Name: Vinayak V Thayil

Roll No:AM.EN.U4CSE21161

```
#importing pandas
import pandas as p;
```

```
#importing csv files
df=p.read_csv('diabetes.csv')
```

```
#print value of df
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]
```

```
#prints last 9 values of df
df.tail(9)
```

| | Pregnancies | Glucose | BloodPressure | SkinThickness | Insulin | BMI | DiabetesPedigreeFunction | Age | Outcome |
|---|---|---|---|---|---|---|---|---|---|
| **759** | 6 | 190 | 92 | 0 | 0 | 35.5 | 0.278 | 66 | 1 |
| **760** | 2 | 88 | 58 | 26 | 16 | 28.4 | 0.766 | 22 | 0 |
| **761** | 9 | 170 | 74 | 31 | 0 | 44.0 | 0.403 | 43 | 1 |
| **762** | 9 | 89 | 62 | 0 | 0 | 22.5 | 0.142 | 33 | 0 |
| **763** | 10 | 101 | 76 | 48 | 180 | 32.9 | 0.171 | 63 | 0 |
| **764** | 2 | 122 | 70 | 27 | 0 | 36.8 | 0.340 | 27 | 0 |
| **765** | 5 | 121 | 72 | 23 | 112 | 26.2 | 0.245 | 30 | 0 |
| **766** | 1 | 126 | 60 | 0 | 0 | 30.1 | 0.349 | 47 | 1 |
| **767** | 1 | 93 | 70 | 31 | 0 | 30.4 | 0.315 | 23 | 0 |

```
#prints first 10 values of df
df.head(9)
```

|   | 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 |

```
#prints statistical values of df
df.describe()
```

|   | Pregnancies | Glucose | BloodPressure | SkinThickness | Insulin | BMI | DiabetesPedigreeFunction | Age | Outcome |
|---|---|---|---|---|---|---|---|---|---|
| **count** | 768.000000 | 768.000000 | 768.000000 | 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 | 0.471876 | 33.240885 | 0.348958 |
| **std** | 3.369578 | 31.972618 | 19.355807 | 15.952218 | 115.244002 | 7.884160 | 0.331329 | 11.760232 | 0.476951 |
| **min** | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.078000 | 21.000000 | 0.000000 |
| **25%** | 1.000000 | 99.000000 | 62.000000 | 0.000000 | 0.000000 | 27.300000 | 0.243750 | 24.000000 | 0.000000 |
| **50%** | 3.000000 | 117.000000 | 72.000000 | 23.000000 | 30.500000 | 32.000000 | 0.372500 | 29.000000 | 0.000000 |
| **75%** | 6.000000 | 140.250000 | 80.000000 | 32.000000 | 127.250000 | 36.600000 | 0.626250 | 41.000000 | 1.000000 |
| **max** | 17.000000 | 199.000000 | 122.000000 | 99.000000 | 846.000000 | 67.100000 | 2.420000 | 81.000000 | 1.000000 |

```
#prints info of df
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    int64  
dtypes: float64(2), int64(7)
memory usage: 54.1 KB
```

```
#returns shape of df
df.shape
```

```
(768, 9)
```

```
#returns columns of df
df.columns
```

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

```
#checking missing values
df.isnull()
```

| | 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 |

```
df2=df.copy()
df2.loc[2:4,'Pregnancies':'Glucose']=None
df2.head(7)
df2.isnull().head(7)
```

| | 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** | True | True | False | False | False | False | False | False | False |
| **3** | True | True | False | False | False | False | False | False | False |
| **4** | True | True | False | False | False | False | False | False | False |
| **5** | False | False | False | False | False | False | False | False | False |
| **6** | False | False | False | False | False | False | False | False | False |

```
df2.isnull().sum()
```

```
Pregnancies                 3
Glucose                     3
BloodPressure               0
SkinThickness               0
Insulin                     0
BMI                         0
DiabetesPedigreeFunction    0
Age                         0
Outcome                     0
dtype: int64
```

```
df2.isnull().sum().sum()
```

```
6
```

```
#prints values from 2 to 4
df[2:4]
```

| | Pregnancies | Glucose | BloodPressure | SkinThickness | Insulin | BMI | DiabetesPedigreeFunction | Age | Outcome |
|---|---|---|---|---|---|---|---|---|---|
| **2** | 8 | 183 | 64 | 0 | 0 | 23.3 | 0.672 | 32 | 1 |
| **3** | 1 | 89 | 66 | 23 | 94 | 28.1 | 0.167 | 21 | 0 |

```
df['Outcome']
```

```
0      1
1      0
2      1
3      0
4      1
      ..
763    0
764    0
765    0
766    1
767    0
Name: Outcome, Length: 768, dtype: int64
```

```
#sorts BMI in descending order
df["BMI"].sort_index(ascending=False).head()
```

```
767    30.4
766    30.1
765    26.2
764    36.8
763    32.9
Name: BMI, dtype: float64
```

```
df[df.index==1]
```

| | Pregnancies | Glucose | BloodPressure | SkinThickness | Insulin | BMI | DiabetesPedigreeFunction | Age | Outcome |
|---|---|---|---|---|---|---|---|---|---|
| **1** | 1 | 85 | 66 | 29 | 0 | 26.6 | 0.351 | 31 | 0 |

```
#prints values of 1,2,3
df[df.index.isin([1,2,3])]
```

| | Pregnancies | Glucose | BloodPressure | SkinThickness | Insulin | BMI | DiabetesPedigreeFunction | Age | Outcome |
|---|---|---|---|---|---|---|---|---|---|
| **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 |

```
#prints values of blood pressure greater than 90
df[df.BloodPressure>90][1:4]
```

| | Pregnancies | Glucose | BloodPressure | SkinThickness | Insulin | BMI | DiabetesPedigreeFunction | Age | Outcome |
|---|---|---|---|---|---|---|---|---|---|
| **10** | 4 | 110 | 92 | 0 | 0 | 37.6 | 0.191 | 30 | 0 |
| **24** | 11 | 143 | 94 | 33 | 146 | 36.6 | 0.254 | 51 | 1 |
| **29** | 5 | 117 | 92 | 0 | 0 | 34.1 | 0.337 | 38 | 0 |

```
#prints mean of df
df.mean()
```

```
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
```

```
#creates new column glucose insulin ratio
df2['Glucose_insulin_ratio']=df2['Glucose']/df2['Insulin']
```

```
#prints count of each value in pregnancies
df2['Pregnancies'].value_counts()
```

```
1.0     134
0.0     110
2.0     103
3.0      75
4.0      68
5.0      57
6.0      50
7.0      45
8.0      37
9.0      28
10.0     24
11.0     11
13.0     10
12.0      9
14.0      2
15.0      1
17.0      1
Name: Pregnancies, dtype: int64
```

```
#1
#stores data in wq
wq=p.read_csv('WineQT.csv')
```

```
#prints wq
print(wq)
```

```
      fixed acidity  volatile acidity  citric acid  residual sugar  chlorides  \
0               7.4             0.700         0.00             1.9      0.076   
1               7.8             0.880         0.00             2.6      0.098   
2               7.8             0.760         0.04             2.3      0.092   
3              11.2             0.280         0.56             1.9      0.075   
4               7.4             0.700         0.00             1.9      0.076   
...             ...               ...          ...             ...        ...   
1138            6.3             0.510         0.13             2.3      0.076   
1139            6.8             0.620         0.08             1.9      0.068   
1140            6.2             0.600         0.08             2.0      0.090   
1141            5.9             0.550         0.10             2.2      0.062   
1142            5.9             0.645         0.12             2.0      0.075   

      free sulfur dioxide  total sulfur dioxide  density    pH  sulphates  \
0                    11.0                  34.0  0.99780  3.51       0.56   
1                    25.0                  67.0  0.99680  3.20       0.68   
2                    15.0                  54.0  0.99700  3.26       0.65   
3                    17.0                  60.0  0.99800  3.16       0.58   
4                    11.0                  34.0  0.99780  3.51       0.56   
...                   ...                   ...      ...   ...        ...   
1138                 29.0                  40.0  0.99574  3.42       0.75   
1139                 28.0                  38.0  0.99651  3.42       0.82   
1140                 32.0                  44.0  0.99490  3.45       0.58   
1141                 39.0                  51.0  0.99512  3.52       0.76   
1142                 32.0                  44.0  0.99547  3.57       0.71   

      alcohol  quality    Id  
0         9.4        5     0  
1         9.8        5     1  
2         9.8        5     2  
3         9.8        6     3  
4         9.4        5     4  
...       ...      ...   ...  
1138     11.0        6  1592  
1139      9.5        6  1593  
1140     10.5        5  1594  
1141     11.2        6  1595  
1142     10.2        5  1597  

[1143 rows x 13 columns]
```

```
#3
#returns shape of wq
wq.shape
```

```
(1143, 13)
```

```
#4
#prints statistical values of wq
wq.describe()
```

| | fixed acidity | volatile acidity | citric acid | residual sugar | chlorides | free sulfur dioxide | t su dic |
|---|---|---|---|---|---|---|---|
| **count** | 1143.000000 | 1143.000000 | 1143.000000 | 1143.000000 | 1143.000000 | 1143.000000 | 1143.00 |
| **mean** | 8.311111 | 0.531339 | 0.268364 | 2.532152 | 0.086933 | 15.615486 | 45.9 |
| **std** | 1.747595 | 0.179633 | 0.196686 | 1.355917 | 0.047267 | 10.250486 | 32.7 |
| **min** | 4.600000 | 0.120000 | 0.000000 | 0.900000 | 0.012000 | 1.000000 | 6.00 |
| **25%** | 7.100000 | 0.392500 | 0.090000 | 1.900000 | 0.070000 | 7.000000 | 21.00 |
| **50%** | 7.900000 | 0.520000 | 0.250000 | 2.200000 | 0.079000 | 13.000000 | 37.00 |
| **75%** | 9.100000 | 0.640000 | 0.420000 | 2.600000 | 0.090000 | 21.000000 | 61.00 |

```
#5
#prints first 5 values of wq
wq.head(5)
```

|  | fixed acidity | volatile acidity | citric acid | residual sugar | chlorides | free sulfur dioxide | total sulfur dioxide | density | pH | sulph |
|---|---|---|---|---|---|---|---|---|---|---|

```
#5
#prints last 5 values of wq
wq.tail(5)
```

|  | fixed acidity | volatile acidity | citric acid | residual sugar | chlorides | free sulfur dioxide | total sulfur dioxide | density | pH | su |
|---|---|---|---|---|---|---|---|---|---|---|
| **1138** | 6.3 | 0.510 | 0.13 | 2.3 | 0.076 | 29.0 | 40.0 | 0.99574 | 3.42 |  |
| **1139** | 6.8 | 0.620 | 0.08 | 1.9 | 0.068 | 28.0 | 38.0 | 0.99651 | 3.42 |  |
| **1140** | 6.2 | 0.600 | 0.08 | 2.0 | 0.090 | 32.0 | 44.0 | 0.99490 | 3.45 |  |
| **1141** | 5.9 | 0.550 | 0.10 | 2.2 | 0.062 | 39.0 | 51.0 | 0.99512 | 3.52 |  |

```
#6
#prints first 7 values of wq
wq.head(7)
```

|  | fixed acidity | volatile acidity | citric acid | residual sugar | chlorides | free sulfur dioxide | total sulfur dioxide | density | pH | sulph |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 7.4 | 0.70 | 0.00 | 1.9 | 0.076 | 11.0 | 34.0 | 0.9978 | 3.51 |  |
| **1** | 7.8 | 0.88 | 0.00 | 2.6 | 0.098 | 25.0 | 67.0 | 0.9968 | 3.20 |  |
| **2** | 7.8 | 0.76 | 0.04 | 2.3 | 0.092 | 15.0 | 54.0 | 0.9970 | 3.26 |  |
| **3** | 11.2 | 0.28 | 0.56 | 1.9 | 0.075 | 17.0 | 60.0 | 0.9980 | 3.16 |  |
| **4** | 7.4 | 0.70 | 0.00 | 1.9 | 0.076 | 11.0 | 34.0 | 0.9978 | 3.51 |  |
| **5** | 7.4 | 0.66 | 0.00 | 1.8 | 0.075 | 13.0 | 40.0 | 0.9978 | 3.51 |  |

```
#7
wq.iloc[0:7,4:7]
```

|  | chlorides | free sulfur dioxide | total sulfur dioxide |
|---|---|---|---|
| **0** | 0.076 | 11.0 | 34.0 |
| **1** | 0.098 | 25.0 | 67.0 |
| **2** | 0.092 | 15.0 | 54.0 |
| **3** | 0.075 | 17.0 | 60.0 |
| **4** | 0.076 | 11.0 | 34.0 |
| **5** | 0.075 | 13.0 | 40.0 |
| **6** | 0.069 | 15.0 | 59.0 |

```
#8
wq.iloc[:,3:6]
```

