

```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
```

```
In [5]: data = pd.read_csv("Diabetes.csv")
```

```
In [8]: data.head()
```

Out[8]:

	Pregnancies	PG Concentration	Diastolic BP	Tri Fold Thick	Serum Ins	BMI	DP Function	Age	Diabetes
0	6	148.0	72.0	35.0	0.0	33.6	0.627	50.0	Sick
1	1	85.0	66.0	29.0	0.0	26.6	0.351	31.0	Healthy
2	8	183.0	64.0	0.0	0.0	23.3	0.672	32.0	Sick
3	1	89.0	66.0	23.0	94.0	28.1	0.167	21.0	Healthy
4	0	137.0	40.0	35.0	168.0	43.1	2.288	33.0	Sick

```
In [7]: data.tail()
```

Out[7]:

	Pregnancies	PG Concentration	Diastolic BP	Tri Fold Thick	Serum Ins	BMI	DP Function	Age	Diabetes
777	Serum Ins: 2- Hour Serum Insulin (mu U/ml)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
778	BMI: Body Mass Index: (weight in kg/ (height...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
779	DP Function: Diabetes Pedigree Function	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
780	Age: Age (years)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
781	Diabetes: Whether or not the person has diabetes	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

```
In [9]: data.shape
```

Out[9]: (782, 9)

In [10]: `data.describe()`

Out[10]:

	PG Concentration	Diastolic BP	Tri Fold Thick	Serum Ins	BMI	DP Function	Age
<b>count</b>	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000
<b>mean</b>	120.894531	69.105469	20.536458	79.799479	31.992578	0.471876	33.240885
<b>std</b>	31.972618	19.355807	15.952218	115.244002	7.884160	0.331329	11.760232
<b>min</b>	0.000000	0.000000	0.000000	0.000000	0.000000	0.078000	21.000000
<b>25%</b>	99.000000	62.000000	0.000000	0.000000	27.300000	0.243750	24.000000
<b>50%</b>	117.000000	72.000000	23.000000	30.500000	32.000000	0.372500	29.000000
<b>75%</b>	140.250000	80.000000	32.000000	127.250000	36.600000	0.626250	41.000000
<b>max</b>	199.000000	122.000000	99.000000	846.000000	67.100000	2.420000	81.000000

In [11]: `data.nunique()`

Out[11]:

Pregnancies	31
PG Concentration	136
Diastolic BP	47
Tri Fold Thick	51
Serum Ins	186
BMI	248
DP Function	517
Age	52
Diabetes	2
dtype:	int64

In [12]: `data['Tri Fold Thick'].unique()`

Out[12]:

```
array([35., 29.,  0., 23., 32., 45., 19., 47., 38., 30., 41., 33., 26
      , 15., 36., 11., 31., 37., 42., 25., 18., 24., 39., 27., 21., 34
      , 10., 60., 13., 20., 22., 28., 54., 40., 51., 56., 14., 17., 50
      , 44., 12., 46., 16.,  7., 52., 43., 48.,  8., 49., 63., 99., na
      n])
```

```
In [13]: data['Age'].unique()
```

```
Out[13]: array([50., 31., 32., 21., 33., 30., 26., 29., 53., 54., 34., 57., 59
    ' ',
    ' ', 51., 27., 41., 43., 22., 38., 60., 28., 45., 35., 46., 56., 37
    ' ',
    ' ', 48., 40., 25., 24., 58., 42., 44., 39., 36., 23., 61., 69., 62
    ' ',
    ' ', 55., 65., 47., 52., 66., 49., 63., 67., 72., 81., 64., 70., 68
    ' ',
    ' ', nan])
```

```
In [14]: data['Diabetes'].unique()
```

```
Out[14]: array(['Sick', 'Healthy', nan], dtype=object)
```

```
In [15]: data.isnull().sum()
```

```
Out[15]: Pregnancies      0
PG Concentration      14
Diastolic BP          14
Tri Fold Thick        14
Serum Ins             14
BMI                   14
DP Function           14
Age                   14
Diabetes              14
dtype: int64
```

```
In [16]: sns.heatmap(corelation, xticklables=corelation.columns, yticklables=co
```

```
-----
-----
NameError                                Traceback (most recent call
last)
<ipython-input-16-f5df89825d15> in <module>
----> 1 sns.heatmap(corelation, xticklables=corelation.columns,
yticklables=corelation.columns)

NameError: name 'corelation' is not defined
```

```
In [17]: student = data.drop(['Tri Fold Thick'], axis=1)
```

