Different way of creating Dataframe

From Empty Data Frame

In [3]:
```
import pandas as pd
```

In [6]:
```
dataset=pd.DataFrame()
dataset['Column']=[1,2,3,4,5]
dataset
```

Out[6]:

	Column
0	1
1	2
2	3
3	4
4	5

In [8]:
```
dataset['Column2']=[5,4,3,2,1]
dataset
```

Out[8]:

	Column	Column2
0	5	5
1	4	4
2	3	3
3	2	2
4	1	1

In [ ]:
```

```

From array of row datas

In [9]:
```
row_data=[['Name 1',30],['Name 2',35],['Name 3',40]]
columns=['Name','Age']

dataset=pd.DataFrame(row_data,columns=columns)
dataset
```

Out[9]:

	Name	Age
0	Name 1	30
1	Name 2	35
2	Name 3	40

In [11]:
```
dataset['Salary']=12000
dataset
```

Out[11]:

	Name	Age	Salary
0	Name 1	30	12000
1	Name 2	35	12000
2	Name 3	40	12000

In [ ]:
```

```

From Array of column datas

In [18]:
```
names=['Name 1','Name 2','Name 3']
slaries=[12000,15000,18000]
data=list(zip(names,slaries))
columns=['Name','Salary']
dataset=pd.DataFrame(data,columns=columns)
dataset
```

Out[18]:

	Name	Salary
0	Name 1	12000
1	Name 2	15000
2	Name 3	18000

In [15]:
```
list(zip(names,slaries))
```

Out[15]:
```
[('Name 1', 12000), ('Name 2', 15000), ('Name 3', 18000)]
```

From Array of column & row datas

In [19]:
```
data=[{'Name':'Name 1','Salary':12000},{'Name':'Name 2','Salary':15000}]
dataset=pd.DataFrame(data)
dataset
```

Out[19]:

	Name	Salary
0	Name 1	12000
1	Name 2	15000

In [ ]:
```

```

Import from excel data

In [24]:
```
dataset=pd.read_excel('BME222.xlsx',skiprows=[0,1,2],index_col=0)
dataset
```

Out[24]:

id	class_1	class_2	class_3	class_4	class_5	class_6	class_7	class_8	class_9	clas
1911001	P	P	P	A	P	P	P	P	P	
1911002	A	P	P	P	P	P	P	P	P	
1911003	P	P	P	P	P	P	P	A	P	
1911004	A	P	P	P	P	P	A	P	P	
1911005	P	P	A	P	P	P	P	A	P	
1911006	P	P	P	P	A	P	A	A	A	
1911007	P	P	P	A	P	P	P	P	P	
1911008	P	P	P	P	P	A	P	P	P	
1911009	A	P	A	A	P	P	P	P	P	
1911010	P	P	P	P	P	P	P	A	P	
1911011	A	P	A	P	P	P	A	P	P	
1911012	P	P	A	P	P	P	P	A	P	
1911013	P	P	P	P	A	P	A	P	P	
1911014	P	P	P	A	P	P	P	P	P	
1911015	P	P	A	P	P	P	P	A	P	
1911016	P	P	A	P	A	P	A	P	P	
1911017	P	P	P	A	P	P	P	P	A	
1911018	P	A	P	P	P	P	P	P	P	
1911019	A	P	P	A	P	P	P	P	A	
1911020	P	P	P	P	P	P	P	P	P	
1911021	P	P	P	P	P	P	P	P	P	
1911022	P	P	A	P	P	P	P	A	P	
1911023	P	P	P	P	P	P	A	A	A	
1911024	P	P	P	A	P	A	P	P	P	
1911025	P	P	A	P	P	P	P	A	P	
1911026	P	P	A	P	A	P	A	A	A	
1911027	P	P	P	P	P	A	P	P	P	
1911028	P	P	P	P	P	P	P	P	P	
1911029	A	P	P	A	P	P	A	P	P	
1911030	P	P	A	P	P	P	P	A	P	

In [ ]:
```

```

In [ ]:

Adding new column in a existing dataframe

In [25]:
```
dataset['class_13']='P'
dataset
```

Out[25]:

id	class_1	class_2	class_3	class_4	class_5	class_6	class_7	class_8	class_9	clas
1911001	P	P	P	A	P	P	P	P	P	
1911002	A	P	P	P	P	P	P	P	P	
1911003	P	P	P	P	P	P	P	A	P	
1911004	A	P	P	P	P	P	A	P	P	
1911005	P	P	A	P	P	P	P	A	P	
1911006	P	P	P	P	A	P	A	A	A	
1911007	P	P	P	A	P	P	P	P	P	
1911008	P	P	P	P	P	A	P	P	P	
1911009	A	P	A	A	P	P	P	P	P	
1911010	P	P	P	P	P	P	P	A	P	
1911011	A	P	A	P	P	P	A	P	P	
1911012	P	P	A	P	P	P	P	A	P	
1911013	P	P	P	P	A	P	A	P	P	
1911014	P	P	P	A	P	P	P	P	P	
1911015	P	P	A	P	P	P	P	A	P	
1911016	P	P	A	P	A	P	A	P	P	
1911017	P	P	P	A	P	P	P	P	A	
1911018	P	A	P	P	P	P	P	P	P	
1911019	A	P	P	A	P	P	P	P	A	
1911020	P	P	P	P	P	P	P	P	P	
1911021	P	P	P	P	P	P	P	P	P	
1911022	P	P	A	P	P	P	P	A	P	
1911023	P	P	P	P	P	P	A	A	A	
1911024	P	P	P	A	P	A	P	P	P	
1911025	P	P	A	P	P	P	P	A	P	
1911026	P	P	A	P	A	P	A	A	A	
1911027	P	P	P	P	P	A	P	P	P	
1911028	P	P	P	P	P	P	P	P	P	
1911029	A	P	P	A	P	P	A	P	P	
1911030	P	P	A	P	P	P	P	A	P	

In [ ]:
```

```

In [ ]:

Adding new row in existing dataframe

In [30]:
```
data=[{'class_1':'P','class_2':'P'}]
index=[1911031]
new_row=pd.DataFrame(data,index=index)
new_row
```

Out[30]:

	class_1	class_2
1911031	P	P

In [34]:
```
dataset=dataset.append(new_row,ignore_index=False)
dataset
```

Out[34]:

	class_1	class_2	class_3	class_4	class_5	class_6	class_7	class_8	class_9	clas
1911001	P	P	P	A	P	P	P	P	P	
1911002	A	P	P	P	P	P	P	P	P	
1911003	P	P	P	P	P	P	P	A	P	
1911004	A	P	P	P	P	P	A	P	P	
1911005	P	P	A	P	P	P	P	A	P	
1911006	P	P	P	P	A	P	A	A	A	
1911007	P	P	P	A	P	P	P	P	P	
1911008	P	P	P	P	P	A	P	P	P	
1911009	A	P	A	A	P	P	P	P	P	
1911010	P	P	P	P	P	P	P	A	P	
1911011	A	P	A	P	P	P	A	P	P	
1911012	P	P	A	P	P	P	P	A	P	
1911013	P	P	P	P	A	P	A	P	P	
1911014	P	P	P	A	P	P	P	P	P	
1911015	P	P	A	P	P	P	P	A	P	
1911016	P	P	A	P	A	P	A	P	P	
1911017	P	P	P	A	P	P	P	P	A	
1911018	P	A	P	P	P	P	P	P	P	
1911019	A	P	P	A	P	P	P	P	A	
1911020	P	P	P	P	P	P	P	P	P	
1911021	P	P	P	P	P	P	P	P	P	
1911022	P	P	A	P	P	P	P	A	P	
1911023	P	P	P	P	P	P	A	A	A	
1911024	P	P	P	A	P	A	P	P	P	
1911025	P	P	A	P	P	P	P	A	P	
1911026	P	P	A	P	A	P	A	A	A	
1911027	P	P	P	P	P	A	P	P	P	
1911028	P	P	P	P	P	P	P	P	P	
1911029	A	P	P	A	P	P	A	P	P	
1911030	P	P	A	P	P	P	P	A	P	
1911031	P	P	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1911031	P	P	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1911031	P	P	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1911031	P	P	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

In [ ]:

In [ ]:

Access Data

In [39]:
```
dataset.loc[1911001:1911010]
```

Out[39]:

	class_1	class_2	class_3	class_4	class_5	class_6	class_7	class_8	class_9	clas
1911001	P	P	P	A	P	P	P	P	P	
1911002	A	P	P	P	P	P	P	P	P	
1911003	P	P	P	P	P	P	P	A	P	
1911004	A	P	P	P	P	P	A	P	P	
1911005	P	P	A	P	P	P	P	A	P	
1911006	P	P	P	P	A	P	A	A	A	
1911007	P	P	P	A	P	P	P	P	P	
1911008	P	P	P	P	P	A	P	P	P	
1911009	A	P	A	A	P	P	P	P	P	
1911010	P	P	P	P	P	P	P	A	P	

In [43]:
```
dataset.class_2  # also
```

Out[43]:
```
1911001    P
1911002    P
1911003    P
1911004    P
1911005    P
1911006    P
1911007    P
1911008    P
1911009    P
1911010    P
1911011    P
1911012    P
1911013    P
1911014    P
1911015    P
1911016    P
1911017    P
1911018    A
1911019    P
1911020    P
1911021    P
1911022    P
1911023    P
1911024    P
1911025    P
1911026    P
1911027    P
1911028    P
1911029    P
1911030    P
1911031    P
1911031    P
1911031    P
1911031    P
Name: class_2, dtype: object
```

Update Data

In [ ]:
```

```

In [45]:
```
dataset['class_1'][1911005]='A'
dataset
```

Out[45]:

	class_1	class_2	class_3	class_4	class_5	class_6	class_7	class_8	class_9	clas
1911001	P	P	P	A	P	P	P	P	P	
1911002	A	P	P	P	P	P	P	P	P	
1911003	P	P	P	P	P	P	P	A	P	
1911004	A	P	P	P	P	P	A	P	P	
1911005	A	P	A	P	P	P	P	A	P	
1911006	P	P	P	P	A	P	A	A	A	
1911007	P	P	P	A	P	P	P	P	P	
1911008	P	P	P	P	P	A	P	P	P	
1911009	A	P	A	A	P	P	P	P	P	
1911010	P	P	P	P	P	P	P	A	P	
1911011	A	P	A	P	P	P	A	P	P	
1911012	P	P	A	P	P	P	P	A	P	
1911013	P	P	P	P	A	P	A	P	P	
1911014	P	P	P	A	P	P	P	P	P	
1911015	P	P	A	P	P	P	P	A	P	
1911016	P	P	A	P	A	P	A	P	P	
1911017	P	P	P	A	P	P	P	P	A	
1911018	P	A	P	P	P	P	P	P	P	
1911019	A	P	P	A	P	P	P	P	A	
1911020	P	P	P	P	P	P	P	P	P	
1911021	P	P	P	P	P	P	P	P	P	
1911022	P	P	A	P	P	P	P	A	P	
1911023	P	P	P	P	P	P	A	A	A	
1911024	P	P	P	A	P	A	P	P	P	
1911025	P	P	A	P	P	P	P	A	P	
1911026	P	P	A	P	A	P	A	A	A	
1911027	P	P	P	P	P	A	P	P	P	
1911028	P	P	P	P	P	P	P	P	P	
1911029	A	P	P	A	P	P	A	P	P	
1911030	P	P	A	P	P	P	P	A	P	
1911031	P	P	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1911031	P	P	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1911031	P	P	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1911031	P	P	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

In [ ]:

Drop Row

In [53]:
```
dataset=dataset.drop([1911001])
dataset
```

Out[53]:

	class_1	class_2	class_3	class_4	class_5	class_6	class_7	class_8	class_9	clas
1911002	A	P	P	P	P	P	P	P	P	
1911003	P	P	P	P	P	P	P	A	P	
1911004	A	P	P	P	P	P	A	P	P	
1911005	A	P	A	P	P	P	P	A	P	
1911006	P	P	P	P	A	P	A	A	A	
1911007	P	P	P	A	P	P	P	P	P	
1911008	P	P	P	P	P	A	P	P	P	
1911009	A	P	A	A	P	P	P	P	P	
1911010	P	P	P	P	P	P	P	A	P	
1911011	A	P	A	P	P	P	A	P	P	
1911012	P	P	A	P	P	P	P	A	P	
1911013	P	P	P	P	A	P	A	P	P	
1911014	P	P	P	A	P	P	P	P	P	
1911015	P	P	A	P	P	P	P	A	P	
1911016	P	P	A	P	A	P	A	P	P	
1911017	P	P	P	A	P	P	P	P	A	
1911018	P	A	P	P	P	P	P	P	P	
1911019	A	P	P	A	P	P	P	P	A	
1911020	P	P	P	P	P	P	P	P	P	
1911021	P	P	P	P	P	P	P	P	P	
1911022	P	P	A	P	P	P	P	A	P	
1911023	P	P	P	P	P	P	A	A	A	
1911024	P	P	P	A	P	A	P	P	P	
1911025	P	P	A	P	P	P	P	A	P	
1911026	P	P	A	P	A	P	A	A	A	
1911027	P	P	P	P	P	A	P	P	P	
1911028	P	P	P	P	P	P	P	P	P	
1911029	A	P	P	A	P	P	A	P	P	
1911030	P	P	A	P	P	P	P	A	P	
1911031	P	P	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1911031	P	P	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1911031	P	P	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1911031	P	P	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

In [ ]:

In [79]:
```
dataset[['class_1','class_2']]
```

Out[79]:

	class_1	class_2
1911002	A	P
1911003	P	P
1911004	A	P
1911005	A	P
1911006	P	P
1911007	P	P
1911008	P	P
1911009	A	P
1911010	P	P
1911011	A	P
1911012	P	P
1911013	P	P
1911014	P	P
1911015	P	P
1911016	P	P
1911017	P	P
1911018	P	A
1911019	A	P
1911020	P	P
1911021	P	P
1911022	P	P
1911023	P	P
1911024	P	P
1911025	P	P
1911026	P	P
1911027	P	P
1911028	P	P
1911029	A	P
1911030	P	P
1911031	P	P
1911031	P	P
1911031	P	P
1911031	P	P

Query in Dataframe

In [80]:
```
dataset[(dataset['class_1']=='P') & (dataset['class_2']=='P')][['class_
1','class_2']]
```

Out[80]:

	class_1	class_2
1911003	P	P
1911006	P	P
1911007	P	P
1911008	P	P
1911010	P	P
1911012	P	P
1911013	P	P
1911014	P	P
1911015	P	P
1911016	P	P
1911017	P	P
1911020	P	P
1911021	P	P
1911022	P	P
1911023	P	P
1911024	P	P
1911025	P	P
1911026	P	P
1911027	P	P
1911028	P	P
1911030	P	P
1911031	P	P
1911031	P	P
1911031	P	P
1911031	P	P

In [ ]:
```

```

Example of query

In [69]:
```
len(list(dataset[dataset['class_1']=='P'].index))
```

Out[69]:
```
26
```

In [75]:
```
list(dataset.loc[1911013])
```

Out[75]:
```
['P', 'P', 'P', 'P', 'A', 'P', 'A', 'P', 'P', 'P', 'A', 'P', 'P']
```

In [74]:
```
list(dataset.loc[1911013]).count('A')
```

Out[74]:
```
3
```

In [ ]:
```

```