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/1i7z9ojSGeKJM06cYTeCDBIC9y4t1ef3i?usp=sharing

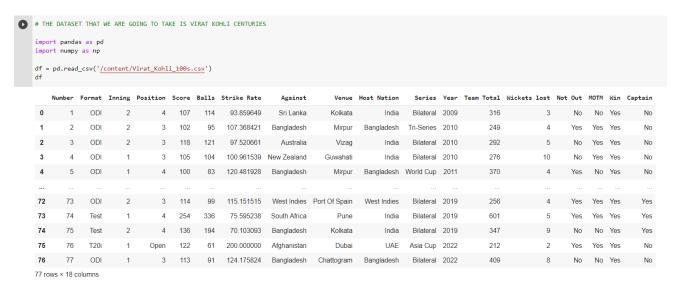
GITHUB LINK OF THE PROJECT:

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:



```
[ ] 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):
                   Non-Null Count Dtype
       Column
                    77 non-null
     0
       Number
                                   int64
                    77 non-null
     1
        Format
                                   object
        Inning
                     77 non-null
                                   int64
        Position
                     77 non-null
                                   object
     4
        Score
                     77 non-null
                                  int64
     5
                    77 non-null
                                   int64
        Balls
     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
                    77 non-null object
     10 Series
     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
                                object
     15 MOTM
                     77 non-null
                                object
     16 Win
                    77 non-null
     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
```

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.