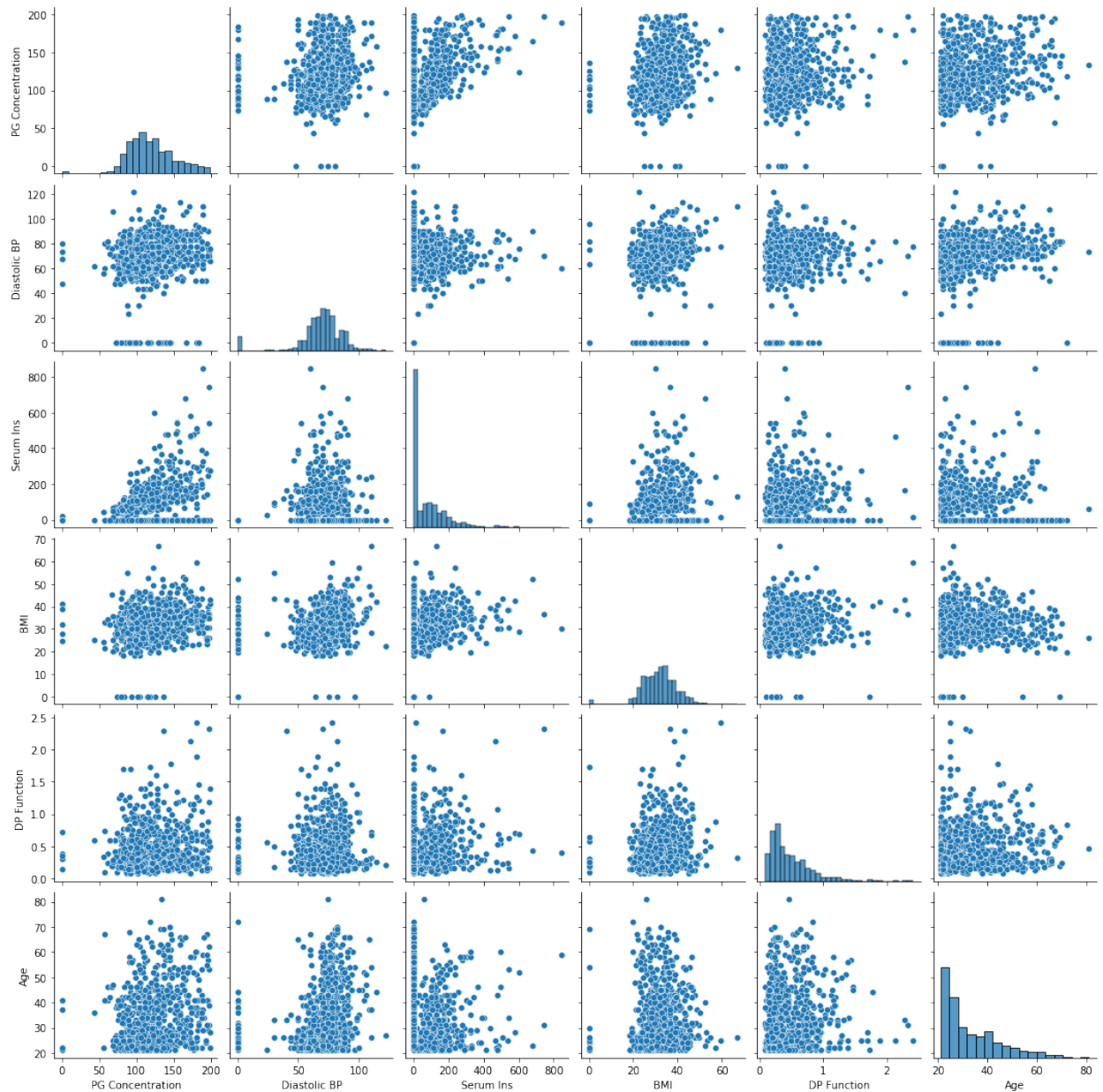
```
In [18]: corelation = student.corr()
```

```
In [19]: data.isnull().sum()
```

```
Out[19]: Pregnancies      0
          PG Concentration 14
          Diastolic BP     14
          Tri Fold Thick   14
          Serum Ins        14
          BMI              14
          DP Function       14
          Age              14
          Diabetes         14
          dtype: int64
```

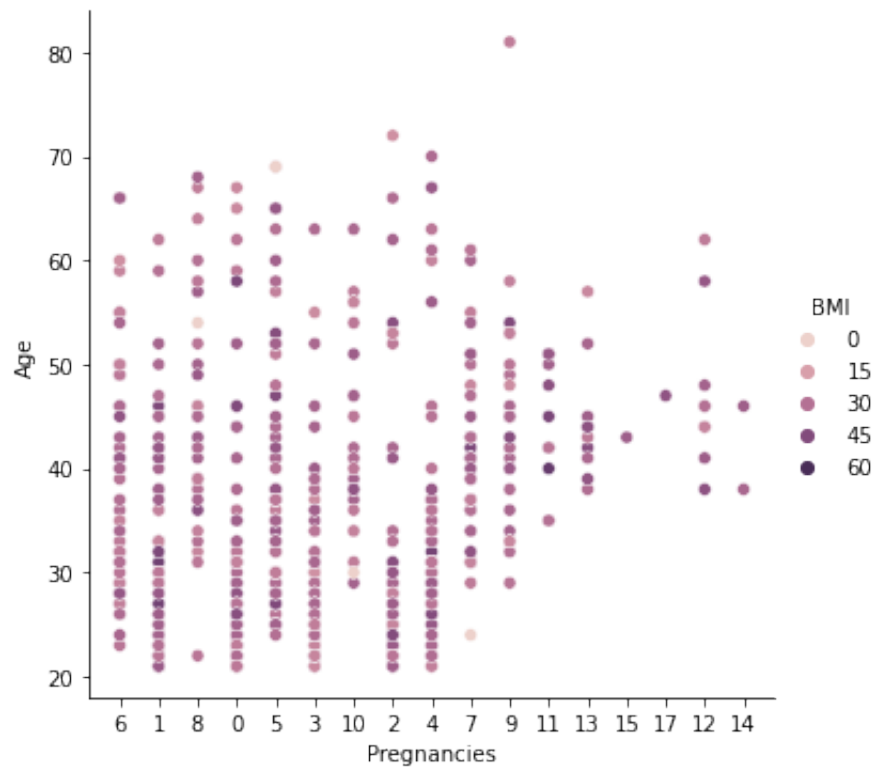
```
In [22]: sns.pairplot(student)
```

```
Out[22]: <seaborn.axisgrid.PairGrid at 0x7f9d78defa00>
```



