```
import numpy as np
import pandas as pd

test1 = pd.read_csv("C:\\Users\\HP\\Desktop\\phy\\health care
diabetes.csv")
test1.head()
```

```
   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  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
```

```
test1.shape
```

```
(768, 9)
```

```
test1.columns
```

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

```
test1.Glucose
```

```
0      148
1       85
2      183
3       89
4      137
      ...
763    101
764    122
765    121
766    126
```

```
767     93
Name: Glucose, Length: 768, dtype: int64

test1.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

test1.BloodPressure

0      72
1      66
2      64
3      66
4      40
       ..
763    76
764    70
765    72
766    60
767    70
Name: BloodPressure, Length: 768, dtype: int64

test1.SkinThickness

0      35
1      29
2       0
3      23
4      35
       ..
763    48
764    27
765    23
766     0
767    31
Name: SkinThickness, Length: 768, dtype: int64

test1.Insulin

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
```

```
test1.isnull().sum()
```

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

```
df = test1.replace(0,'NAN')
df.head()
```

```
  Pregnancies Glucose BloodPressure SkinThickness Insulin   BMI  \
0           6     148            72            35     NAN  33.6
1           1      85            66            29     NAN  26.6
2           8     183            64           NAN     NAN  23.3
3           1      89            66            23      94  28.1
4         NAN     137            40            35     168  43.1

   DiabetesPedigreeFunction  Age Outcome
0                     0.627   50       1
1                     0.351   31     NAN
2                     0.672   32       1
3                     0.167   21     NAN
4                     2.288   33       1
```

```
df.describe()
```

```
       DiabetesPedigreeFunction         Age
count                768.000000  768.000000
mean                   0.471876   33.240885
std                    0.331329   11.760232
min                    0.078000   21.000000
25%                    0.243750   24.000000
50%                    0.372500   29.000000
```

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

```
test1.describe()
```

```
       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
```

```
df1 = df.replace('NAN',np.nan)
df1.head()
```

```
   Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin
BMI  \
0          6.0    148.0           72.0           35.0      NaN  33.6

1          1.0     85.0           66.0           29.0      NaN  26.6

2          8.0    183.0           64.0            NaN      NaN  23.3

3          1.0     89.0           66.0           23.0     94.0  28.1

4          NaN    137.0           40.0           35.0    168.0  43.1
```

```
   DiabetesPedigreeFunction  Age  Outcome
0                     0.627   50      1.0
1                     0.351   31      NaN
2                     0.672   32      1.0
3                     0.167   21      NaN
4                     2.288   33      1.0
```

```
df1.isnull().sum()
```

```
Pregnancies                 111
Glucose                       5
BloodPressure                35
SkinThickness               227
Insulin                     374
BMI                          11
DiabetesPedigreeFunction      0
Age                           0
Outcome                     500
dtype: int64
```

```
df2= df1.interpolate()
df2.head()
```

```
   Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin
BMI  \
0          6.0    148.0           72.0           35.0      NaN  33.6

1          1.0     85.0           66.0           29.0      NaN  26.6

2          8.0    183.0           64.0           26.0      NaN  23.3

3          1.0     89.0           66.0           23.0     94.0  28.1

4          3.0    137.0           40.0           35.0    168.0  43.1


   DiabetesPedigreeFunction  Age  Outcome
0                     0.627   50      1.0
1                     0.351   31      1.0
2                     0.672   32      1.0
3                     0.167   21      1.0
4                     2.288   33      1.0
```

```
x = df2['Insulin'].mean()
df2['Insulin'].fillna(x,inplace = True)
print(df2)
```

```
     Pregnancies  Glucose  BloodPressure  SkinThickness     Insulin
BMI  \
0            6.0    148.0           72.0           35.0  159.045098
33.6
```

```
1            1.0     85.0           66.0           29.0  159.045098
26.6
2            8.0    183.0           64.0           26.0  159.045098
23.3
3            1.0     89.0           66.0           23.0   94.000000
28.1
4            3.0    137.0           40.0           35.0  168.000000
43.1
..           ...      ...            ...            ...         ...
...
763         10.0    101.0           76.0           48.0  180.000000
32.9
764          2.0    122.0           70.0           27.0  146.000000
36.8
765          5.0    121.0           72.0           23.0  112.000000
26.2
766          1.0    126.0           60.0           27.0  112.000000
30.1
767          1.0     93.0           70.0           31.0  112.000000
30.4

     DiabetesPedigreeFunction  Age  Outcome
0                       0.627   50      1.0
1                       0.351   31      1.0
2                       0.672   32      1.0
3                       0.167   21      1.0
4                       2.288   33      1.0
..                        ...  ...      ...
763                     0.171   63      1.0
764                     0.340   27      1.0
765                     0.245   30      1.0
766                     0.349   47      1.0
767                     0.315   23      1.0

[768 rows x 9 columns]
```

df2.head()

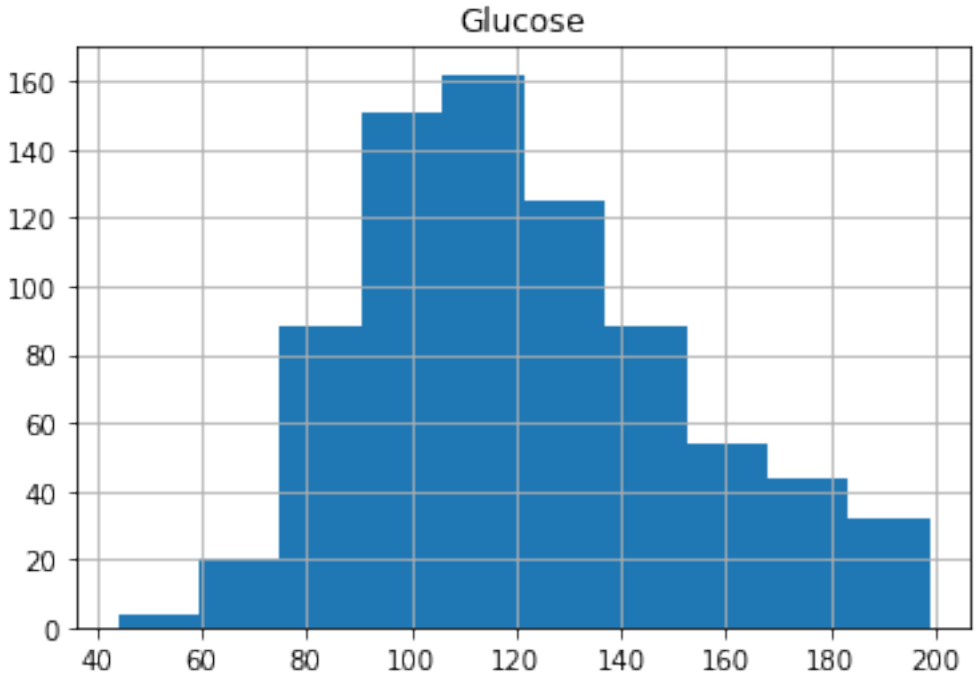
```
   Pregnancies  Glucose  BloodPressure  SkinThickness     Insulin
BMI  \
0          6.0    148.0           72.0           35.0  159.045098
33.6
1          1.0     85.0           66.0           29.0  159.045098
26.6
2          8.0    183.0           64.0           26.0  159.045098
23.3
3          1.0     89.0           66.0           23.0   94.000000
28.1
4          3.0    137.0           40.0           35.0  168.000000
43.1
```

```
   DiabetesPedigreeFunction  Age  Outcome
0                     0.627   50      1.0
1                     0.351   31      1.0
2                     0.672   32      1.0
3                     0.167   21      1.0
4                     2.288   33      1.0
```

```
import matplotlib.pyplot as plt
from matplotlib import style
%matplotlib inline
```

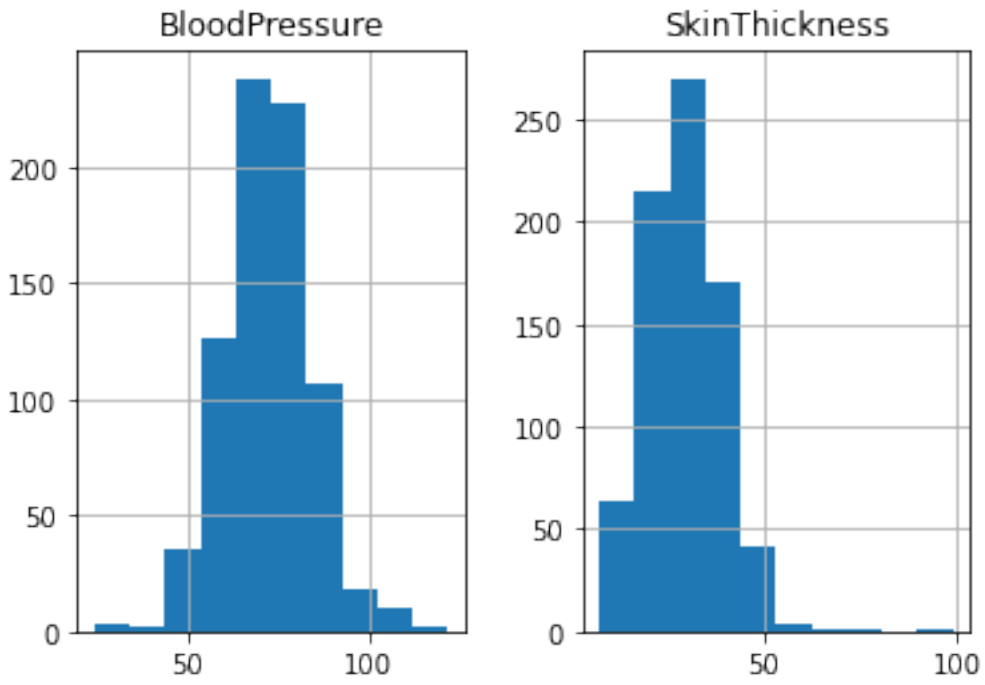
```
df2.hist(column = 'Glucose')
```

```
array([[<AxesSubplot:title={'center':'Glucose'}>]], dtype=object)
```



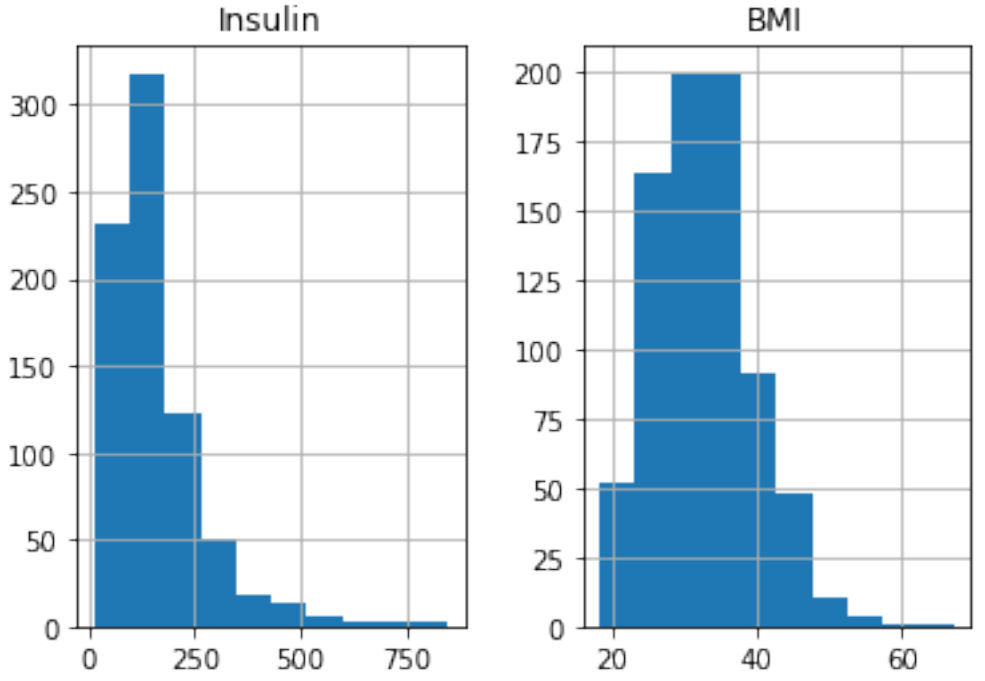

```
df2.hist(column = ['BloodPressure','SkinThickness'])
```

```
array([[<AxesSubplot:title={'center':'BloodPressure'}>,
        <AxesSubplot:title={'center':'SkinThickness'}>]],
dtype=object)
```




```
df2.hist(column = ['Insulin','BMI'])
```

```
array([[<AxesSubplot:title={'center':'Insulin'}>,
        <AxesSubplot:title={'center':'BMI'}>]], dtype=object)
```




```
type(df2)
```

```
pandas.core.frame.DataFrame
```

```
print(df2.dtypes)
```

```
Pregnancies                 float64
Glucose                     float64
BloodPressure               float64
SkinThickness               float64
Insulin                     float64
BMI                         float64
DiabetesPedigreeFunction    float64
Age                           int64
Outcome                     float64
dtype: object
```

```
df3 =
['float64','float64','float64','float64','float64','float64','float64'
,'float64','int64']
print(df3)
```

```
['float64', 'float64', 'float64', 'float64', 'float64', 'float64',
'float64', 'float64', 'int64']
```

```
plt.hist(df3)
```

```
(array([8., 0., 0., 0., 0., 0., 0., 0., 0., 1.]),
 array([0. , 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1. ]),
 <BarContainer object of 10 artists>)
```