

```
In [1]: # Experiment No: 4

In [2]: # Aim:Data Manipulation using Pandas

In [3]: # Name: Sakshi Padmakar Yeole

In [4]: # Class: 3rd year(B)

In [5]: # Roll No: 69

In [6]: # Date:26th July 2024

In [7]: #importing the basic library
import pandas as pd

In [8]: import os

In [9]: os.getcwd()

Out[9]: 'C:\\Users\\hp'
```

```
In [11]: os.chdir('C:\\Users\\hp\\OneDrive\\Desktop')
```

```
In [12]: data=pd.read_csv("diabetes.csv")
```

```
In [13]: data.head()
```

Out[13]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

```
In [14]: data.tail()
```

Out[14]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
763	10	101	76	48	180	32.9	0.171	63	0
764	2	122	70	27	0	36.8	0.340	27	0
765	5	121	72	23	112	26.2	0.245	30	0
766	1	126	60	0	0	30.1	0.349	47	1
767	1	93	70	31	0	30.4	0.315	23	0

```
In [15]: data.head(50)
```

Out[15]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1
5	5	116	74	0	0	25.6	0.201	30	0

6	3	78	50	32	88	31.0	0.248	26	1
7	10	115	0	0	0	35.3	0.134	29	0
8	2	197	70	45	543	30.5	0.158	53	1
9	8	125	96	0	0	0.0	0.232	54	1
10	4	110	92	0	0	37.6	0.191	30	0
11	10	168	74	0	0	38.0	0.537	34	1
12	10	139	80	0	0	27.1	1.441	57	0
13	1	189	60	23	846	30.1	0.398	59	1
14	5	166	72	19	175	25.8	0.587	51	1
15	7	100	0	0	0	30.0	0.484	32	1
16	0	118	84	47	230	45.8	0.551	31	1
17	7	107	74	0	0	29.6	0.254	31	1
18	1	103	30	38	83	43.3	0.183	33	0
19	1	115	70	30	96	34.6	0.529	32	1
20	3	126	88	41	235	39.3	0.704	27	0
21	8	99	84	0	0	35.4	0.388	50	0
22	7	196	90	0	0	39.8	0.451	41	1
23	9	119	80	35	0	29.0	0.263	29	1
24	11	143	94	33	146	36.6	0.254	51	1
25	10	125	70	26	115	31.1	0.205	41	1
26	7	147	76	0	0	39.4	0.257	43	1
27	1	97	66	15	140	23.2	0.487	22	0
28	13	145	82	19	110	22.2	0.245	57	0
29	5	117	92	0	0	34.1	0.337	38	0
30	5	109	75	26	0	36.0	0.546	60	0
31	3	158	76	36	245	31.6	0.851	28	1
32	3	88	58	11	54	24.8	0.267	22	0
33	6	92	92	0	0	19.9	0.188	28	0
34	10	122	78	31	0	27.6	0.512	45	0
35	4	103	60	33	192	24.0	0.966	33	0
36	11	138	76	0	0	33.2	0.420	35	0
37	9	102	76	37	0	32.9	0.665	46	1
38	2	90	68	42	0	38.2	0.503	27	1
39	4	111	72	47	207	37.1	1.390	56	1
40	3	180	64	25	70	34.0	0.271	26	0
41	7	133	84	0	0	40.2	0.696	37	0
42	7	106	92	18	0	22.7	0.235	48	0
43	9	171	110	24	240	45.4	0.721	54	1
44	7	159	64	0	0	27.4	0.294	40	0
45	0	180	66	39	0	42.0	1.893	25	1
46	1	146	56	0	0	29.7	0.564	29	0
47	2	71	70	27	0	28.0	0.586	22	0
48	7	103	66	32	0	39.1	0.344	31	1
49	7	105	0	0	0	0.0	0.305	24	0

In [16]: `data.shape`

Out[16]: (768, 9)

In [17]: `data.size`

Out[17]: 6912

In [19]:

data.columns

Out[19]:

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

In [20]:

data.head()

Out[20]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

In [21]:

data.drop(labels="Age",axis=1)

Out[21]:

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

768 rows × 8 columns

In [22]:

data.drop(labels=["Age","Glucose"],axis=1)

Out[22]:

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

768 rows × 7 columns

In [23]:

data.head(10)

Out[23]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1

1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1
5	5	116	74	0	0	25.6	0.201	30	0
6	3	78	50	32	88	31.0	0.248	26	1
7	10	115	0	0	0	35.3	0.134	29	0
8	2	197	70	45	543	30.5	0.158	53	1
9	8	125	96	0	0	0.0	0.232	54	1

In [24]:

data.head(10)

Out[24]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1
5	5	116	74	0	0	25.6	0.201	30	0
6	3	78	50	32	88	31.0	0.248	26	1
7	10	115	0	0	0	35.3	0.134	29	0
8	2	197	70	45	543	30.5	0.158	53	1
9	8	125	96	0	0	0.0	0.232	54	1

In [25]:

data.drop(labels=[2,3],axis=0)

Out[25]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
4	0	137	40	35	168	43.1	2.288	33	1
5	5	116	74	0	0	25.6	0.201	30	0
6	3	78	50	32	88	31.0	0.248	26	1
...
763	10	101	76	48	180	32.9	0.171	63	0
764	2	122	70	27	0	36.8	0.340	27	0
765	5	121	72	23	112	26.2	0.245	30	0
766	1	126	60	0	0	30.1	0.349	47	1
767	1	93	70	31	0	30.4	0.315	23	0

766 rows × 9 columns

In []:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js