pima-indian

July 19, 2024

Name : Mayuresh Bhiva Dhuri
Class : TY.BTECH.CSE
Panel : F
PRN : 1032222712

[11]:
```
import pandas as pd

df = pd.read_csv('diabetes.csv')
print(df)
```

```
     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 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]
```

[8]:
```
df.head()
```

[8]:
```
   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 Outcome0 0.627 50                        1

1 0.351 31 02 0.672 32                                                1

3 0.167 21 04 2.288 33                                                1
```

[9]:
```
df.tail()
```

[9]:
```
     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
```

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

[10]:

	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			
11	10	168	74	0	0	38.0			
12	10	139	80	0	0	27.1			
13	1	189	60	23	846	30.1			
14	5	166	72	19	175	25.8			
15	7	100	0	0	0	30.0			
16	0	118	84	47	230	45.8			
17	7	107	74	0	0	29.6			

```
10                  0.191   30    0
11                  0.537   34    1
12                  1.441   57    0
13                  0.398   59    1
14                  0.587   51    1
15                  0.484   32    1
16                  0.551   31    1
17                  0.254   31    1
```

[12]:
```
df.info()
```

```
<class
'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to
767 Data columns (total 9
columns):
#   Column                 Non-Null Count Dtype
--- ------                 -------------- -----
0   Pregnancies            768 non-null  int64
1   Glucose                768 non-null  int64
2   BloodPressure          768 non-null  int64
3   SkinThickness          768 non-null  int64
4   Insulin                768 non-null  int64
5   BMI                    768 non-null  float64
6  DiabetesPedigreeFunction 768 non-null float64
7   Age                    768 non-null  int64
 8   Outcome    768 non-null dtypes:    int64
float64(2), int64(7) memory usage:
54.1 KB
```

[46]:
```
df.describe()
```

[46]:
```
       Pregnancies  Glucose BloodPressure SkinThickness   Insulin \
count 768.000000 768.000000   768.000000    768.000000 768.000000
mean    3.845052 120.894531    69.105469     20.536458 79.799479
std     3.369578 31.972618     19.355807     15.952218 115.244002
min     0.000000   0.000000     0.000000    0.000000    0.000000
25%     1.000000 99.000000     62.000000    0.000000    0.000000
50%     3.000000 117.000000    72.000000     23.000000 30.500000
75%     6.000000 140.250000    80.000000     32.000000 127.250000
max    17.000000 199.000000    122.000000     99.000000 846.000000

            BMI DiabetesPedigreeFunction     Age     Outcome
count 768.000000            768.000000 768.000000 768.000000
```

```
mean   31.992578                  0.471876 33.240885   0.348958
std     7.884160                  0.331329 11.760232   0.476951
min     0.000000                  0.078000 21.000000   0.000000
25%    27.300000                  0.243750 24.000000   0.000000
50%    32.000000                  0.372500 29.000000   0.000000
75%    36.600000                  0.626250 41.000000   1.000000
max    67.100000                  2.420000 81.000000   1.000000
```

[15]:
```
print(df.shape)
```

```
(768, 9)
```

[25]:
```
print(df['BMI'])
```

```
0      33.6
1      26.6
2      23.3
3      28.1
4      43.1
      ...
763    32.9
764    36.8
765    26.2
766    30.1
767    30.4
Name: BMI, Length: 768, dtype: float64
```

[37]:
```
df.iloc[:,4]
```

```
[37]: 0      0
1      0
2      0
3     94
4    168
...
763    180
764      0
765    112
766      0
767      0
```

```
Name: Insulin, Length: 768, dtype: int64
```

[43]:
```
feat_5 = df.columns[4]
print(feat_5)
```

```
Insulin
```

[45]:
```

```

[45]:
```
       Pregnancies  Glucose  BloodPressure  SkinThickness     Insulin \
count   768.000000 768.000000     768.000000     768.000000 768.000000
mean      3.845052 120.894531      69.105469      20.536458  79.799479
std       3.369578  31.972618      19.355807      15.952218 115.244002
min       0.000000   0.000000       0.000000       0.000000   0.000000
25%       1.000000  99.000000      62.000000       0.000000   0.000000
50%       3.000000 117.000000      72.000000      23.000000  30.500000
75%       6.000000 140.250000      80.000000      32.000000 127.250000
max      17.000000 199.000000     122.000000      99.000000 846.000000

              BMI  DiabetesPedigreeFunction         Age     Outcome
count  768.000000                768.000000  768.000000  768.000000
mean    31.992578                  0.471876   33.240885    0.348958
std      7.884160                  0.331329   11.760232    0.476951
min      0.000000                  0.078000   21.000000    0.000000
25%     27.300000                  0.243750   24.000000    0.000000
50%     32.000000                  0.372500   29.000000    0.000000
75%     36.600000                  0.626250   41.000000    1.000000
max     67.100000                  2.420000   81.000000    1.000000
```

[48]:
```
df.mean()
```

[48]:
```
Pregnancies 3.845052 Glucose
 120.894531
BloodPressure              69.105469
SkinThickness              20.536458
Insulin                    79.799479
BMI                        31.992578
DiabetesPedigreeFunction    0.471876
Age                        33.240885
Outcome                     0.348958
dtype: float64
```

[49]:
```
df.mean(axis=1)
```

[49]:
```
0    38.469667
1    26.550111
2    34.663556
3    35.807444
```

```
4     51.043111
        …
763   56.785667
764   31.682222
765   43.271667
766   29.494333
767   27.635000
Length: 768, dtype: float64
```

[78]:
```
df.median()
```

```
[78]: Pregnancies 3.0000 Glucose 117.0000
BloodPressure              72.0000
SkinThickness              23.0000
Insulin                    30.5000
BMI                        32.0000
DiabetesPedigreeFunction    0.3725
Age                        29.0000
Outcome                     0.0000
dtype: float64
```

[79]:
```
df.mean(axis=1)
```

```
[79]: 0    38.469667
1     26.550111
2     34.663556
3     35.807444
4     51.043111
        …
763   56.785667
764   31.682222
765   43.271667
766   29.494333
767   27.635000
Length: 768, dtype: float64
```

[63]:
```
x = df[["BMI","Age"]]
print(x.mean())
```

```
BMI   31.992578
Age   33.240885
dtype: float64
```

[65]:
```
df.mode()
```

```
[65]: Pregnancies Glucose BloodPressure SkinThickness Insulin BMI \
0         1.0    99    70.0  0.0   0.0 32.0
1         NaN    100   NaN   NaN   NaN   NaN
```

```
   DiabetesPedigreeFunction Age Outcome
0                     0.254 22.0    0.0
1                     0.258   NaN   NaN
```

[66]:
```
df.mode(axis=1)
```

[66]:
```
       0      1      2    3    4    5    6      7      8
0    0.0  0.627  1.000  6.0 33.6 35.0 50.0   72.0 148.0
1    0.0    NaN    NaN  NaN  NaN  NaN  NaN    NaN    NaN
2    0.0    NaN    NaN  NaN  NaN  NaN  NaN    NaN    NaN
3    0.0  0.167  1.000 21.0 23.0 28.1 66.0   89.0   94.0
4    0.0  1.000  2.288 33.0 35.0 40.0 43.1 137.0 168.0
..   …      …      …   …    …    …    …      …      …
763  0.0  0.171 10.000 32.9 48.0 63.0 76.0 101.0 180.0
764  0.0 27.000    NaN  NaN  NaN  NaN  NaN    NaN    NaN
765  0.0  0.245  5.000 23.0 26.2 30.0 72.0 112.0 121.0
766  0.0  1.000    NaN  NaN  NaN  NaN  NaN    NaN    NaN
767  0.0    NaN    NaN  NaN  NaN  NaN  NaN    NaN    NaN

[768 rows x 9 columns]
```

[67]:
```
df.std()
```

[67]:
```
Pregnancies 3.369578 Glucose
31.972618
BloodPressure             19.355807
SkinThickness    15.952218 Insulin
115.244002
BMI                        7.884160
DiabetesPedigreeFunction   0.331329
Age                       11.760232
Outcome                    0.476951
dtype: float64
```

[68]:
```
df.std(axis=1)
```

[68]:
```
0    48.296112
1    31.119744
2    59.585320
3    37.639873
4    60.541569
       …
763  57.890988
764  41.180224
765  46.980802
766  42.959124
767  33.710199
Length: 768, dtype: float64
```

[76]:
```
x = df[["BMI","Glucose"]]

print(x.std())
```

```
BMI        7.884160
Glucose    31.972618
dtype: float64
```

[81]:
```
df.isnull()
```

[81]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False
..	…	…	…	…	…	…	…	…	…
763	False	False	False	False	False	False	False	False	False
764	False	False	False	False	False	False	False	False	False
765	False	False	False	False	False	False	False	False	False
766	False	False	False	False	False	False	False	False	False
767	False	False	False	False	False	False	False	False	False

```
766                 False False    False
767                 False False    False

[768 rows x 9 columns]
```

[82]:
```
x = df["BMI"]
print(x.isnull())
```

```
0      False
1      False
2      False
3      False
4      False
       …
763    False
764    False
765    False
766    False
767    False
Name: BMI, Length: 768, dtype: bool
```

[ ]:
```

```