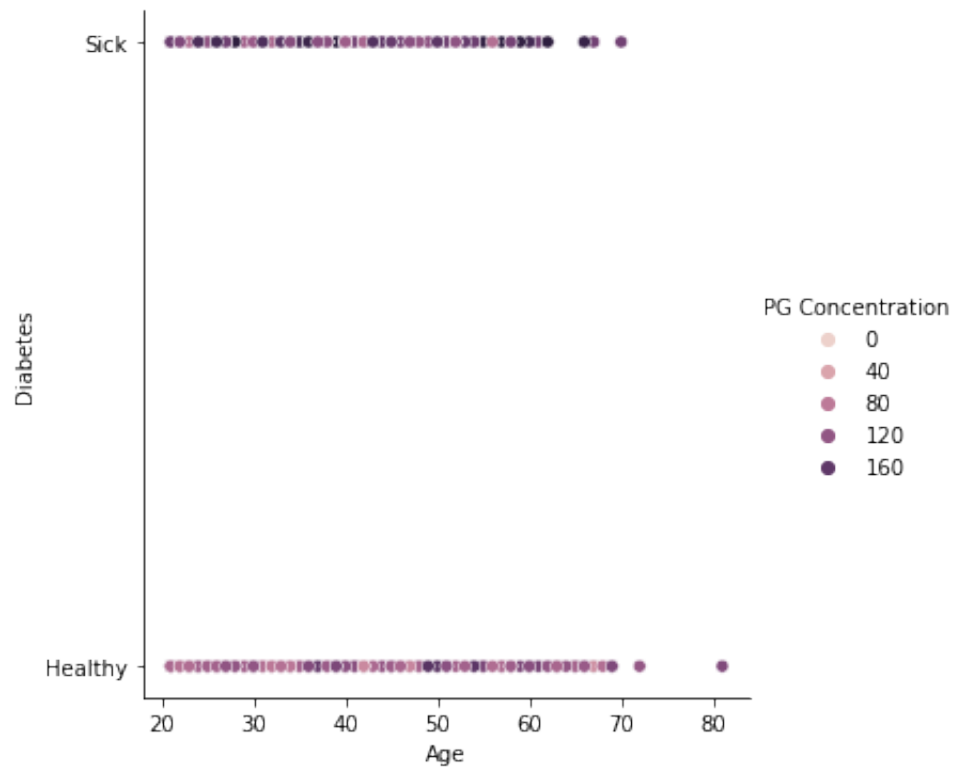
```
In [23]: sns.relplot(x= 'Pregnancies', y= 'Age', hue= 'BMI', data= student)
```

```
Out[23]: <seaborn.axisgrid.FacetGrid at 0x7f9d7a227a30>
```



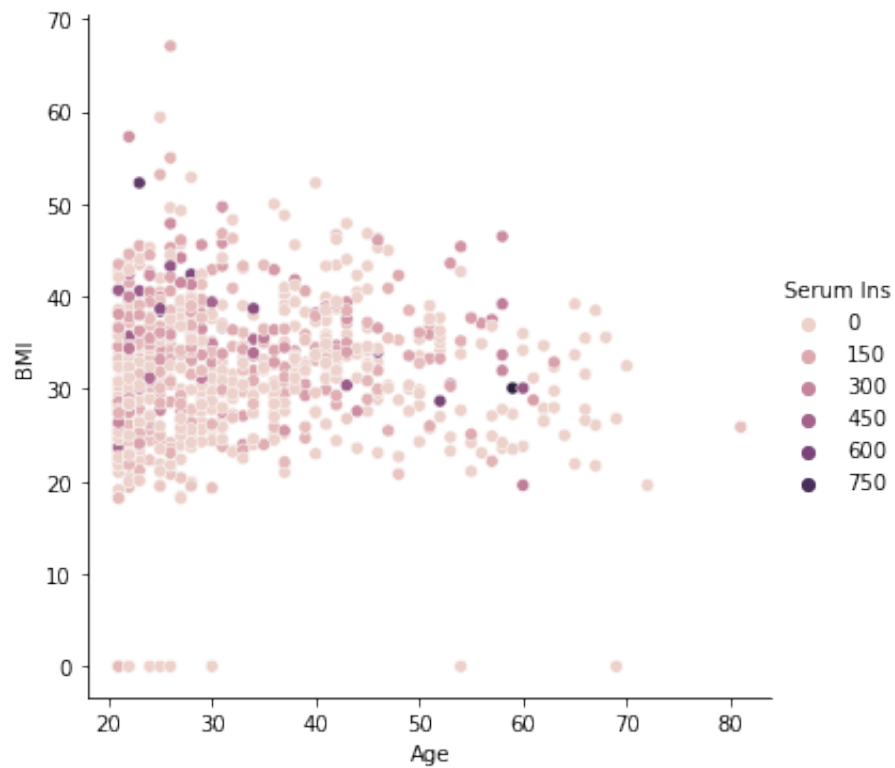
```
In [24]: sns.relplot(x= 'Age', y= 'Diabetes', hue= 'PG Concentration', data= st
```

```
Out[24]: <seaborn.axisgrid.FacetGrid at 0x7f9d7a0ddb50>
```



```
In [25]: sns.relplot(x= 'Age', y= 'BMI', hue= 'Serum Ins', data= student)
```

```
Out[25]: <seaborn.axisgrid.FacetGrid at 0x7f9d7a16a3d0>
```



```
In [ ]:
```