In [4]:
```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

In [5]:
```
hd=pd.read_csv("diabetes.csv")
hd
```

Out[5]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedig
0	2	138	62	35	0	33.6	
1	0	84	82	31	125	38.2	
2	0	145	0	0	0	44.2	
3	0	135	68	42	250	42.3	
4	1	139	62	41	480	40.7	
...	...	...	...	...	...	...	
1995	2	75	64	24	55	29.7	
1996	8	179	72	42	130	32.7	
1997	6	85	78	0	0	31.2	
1998	0	129	110	46	130	67.1	
1999	2	81	72	15	76	30.1	

2000 rows × 9 columns

In [8]:
```
hd.columns
```

Out[8]:
```
Index(['Pregnancies', 'Glucose', 'BloodPressure', 'SkinThickness', 'Insulin',
       'BMI', 'DiabetesPedigreeFunction', 'Age', 'Outcome'],
      dtype='object')
```

In [10]:
```
hd.index
```

Out[10]:
```
RangeIndex(start=0, stop=2000, step=1)
```

In [74]:
```
len(hd)
```

Out[74]:
```
2000
```

Describing Data

In [13]:
```
hd.describe()
```

Out[13]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BM
count	2000.000000	2000.000000	2000.000000	2000.000000	2000.000000	2000.00000
mean	3.703500	121.182500	69.145500	20.935000	80.254000	32.19300
std	3.306063	32.068636	19.188315	16.103243	111.180534	8.14990
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000
25%	1.000000	99.000000	63.500000	0.000000	0.000000	27.37500
50%	3.000000	117.000000	72.000000	23.000000	40.000000	32.30000
75%	6.000000	141.000000	80.000000	32.000000	130.000000	36.80000
max	17.000000	199.000000	122.000000	110.000000	744.000000	80.60000

In [15]:
```
hd.dtypes
```

Out[15]:
```
Pregnancies                   int64
Glucose                       int64
BloodPressure                 int64
SkinThickness                 int64
Insulin                       int64
BMI                         float64
DiabetesPedigreeFunction    float64
Age                           int64
Outcome                       int64
dtype: object
```

In [17]:
```
hd.std()
```

Out[17]:
```
Pregnancies                   3.306063
Glucose                      32.068636
BloodPressure                19.188315
SkinThickness                16.103243
Insulin                     111.180534
BMI                           8.149901
DiabetesPedigreeFunction      0.323553
Age                          11.786423
Outcome                       0.474498
dtype: float64
```

In [19]:
```
hd.mean()
```

Out[19]:
```
Pregnancies                   3.70350
Glucose                     121.18250
BloodPressure                69.14550
SkinThickness                20.93500
Insulin                      80.25400
BMI                          32.19300
DiabetesPedigreeFunction      0.47093
Age                          33.09050
Outcome                       0.34200
dtype: float64
```

In [21]:
```
hd.mode()
```

Out[21]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeF
0	1.0	99.0	74.0	0.0	0.0	31.2	
1	NaN	NaN	NaN	NaN	NaN	32.0	

In [23]:
```
hd.var()
```

Out[23]:
```
Pregnancies                   10.930053
Glucose                     1028.397392
BloodPressure                368.191425
SkinThickness                259.314432
Insulin                    12361.111040
BMI                           66.420881
DiabetesPedigreeFunction       0.104686
Age                          138.919770
Outcome                        0.225149
dtype: float64
```

In [25]:
```
hd.sum()
```

Out[25]:
```
Pregnancies                  7407.00
Glucose                    242365.00
BloodPressure              138291.00
SkinThickness               41870.00
Insulin                    160508.00
BMI                         64386.00
DiabetesPedigreeFunction      941.86
Age                         66181.00
Outcome                       684.00
dtype: float64
```

In [29]:
```
hd.tail(9)
```

Out[29]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedig
1991	6	102	82	0	0	30.8	
1992	6	134	70	23	130	35.4	
1993	2	87	0	23	0	28.9	
1994	1	79	60	42	48	43.5	
1995	2	75	64	24	55	29.7	
1996	8	179	72	42	130	32.7	
1997	6	85	78	0	0	31.2	
1998	0	129	110	46	130	67.1	
1999	2	81	72	15	76	30.1	

In [31]:
```
hd.head(3)
```

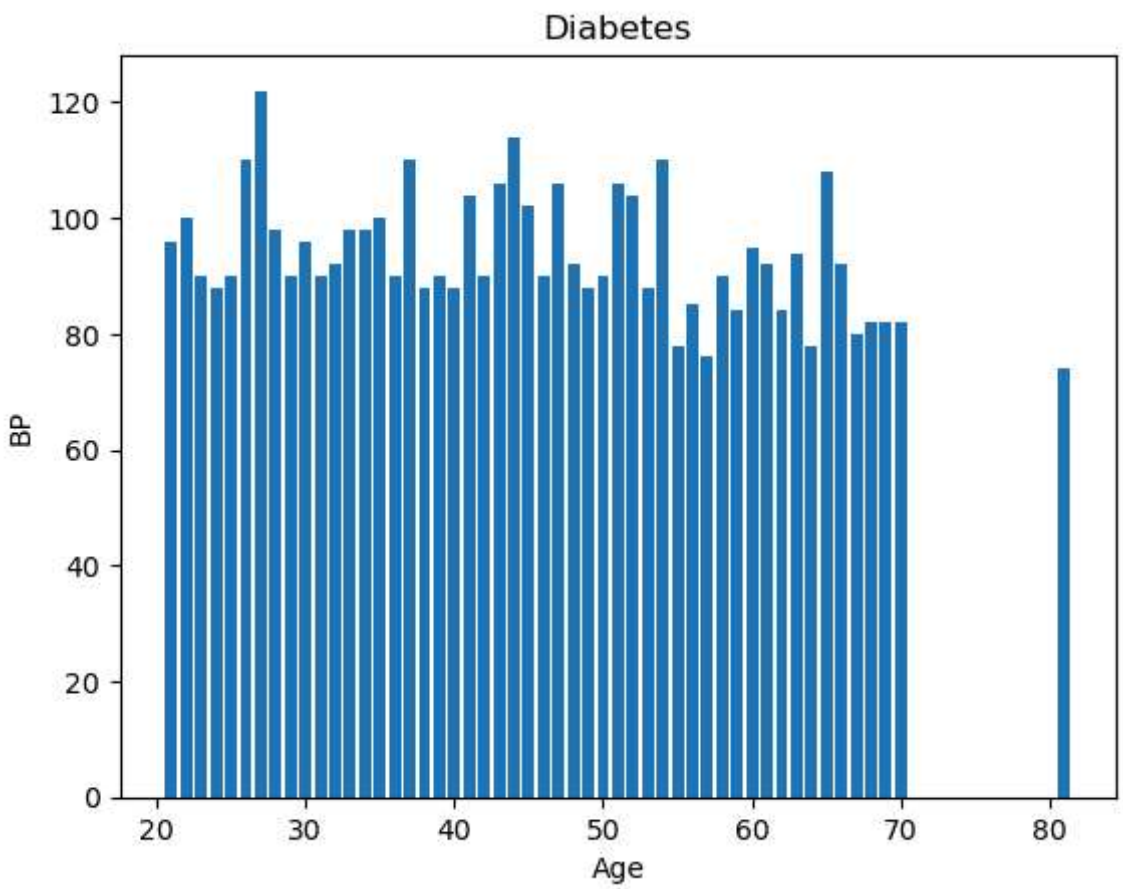
Out[31]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeF
0	2	138	62	35	0	33.6	
1	0	84	82	31	125	38.2	
2	0	145	0	0	0	44.2	

In [33]:
```
hd['Pregnancies']
```

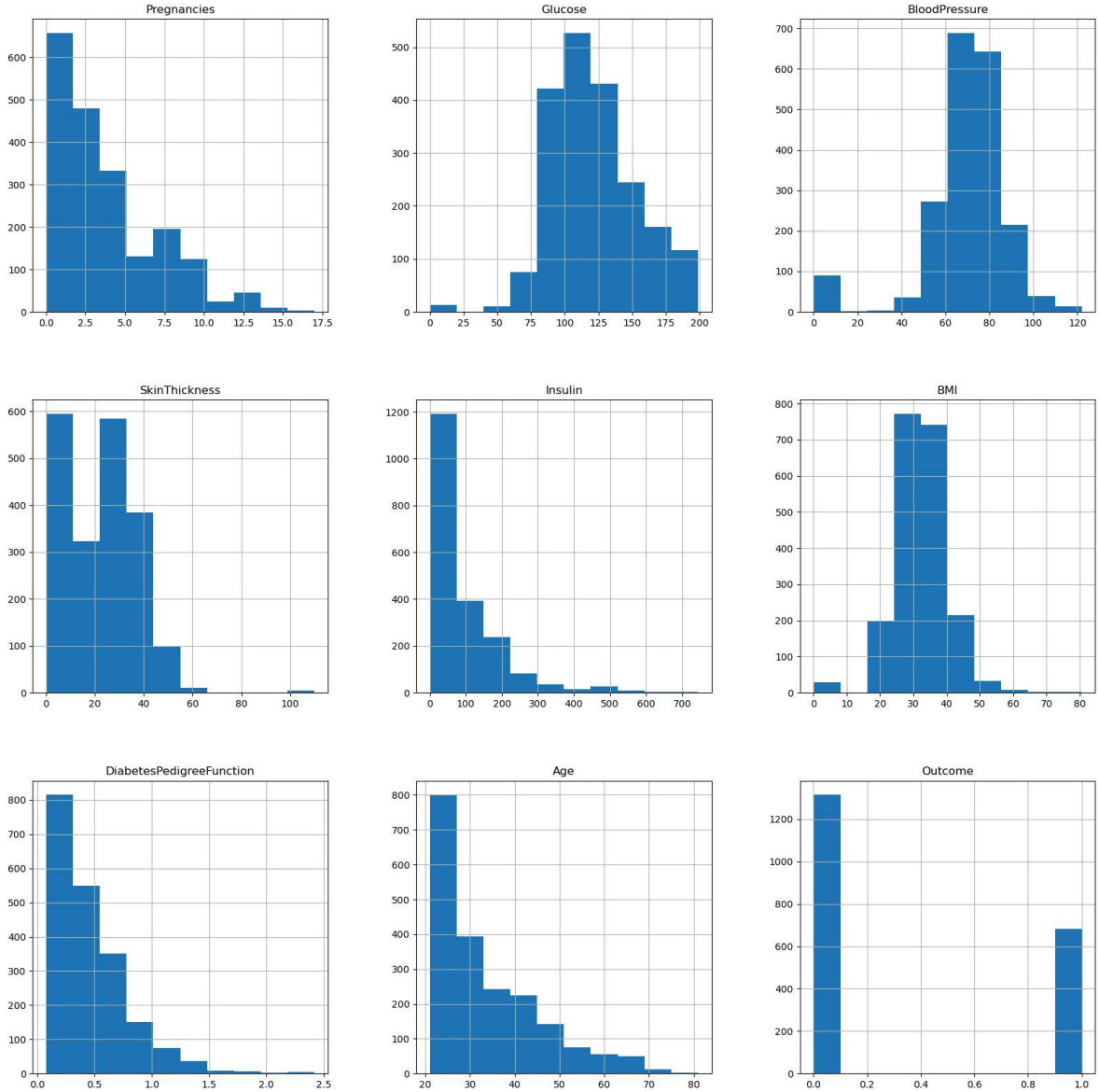
Out[33]:
```
0       2
1       0
2       0
3       0
4       1
       ..
1995    2
1996    8
1997    6
1998    0
1999    2
Name: Pregnancies, Length: 2000, dtype: int64
```

In [35]:
```
x=hd['Age']
y=hd['BloodPressure']
plt.bar(x,y)
plt.title("Diabetes")
plt.xlabel("Age")
plt.ylabel("BP")
plt.show()
```



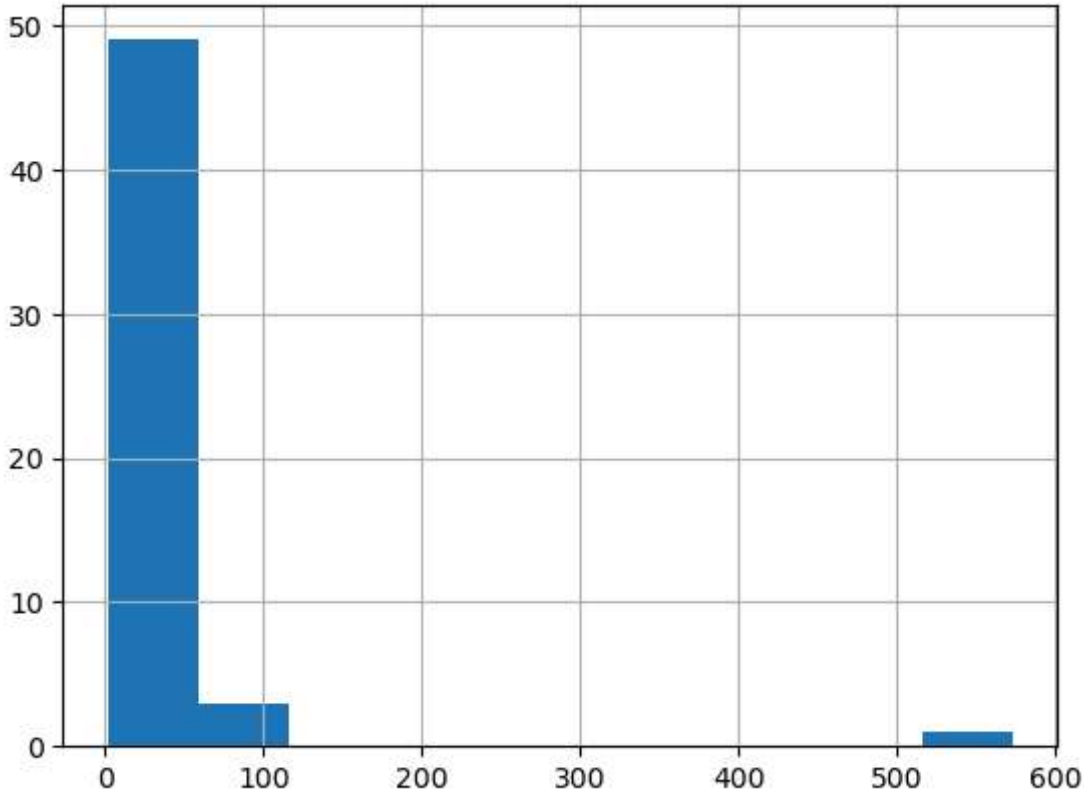
In [36]:
```
hd.hist(figsize=(20,20))
```

Out[36]:
```
array([[<Axes: title={'center': 'Pregnancies'}>,
        <Axes: title={'center': 'Glucose'}>,
        <Axes: title={'center': 'BloodPressure'}>],
       [<Axes: title={'center': 'SkinThickness'}>,
        <Axes: title={'center': 'Insulin'}>,
        <Axes: title={'center': 'BMI'}>],
       [<Axes: title={'center': 'DiabetesPedigreeFunction'}>,
        <Axes: title={'center': 'Age'}>,
        <Axes: title={'center': 'Outcome'}>]], dtype=object)
```



In [37]:
```
hd['SkinThickness'].value_counts().hist()
```

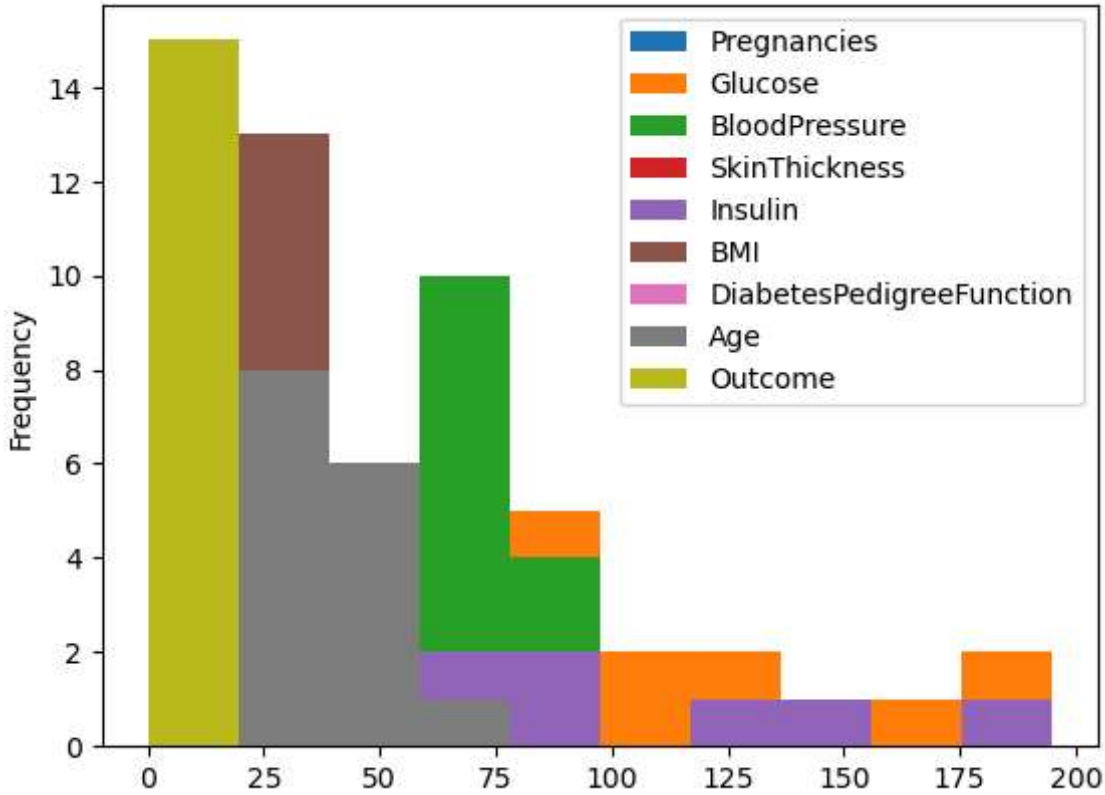
Out[37]:
```
<Axes: >
```



iloc

In [39]:
```
hd.iloc[10:25].plot.hist()
```

Out[39]: <Axes: ylabel='Frequency'>



Conditional Filtering

In [41]:
```
hd[hd["Insulin"]==0]
```

Out[41]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedig
0	2	138	62	35	0	33.6	
2	0	145	0	0	0	44.2	
6	4	99	72	17	0	25.6	
7	8	194	80	0	0	26.1	
9	2	89	90	30	0	33.5	
...	...	...	...	...	...	...	
1988	4	120	68	0	0	29.6	
1989	4	110	66	0	0	31.9	
1991	6	102	82	0	0	30.8	
1993	2	87	0	23	0	28.9	
1997	6	85	78	0	0	31.2	

956 rows × 9 columns

In [42]:
```
hd[hd["BMI"]>=29]
```

Out[42]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedig
0	2	138	62	35	0	33.6	
1	0	84	82	31	125	38.2	
2	0	145	0	0	0	44.2	
3	0	135	68	42	250	42.3	
4	1	139	62	41	480	40.7	
...	...	...	...	...	...	...	
