

Data Acquisition

```
In [2]: #aim: To perform operation on Data acquisition using pandas.
```

```
In [3]: #Name:Sahil A. Bankar  
#Roll no:04  
#Sec:B  
#Subject:ET1  
#Date:24/07/2025
```

```
In [4]: #importing the basic Library  
import pandas as pd
```

```
In [5]: import os
```

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

```
Out[6]: 'C:\\Users\\DELL'
```

```
In [8]: os.chdir('C:\\Users\\DELL\\Desktop')
```

```
In [10]: data=pd.read_csv("diabetes.csv")
```

```
In [13]: data.head(30)
```

```
Out[13]: Pregnancies Glucose BloodPressure SkinThickness Insulin BMI DiabetesPedigreeFunction Age  
0 6 148 72 35 0 33.6 0.627 50  
1 1 85 66 29 0 26.6 0.351 31  
2 8 183 64 0 0 23.3 0.672 39  
3 1 89 66 23 94 28.1 0.167 21  
4 0 137 40 35 168 43.1 2.288 31  
5 5 116 74 0 0 25.6 0.201 31  
6 3 78 50 32 88 31.0 0.248 21  
7 10 115 0 0 0 35.3 0.134 29  
8 2 197 70 45 543 30.5 0.158 53  
9 8 125 96 0 0 0.0 0.232 54  
10 4 110 92 0 0 37.6 0.191 31  
11 10 168 74 0 0 38.0 0.537 34  
12 10 139 80 0 0 27.1 1.441 51  
13 1 189 60 23 846 30.1 0.398 51  
14 5 166 72 19 175 25.8 0.587 51  
15 7 100 0 0 0 30.0 0.484 31  
16 0 118 84 47 230 45.8 0.551 31  
17 7 107 74 0 0 29.6 0.254 31
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
18	1	103	30	38	83	43.3		0.183 33
19	1	115	70	30	96	34.6		0.529 31
20	3	126	88	41	235	39.3		0.704 29
21	8	99	84	0	0	35.4		0.388 50
22	7	196	90	0	0	39.8		0.451 41
23	9	119	80	35	0	29.0		0.263 28
24	11	143	94	33	146	36.6		0.254 51
25	10	125	70	26	115	31.1		0.205 40
26	7	147	76	0	0	39.4		0.257 41
27	1	97	66	15	140	23.2		0.487 24
28	13	145	82	19	110	22.2		0.245 51
29	5	117	92	0	0	34.1		0.337 38

In [14]: `data.tail(30)`

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
738	2	99	60	17	160	36.6		0.453 31
739	1	102	74	0	0	39.5		0.293 31
740	11	120	80	37	150	42.3		0.785 31
741	3	102	44	20	94	30.8		0.400 31
742	1	109	58	18	116	28.5		0.219 31
743	9	140	94	0	0	32.7		0.734 31
744	13	153	88	37	140	40.6		1.174 31
745	12	100	84	33	105	30.0		0.488 31
746	1	147	94	41	0	49.3		0.358 31
747	1	81	74	41	57	46.3		1.096 31
748	3	187	70	22	200	36.4		0.408 31
749	6	162	62	0	0	24.3		0.178 31
750	4	136	70	0	0	31.2		1.182 31
751	1	121	78	39	74	39.0		0.261 31
752	3	108	62	24	0	26.0		0.223 31
753	0	181	88	44	510	43.3		0.222 31
754	8	154	78	32	0	32.4		0.443 31
755	1	128	88	39	110	36.5		1.057 31
756	7	137	90	41	0	32.0		0.391 31
757	0	123	72	0	0	36.3		0.258 31
758	1	106	76	0	0	37.5		0.197 31

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
759	6	190	92	0	0	35.5		0.278
760	2	88	58	26	16	28.4		0.766
761	9	170	74	31	0	44.0		0.403
762	9	89	62	0	0	22.5		0.142
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 [15]: `data.size`

Out[15]: 6912

In [17]: `data.describe()`

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction
count	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	
mean	3.845052	120.894531	69.105469	20.536458	79.799479	31.992578	
std	3.369578	31.972618	19.355807	15.952218	115.244002	7.884160	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	1.000000	99.000000	62.000000	0.000000	0.000000	27.300000	
50%	3.000000	117.000000	72.000000	23.000000	30.500000	32.000000	
75%	6.000000	140.250000	80.000000	32.000000	127.250000	36.600000	
max	17.000000	199.000000	122.000000	99.000000	846.000000	67.100000	

Conclusion:

In this practical, I acquired a CSV file and performed basic data exploration using head, tail, size, describe(), and shape functions. These operations provided an initial understanding of the dataset's structure and key statistics. This

In []: