Colombo (RPS)   27-Aug-08
4    31  46   3   1  ...     2  v Sri Lanka  Colombo (RPS)   29-Aug-08

[5 rows x 11 columns]
Runs          0
BF            0
4s            0
6s            0
SR            0
Pos           0
Dismissal     0
Inns          0
Opposition    0
Ground        0
Start Date    0
dtype: int64
6184
46.84848484848485
     Runs  BF  4s  6s  ...  Inns     Opposition         Ground  Start Date
8      27  19   4   0  ...     1    v Sri Lanka         Rajkot   15-Dec-09
32    100  83   8   2  ...     1   v Bangladesh          Dhaka   19-Feb-11
56     23  11   3   0  ...     1  v West Indies         Indore    8-Dec-11
76     43  34   4   1  ...     1      v England     Birmingham   23-Jun-13
78    102  83  13   2  ...     1  v West Indies  Port of Spain    5-Jul-13
83    100  52   8   7  ...     2    v Australia         Jaipur   16-Oct-13
85    115  66  18   1  ...     2    v Australia         Nagpur   30-Oct-13
93     78  65   7   2  ...     2  v New Zealand       Hamilton   22-Jan-14
130     8   5   2   0  ...     1      v England        Cuttack   19-Jan-17

[9 rows x 11 columns]
```

So, this is how we can analyze the performance of Virat Kohli or any other cricket batsman in the world.

IMPLEMENTATION

```
import pandas as pd import
plotly.express as px import
plotly.graph_objects as go
```

```
data = pd.read_csv("Virat_Kohli.csv")
print(data.head())

#to check if we have null values
print(data.isnull().sum())

#to get Total runs print(data["Runs"].sum())

#to get Average Runs  print(data["Runs"].mean())


""" TO GET THE RUNS SCORED """
matches = data.index
fig = px.line(data, x=matches, y="Runs",
title='Runs Scored by Virat Kohli Between 18-Aug-08 -
22-Jan-17')
fig.write_html('first_figure.html', auto_open=True)



""" ALL THE MATCHED AT DIFFERENT BATTING POSITIONS
"""

data["Pos"] = data["Pos"].map({3.0: "Batting At 3",
4.0: "Batting At 4", 2.0: "Batting At 2",
                               1.0: "Batting At 1",
7.0:"Batting At 7", 5.0:"Batting At 5",
                               6.0: "batting At 6"})

Pos = data["Pos"].value_counts()
label = Pos.index counts =
Pos.values
colors = ['gold','lightgreen', "pink", "blue",
"skyblue", "cyan", "orange"]

fig1 = go.Figure(data=[go.Pie(labels=label,
values=counts)])
fig1.update_layout(title_text='Number of Matches At
Different Batting Positions')
fig1.update_traces(hoverinfo='label+percent',
textinfo='value', textfont_size=30,
marker=dict(colors=colors, line=dict(color='black',
width=3)))
```

```
fig1.write_html('second_figure.html', auto_open=True)



""" TOTAL RUNS BY VIRAT KOHLI """
label = data["Pos"] counts =
data["Runs"]
colors = ['gold','lightgreen', "pink", "blue",
"skyblue", "cyan", "orange"]

fig2 = go.Figure(data=[go.Pie(labels=label,
values=counts)])
fig2.update_layout(title_text='Runs By Virat Kohli At
Different Batting Positions')
fig2.update_traces(hoverinfo='label+percent',
textinfo='value', textfont_size=30,
marker=dict(colors=colors, line=dict(color='black',
width=3)))
fig2.write_html('third_figure.html', auto_open=True)



""" NUMBER OF CETURIES IN FIRST INNINGS """
                          centuries =
data.query("Runs >= 100") fig3 =
px.bar(centuries, x=centuries["Inns"], y =
centuries["Runs"],
                color = centuries["Runs"],
title="Centuries By Virat Kohli in First Innings Vs.
Second Innings")
fig3.write_html('fourth_figure.html', auto_open=True)



""" SCORES AGAINST TEAMS """

fig4 = px.bar(data, x=data["Opposition"], y =
data["Runs"], color = data["Runs"],
title="Most Runs Against Teams")
fig4.write_html('fifth_figure.html', auto_open=True)
```

```
#dataset for all the matches played by Vurat Kohli
where strike rate is more than 120 strike_rate =
data.query("SR >= 120") print(strike_rate)

fig5 = px.bar(strike_rate, x = strike_rate["Inns"],
y = strike_rate["SR"],                  color =
strike_rate["SR"],
            title="Virat Kohli's High Strike Rates in
First Innings Vs. Second Innings")
fig5.write_html('sixth_figure.html', auto_open=True)



""" RUNS SCORED AND FOURS PLAYED BY HIM IN EACH
INNINGS """
fig6 = px.scatter(data_frame = data, x="Runs",
y="4s", size="SR", trendline="ols",
                    title="Relationship Between Runs
Scored and Fours")
fig6.write_html('seventh_figure.html',
auto_open=True)



""" RELATIONSHIP BETWEEN RUNS SCORED AND SIXES """

fig7 = px.scatter(data_frame = data, x="Runs",
y="6s", size="SR", trendline="ols",
                    title= "Relationship Between Runs
Scored and Sixes")
fig7.write_html('eighth_figure.html', auto_open=True)
```

CONCLUSION

So, this is how you can perform a player's performance analysis using the Python programming language. Analyzing a player's performance is one of the use cases of Data Science in sports analytics.

REFERENCES

[1] plotly.graph_objects.Figure

[2] Matplotlib-vs-plotly-js

[3] Pandas Tutorial

[4] Numpy Tutorial

[5] Python Plotly Tutorial- GeeksforGeeks

[6] Plotly Tutorial 2021