1995	2	75	64	24	55	29.7	
1996	8	179	72	42	130	32.7	
1997	6	85	78	0	0	31.2	
1998	0	129	110	46	130	67.1	
1999	2	81	72	15	76	30.1	

1325 rows × 9 columns

Multiplying new col with applying some maths

In [45]:
```
hd['New']=hd['Glucose']*hd['Insulin']
hd
```

Out[45]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedig
0	2	138	62	35	0	33.6	
1	0	84	82	31	125	38.2	
2	0	145	0	0	0	44.2	
3	0	135	68	42	250	42.3	
4	1	139	62	41	480	40.7	
...	...	...	...	...	...	...	
1995	2	75	64	24	55	29.7	
1996	8	179	72	42	130	32.7	
1997	6	85	78	0	0	31.2	
1998	0	129	110	46	130	67.1	
1999	2	81	72	15	76	30.1	

2000 rows × 10 columns

col drop method

In [47]:
```
hd.drop('New',axis=1,inplace=True)
hd
```

Out[47]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedig
0	2	138	62	35	0	33.6	
1	0	84	82	31	125	38.2	
2	0	145	0	0	0	44.2	
3	0	135	68	42	250	42.3	
4	1	139	62	41	480	40.7	
...	...	...	...	...	...	...	
1995	2	75	64	24	55	29.7	
1996	8	179	72	42	130	32.7	
1997	6	85	78	0	0	31.2	
1998	0	129	110	46	130	67.1	
1999	2	81	72	15	76	30.1	

2000 rows × 9 columns

cross tab

In [52]:
```
hd.nunique()
```

Out[52]:
```
Pregnancies                 17
Glucose                    136
BloodPressure               47
SkinThickness               53
Insulin                    182
BMI                        247
DiabetesPedigreeFunction   505
Age                         52
Outcome                      2
dtype: int64
```

In [58]:
```
pd.crosstab(hd['DiabetesPedigreeFunction'],hd['Outcome'])
```

Out[58]:

Outcome	0	1
DiabetesPedigreeFunction		
0.078	2	0
0.084	2	0
0.085	5	0
0.088	3	3
0.089	2	0
...	...	...
1.781	2	0
1.893	0	2
2.137	0	3
2.329	2	0
2.420	0	3

505 rows × 2 columns

SP Graphs

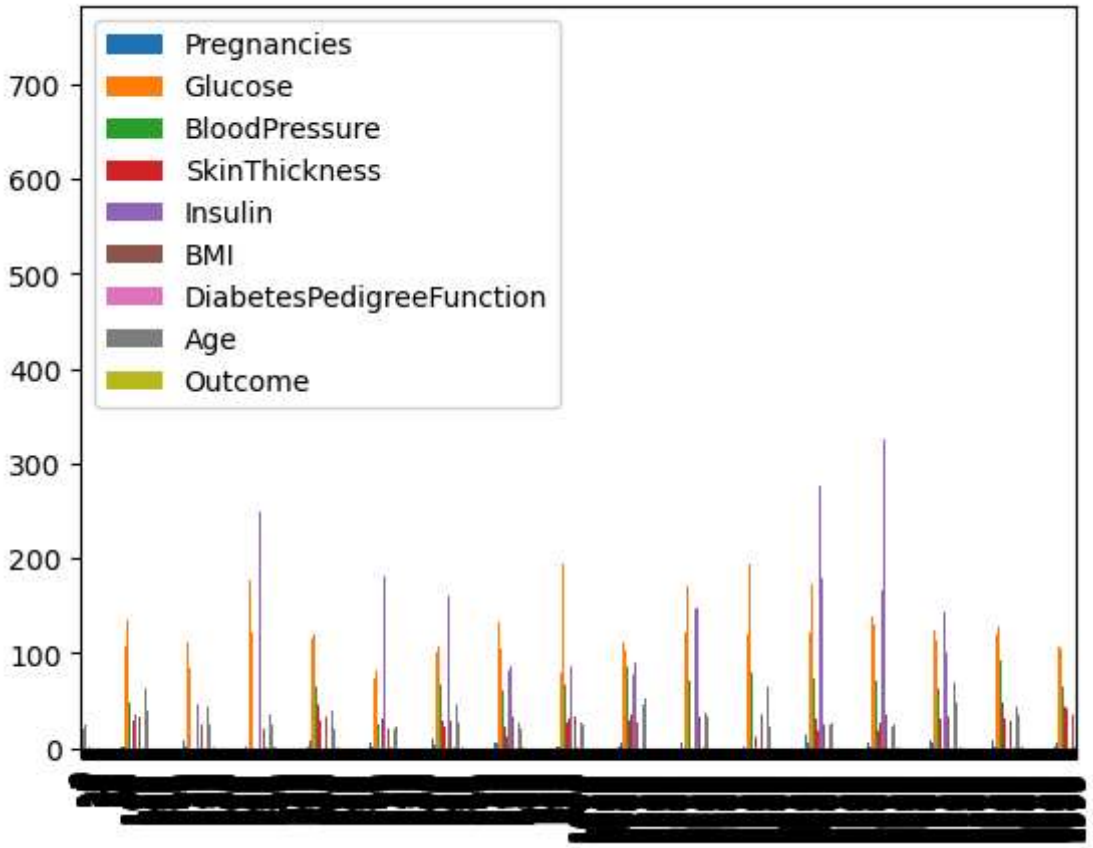
In [61]:

```
hd.plot(kind='bar')
hd
```

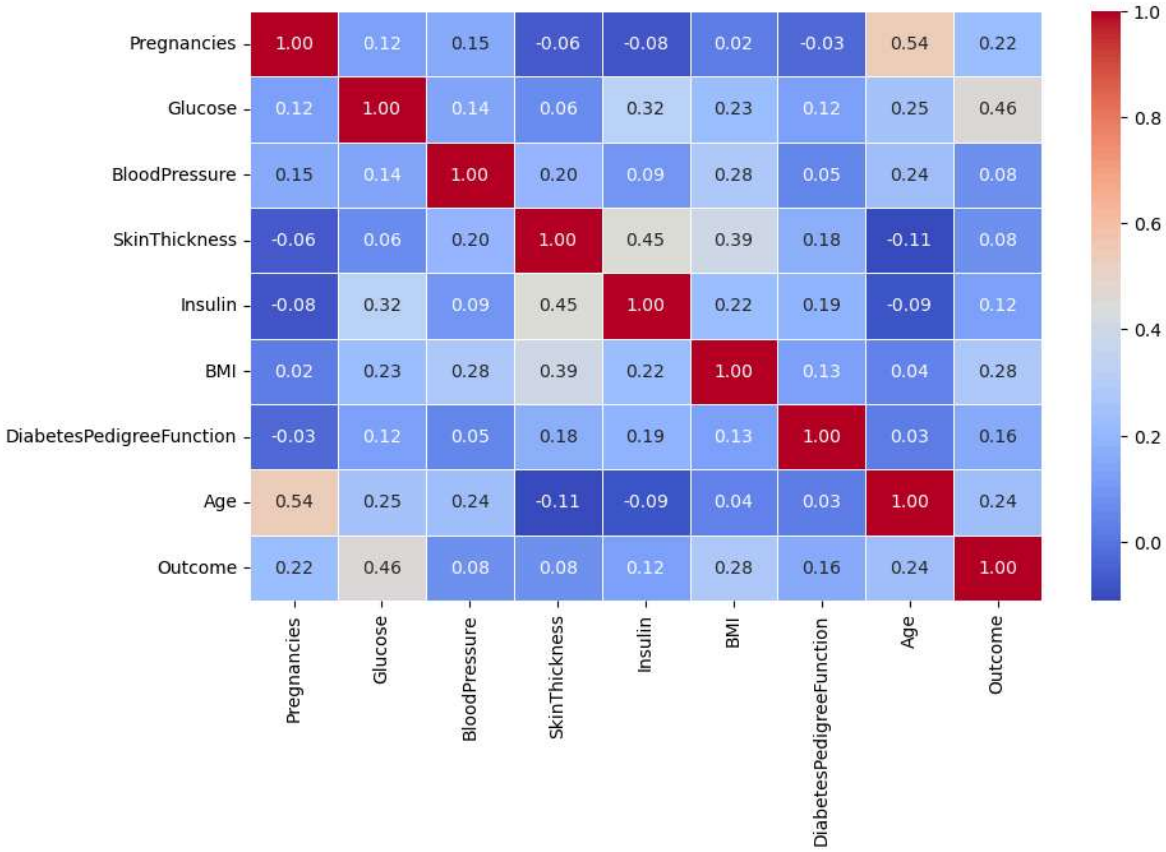
Out[61]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedig
0	2	138	62	35	0	33.6	
1	0	84	82	31	125	38.2	
2	0	145	0	0	0	44.2	
3	0	135	68	42	250	42.3	
4	1	139	62	41	480	40.7	
...	...	...	...	...	...	...	
1995	2	75	64	24	55	29.7	
1996	8	179	72	42	130	32.7	
1997	6	85	78	0	0	31.2	
1998	0	129	110	46	130	67.1	
1999	2	81	72	15	76	30.1	

2000 rows × 9 columns

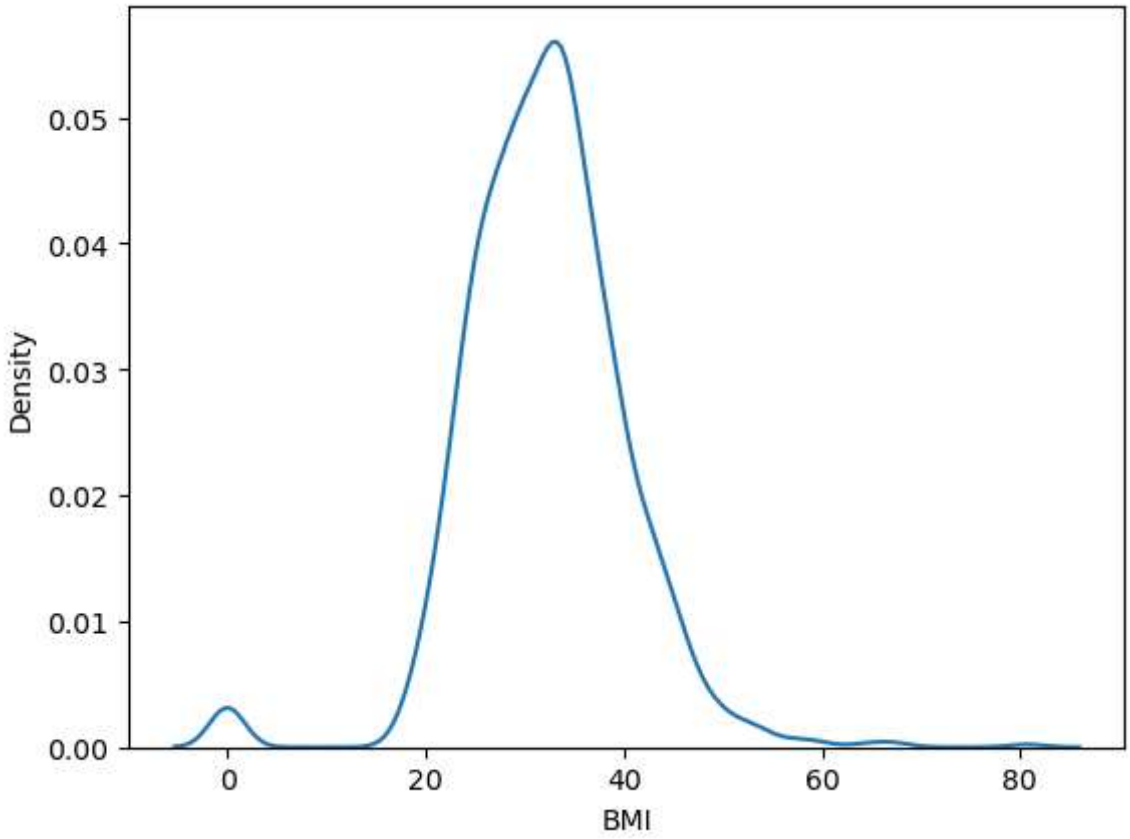


In [62]:
```
import seaborn as sns
correlation=hd.corr()
plt.figure(figsize=(10,6))
sns.heatmap(correlation,annot=True,cmap='coolwarm',
            fmt='.2f',linewidths=0.5,cbar=True,)
plt.show()
```

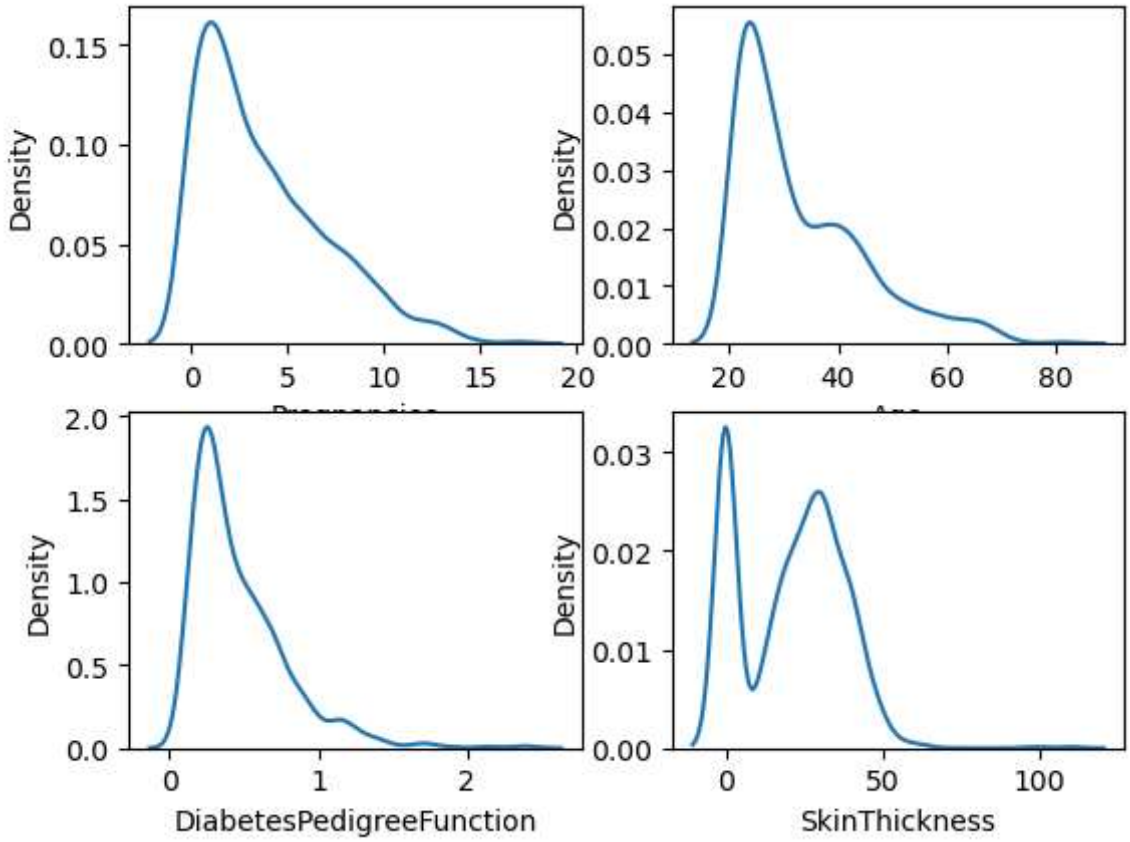


In [63]:
```
sns.kdeplot(hd["BMI"])
```

Out[63]: <Axes: xlabel='BMI', ylabel='Density'>



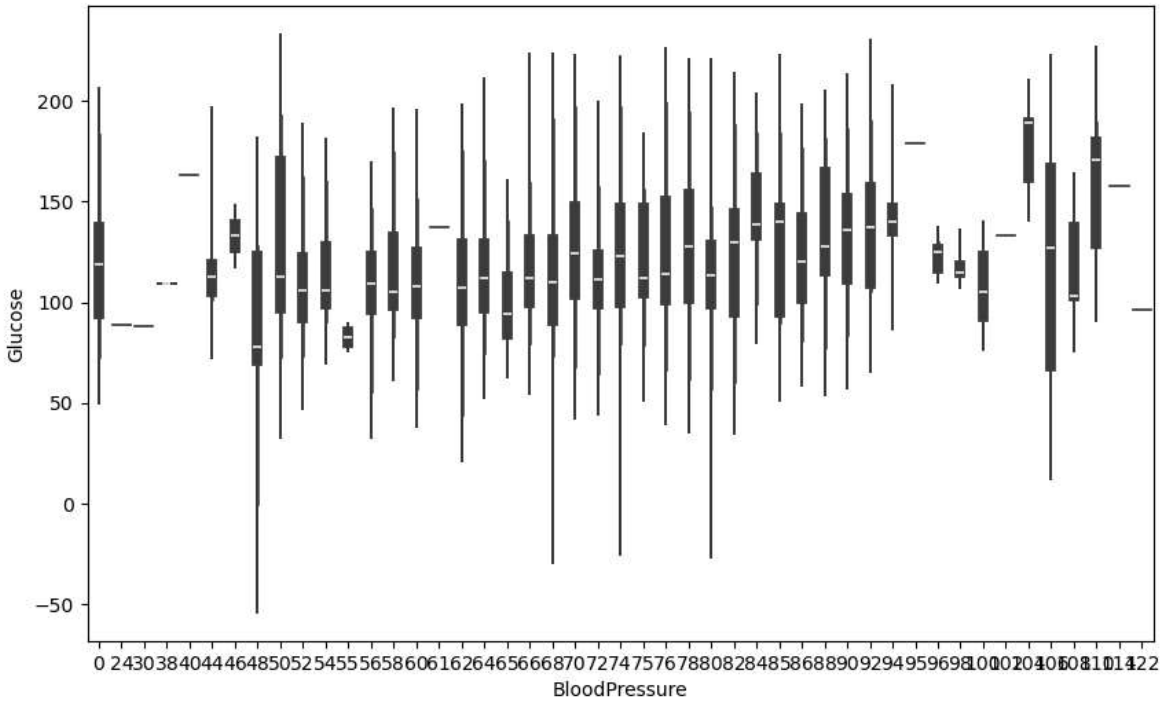
In [64]:
```
plt.subplot(2,2,1)
sns.kdeplot(hd["Pregnancies"])

plt.subplot(2,2,2)
sns.kdeplot(hd["Age"])

plt.subplot(2,2,3)
sns.kdeplot(hd["DiabetesPedigreeFunction"])

plt.subplot(2,2,4)
sns.kdeplot(hd["SkinThickness"])
plt.show()
```



Violin Chart

In [68]:
```
plt.figure(figsize=(10,6))
sns.violinplot (x=hd["BloodPressure"],y=hd["Glucose"],data=hd)
```

Out[68]: <Axes: xlabel='BloodPressure', ylabel='Glucose'>



In [ ]: