

# DATA MANIPULATION

*#Aim : To perform operation of Data Manipulation on Data set.*

*# Name : Samruddhi Umap*

*# Roll no. :67*

*# Sec : C*

*# Subject : ET1*

*#date: 02/08/2024*

```
import pandas as pd
```

```
import os
```

```
os.getcwd()
```

```
'C:\\Users\\HP'
```

```
os.chdir("C:\\Users\\HP\\Desktop")
```

```
data = pd.read_csv("diabetes.csv")
```

```
data.head(10)
```

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

	DiabetesPedigreeFunction	Age	Outcome
0	0.627	50	1
1	0.351	31	0

2	0.672	32	1
3	0.167	21	0
4	2.288	33	1
5	0.201	30	0
6	0.248	26	1
7	0.134	29	0
8	0.158	53	1
9	0.232	54	1

data.tail

<bound method NDataFrame.tail of SkinThickness	Insulin	BMI \	Pregnancies	Glucose	BloodPressure
0	6	148	72	35	0 33.6
1	1	85	66	29	0 26.6
2	8	183	64	0	0 23.3
3	1	89	66	23	94 28.1
4	0	137	40	35	168 43.1
..	...	...	...	...	...
763	10	101	76	48	180 32.9
764	2	122	70	27	0 36.8
765	5	121	72	23	112 26.2
766	1	126	60	0	0 30.1
767	1	93	70	31	0 30.4

	DiabetesPedigreeFunction	Age	Outcome
0	0.627	50	1
1	0.351	31	0
2	0.672	32	1
3	0.167	21	0
4	2.288	33	1
..	...	...	...
763	0.171	63	0
764	0.340	27	0
765	0.245	30	0
766	0.349	47	1
767	0.315	23	0

[768 rows x 9 columns]>

```
data.shape
(768, 9)
data.size
6912
data.ndim
2
data.columns
Index(['Pregnancies', 'Glucose', 'BloodPressure', 'SkinThickness',
       'Insulin',
       'BMI', 'DiabetesPedigreeFunction', 'Age', 'Outcome'],
      dtype='object')
```

```
data.head()
```

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

	DiabetesPedigreeFunction	Age	Outcome
0	0.627	50	1
1	0.351	31	0
2	0.672	32	1
3	0.167	21	0
4	2.288	33	1

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

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI
0	6	148	72	35	0	33.6
1	1	85	66	29	0	26.6
2	8	183	64	0	0	23.3
3	1	89	66	23	94	28.1

4	0	137	40	35	168	43.1
..	...	...	...	...	...	...
763	10	101	76	48	180	32.9
764	2	122	70	27	0	36.8
765	5	121	72	23	112	26.2
766	1	126	60	0	0	30.1
767	1	93	70	31	0	30.4

	DiabetesPedigreeFunction	Outcome
0	0.627	1
1	0.351	0
2	0.672	1
3	0.167	0
4	2.288	1
..	...	...
763	0.171	0
764	0.340	0
765	0.245	0
766	0.349	1
767	0.315	0

[768 rows x 8 columns]

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

	Pregnancies	BloodPressure	SkinThickness	Insulin	BMI	\
0	6	72	35	0	33.6	
1	1	66	29	0	26.6	
2	8	64	0	0	23.3	
3	1	66	23	94	28.1	
4	0	40	35	168	43.1	
..	...	...	...	...	...	...
763	10	76	48	180	32.9	
764	2	70	27	0	36.8	
765	5	72	23	112	26.2	
766	1	60	0	0	30.1	
767	1	70	31	0	30.4	

	DiabetesPedigreeFunction	Outcome
0	0.627	1
1	0.351	0
2	0.672	1
3	0.167	0
4	2.288	1

```

..      ...
763      0.171      0
764      0.340      0
765      0.245      0
766      0.349      1
767      0.315      0

```

```
[768 rows x 7 columns]
```

```
data.drop(labels = 1,axis =0)
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI
\						
0	6	148	72	35	0	33.6
2	8	183	64	0	0	23.3
3	1	89	66	23	94	28.1
4	0	137	40	35	168	43.1
5	5	116	74	0	0	25.6
..	...	...	...	...	...	...
763	10	101	76	48	180	32.9
764	2	122	70	27	0	36.8
765	5	121	72	23	112	26.2
766	1	126	60	0	0	30.1
767	1	93	70	31	0	30.4

	DiabetesPedigreeFunction	Age	Outcome
0	0.627	50	1
2	0.672	32	1
3	0.167	21	0
4	2.288	33	1
5	0.201	30	0
..	...	...	...
763	0.171	63	0
764	0.340	27	0
765	0.245	30	0
766	0.349	47	1
767	0.315	23	0

```
[767 rows x 9 columns]
```

```
data.head(10)
```

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

	DiabetesPedigreeFunction	Age	Outcome
0	0.627	50	1
1	0.351	31	0
2	0.672	32	1
3	0.167	21	0
4	2.288	33	1
5	0.201	30	0
6	0.248	26	1
7	0.134	29	0
8	0.158	53	1
9	0.232	54	1

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

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI
\						
0	6	148	72	35	0	33.6
1	1	85	66	29	0	26.6
4	0	137	40	35	168	43.1
5	5	116	74	0	0	25.6
6	3	78	50	32	88	31.0

..	...	...	...	...	...	...
763	10	101	76	48	180	32.9
764	2	122	70	27	0	36.8
765	5	121	72	23	112	26.2
766	1	126	60	0	0	30.1
767	1	93	70	31	0	30.4

	DiabetesPedigreeFunction	Age	Outcome
0	0.627	50	1
1	0.351	31	0
4	2.288	33	1
5	0.201	30	0
6	0.248	26	1
...	...	...	...
763	0.171	63	0
764	0.340	27	0
765	0.245	30	0
766	0.349	47	1
767	0.315	23	0

```
[766 rows x 9 columns]
```