⌄ Project Title : Virat Kohli Performance Analysis using Python

---

Abstract of the project:

The project "Virat Kohli Performance Analysis using Python" aims to analyze the batting performance of the acclaimed Indian cricketer Virat Kohli in One Day Internationals (ODIs). Using Python and popular data analysis libraries, the project explores key aspects such as runs scored, balls faced, boundaries hit, strike rate, batting positions, dismissal types, opposition teams, and playing grounds. The methodology involves data cleaning, preprocessing, and visualization using Python libraries such as pandas, matplotlib, and Plotly. The project provides insightful visualizations and an in-depth analysis of Virat Kohli's performance, making it valuable for those interested in sports analytics and data science involving cricket statistics

Analyzing a player's performance is one of the use cases of Data Science in sports analytics.

Virat Kohli is one of the most famous cricketers in the world. So it will be a great data science project if we analyze the batting performance of Virat Kohli over the years. So if you want to learn how to analyze Virat Kohl's performance, this project is for you. In this project, I will take you through the task of Virat Kohli Performance Analysis using Python.

⌄ Case Study

Here you are given a dataset of all the ODI matches played by Virat Kohli from 18 August 2008 to 22 January 2017. You are required to analyze the performance of Virat Kohli in ODI matches.

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

- Runs: Runs in the match
- BF: Balls faced in the match
- 4s: number of 4s in a match
- 6s: number of 6s in a match
- SR: Strike Rate in the match
- Dismissal: How Virat Kohli got out in the match
- Inns: 1st and 2nd innings
- Opposition: Who was the opponent of India
- Ground: Venue of the match
- Start Date: Date of the match

Now let's start with the task of Virat Kohli performance analysis using Python. I will start this task by importing the necessary Python libraries and the dataset:

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

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

```
      Runs  BF  4s  6s     SR  Pos Dismissal  Inns   Opposition        Ground  \
  0     12  22   1   0  54.54  2.0       lbw     1  v Sri Lanka      Dambulla
  1     37  67   6   0  55.22  2.0    caught     2  v Sri Lanka      Dambulla
  2     25  38   4   0  65.78  1.0   run out     1  v Sri Lanka  Colombo (RPS)
  3     54  66   7   0  81.81  1.0    bowled     1  v Sri Lanka  Colombo (RPS)
  4     31  46   3   1  67.39  1.0       lbw     2  v Sri Lanka  Colombo (RPS)

      Start Date
  0    18-Aug-08
  1    20-Aug-08
  2    24-Aug-08
  3    27-Aug-08
  4    29-Aug-08
```

Let's have a look at whether this dataset contains any null values or not before moving forward:

```
print(data.isnull().sum())
```

```
  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
```

The dataset contains matches played by Virat Kohli between 18 August 2008 and 22 January 2017. So let's have a look at the total runs scored by Virat Kohli:

```
# Total Runs Between 18 Aug 08   22 Jan 17
```

```
# Total Runs Between 18-Aug-08 - 22-Jan-17
data["Runs"].sum()
```

    6184

Now let's have a look at the average of Virat Kohli during

```
# Average Runs Between 18-Aug-08 - 22-Jan-17
data["Runs"].mean()
```

    46.84848484848485

In ODIs, the batting average of 35-37 is considered a good average. So Virat Kohl's batting average is good. Now let's have a look at the trend of runs scored by Virat Kohli in his career from 18 August 2008 to 22 January 2017:

```
matches = data.index
figure = px.line(data, x=matches, y="Runs",
                 title='Runs Scored by Virat Kohli Between 18-Aug-08 - 22-Jan-17')
figure.show()
```

Runs Scored by Virat Kohli Between 18-Aug-08 - 22-Jan-17
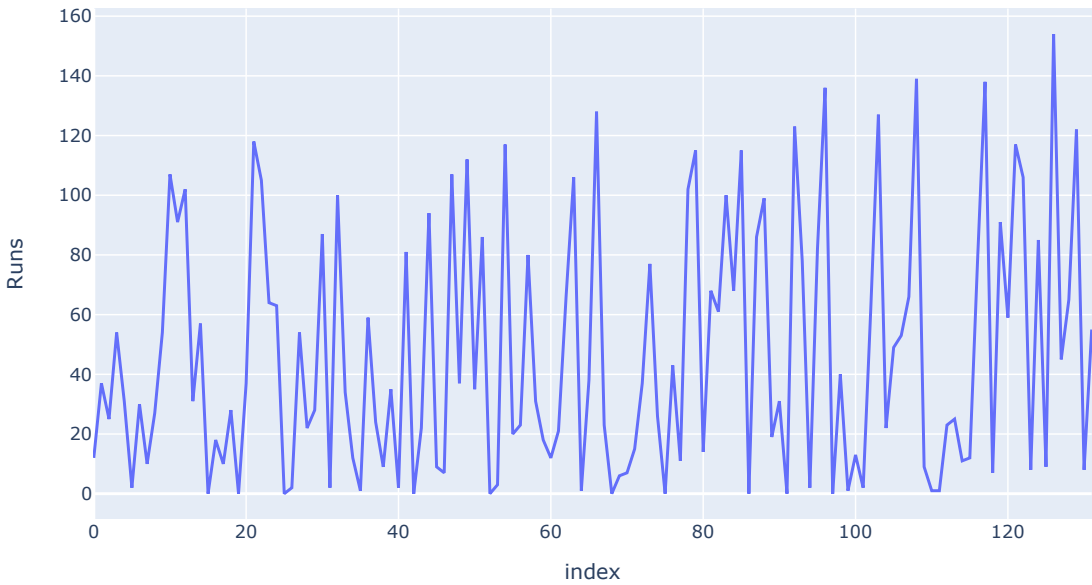
In so many innings played by Virat Kohli, he scored over 100 or close to it. That is a good sign of consistency. Now let's see all the batting positions played by Virat Kohli:
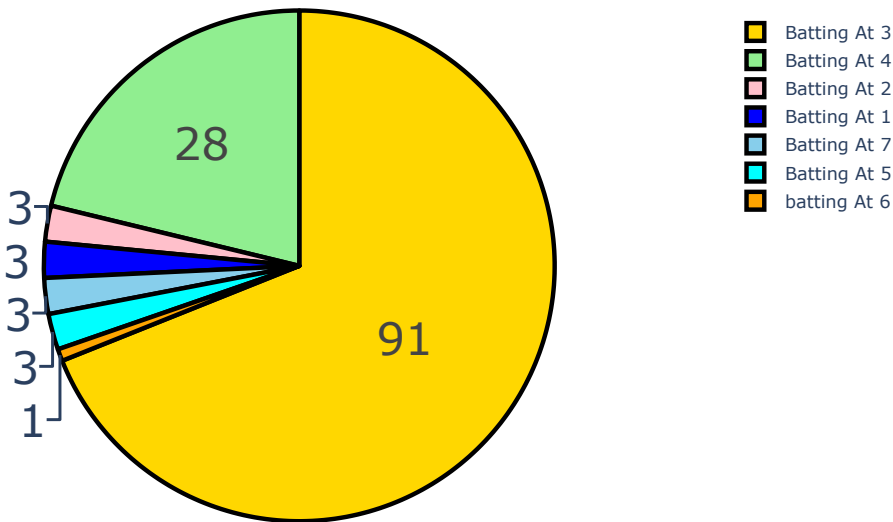
```
# Batting Positions
data["Pos"] = data["Pos"].map({3.0: "Batting At 3", 4.0: "Batting At 4", 2.0: "Batting At 2",
                               1.0: "Batting At 1", 7.0:"Batting At 7", 5.0:"Batting At 5",
                               6.0: "batting At 6"})

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

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

### Number of Matches At Different Batting Positions

In more than 68% of all the innings played by Virat Kohli, he batted in the third position. Now let's have a look at the total runs scored by Virat Kohli in different positions:

```
label = data["Pos"]
counts = data["Runs"]
colors = ['gold','lightgreen', "pink", "blue", "skyblue", "cyan", "orange"]

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

Runs By Virat Kohli At Different Batting Positions
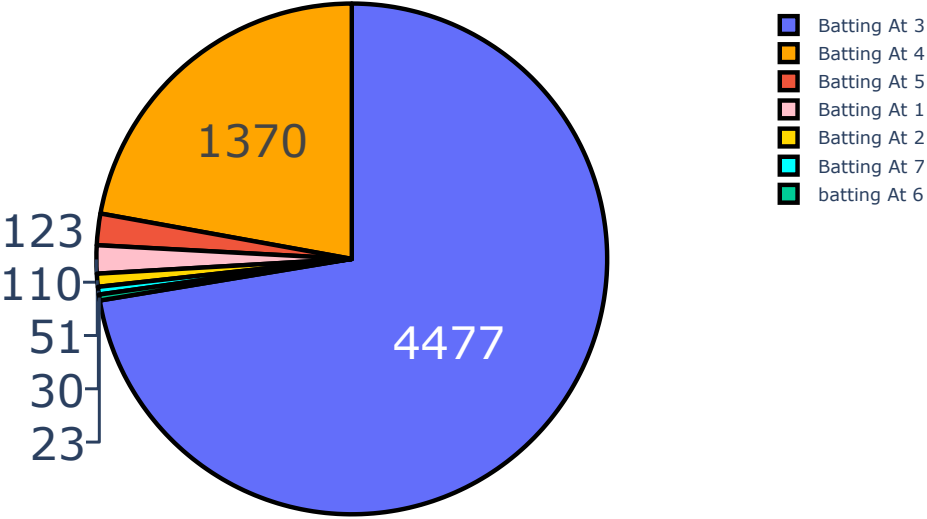
Taking snapshot - this ✕
may take a few
seconds

Snapshot succeeded - ✕
newplot.png

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.

Now let's have a look at the number of centuries scored by Virat Kohli while batting in the first innings and second innings:

```
centuries = data.query("Runs >= 100")
figure = px.bar(centuries, x=centuries["Inns"], y = centuries["Runs"],
                color = centuries["Runs"],
                title="Centuries By Virat Kohli in First Innings Vs. Second Innings")
figure.show()
```

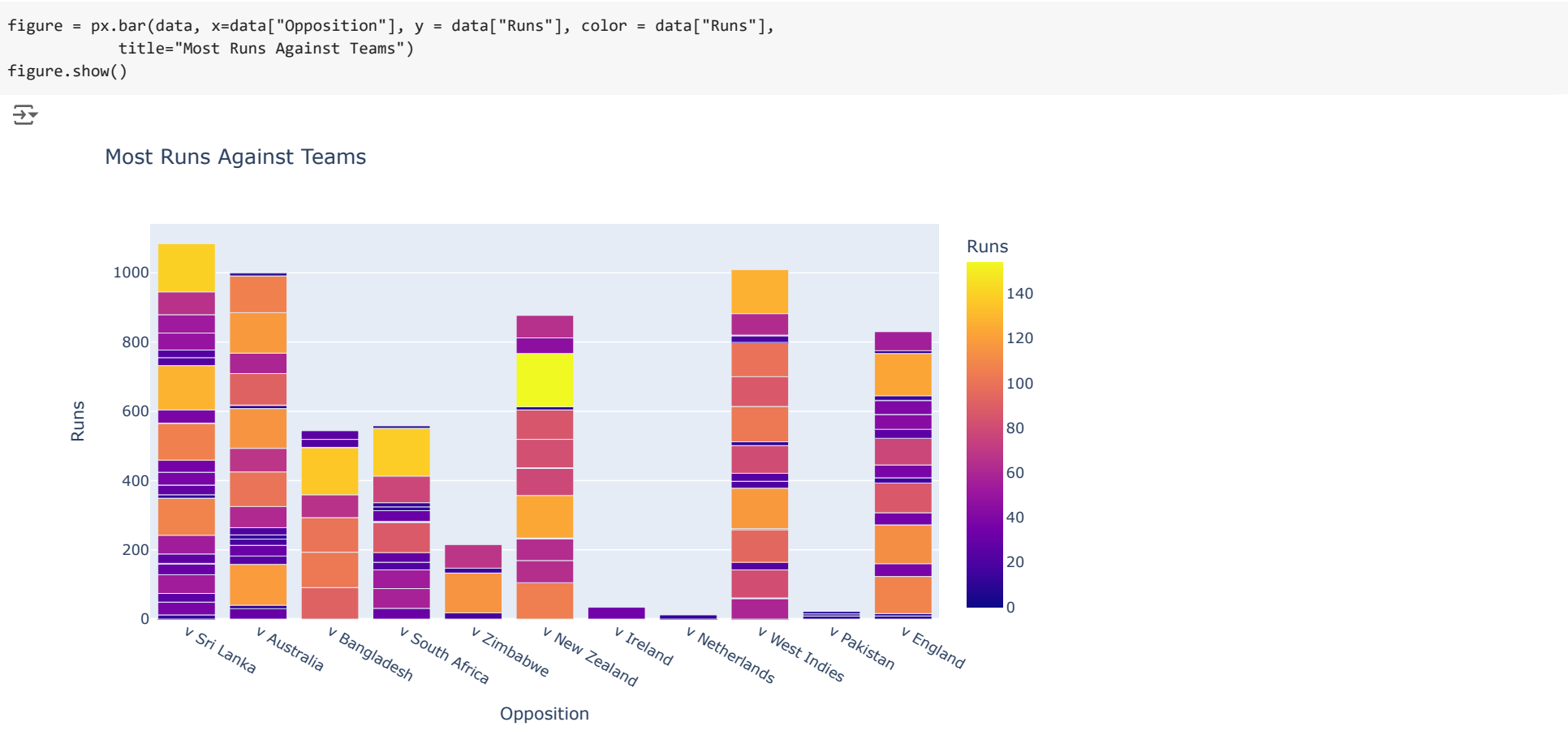### Centuries By Virat Kohli in First Innings Vs. Second Innings
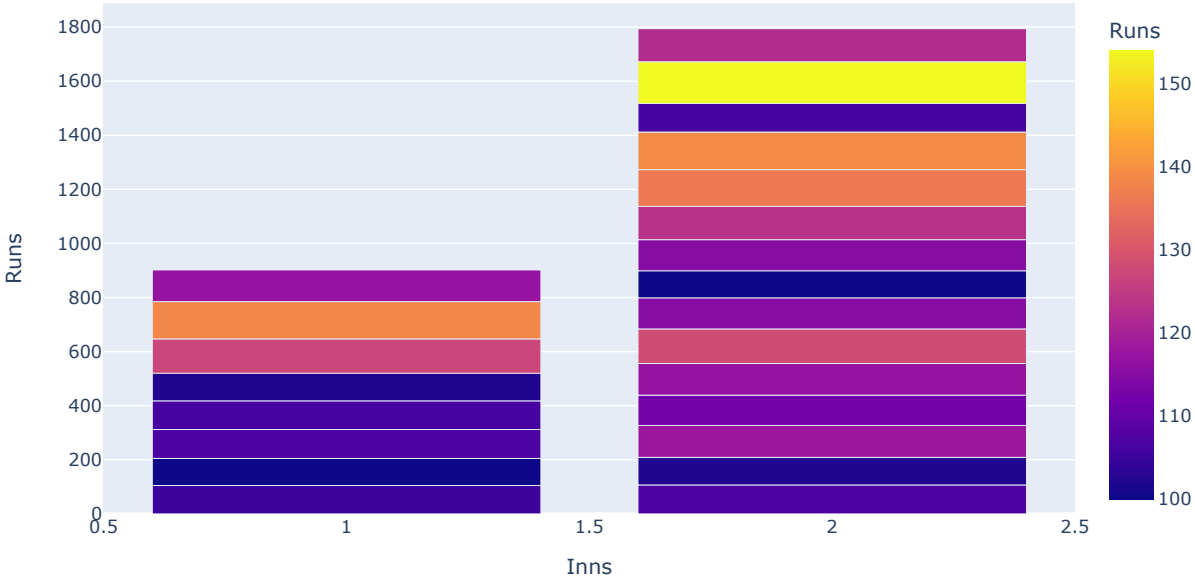
So most of the centuries are scored while batting in the second innings. By this, we can say that Virat Kohli likes chasing scores. Now let's have a look at the kind of dismissals Virat Kohli faced most of the time:

```
# Dismissals of Virat Kohli
dismissal = data["Dismissal"].value_counts()
label = dismissal.index
counts = dismissal.values
colors = ['gold','lightgreen', "pink", "blue", "skyblue", "cyan", "orange"]

fig = go.Figure(data=[go.Pie(labels=label, values=counts)])
fig.update_layout(title_text='Dismissals of Virat Kohli')
fig.update_traces(hoverinfo='label+percent', textinfo='value', textfont_size=30,
                  marker=dict(colors=colors, line=dict(color='black', width=3)))
fig.show()
```
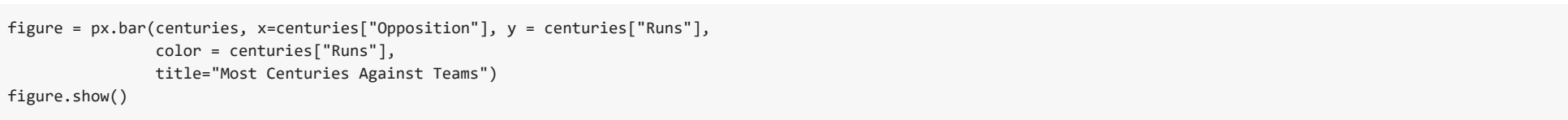
Dismissals of Virat Kohli
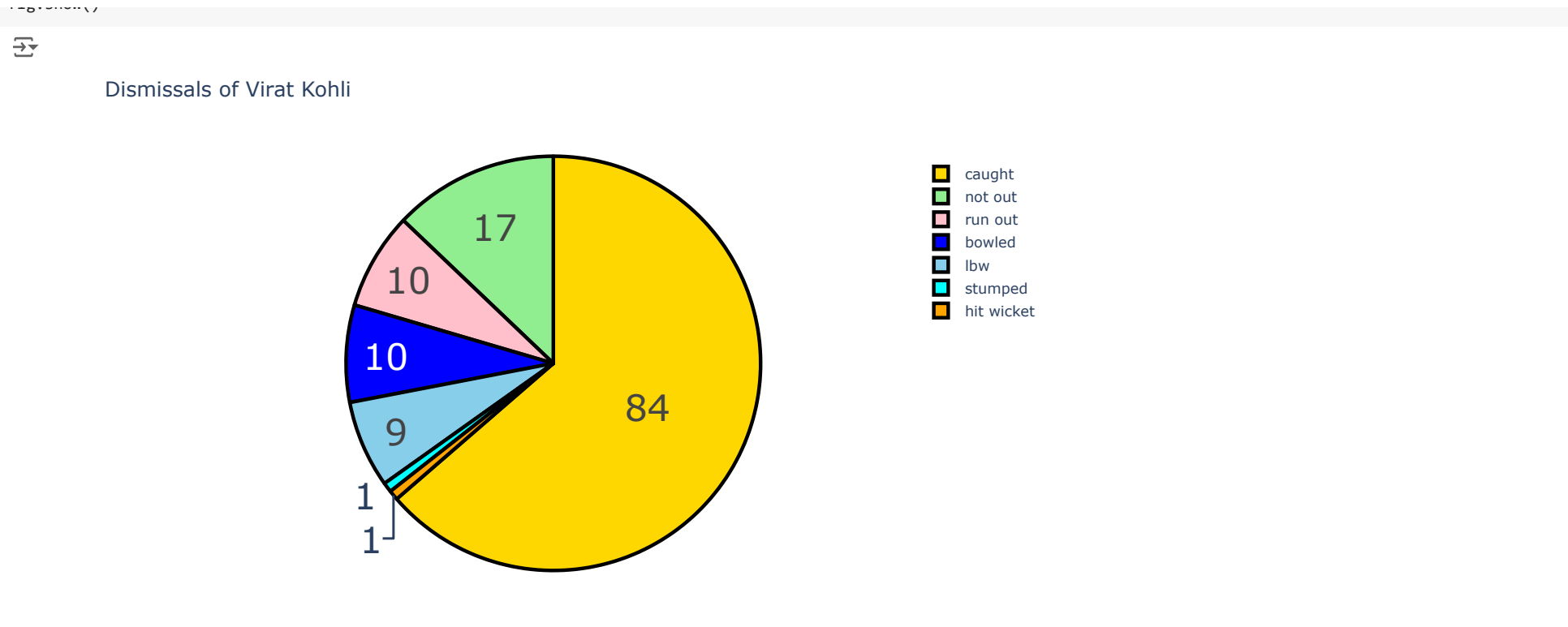
So most of the time, Virat Kohli gets out by getting caught by the fielder or the keeper. Now let's have a look at against which team Virat Kohli scored most of his runs:

```
figure = px.bar(data, x=data["Opposition"], y = data["Runs"], color = data["Runs"],
            title="Most Runs Against Teams")
figure.show()
```
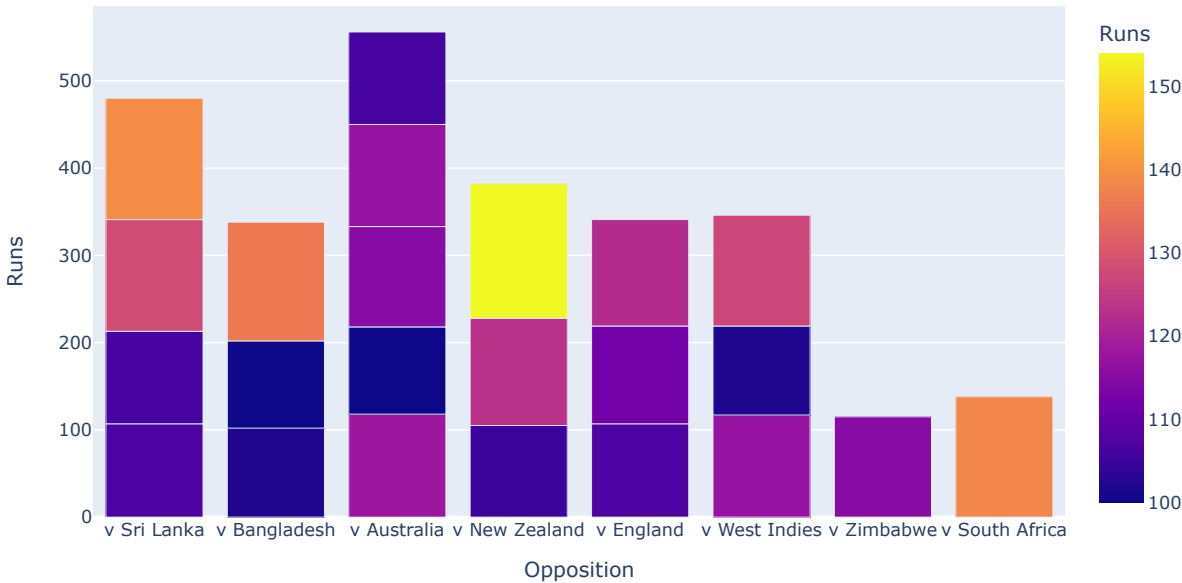
Most Runs Against Teams

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. Now let's have a look at against which team Virat Kohli scored most of his centuries:

```
figure = px.bar(centuries, x=centuries["Opposition"], y = centuries["Runs"],
                color = centuries["Runs"],
                title="Most Centuries Against Teams")
figure.show()
```

Most Centuries Against Teams

So, most of the centuries scored by Virat Kohli were against Australia. Now let's analyze Virat Kohli's strike rate. To analyze Virat Kohli's strike rate, I will create a new dataset of all the matches played by Virat Kohli where his strike rate was more than 120:

```
strike_rate = data.query("SR >= 120")
print(strike_rate)
```

```
     Runs  BF  4s  6s     SR         Pos Dismissal  Inns    Opposition  \
8      27  19   4   0  142.10  Batting At 7    bowled     1   v Sri Lanka
32    100  83   8   2  120.48  Batting At 4   not out     1  v Bangladesh
56     23  11   3   0  209.09  batting At 6   not out     1  v West Indies
76     43  34   4   1  126.47  Batting At 3    caught     1      v England
78    102  83  13   2  122.89  Batting At 3    caught     1  v West Indies
83    100  52   8   7  192.30  Batting At 3   not out     2    v Australia
85    115  66  18   1  174.24  Batting At 3   not out     2    v Australia
93     78  65   7   2  120.00  Batting At 3    caught     2  v New Zealand
130     8   5   2   0  160.00  Batting At 3    caught     1      v England

           Ground Start Date
8          Rajkot  15-Dec-09
32          Dhaka  19-Feb-11
56         Indore   8-Dec-11
76     Birmingham  23-Jun-13
78   Port of Spain   5-Jul-13
83         Jaipur  16-Oct-13
85         Nagpur  30-Oct-13
93       Hamilton  22-Jan-14
130        Cuttack  19-Jan-17
```
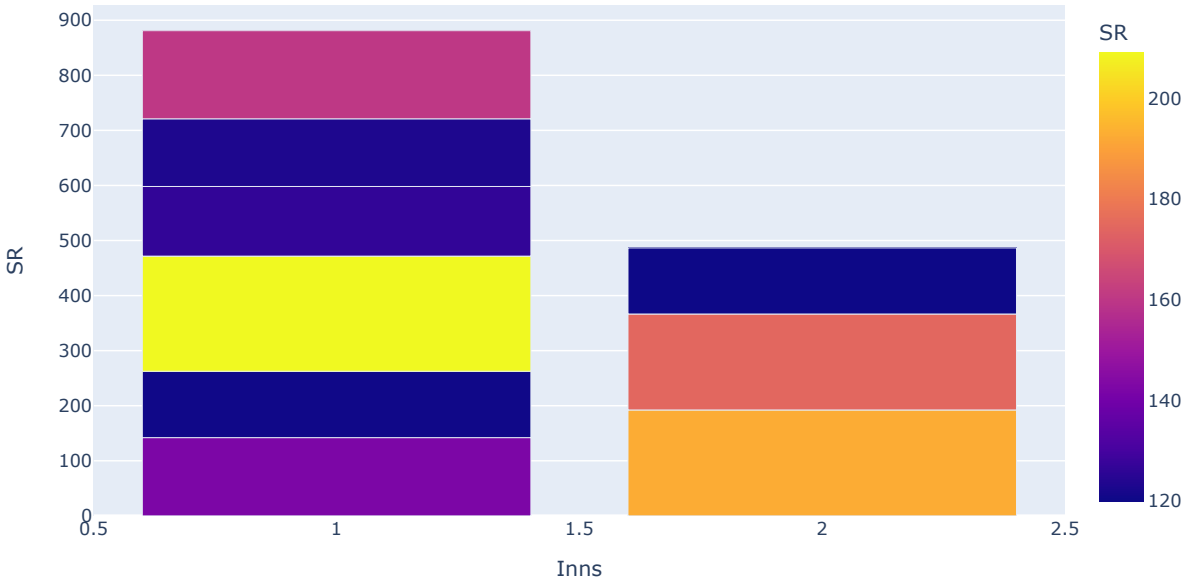
Now let's see whether Virat Kohli plays with high strike rates in the first innings or second innings:

```
figure = 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")
figure.show()
```

### Virat Kohli's High Strike Rates in First Innings Vs. Second Innings

So according to the above figure, Virat Kohli likes playing more aggressively in the first innings compared to the second innings. Now let's see the relationship between runs scored by Virat Kohli and fours played by him in each innings:

```
figure = px.scatter(data_frame = data, x="Runs",
```

```
                    y="4s", size="SR", trendline="ols",
                    title="Relationship Between Runs Scored and Fours")
figure.show()
```

Relationship Between Runs Scored and Fours
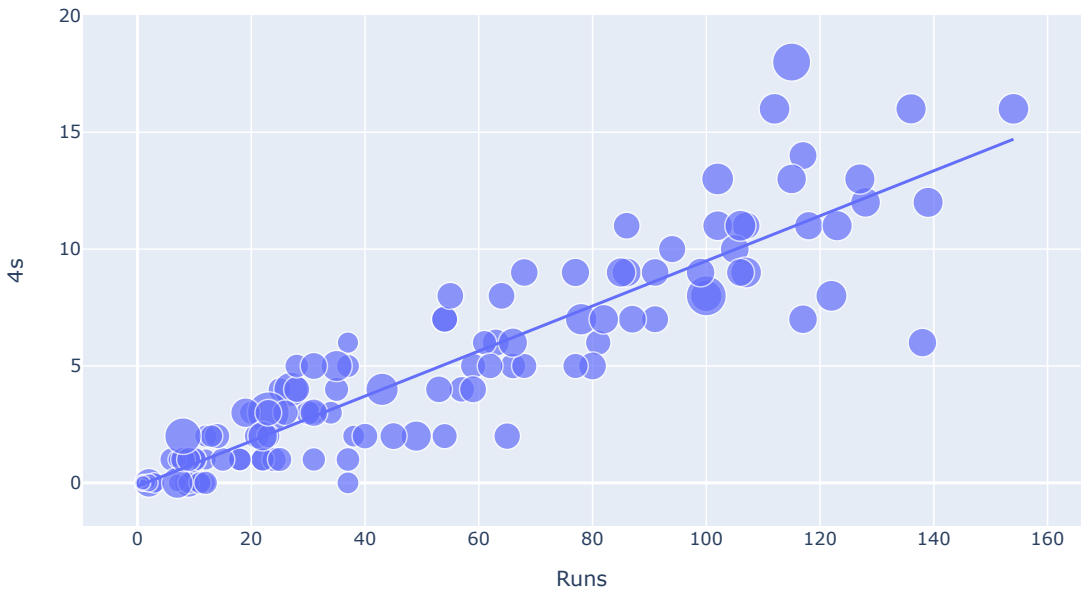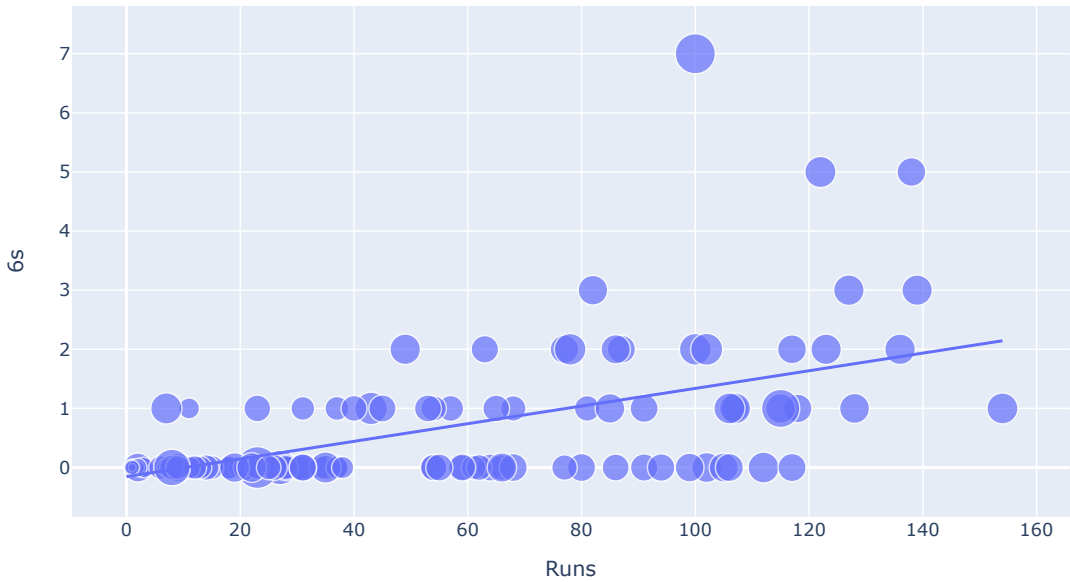

There is a linear relationship. It means that Virat Kohli likes playing fours. The more runs he scores in the innings, the more fours he plays. Let's see if there is some relationship with the sixes:

```
figure = px.scatter(data_frame = data, x="Runs",
                    y="6s", size="SR", trendline="ols",
                    title= "Relationship Between Runs Scored and Sixes")
figure.show()
```

Relationship Between Runs Scored and Sixes

There is no strong linear relationship here. It means Virat Kohli likes playing fours more than sixes. So this is how you can analyze the performance of Virat Kohli or any other cricketer in the world.

⌄ Summary

So this is how you can perform Virat Kohli performance analysis using the Python programming language not only him you can perform this analysis on any player to understand his performance. Analyzing a player's performance is one of the use cases of Data Science in sports analytics.