

# Data Manipulation using pandas

In [36]: `#Exp 4`

In [4]: `#Name: Leena Rajeshwar Kale  
#Roll No.:71  
#Sec: C  
#Subject:ET - 1`

In [1]: `import pandas as pd`

In [2]: `import os`

In [3]: `os.getcwd()`

Out[3]: `'C:\\Users\\dishu'`

In [4]: `os.chdir("C:\\Users\\dishu")`

In [5]: `df=pd.read_csv("diabetes.csv")`

In [23]: `df.head()`

Out[23]:

|   | Pregnancies | Glucose | BloodPressure | SkinThickness | Insulin | BMI  | DiabetesPedigreeFunction | A |
|---|-------------|---------|---------------|---------------|---------|------|--------------------------|---|
| 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 [25]: `df.tail()`

Out[25]:

|     | Pregnancies | Glucose | BloodPressure | SkinThickness | Insulin | BMI  | DiabetesPedigreeFunction |  |
|-----|-------------|---------|---------------|---------------|---------|------|--------------------------|--|
| 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                    |  |

In [27]: `df.shape`

Out[27]: `(768, 9)`

In [29]: `df.size`

Out[29]: 6912

In [31]: `df.ndim`

Out[31]: 2

In [33]: `df.isnull().sum()` *# Summarize the number of missing values per column*

Out[33]:

|                          |   |
|--------------------------|---|
| Pregnancies              | 0 |
| Glucose                  | 0 |
| BloodPressure            | 0 |
| SkinThickness            | 0 |
| Insulin                  | 0 |
| BMI                      | 0 |
| DiabetesPedigreeFunction | 0 |
| Age                      | 0 |
| Outcome                  | 0 |
| dtype: int64             |   |

In [ ]: