November 1, 2023

1 Data Manipulation

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

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

[3]:
```
#Name: Vedant Wankhade
#Roll no.: 74
#Sec: B
#Subject: Data Science and Statistics (Lab 1)
```

[4]:
```
import pandas as pd
```

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

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

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

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

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

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

[9]:

	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
```

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

[10]:
```
     Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin   BMI  \
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
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
```

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

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

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

[12]:
```
6912
```

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

[13]:
```
2
```

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

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

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

[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

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

[16]:

	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 x 8 columns]

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

[17]:
```
     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]
```

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

[18]:
```
   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
```

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

[19]:
```
     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]
```