

```
In [1]: ## How to get descriptive statistics of a Pandas DataFrame
```

```
In [2]: print()
print(format('How to get descriptive statistics of a Pandas DataFrame', '^82'))
import warnings
warnings.filterwarnings("ignore")

# load libraries
import pandas as pd

# Create dataframe
data = {'name': ['Jason', 'Molly', 'Tina', 'Jake', 'Amy'],
        'age': [42, 52, 36, 24, 73],
        'preTestScore': [4, 24, 31, 2, 3],
        'postTestScore': [25, 94, 57, 62, 70]}
df = pd.DataFrame(data, columns = ['name', 'age', 'preTestScore', 'postTestScore'])

print(); print(df)
print(); print(df.info())
```

*****How to get descriptive statistics of a Pandas DataFrame*****

	name	age	preTestScore	postTestScore
0	Jason	42	4	25
1	Molly	52	24	94
2	Tina	36	31	57
3	Jake	24	2	62
4	Amy	73	3	70

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5 entries, 0 to 4
Data columns (total 4 columns):
name          5 non-null object
age           5 non-null int64
preTestScore  5 non-null int64
postTestScore 5 non-null int64
dtypes: int64(3), object(1)
memory usage: 240.0+ bytes
None
```

```
In [3]: # The sum of all the ages
print(); print(df['age'].sum())
```

227

```
In [4]: # Mean preTestScore
print(); print(df['preTestScore'].mean())
```

12.8

```
In [5]: # Cumulative sum of preTestScores, moving from the rows from the top
print(); print(df['preTestScore'].cumsum())
```

0	4
1	28
2	59
3	61
4	64

Name: preTestScore, dtype: int64

```
In [6]: # Summary statistics on preTestScore
print(); print(df['preTestScore'].describe())
```

count	5.000000
mean	12.800000
std	13.663821
min	2.000000
25%	3.000000
50%	4.000000
75%	24.000000
max	31.000000

Name: preTestScore, dtype: float64

```
In [7]: # Count the number of non-NA values
print(); print(df['preTestScore'].count())
```

5

```
In [8]: # Minimum value of preTestScore
print(); print(df['preTestScore'].min())
```

2

```
In [9]: # Maximum value of preTestScore
print(); print(df['preTestScore'].max())
```

31

```
In [10]: # Median value of preTestScore
print(); print(df['preTestScore'].median())
```

4.0

```
In [11]: # Sample variance of preTestScore values
print(); print(df['preTestScore'].var())
```

186.7

```
In [12]: # Sample standard deviation of preTestScore values
print(); print(df['preTestScore'].std())
```

13.663820841916802

```
In [13]: # Skewness of preTestScore values
print(); print(df['preTestScore'].skew())
```

0.7433452457326751

```
In [14]: # Kurtosis of preTestScore values
print(); print(df['preTestScore'].kurt())
```

-2.4673543738411547

```
In [15]: # Correlation Matrix Of Values
print(); print(df.corr())
```

	age	preTestScore	postTestScore
age	1.000000	-0.105651	0.328852
preTestScore	-0.105651	1.000000	0.378039
postTestScore	0.328852	0.378039	1.000000

```
In [16]: # Covariance Matrix Of Values
print(); print(df.cov())
```

	age	preTestScore	postTestScore
age	340.80	-26.65	151.20
preTestScore	-26.65	186.70	128.65
postTestScore	151.20	128.65	620.30

```
In [ ]:
```