In [1]:

```
# Pandas
import numpy as np
import pandas as pd
```

In [2]:

```
# import data set

df = pd.read_csv('sets/data.csv')
df
```

Out[2]:

	ID	Name	Industry	Inception	Revenue	Expenses	Profit	Growth
0	1	Lamtone	IT Services	2009	$11,757,018	6,482,465 Dollars	5274553	30%
1	2	Stripfind	Financial Services	2010	$12,329,371	916,455 Dollars	11412916	20%
2	3	Canecorporation	Health	2012	$10,597,009	7,591,189 Dollars	3005820	7%
3	4	Mattouch	IT Services	2013	$14,026,934	7,429,377 Dollars	6597557	26%
4	5	Techdrill	Health	2009	$10,573,990	7,435,363 Dollars	3138627	8%
5	6	Techline	Health	2006	$13,898,119	5,470,303 Dollars	8427816	23%
6	7	Cityace	Health	2010	$9,254,614	6,249,498 Dollars	3005116	6%
7	8	Kayelectronics	Health	2009	$9,451,943	3,878,113 Dollars	5573830	4%
8	9	Ganzlax	IT Services	2011	$14,001,180	916,455 Dollars	11901180	18%
9	10	Trantraxlax	Government Services	2011	$11,088,336	5,635,276 Dollars	5453060	7%

In [3]:

```
# na_filter
# na_filter will boost ur application ,telling no syntax of missing value

df = pd.read_csv('sets/data.csv',na_filter=False)
df
```

Out[3]:

	ID	Name	Industry	Inception	Revenue	Expenses	Profit	Growth
0	1	Lamtone	IT Services	2009	$11,757,018	6,482,465 Dollars	5274553	30%
1	2	Stripfind	Financial Services	2010	$12,329,371	916,455 Dollars	11412916	20%
2	3	Canecorporation	Health	2012	$10,597,009	7,591,189 Dollars	3005820	7%
3	4	Mattouch	IT Services	2013	$14,026,934	7,429,377 Dollars	6597557	26%
4	5	Techdrill	Health	2009	$10,573,990	7,435,363 Dollars	3138627	8%
5	6	Techline	Health	2006	$13,898,119	5,470,303 Dollars	8427816	23%
6	7	Cityace	Health	2010	$9,254,614	6,249,498 Dollars	3005116	6%
7	8	Kayelectronics	Health	2009	$9,451,943	3,878,113 Dollars	5573830	4%
8	9	Ganzlax	IT Services	2011	$14,001,180	916,455 Dollars	11901180	18%
9	10	Trantraxlax	Government Services	2011	$11,088,336	5,635,276 Dollars	5453060	7%

In [ ]:

```

```

In [4]:

```
df = pd.read_csv('sets/data.csv',na_filter=False)
df
```

Out[4]:

	ID	Name	Industry	Inception	Revenue	Expenses	Profit	Growth
0	1	Lamtone	IT Services	2009	$11,757,018	6,482,465 Dollars	5274553	30%
1	2	Stripfind	Financial Services	2010	$12,329,371	916,455 Dollars	11412916	20%
2	3	Canecorporation	Health	2012	$10,597,009	7,591,189 Dollars	3005820	7%
3	4	Mattouch	IT Services	2013	$14,026,934	7,429,377 Dollars	6597557	
4	5	Techdrill	Health		$10,573,990	7,435,363 Dollars	3138627	8%
5	6	Techline	Health	2006	$13,898,119	5,470,303 Dollars	8427816	23%
6	7	Cityace			$9,254,614	6,249,498 Dollars	3005116	6%
7	8	Kayelectronics		2009	$9,451,943	3,878,113 Dollars	5573830	4%
8	9	Ganzlax	IT Services	2011	$14,001,180		11901180	18%
9	10	Trantraxlax	Government Services	2011	$11,088,336	5,635,276 Dollars	5453060	7%

In [ ]:

```
```

In [ ]:

```
```

In [5]:

```
# want to know how many missing value is present
df2 = pd.read_csv('sets/data.csv')
df2
```

Out[5]:

	ID	Name	Industry	Inception	Revenue	Expenses	Profit	Growth
0	1	Lamtone	IT Services	2009.0	$11,757,018	6,482,465 Dollars	5274553	30%
1	2	Stripfind	Financial Services	2010.0	$12,329,371	916,455 Dollars	11412916	20%
2	3	Canecorporation	Health	2012.0	$10,597,009	7,591,189 Dollars	3005820	7%
3	4	Mattouch	IT Services	2013.0	$14,026,934	7,429,377 Dollars	6597557	NaN
4	5	Techdrill	Health	NaN	$10,573,990	7,435,363 Dollars	3138627	8%
5	6	Techline	Health	2006.0	$13,898,119	5,470,303 Dollars	8427816	23%
6	7	Cityace	NaN	NaN	$9,254,614	6,249,498 Dollars	3005116	6%
7	8	Kayelectronics	NaN	2009.0	$9,451,943	3,878,113 Dollars	5573830	4%
8	9	Ganzlax	IT Services	2011.0	$14,001,180	NaN	11901180	18%
9	10	Trantraxlax	Government Services	2011.0	$11,088,336	5,635,276 Dollars	5453060	7%

In [6]:

```
# isnull => will give value in True/ False Format
# if value is missing it will show True else false
df2.isnull()
```

Out[6]:

	ID	Name	Industry	Inception	Revenue	Expenses	Profit	Growth
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	True
4	False	False	False	True	False	False	False	False
5	False	False	False	False	False	False	False	False
6	False	False	True	True	False	False	False	False
7	False	False	True	False	False	False	False	False
8	False	False	False	False	False	True	False	False
9	False	False	False	False	False	False	False	False

In [7]:

```
# how many values are missing from particular column
df2.isnull().sum()
```

Out[7]:

```
ID           0
Name         0
Industry     2
Inception    2
Revenue      0
Expenses     1
Profit       0
Growth       1
dtype: int64
```

In [8]:

```
# total missing values
df2.isnull().sum().sum()
```

Out[8]:

```
6
```

In [ ]:

```

```

In [10]:

```
# to know hoe many data is present in data set
# notnull()
# notnull => it will tell fill data value

df2
```

Out[10]:

	ID	Name	Industry	Inception	Revenue	Expenses	Profit	Growth
0	1	Lamtone	IT Services	2009.0	$11,757,018	6,482,465 Dollars	5274553	30%
1	2	Stripfind	Financial Services	2010.0	$12,329,371	916,455 Dollars	11412916	20%
2	3	Canecorporation	Health	2012.0	$10,597,009	7,591,189 Dollars	3005820	7%
3	4	Mattouch	IT Services	2013.0	$14,026,934	7,429,377 Dollars	6597557	NaN
4	5	Techdrill	Health	NaN	$10,573,990	7,435,363 Dollars	3138627	8%
5	6	Techline	Health	2006.0	$13,898,119	5,470,303 Dollars	8427816	23%
6	7	Cityace	NaN	NaN	$9,254,614	6,249,498 Dollars	3005116	6%
7	8	Kayelectronics	NaN	2009.0	$9,451,943	3,878,113 Dollars	5573830	4%
8	9	Ganzlax	IT Services	2011.0	$14,001,180	NaN	11901180	18%
9	10	Trantraxlax	Government Services	2011.0	$11,088,336	5,635,276 Dollars	5453060	7%

In [11]:

```
# if value are present then it will show True else False
df2.notnull()
```

Out[11]:

	ID	Name	Industry	Inception	Revenue	Expenses	Profit	Growth
0	True	True	True	True	True	True	True	True
1	True	True	True	True	True	True	True	True
2	True	True	True	True	True	True	True	True
3	True	True	True	True	True	True	True	False
4	True	True	True	False	True	True	True	True
5	True	True	True	True	True	True	True	True
6	True	True	False	False	True	True	True	True
7	True	True	False	True	True	True	True	True
8	True	True	True	True	True	False	True	True
9	True	True	True	True	True	True	True	True

In [12]:

```
df2.notnull().sum()
```

Out[12]:

```
ID           10
Name         10
Industry      8
Inception     8
Revenue      10
Expenses      9
Profit       10
Growth        9
dtype: int64
```

In [13]:

```
df2.notnull().sum().sum()
```

Out[13]:

```
74
```

In [ ]:

```

```

In [15]:

```
# missing values in series
data1 = pd.Series([10,np.nan,20,30,np.nan,40,50,60,np.NaN,70])
data1
```

Out[15]:

```
0    10.0
1     NaN
2    20.0
3    30.0
4     NaN
5    40.0
6    50.0
7    60.0
8     NaN
9    70.0
dtype: float64
```

In [17]:

```
data1.isnull()
```

Out[17]:

```
0    False
1     True
2    False
3    False
4     True
5    False
6    False
7    False
8     True
9    False
dtype: bool
```

In [18]:

```
data1.isnull().sum()
```

Out[18]:

```
3
```

In [ ]:

```

```

In [19]:

```
data1.notnull()
```

Out[19]:

```
0     True
1    False
2     True
3     True
4    False
5     True
6     True
7     True
8    False
9     True
dtype: bool
```

In [20]:

```
data1.notnull().sum()
```

Out[20]:

```
7
```