In [1]:

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

In [3]:

```
# Missing Values
# import data set

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

```
# missing value will be displayed as NAN
df = pd.read_csv('sets/data.csv')
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	NaN
4	5	Techdrill	Health	2009	$10,573,990	NaN	3138627	8%
5	6	Techline	NaN	2006	$13,898,119	5,470,303 Dollars	8427816	23%
6	7	Cityace	NaN	2010	$9,254,614	6,249,498 Dollars	3005116	6%
7	8	Kayelectronics	Health	2009	$9,451,943	3,878,113 Dollars	5573830	NaN
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 [ ]:

```
# N/A          -NaN      null
# N/A N/A      -nan      n/a
# NA            N/A      nan
```

In [6]:

```
# missing value will be displayed as NAN
df = pd.read_csv('sets/data.csv')
df
```

Out[6]:

	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	NaN
4	5	Techdrill	Health	2009	$10,573,990	NaN	3138627	8%
5	6	Techline	NaN	2006	$13,898,119	5,470,303 Dollars	8427816	23%
6	7	Cityace	NaN	2010	$9,254,614	6,249,498 Dollars	3005116	6%
7	8	Kayelectronics	Health	2009	$9,451,943	3,878,113 Dollars	5573830	NaN
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 [7]:

```
# missing value will be displayed as NAN
df = pd.read_csv('sets/data.csv')
df
```

Out[7]:

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

In [ ]:

```

```

In [8]:

```
# missing value will be displayed as NAN
df = pd.read_csv('sets/data.csv')
df
```

Out[8]:

	ID	Name	Status	Industry	Inception	Revenue	Expenses	Profit	Gro
0	1	Lamtone	NaN	IT Services	2009.0	$11,757,018	6,482,465 Dollars	5274553.0	3
1	2	Stripfind	NaN	Financial Services	2010.0	$12,329,371	916,455 Dollars	11412916.0	2
2	3	Canecorporation	NaN	Health	2012.0	$10,597,009	7,591,189 Dollars	3005820.0	
3	4	Mattouch	NaN	IT Services	2013.0	$14,026,934	7,429,377 Dollars	6597557.0	N
4	5	Techdrill	NaN	Health	2009.0	$10,573,990	NaN	3138627.0	
5	6	Techline	NaN	NaN	2006.0	$13,898,119	5,470,303 Dollars	8427816.0	2
6	7	Cityace	NaN	NaN	2010.0	$9,254,614	6,249,498 Dollars	3005116.0	
7	8	Kayelectronics	NaN	Health	2009.0	$9,451,943	3,878,113 Dollars	5573830.0	N
8	9	Ganzlax	NaN	IT Services	2011.0	$14,001,180	916,455 Dollars	11901180.0	1
9	10	Trantraxlax	NaN	Government Services	2011.0	$11,088,336	5,635,276 Dollars	5453060.0	
10	11	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N

In [ ]:

```

```

In [ ]:

```

```

In [9]:

```
# Some time your data set contains diff values
df = pd.read_csv('sets/data.csv')
df
```

Out[9]:

	ID	Name	Status	Industry	Inception	Revenue	Expenses	Profit	Gro
0	1	Lamtone	NaN	IT Services	2009.0	$11,757,018	6,482,465 Dollars	5274553.0	3
1	2	Stripfind	NaN	Financial Services	2010.0	$12,329,371	916,455 Dollars	11412916.0	2
2	3	Canecorporation	NaN	Health	2012.0	$10,597,009	7,591,189 Dollars	3005820.0	
3	4	Mattouch	NaN	IT Services	2013.0	$14,026,934	7,429,377 Dollars	6597557.0	noc
4	5	Techdrill	NaN	Health	2009.0	$10,573,990	NaN	3138627.0	
5	6	Techline	NaN	nodata	2006.0	$13,898,119	5,470,303 Dollars	8427816.0	2
6	7	Cityace	NaN	nodata	2010.0	$9,254,614	6,249,498 Dollars	3005116.0	
7	8	Kayelectronics	NaN	Health	2009.0	$9,451,943	3,878,113 Dollars	5573830.0	noc
8	9	Ganzlax	NaN	IT Services	2011.0	$14,001,180	916,455 Dollars	11901180.0	1
9	10	Trantraxlax	NaN	Government Services	2011.0	$11,088,336	5,635,276 Dollars	5453060.0	
10	11	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N

In [10]:

```
# change nodata string to NaN
df = pd.read_csv('sets/data.csv', na_values='nodata')
df
```

Out[10]:

	ID	Name	Status	Industry	Inception	Revenue	Expenses	Profit	Gro
0	1	Lamtone	NaN	IT Services	2009.0	$11,757,018	6,482,465 Dollars	5274553.0	3
1	2	Stripfind	NaN	Financial Services	2010.0	$12,329,371	916,455 Dollars	11412916.0	2
2	3	Canecorporation	NaN	Health	2012.0	$10,597,009	7,591,189 Dollars	3005820.0	
3	4	Mattouch	NaN	IT Services	2013.0	$14,026,934	7,429,377 Dollars	6597557.0	N
4	5	Techdrill	NaN	Health	2009.0	$10,573,990	NaN	3138627.0	
5	6	Techline	NaN	NaN	2006.0	$13,898,119	5,470,303 Dollars	8427816.0	2
6	7	Cityace	NaN	NaN	2010.0	$9,254,614	6,249,498 Dollars	3005116.0	
7	8	Kayelectronics	NaN	Health	2009.0	$9,451,943	3,878,113 Dollars	5573830.0	N
8	9	Ganzlax	NaN	IT Services	2011.0	$14,001,180	916,455 Dollars	11901180.0	1
9	10	Trantraxlax	NaN	Government Services	2011.0	$11,088,336	5,635,276 Dollars	5453060.0	
10	11	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N

In [ ]:

```
```

In [11]:

```
# Some time your data set contains diff values
df = pd.read_csv('sets/data.csv')
df
```

Out[11]:

	ID	Name	Status	Industry	Inception	Revenue	Expenses	Profit	Gro
0	1	Lamtone	NaN	IT Services	2009.0	$11,757,018	6,482,465 Dollars	5274553.0	3
1	2	Stripfind	NaN	Financial Services	2010.0	$12,329,371	916,455 Dollars	11412916.0	2
2	3	Canecorporation	NaN	Health	2012.0	$10,597,009	7,591,189 Dollars	3005820.0	
3	4	Mattouch	NaN	IT Services	2013.0	$14,026,934	7,429,377 Dollars	6597557.0	noc
4	5	Techdrill	NaN	Health	2009.0	$10,573,990	NaN	3138627.0	
5	6	Techline	NaN	missing Data	2006.0	$13,898,119	5,470,303 Dollars	8427816.0	2
6	7	Cityace	NaN	missing Data	2010.0	$9,254,614	6,249,498 Dollars	3005116.0	
7	8	Kayelectronics	NaN	Health	2009.0	$9,451,943	3,878,113 Dollars	5573830.0	noc
8	9	Ganzlax	NaN	IT Services	2011.0	$14,001,180	916,455 Dollars	11901180.0	1
9	10	Trantraxlax	NaN	Government Services	2011.0	$11,088,336	5,635,276 Dollars	5453060.0	
10	11	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N

In [13]:

```
# change nodata,missing data string to NaN
df = pd.read_csv('sets/data.csv', na_values=['nodata','missing Data'])
df
```

Out[13]:

	ID	Name	Status	Industry	Inception	Revenue	Expenses	Profit	Gro
0	1	Lamtone	NaN	IT Services	2009.0	$11,757,018	6,482,465 Dollars	5274553.0	3
1	2	Stripfind	NaN	Financial Services	2010.0	$12,329,371	916,455 Dollars	11412916.0	2
2	3	Canecorporation	NaN	Health	2012.0	$10,597,009	7,591,189 Dollars	3005820.0	
3	4	Mattouch	NaN	IT Services	2013.0	$14,026,934	7,429,377 Dollars	6597557.0	N
4	5	Techdrill	NaN	Health	2009.0	$10,573,990	NaN	3138627.0	
5	6	Techline	NaN	NaN	2006.0	$13,898,119	5,470,303 Dollars	8427816.0	2
6	7	Cityace	NaN	NaN	2010.0	$9,254,614	6,249,498 Dollars	3005116.0	
7	8	Kayelectronics	NaN	Health	2009.0	$9,451,943	3,878,113 Dollars	5573830.0	N
8	9	Ganzlax	NaN	IT Services	2011.0	$14,001,180	916,455 Dollars	11901180.0	1
9	10	Trantraxlax	NaN	Government Services	2011.0	$11,088,336	5,635,276 Dollars	5453060.0	
10	11	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N

In [ ]:

In [14]:

```
# Some time your data set contains diff values
df = pd.read_csv('sets/data.csv')
df
```

Out[14]:

	ID	Name	Status	Industry	Inception	Revenue	Expenses	Profit	Grow
0	1	Lamtone	yes	IT Services	2009	$11,757,018	6,482,465 Dollars	5274553	30
1	2	Stripfind	yes	Financial Services	2010	$12,329,371	916,455 Dollars	11412916	20
2	3	Canecorporation	yes	Health	2012	$10,597,009	7,591,189 Dollars	3005820	7
3	4	Mattouch	yes	IT Services	2013	$14,026,934	7,429,377 Dollars	6597557	26
4	5	Techdrill	yes	Health	2009	$10,573,990	7,435,363 Dollars	3138627	8
5	6	Techline	yes	Health	2006	$13,898,119	5,470,303 Dollars	8427816	23
6	7	Cityace	yes	Health	2010	$9,254,614	6,249,498 Dollars	3005116	6
7	8	Kayelectronics	yes	Health	2009	$9,451,943	3,878,113 Dollars	5573830	4
8	9	Ganzlax	yes	IT Services	2011	$14,001,180	916,455 Dollars	11901180	18
9	10	Trantraxlax	yes	Government Services	2011	$11,088,336	5,635,276 Dollars	5453060	7
10	11	no	no	no	no	no	no	no	r

