

```
In [5]: import pandas as pd
import numpy as np
df=pd.read_csv("D://DataSet//cricket//t20.csv")
```

```
In [6]: df
```

Out[6]:

	Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100	50	0
0	0	V Kohli (INDIA)	2010-2019	75	70	20	2633	94*	52.66	1907	138.07	0	24	2
1	1	RG Sharma (INDIA)	2007-2019	104	96	14	2633	118	32.1	1905	138.21	4	19	6
2	2	MJ Guptill (NZ)	2009-2019	83	80	7	2436	105	33.36	1810	134.58	2	15	2
3	3	Shoaib Malik (ICC/PAK)	2006-2019	111	104	30	2263	75	30.58	1824	124.06	0	7	1
4	4	BB McCullum (NZ)	2005-2015	71	70	10	2140	123	35.66	1571	136.21	2	13	3
...
2001	1	SG Whittingham (SCOT)	2018-2018	3	-	-	-	-	-	-	-	-	-	-
2002	2	LJ Woodcock (NZ)	2010-2011	3	-	-	-	-	-	-	-	-	-	-
2003	3	Zamir Khan (AFG)	2012-2012	1	-	-	-	-	-	-	-	-	-	-
2004	4	S Zargar (Mex)	2019-2019	1	-	-	-	-	-	-	-	-	-	-
2005	5	Ziaur Rahman (AFG)	2019-2019	1	-	-	-	-	-	-	-	-	-	-

2006 rows × 17 columns



```
In [7]: df.head()
```

Out[7]:

	Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100	50	0	4s
0	0	V Kohli (INDIA)	2010-2019	75	70	20	2633	94*	52.66	1907	138.07	0	24	2	247
1	1	RG Sharma (INDIA)	2007-2019	104	96	14	2633	118	32.1	1905	138.21	4	19	6	234
2	2	MJ Guptill (NZ)	2009-2019	83	80	7	2436	105	33.36	1810	134.58	2	15	2	215
3	3	Shoaib Malik	2006-2019	111	104	30	2263	75	30.58	1824	124.06	0	7	1	186

Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100	50	0	4s
(ICC/PAK)														
4	BB McCullum (NZ)	2005- 2015	71	70	10	2140	123	35.66	1571	136.21	2	13	3	199

In [8]: df.describe()

	Unnamed: 0	Mat	Unnamed: 15
count	2006.000000	2006.000000	0.0
mean	24.434197	11.231805	NaN
std	14.463176	14.923332	NaN
min	0.000000	1.000000	NaN
25%	12.000000	3.000000	NaN
50%	24.000000	5.000000	NaN
75%	37.000000	13.000000	NaN
max	49.000000	111.000000	NaN

In [9]: df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2006 entries, 0 to 2005
Data columns (total 17 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Unnamed: 0            2006 non-null  int64
1   Player                2006 non-null  object
2   Span                  2006 non-null  object
3   Mat                   2006 non-null  int64
4   Inns                  2006 non-null  object
5   NO                    2006 non-null  object
6   Runs                  2006 non-null  object
7   HS                    2006 non-null  object
8   Ave                   2006 non-null  object
9   BF                    2006 non-null  object
10  SR                     2006 non-null  object
11  100                    2006 non-null  object
12  50                      2006 non-null  object
13  0                       2006 non-null  object
14  4s                     2006 non-null  object
15  6s                     2006 non-null  object
16  Unnamed: 15           0 non-null     float64
dtypes: float64(1), int64(2), object(14)
memory usage: 156.8+ KB
```

In [13]: df["Inns"].value_counts()

1	328
2	268
3	222
4	202
-	152
...	
104	1
43	1

```
84      1
52      1
51      1
Name: Inns, Length: 77, dtype: int64
```

```
In [25]: df[df['HS']>'100']
```

Out[25]:

	Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100	50	0
0	0	V Kohli (INDIA)	2010-2019	75	70	20	2633	94*	52.66	1907	138.07	0	24	2
1	1	RG Sharma (INDIA)	2007-2019	104	96	14	2633	118	32.1	1905	138.21	4	19	6
2	2	MJ Guptill (NZ)	2009-2019	83	80	7	2436	105	33.36	1810	134.58	2	15	2
3	3	Shoaib Malik (ICC/PAK)	2006-2019	111	104	30	2263	75	30.58	1824	124.06	0	7	1
4	4	BB McCullum (NZ)	2005-2015	71	70	10	2140	123	35.66	1571	136.21	2	13	3
...
1679	29	Waqas Raja (Fin)	2019-2019	1	1	1	2	2*	-	2	100.0	0	0	0
1680	30	UWMBCA Welegedara (SL)	2010-2010	2	1	1	2	2*	-	3	66.66	0	0	0
1681	31	BTJ Wheal (SCOT)	2016-2017	5	2	2	2	2*	-	2	100.0	0	0	0
1682	32	UT Yadav (INDIA)	2012-2019	7	1	0	2	2	2.00	4	50.0	0	0	0
1683	33	Zaheer Maqsood (UAE)	2015-2016	3	1	0	2	2	2.00	5	40.0	0	0	0

1639 rows × 17 columns



```
In [26]: df.corr()
```

Out[26]:

	Unnamed: 0	Mat	Unnamed: 15
Unnamed: 0	1.000000	-0.038502	NaN
Mat	-0.038502	1.000000	NaN
Unnamed: 15	NaN	NaN	NaN

```
In [27]: df['100'].value_counts()
```

```
Out[27]: 0    1812
-     152
1      33
2       6
3       2
```

```
4      1
Name: 100, dtype: int64
```

```
In [31]: df.median()
```

```
Out[31]: Unnamed: 0      24.0
Mat      5.0
Unnamed: 15      NaN
dtype: float64
```

```
In [32]: df.max()
```

```
Out[32]: Unnamed: 0      49
Player      Zulqarnain Haider (PAK)
Span      2019-2019
Mat      111
Inns      96
NO      9
Runs      990
HS      99*
Ave      96.0
BF      99
SR      99.54
100      4
50      9
0      9
4s      96
6s      96
Unnamed: 15      NaN
dtype: object
```

```
In [33]: df.min()
```

```
Out[33]: Unnamed: 0      0
Player      A Ahmadhel ()
Span      2005-2005
Mat      1
Inns      -
NO      -
Runs      -
HS      -
Ave      -
BF      -
SR      -
100      -
50      -
0      -
4s      -
6s      -
Unnamed: 15      NaN
dtype: object
```

```
In [34]: df.dropna(axis=0)
```

Out[34]:

Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100	50	0	4s	6s	Unnamed: 15
0																

```
In [36]: df.dropna(axis=1)
```

Out[36]:

Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100	50	0	
0	0	V Kohli (INDIA)	2010-2019	75	70	20	2633	94*	52.66	1907	138.07	0	24	2

	Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100	50	0
1	1	RG Sharma (INDIA)	2007-2019	104	96	14	2633	118	32.1	1905	138.21	4	19	6
2	2	MJ Guptill (NZ)	2009-2019	83	80	7	2436	105	33.36	1810	134.58	2	15	2
3	3	Shoaib Malik (ICC/PAK)	2006-2019	111	104	30	2263	75	30.58	1824	124.06	0	7	1
4	4	BB McCullum (NZ)	2005-2015	71	70	10	2140	123	35.66	1571	136.21	2	13	3
...
2001	1	SG Whittingham (SCOT)	2018-2018	3	-	-	-	-	-	-	-	-	-	-
2002	2	LJ Woodcock (NZ)	2010-2011	3	-	-	-	-	-	-	-	-	-	-
2003	3	Zamir Khan (AFG)	2012-2012	1	-	-	-	-	-	-	-	-	-	-
2004	4	S Zargar (Mex)	2019-2019	1	-	-	-	-	-	-	-	-	-	-
2005	5	Ziaur Rahman (AFG)	2019-2019	1	-	-	-	-	-	-	-	-	-	-

2006 rows × 16 columns

In [39]:

df["100"].notna()

Out[39]:

0	True
1	True
2	True
3	True
4	True
...	
2001	True
2002	True
2003	True
2004	True
2005	True

Name: 100, Length: 2006, dtype: bool

In [40]:

df.isnull()

Out[40]:

	Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100	50	0
0	False	False	False	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False	False	False	False

	Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100	50	0
4	False	False	False	False	False	False	False	False	False	False	False	False	False	False
...
2001	False	False	False	False	False	False	False	False	False	False	False	False	False	False
2002	False	False	False	False	False	False	False	False	False	False	False	False	False	False
2003	False	False	False	False	False	False	False	False	False	False	False	False	False	False
2004	False	False	False	False	False	False	False	False	False	False	False	False	False	False
2005	False	False	False	False	False	False	False	False	False	False	False	False	False	False

2006 rows × 17 columns

In [41]:

df.notnull()

	Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100	50	0	4s
0	True	True	True	True	True	True	True	True	True	True	True	True	True	True	True
1	True	True	True	True	True	True	True	True	True	True	True	True	True	True	True
2	True	True	True	True	True	True	True	True	True	True	True	True	True	True	True
3	True	True	True	True	True	True	True	True	True	True	True	True	True	True	True
4	True	True	True	True	True	True	True	True	True	True	True	True	True	True	True
...
2001	True	True	True	True	True	True	True	True	True	True	True	True	True	True	True
2002	True	True	True	True	True	True	True	True	True	True	True	True	True	True	True
2003	True	True	True	True	True	True	True	True	True	True	True	True	True	True	True
2004	True	True	True	True	True	True	True	True	True	True	True	True	True	True	True
2005	True	True	True	True	True	True	True	True	True	True	True	True	True	True	True

2006 rows × 17 columns

In [42]:

lst_data=[[1,2,3],[3,4,np.nan],[5,6,np.nan],[np.nan,np.nan,np.nan]]

In [43]:

type(lst_data)

Out[43]: list

In [44]:

df=pd.DataFrame(lst_data)
df.head()

	0	1	2
0	1.0	2.0	3.0
1	3.0	4.0	NaN

	0	1	2
2	5.0	6.0	NaN
3	NaN	NaN	NaN

In [45]: `df.dropna(axis=0)`

Out[45]:

	0	1	2
0	1.0	2.0	3.0

In [47]: `df.dropna(axis=1)`

Out[47]:

0
1
2
3

In [48]: `df = pd.DataFrame(np.random.randn(5, 3), index=['a', 'c', 'e', 'f', 'h'],
columns=['one', 'two', 'three'])
df.head()`

Out[48]:

	one	two	three
a	-0.069808	1.270674	1.625797
c	0.903771	-1.910739	1.352563
e	-0.570173	-0.108705	-1.371389
f	0.951345	0.563916	0.723534
h	0.058664	0.521628	-0.149008

In [49]: `df2=df.reindex(['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h'])`

In [50]: `df2`

Out[50]:

	one	two	three
a	-0.069808	1.270674	1.625797
b	NaN	NaN	NaN
c	0.903771	-1.910739	1.352563
d	NaN	NaN	NaN
e	-0.570173	-0.108705	-1.371389
f	0.951345	0.563916	0.723534
g	NaN	NaN	NaN
h	0.058664	0.521628	-0.149008

In [51]: `df2.dropna(axis=0)`

```
Out[51]:
```

	one	two	three
a	-0.069808	1.270674	1.625797
c	0.903771	-1.910739	1.352563
e	-0.570173	-0.108705	-1.371389
f	0.951345	0.563916	0.723534
h	0.058664	0.521628	-0.149008

```
In [52]: pd.isna(df2['one'])
```

```
Out[52]: a    False
b     True
c    False
d     True
e    False
f    False
g     True
h    False
Name: one, dtype: bool
```

```
In [53]: df2['one'].notna()
```

```
Out[53]: a     True
b    False
c     True
d    False
e     True
f     True
g    False
h     True
Name: one, dtype: bool
```

```
In [54]: df2.fillna('Missing')
```

```
Out[54]:
```

	one	two	three
a	-0.0698077	1.27067	1.6258
b	Missing	Missing	Missing
c	0.903771	-1.91074	1.35256
d	Missing	Missing	Missing
e	-0.570173	-0.108705	-1.37139
f	0.951345	0.563916	0.723534
g	Missing	Missing	Missing
h	0.0586644	0.521628	-0.149008

```
In [55]: df2['one'].values
```

```
Out[55]: array([-0.0698077 ,          nan,  0.90377137,          nan, -0.57017277,
                0.95134521,          nan,  0.05866444])
```

```
In [ ]:
```