|  | residual sugar | chlorides | free sulfur dioxide |
|---|---|---|---|
| **0** | 1.9 | 0.076 | 11.0 |
| **1** | 2.6 | 0.098 | 25.0 |

```
#9
#checks for null values in wq
wq.isnull().sum()
```

```
fixed acidity           0
volatile acidity        0
citric acid             0
residual sugar          0
chlorides               0
free sulfur dioxide     0
total sulfur dioxide    0
density                 0
pH                      0
sulphates               0
alcohol                 0
quality                 0
Id                      0
dtype: int64
```

```
#10
#renames column ph to ph_value
wq.rename(columns={'pH':'potential of hydrogen'},inplace=True)


print(wq['potential of hydrogen'])
```

```
0       3.51
1       3.20
2       3.26
3       3.16
4       3.51
        ... 
1138    3.42
1139    3.42
1140    3.45
1141    3.52
1142    3.57
Name: potential of hydrogen, Length: 1143, dtype: float64
```

```
#11
wq['total_free_ratio'] = wq['total sulfur dioxide'] / wq['free sulfur dioxide']
print(wq.columns)
```

```
Index(['fixed acidity', 'volatile acidity', 'citric acid', 'residual sugar',
       'chlorides', 'free sulfur dioxide', 'total sulfur dioxide', 'density',
       'potential of hydrogen', 'sulphates', 'alcohol', 'quality', 'Id',
       'total_free_ratio'],
      dtype='object')
```

```
#12
aggrgatedata=wq.groupby('quality').mean()
aggrgatedata.head(5)
```

|  | fixed acidity | volatile acidity | citric acid | residual sugar | chlorides | free sulfur dioxide | total sulfur dioxide | density |
|---|---|---|---|---|---|---|---|---|
| **quality** |  |  |  |  |  |  |  |  |
| **3** | 8.450000 | 0.897500 | 0.211667 | 2.666667 | 0.105333 | 8.166667 | 24.500000 | 0.997682 |
| **4** | 7.809091 | 0.700000 | 0.165758 | 2.566667 | 0.094788 | 14.848485 | 40.606061 | 0.996669 |
| **5** | 8.161077 | 0.585280 | 0.240124 | 2.540476 | 0.091770 | 16.612836 | 55.299172 | 0.997073 |
| **6** | 8.317749 | 0.504957 | 0.263680 | 2.444805 | 0.085281 | 15.215368 | 39.941558 | 0.996610 |

```
#13
slicedframe = (wq['total_free_ratio'] > 2.7) & (wq['total_free_ratio'] < 3.2)
```

```
sf=wq[slicedframe]
print(sf.head(5))
```

```
    fixed acidity  volatile acidity  citric acid  residual sugar  chlorides  \
0             7.4              0.70         0.00             1.9      0.076   
4             7.4              0.70         0.00             1.9      0.076   
5             7.4              0.66         0.00             1.8      0.075   
12            8.5              0.28         0.56             1.8      0.092   
14            7.6              0.39         0.31             2.3      0.082   

    free sulfur dioxide  total sulfur dioxide  density  potential of hydrogen  \
0                  11.0                  34.0   0.9978                   3.51   
4                  11.0                  34.0   0.9978                   3.51   
5                  13.0                  40.0   0.9978                   3.51   
12                 35.0                 103.0   0.9969                   3.30   
14                 23.0                  71.0   0.9982                   3.52   

    sulphates  alcohol  quality  Id  total_free_ratio  
0        0.56      9.4        5   0          3.090909  
4        0.56      9.4        5   4          3.090909  
5        0.56      9.4        5   5          3.076923  
12       0.75     10.5        7  16          2.942857  
14       0.65      9.7        5  21          3.086957  
```

```
#14
wq.drop(wq.columns[2],axis=1)
```

| | fixed acidity | volatile acidity | residual sugar | chlorides | free sulfur dioxide | total sulfur dioxide | density | potential of hydrogen | sulpl |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 7.4 | 0.700 | 1.9 | 0.076 | 11.0 | 34.0 | 0.99780 | 3.51 | |
| **1** | 7.8 | 0.880 | 2.6 | 0.098 | 25.0 | 67.0 | 0.99680 | 3.20 | |
| **2** | 7.8 | 0.760 | 2.3 | 0.092 | 15.0 | 54.0 | 0.99700 | 3.26 | |
| **3** | 11.2 | 0.280 | 1.9 | 0.075 | 17.0 | 60.0 | 0.99800 | 3.16 | |
| **4** | 7.4 | 0.700 | 1.9 | 0.076 | 11.0 | 34.0 | 0.99780 | 3.51 | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | |
| **1138** | 6.3 | 0.510 | 2.3 | 0.076 | 29.0 | 40.0 | 0.99574 | 3.42 | |
| **1139** | 6.8 | 0.620 | 1.9 | 0.068 | 28.0 | 38.0 | 0.99651 | 3.42 | |
| **1140** | 6.2 | 0.600 | 2.0 | 0.090 | 32.0 | 44.0 | 0.99490 | 3.45 | |
| **1141** | 5.9 | 0.550 | 2.2 | 0.062 | 39.0 | 51.0 | 0.99512 | 3.52 | |
| **1142** | 5.9 | 0.645 | 2.0 | 0.075 | 32.0 | 44.0 | 0.99547 | 3.57 | |

```
#15
print(wq.columns[0])
print((wq.sort_values(by=wq.columns[0])).head(5))
```

```
fixed acidity
     fixed acidity  volatile acidity  citric acid  residual sugar  chlorides  \
32             4.6              0.52         0.15             2.1      0.054   
589            4.9              0.42         0.00             2.1      0.048   
396            5.0              1.04         0.24             1.6      0.050   
935            5.0              0.74         0.00             1.2      0.041   
898            5.0              0.38         0.01             1.6      0.048   

     free sulfur dioxide  total sulfur dioxide  density  \
32                   8.0                  65.0  0.99340   
589                 16.0                  42.0  0.99154   
396                 32.0                  96.0  0.99340   
935                 16.0                  46.0  0.99258   
898                 26.0                  60.0  0.99084   

     potential of hydrogen  sulphates  alcohol  quality    Id  \
32                    3.90       0.56     13.1        4    45   
589                   3.71       0.74     14.0        7   821   
396                   3.74       0.62     11.5        5   553   
935                   4.01       0.59     12.5        6  1321   
898                   3.70       0.75     14.0        6  1270   

     total_free_ratio  
32           8.125000  
```

```
589          2.625000
396          3.000000
935          2.875000
898          2.307692
```

```
#15
print(wq.columns[0])
print((wq.sort_values(by=wq.columns[0],ascending=False)).head(5))
```

```
fixed acidity
     fixed acidity  volatile acidity  citric acid  residual sugar  chlorides  \
462           15.9             0.360         0.65             7.5      0.096   
311           15.6             0.685         0.76             3.7      0.100   
399           15.6             0.645         0.49             4.2      0.095   
397           15.5             0.645         0.49             4.2      0.095   
171           15.0             0.210         0.44             2.2      0.075   

     free sulfur dioxide  total sulfur dioxide  density  \
462                 22.0                  71.0  0.99760   
311                  6.0                  43.0  1.00320   
399                 10.0                  23.0  1.00315   
397                 10.0                  23.0  1.00315   
171                 10.0                  24.0  1.00005   

     potential of hydrogen  sulphates  alcohol  quality   Id  total_free_ratio  
462                   2.98       0.84     14.9        5  652          3.227273  
311                   2.95       0.68     11.2        7  442          7.166667  
399                   2.92       0.74     11.1        5  557          2.300000  
397                   2.92       0.74     11.1        5  554          2.300000  
171                   3.07       0.84      9.2        7  244          2.400000  
```

```
#16
s=wq[wq.columns[2]].sum()
print(s)
```

```
306.74
```

```
#17
min=wq[wq.columns[3]].min()
max=wq[wq.columns[3]].max()
print(max,min)
```

```
15.5 0.9
```

```
#18
wq.rename(columns={"potential of hydrogen":"ph"},inplace=True)
print(wq.columns)
```

```
Index(['fixed acidity', 'volatile acidity', 'citric acid', 'residual sugar',
       'chlorides', 'free sulfur dioxide', 'total sulfur dioxide', 'density',
       'ph', 'sulphates', 'alcohol', 'quality', 'Id', 'total_free_ratio'],
      dtype='object')
```

```
#20
wq.to_csv("balance.csv",index=True,header=True);
```

```
#21
wq.fillna(0);
print(wq.isnull().sum());
```

```
fixed acidity           0
volatile acidity        0
citric acid             0
residual sugar          0
chlorides               0
free sulfur dioxide     0
total sulfur dioxide    0
density                 0
ph                      0
sulphates               0
alcohol                 0
quality                 0
Id                      0
total_free_ratio        0
dtype: int64
```

```
wq1=wq.iloc[:600];
```

```
wq2=wq.iloc[600:];
print(len(wq1))
print(len(wq2))
```

```
600
543
```

```
index = wq[wq.columns[2]].idxmax()
row = df.iloc[index];
print(row)
```

```
Pregnancies                  1.000
Glucose                     81.000
BloodPressure               72.000
SkinThickness               18.000
Insulin                     40.000
BMI                         26.600
DiabetesPedigreeFunction     0.283
Age                         24.000
Outcome                      0.000
Name: 103, dtype: float64
```