

MINOR PROJECT – 1

EXPLORATORY DATA ANALYSIS ON THE
DATASET VIRAT KOHLI CENTURIES

NAME : M SRINIVAS

COLLEGE : IIT BHUBANESWAR

YEAR : 2nd YEAR

BRANCH : COMPUTER SCIENCE

PROJECT LINK:

<https://colab.research.google.com/drive/li7z9ojSGeKJMO6cYTeCDBIC9y4tlef3i?usp=sharing>

GITHUB LINK OF THE PROJECT:

<https://github.com/srinivas1667/RINEX-PROJECTS/tree/main/mini%20project%201>

LINK OF THE DATASET SOURCE:

<https://www.kaggle.com/datasets/suyashkapil/virat-kohli-100s>

SCREENSHOTS OF THE CODE:

```
# THE DATASET THAT WE ARE GOING TO TAKE IS VIRAT KOHLI CENTURIES

import pandas as pd
import numpy as np

df = pd.read_csv('/content/Virat_Kohli_100s.csv')
df
```

	Number	Format	Inning	Position	Score	Balls	Strike Rate	Against	Venue	Host Nation	Series	Year	Team Total	Wickets lost	Not Out	MOTM	Win	Captain
0	1	ODI	2	4	107	114	93.859649	Sri Lanka	Kolkata	India	Bilateral	2009	316	3	No	No	Yes	No
1	2	ODI	2	3	102	95	107.368421	Bangladesh	Mirpur	Bangladesh	Tri-Series	2010	249	4	Yes	Yes	Yes	No
2	3	ODI	2	3	118	121	97.520661	Australia	Vizag	India	Bilateral	2010	292	5	No	Yes	Yes	No
3	4	ODI	1	3	105	104	100.961539	New Zealand	Guwahati	India	Bilateral	2010	276	10	No	Yes	Yes	No
4	5	ODI	1	4	100	83	120.481928	Bangladesh	Mirpur	Bangladesh	World Cup	2011	370	4	Yes	No	Yes	No
...
72	73	ODI	2	3	114	99	115.151515	West Indies	Port Of Spain	West Indies	Bilateral	2019	256	4	Yes	Yes	Yes	Yes
73	74	Test	1	4	254	336	75.595238	South Africa	Pune	India	Bilateral	2019	601	5	Yes	Yes	Yes	Yes
74	75	Test	2	4	136	194	70.103093	Bangladesh	Kolkata	India	Bilateral	2019	347	9	No	No	Yes	Yes
75	76	T20i	1	Open	122	61	200.000000	Afghanistan	Dubai	UAE	Asia Cup	2022	212	2	Yes	Yes	Yes	No
76	77	ODI	1	3	113	91	124.175824	Bangladesh	Chattogram	Bangladesh	Bilateral	2022	409	8	No	No	Yes	No

77 rows x 18 columns

```
[ ] df.info() # shows the info about the data frame
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 77 entries, 0 to 76
Data columns (total 18 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Number                77 non-null    int64
1   Format                77 non-null    object
2   Inning                77 non-null    int64
3   Position              77 non-null    object
4   Score                 77 non-null    int64
5   Balls                 77 non-null    int64
6   Strike Rate           77 non-null    float64
7   Against               77 non-null    object
8   Venue                 77 non-null    object
9   Host Nation           77 non-null    object
10  Series                77 non-null    object
11  Year                  77 non-null    int64
12  Team Total            77 non-null    int64
13  Wickets lost          77 non-null    int64
14  Not Out               77 non-null    object
15  MOTM                  77 non-null    object
16  Win                   77 non-null    object
17  Captain               77 non-null    object
dtypes: float64(1), int64(7), object(10)
memory usage: 11.0+ KB
```

```
[ ] df.size # shows the no.of elements in the dataframe
```

```
1386
```

```
[ ] df.shape # shows the no.of rows and coloumns
```

```
(77, 18)
```

CONCLUSION 1:

How many times was Kohli MOTM in the matches he scored hundred?

```
[ ] # CONCLUSION 1 - How many times was kohli MOTM in the matches he scored hundred
```

```
a1 = np.sum(df['MOTM']=='Yes')
a1
```

```
44
```

CONCLUSION 2:

How many matches was Kohli not out scoring a century?

```
[ ] # CONCLUSION 2 - How many matches was kolhi notout scoring a century
a2 = np.sum(df['Not Out']=='Yes')
a2

22
```

CONCLUSION 3:

What was the highest strike rate of Kohli scoring a century?

```
[ ] # CONCLUSION 3 - What was the highest strike rate
a3 = np.max(df['Strike Rate'])
a3

226.0
```

CONCLUSION 4:

In how many matches did Kohli come at no.3 and India won the game?

```
[ ] # CONCLUSION 4 - In how many matches did kohli come at no.3 and won the game
a4 = np.sum((df['Position']=='3')&(df['Win']=='Yes'))
a4

30
```

CONCLUSION 5:

What's the count of wins/losses of team India when Kohli scored century based on the innings number

```
[ ] # CONCLUSION 5 - COUNT OF THE WINS BASED ON THE Innings number
df.groupby(['Inning', 'Win']).size()
```

```
Inning  Win
1        Drawn      1
        No         4
        Tie        1
        Yes       25
2        Drawn      3
        No        10
        Yes       28
3        Drawn      1
        Yes        2
4        Drawn      1
        No         1
dtype: int64
```

CONCLUSION 6:

The no.of centuries Kohli scored in each venue.

```
# CONCLUSION - 6 : THE NO OF CENTURIES IN EVERY VENUE
import matplotlib.pyplot as plt

plt.figure(figsize=(30,10))
df['Venue'].value_counts().plot(kind='bar')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f8e41e421c0>