In [15]:

```
# convert yes or no into NaN
df = pd.read_csv('sets/data.csv',na_values={'Status':'yes'})
df
```

Out[15]:

	ID	Name	Status	Industry	Inception	Revenue	Expenses	Profit	Grow
0	1	Lamtone	NaN	IT Services	2009	$11,757,018	6,482,465 Dollars	5274553	30
1	2	Stripfind	NaN	Financial Services	2010	$12,329,371	916,455 Dollars	11412916	20
2	3	Canecorporation	NaN	Health	2012	$10,597,009	7,591,189 Dollars	3005820	7
3	4	Mattouch	NaN	IT Services	2013	$14,026,934	7,429,377 Dollars	6597557	26
4	5	Techdrill	NaN	Health	2009	$10,573,990	7,435,363 Dollars	3138627	8
5	6	Techline	NaN	Health	2006	$13,898,119	5,470,303 Dollars	8427816	23
6	7	Cityace	NaN	Health	2010	$9,254,614	6,249,498 Dollars	3005116	6
7	8	Kayelectronics	NaN	Health	2009	$9,451,943	3,878,113 Dollars	5573830	4
8	9	Ganzlax	NaN	IT Services	2011	$14,001,180	916,455 Dollars	11901180	18
9	10	Trantraxlax	NaN	Government Services	2011	$11,088,336	5,635,276 Dollars	5453060	7
10	11	no	no	no	no	no	no	no	r

In [17]:

```
# convert yes or no into NaN
df1 = pd.read_csv('sets/data.csv',na_values='no')
df1
```

Out[17]:

	ID	Name	Status	Industry	Inception	Revenue	Expenses	Profit	Gro
0	1	Lamtone	yes	IT Services	2009.0	$11,757,018	6,482,465 Dollars	5274553.0	3
1	2	Stripfind	yes	Financial Services	2010.0	$12,329,371	916,455 Dollars	11412916.0	2
2	3	Canecorporation	yes	Health	2012.0	$10,597,009	7,591,189 Dollars	3005820.0	
3	4	Mattouch	yes	IT Services	2013.0	$14,026,934	7,429,377 Dollars	6597557.0	2
4	5	Techdrill	yes	Health	2009.0	$10,573,990	7,435,363 Dollars	3138627.0	
5	6	Techline	yes	Health	2006.0	$13,898,119	5,470,303 Dollars	8427816.0	2
6	7	Cityace	yes	Health	2010.0	$9,254,614	6,249,498 Dollars	3005116.0	
7	8	Kayelectronics	yes	Health	2009.0	$9,451,943	3,878,113 Dollars	5573830.0	
8	9	Ganzlax	yes	IT Services	2011.0	$14,001,180	916,455 Dollars	11901180.0	1
9	10	Trantraxlax	yes	Government Services	2011.0	$11,088,336	5,635,276 Dollars	5453060.0	
10	11	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N

In [ ]:

```
```

In [19]:

```
# convert multiple columns
df1 = pd.read_csv('sets/data.csv')
df1
```

Out[19]:

	ID	Name	Smoke	Status	Industry	Inception	Revenue	Expenses	Profit
0	1	Lamtone	no	yes	IT Services	2009	$11,757,018	6,482,465 Dollars	5274553
1	2	Stripfind	no	yes	Financial Services	2010	$12,329,371	916,455 Dollars	11412916
2	3	Canecorporation	no	yes	Health	2012	$10,597,009	7,591,189 Dollars	3005820
3	4	Mattouch	no	yes	IT Services	2013	$14,026,934	7,429,377 Dollars	6597557
4	5	Techdrill	no	yes	Health	2009	$10,573,990	7,435,363 Dollars	3138627
5	6	Techline	no	yes	Health	2006	$13,898,119	5,470,303 Dollars	8427816
6	7	Cityace	no	yes	Health	2010	$9,254,614	6,249,498 Dollars	3005116
7	8	Kayelectronics	no	yes	Health	2009	$9,451,943	3,878,113 Dollars	5573830
8	9	Ganzlax	no	yes	IT Services	2011	$14,001,180	916,455 Dollars	11901180
9	10	Trantraxlax	no	yes	Government Services	2011	$11,088,336	5,635,276 Dollars	5453060

In [21]:

```
df1 = pd.read_csv('sets/data.csv',na_values={'Smoke':'no','Status':'yes'})
df1
```

Out[21]:

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

In [ ]:

```

```

In [ ]:

```

```

In [22]:

```
# Keep Default Value
df2 = pd.read_csv('sets/data.csv')
df2
```

Out[22]:

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

In [24]:

```
# It Will Show Default Value
df2 = pd.read_csv('sets/data.csv',keep_default_na=False)
df2
```

Out[24]:

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

In [25]:

```
# It replace Default Value with NaN
df2 = pd.read_csv('sets/data.csv',keep_default_na=True)
df2
```

Out[25]:

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

In [ ]:

In [ ]:

In [ ]:

In [ ]: