

Data Manipulation

In [1]: `#Exp no.: 2`

In [2]: `#Aim: Data Manipulation`

In [3]: `#Name: Swapnil Rahul Wankhade
#Roll no.: 73
#Sec: B
#Subject: Data Science and Statistics (Lab 1)`

In [4]: `import pandas as pd`

In [5]: `import os`

In [6]: `os.getcwd()`

Out[6]: `'C:\\Users\\hp\\Downloads'`

In [7]: `os.chdir('C:\\Users\\hp\\Desktop')`

In [8]: `data=pd.read_csv('diabetes.csv')`

In [9]: `data.head(10)`

Out[9]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction
0	6	148	72	35	0	33.6	0.627
1	1	85	66	29	0	26.6	0.351
2	8	183	64	0	0	23.3	0.672
3	1	89	66	23	94	28.1	0.167
4	0	137	40	35	168	43.1	2.288
5	5	116	74	0	0	25.6	0.201
6	3	78	50	32	88	31.0	0.248
7	10	115	0	0	0	35.3	0.134
8	2	197	70	45	543	30.5	0.158
9	8	125	96	0	0	0.0	0.232

```
In [10]: data.tail()
```

```
Out[10]:
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction
763	10	101	76	48	180	32.9	0.17
764	2	122	70	27	0	36.8	0.34
765	5	121	72	23	112	26.2	0.24
766	1	126	60	0	0	30.1	0.34
767	1	93	70	31	0	30.4	0.31

```
In [11]: data.shape
```

```
Out[11]: (768, 9)
```

```
In [12]: data.size
```

```
Out[12]: 6912
```

```
In [13]: data.ndim
```

```
Out[13]: 2
```

```
In [14]: data.columns
```

```
Out[14]: Index(['Pregnancies', 'Glucose', 'BloodPressure', 'SkinThickness', 'Insulin',  
               'BMI', 'DiabetesPedigreeFunction', 'Age', 'Outcome'],  
              dtype='object')
```

```
In [15]: data.head()
```

```
Out[15]:
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction
0	6	148	72	35	0	33.6	0.627
1	1	85	66	29	0	26.6	0.351
2	8	183	64	0	0	23.3	0.672
3	1	89	66	23	94	28.1	0.167
4	0	137	40	35	168	43.1	2.288

```
In [16]: data.drop(labels="Age",axis=1)
```

```
Out[16]:
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction
0	6	148	72	35	0	33.6	0.627
1	1	85	66	29	0	26.6	0.351
2	8	183	64	0	0	23.3	0.672
3	1	89	66	23	94	28.1	0.167
4	0	137	40	35	168	43.1	2.288
...
763	10	101	76	48	180	32.9	0.171
764	2	122	70	27	0	36.8	0.340
765	5	121	72	23	112	26.2	0.245
766	1	126	60	0	0	30.1	0.349
767	1	93	70	31	0	30.4	0.315

768 rows × 8 columns



```
In [17]: data.drop(labels=["Age","Glucose"],axis=1)
```

```
Out[17]:
```

	Pregnancies	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Outcome
0	6	72	35	0	33.6	0.627	
1	1	66	29	0	26.6	0.351	
2	8	64	0	0	23.3	0.672	
3	1	66	23	94	28.1	0.167	
4	0	40	35	168	43.1	2.288	
...
763	10	76	48	180	32.9	0.171	
764	2	70	27	0	36.8	0.340	
765	5	72	23	112	26.2	0.245	
766	1	60	0	0	30.1	0.349	
767	1	70	31	0	30.4	0.315	

768 rows × 7 columns



```
In [18]: data.head(10)
```

```
Out[18]:
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction
0	6	148	72	35	0	33.6	0.627
1	1	85	66	29	0	26.6	0.351
2	8	183	64	0	0	23.3	0.672
3	1	89	66	23	94	28.1	0.167
4	0	137	40	35	168	43.1	2.288
5	5	116	74	0	0	25.6	0.201
6	3	78	50	32	88	31.0	0.248
7	10	115	0	0	0	35.3	0.134
8	2	197	70	45	543	30.5	0.158
9	8	125	96	0	0	0.0	0.232

```
In [19]: data.drop(labels=[2,3],axis=0)
```

```
Out[19]:
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction
0	6	148	72	35	0	33.6	0.627
1	1	85	66	29	0	26.6	0.351
4	0	137	40	35	168	43.1	2.288
5	5	116	74	0	0	25.6	0.201
6	3	78	50	32	88	31.0	0.248
...
763	10	101	76	48	180	32.9	0.171
764	2	122	70	27	0	36.8	0.342
765	5	121	72	23	112	26.2	0.242
766	1	126	60	0	0	30.1	0.342
767	1	93	70	31	0	30.4	0.341

766 rows × 9 columns