practical-2

April 15, 2025

[2]:
```
import pandas as pd
```

[3]:
```
from pandas import read_csv
```

[4]:
```
ssd = read_csv('StudentsPerformance_m.csv', header=None)
```

[5]:
```
print(ssd.describe())
```

```
              0        1             2         3     4    5    6    7
count      1001     1001          1001      1001  1001  999  997  999
unique        3        6             7         3     3   82   73   78
top      female  group C  some college  standard  none   65   72   74
freq        518      319           226       645   642   36   34   35
```

[7]:
```
from pandas import read_csv
```

[8]:
```
dataset = read_csv('pima_indians_diabetes.csv', header=None)
```

[9]:
```
print(dataset.describe())
```

```
                0           1           2           3           4           5  \
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

                6           7           8
count  768.000000  768.000000  768.000000
mean     0.471876   33.240885    0.348958
std      0.331329   11.760232    0.476951
min      0.078000   21.000000    0.000000
25%      0.243750   24.000000    0.000000
50%      0.372500   29.000000    0.000000
```

```
75%       0.626250   41.000000    1.000000
max       2.420000   81.000000    1.000000
```

[10]:
```
dataset.shape
```

[10]:
```
(768, 9)
```

[11]:
```
print(dataset.head(20))
```

```
     0    1   2   3    4     5      6   7  8
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  0.191  30  0
11  10  168  74   0    0  38.0  0.537  34  1
12  10  139  80   0    0  27.1  1.441  57  0
13   1  189  60  23  846  30.1  0.398  59  1
14   5  166  72  19  175  25.8  0.587  51  1
15   7  100   0   0    0  30.0  0.484  32  1
16   0  118  84  47  230  45.8  0.551  31  1
17   7  107  74   0    0  29.6  0.254  31  1
18   1  103  30  38   83  43.3  0.183  33  0
19   1  115  70  30   96  34.6  0.529  32  1
```

[12]:
```
dataset.shape
```

[12]:
```
(768, 9)
```

[13]:
```
dataset.head()
```

[13]:
```
   0    1   2   3    4     5      6   7  8
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
```

[14]:
```
num_missing = (dataset[[1,2,3,4,5]] == 0).sum()
```

[15]:
```
print(num_missing)
```

```
1      5
2     35
3    227
4    374
5     11
dtype: int64
```

[16]:
```
from numpy import nan
```

[17]:
```
dataset = read_csv('pima_indians_diabetes.csv', header=None)
```

[18]:
```
dataset[[1,2,3,4,5]] = dataset[[1,2,3,4,5]].replace(0, nan)
```

[19]:
```
print(dataset.isnull().sum())
```

```
0      0
1      5
2     35
3    227
4    374
5     11
6      0
7      0
8      0
dtype: int64
```

[20]:
```
dataset[[1,2,3,4,5]] = dataset[[1,2,3,4,5]].replace(0, nan)
```

[21]:
```
print(dataset.head(20))
```

```
     0      1     2     3      4     5      6   7  8
0    6  148.0  72.0  35.0    NaN  33.6  0.627  50  1
1    1   85.0  66.0  29.0    NaN  26.6  0.351  31  0
2    8  183.0  64.0   NaN    NaN  23.3  0.672  32  1
3    1   89.0  66.0  23.0   94.0  28.1  0.167  21  0
4    0  137.0  40.0  35.0  168.0  43.1  2.288  33  1
5    5  116.0  74.0   NaN    NaN  25.6  0.201  30  0
6    3   78.0  50.0  32.0   88.0  31.0  0.248  26  1
7   10  115.0   NaN   NaN    NaN  35.3  0.134  29  0
8    2  197.0  70.0  45.0  543.0  30.5  0.158  53  1
9    8  125.0  96.0   NaN    NaN   NaN  0.232  54  1
10   4  110.0  92.0   NaN    NaN  37.6  0.191  30  0
11  10  168.0  74.0   NaN    NaN  38.0  0.537  34  1
12  10  139.0  80.0   NaN    NaN  27.1  1.441  57  0
13   1  189.0  60.0  23.0  846.0  30.1  0.398  59  1
14   5  166.0  72.0  19.0  175.0  25.8  0.587  51  1
15   7  100.0   NaN   NaN    NaN  30.0  0.484  32  1
16   0  118.0  84.0  47.0  230.0  45.8  0.551  31  1
```

```
17   7  107.0  74.0   NaN    NaN  29.6  0.254  31  1
18   1  103.0  30.0  38.0   83.0  43.3  0.183  33  0
19   1  115.0  70.0  30.0   96.0  34.6  0.529  32  1
```

[ ]: