In [3]:
```python
import pandas as pd
import seaborn as sns
```
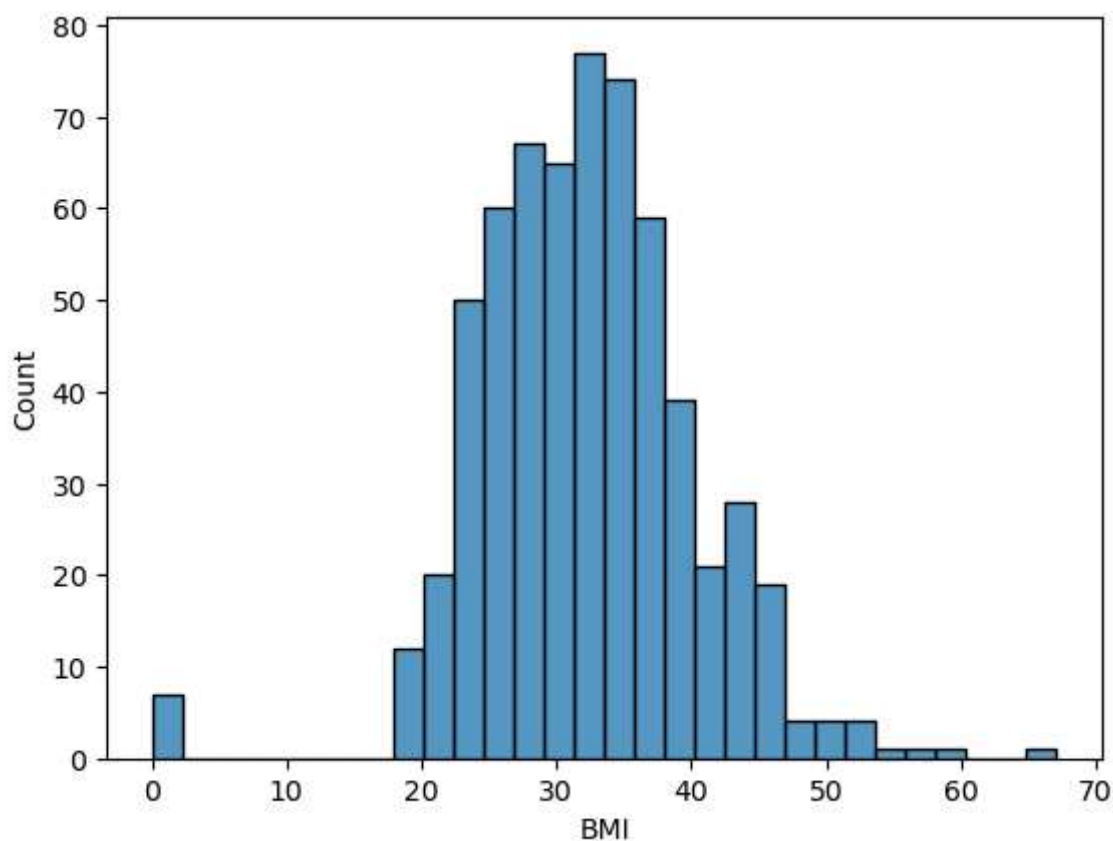
In [85]:
```python
data=pd.read_csv("C:\\Users\\Gouri\\Downloads\\diabetes_model.csv")
data.head()
```

Out[85]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeF
0	5	77	82	41	42	35.8	
1	9	122	56	0	0	33.3	
2	0	113	76	0	0	33.3	
3	1	139	62	41	480	40.7	
4	10	161	68	23	132	25.5	

In [24]:
```python
sns.histplot(data["BMI"])
```

Out[24]: <Axes: xlabel='BMI', ylabel='Count'>

In [15]:
```python
sns.histplot(data["Glucose"])
```
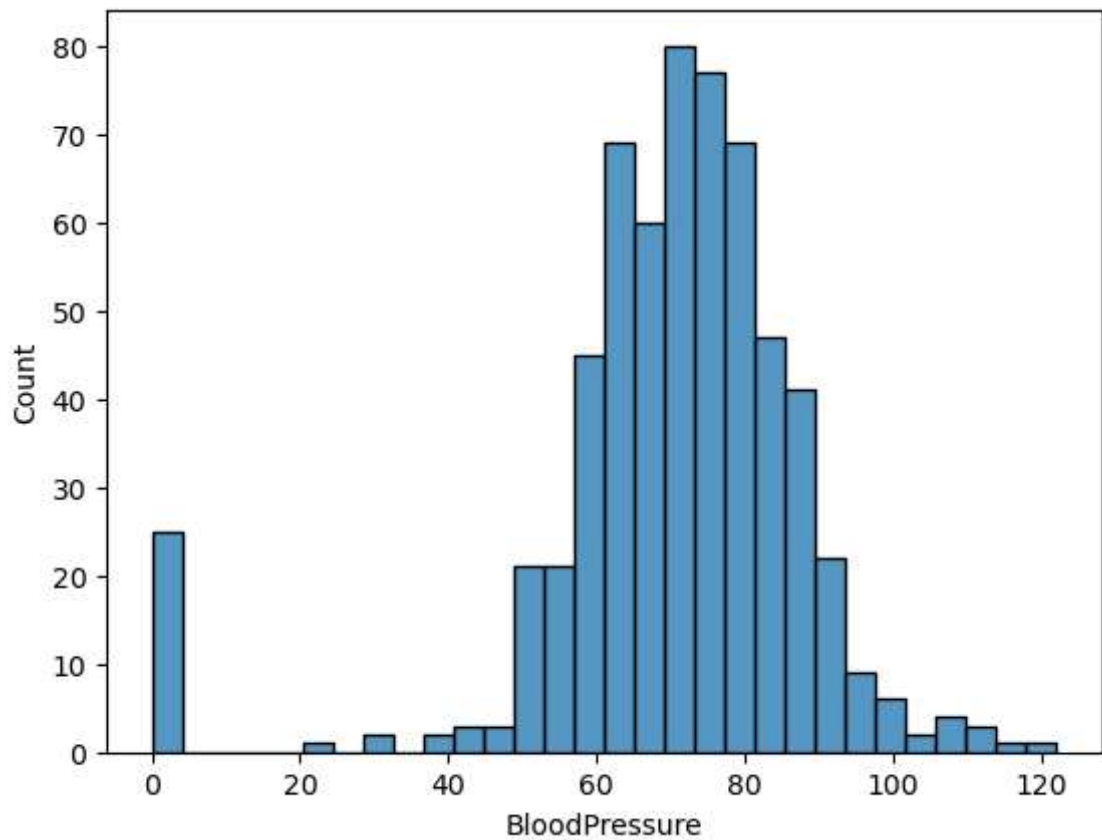
Out[15]: <Axes: xlabel='Glucose', ylabel='Count'>
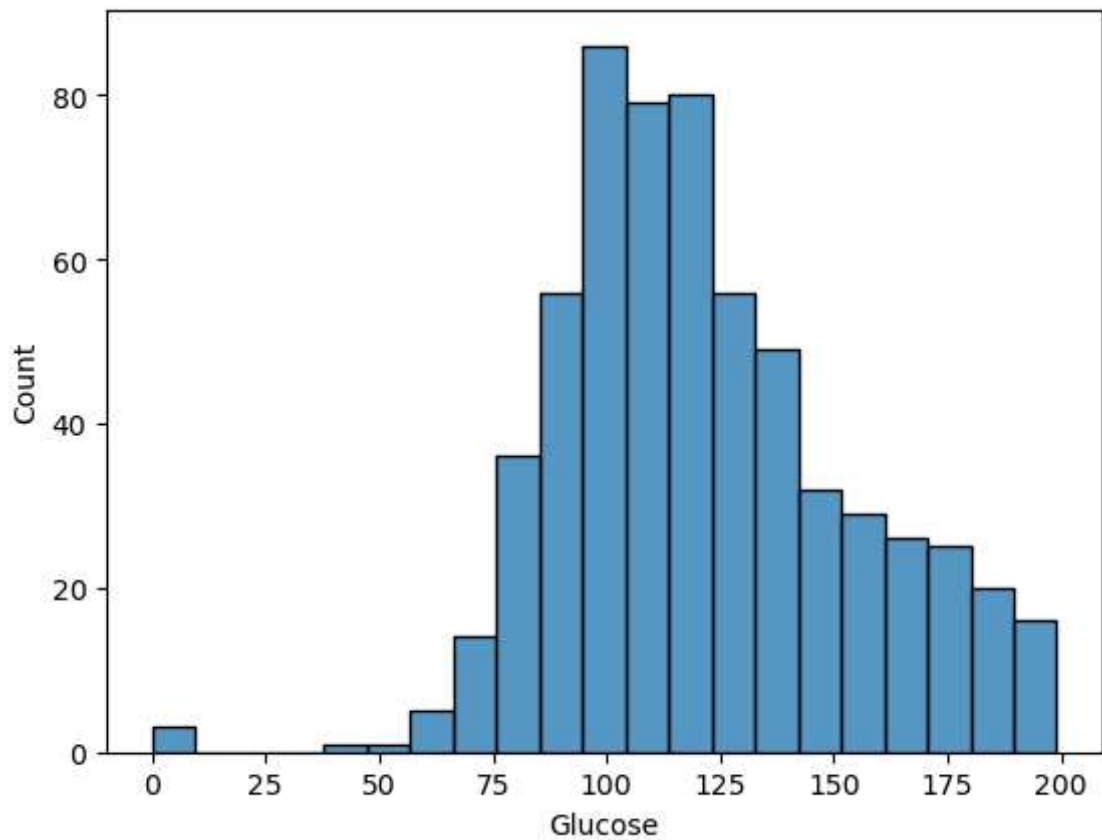
```
In [17]:  sns.histplot(data["BloodPressure"])
```
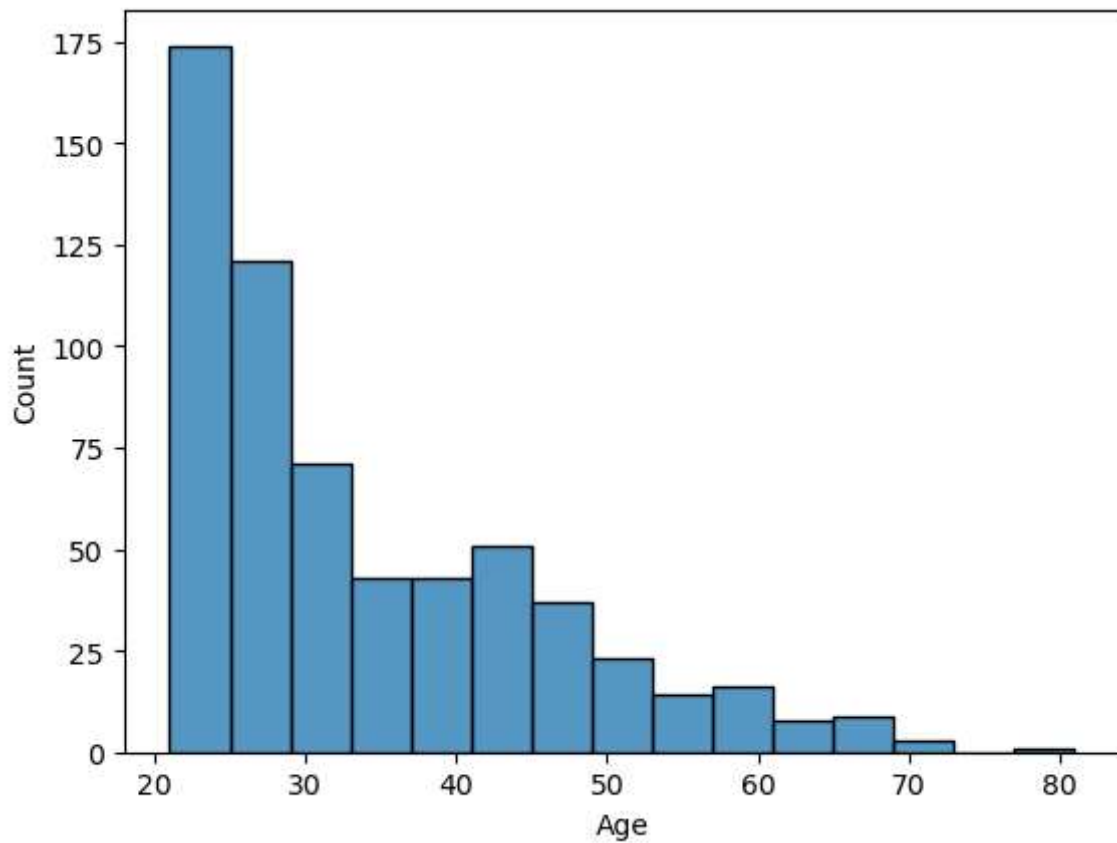
```
Out[17]:  <Axes: xlabel='BloodPressure', ylabel='Count'>
```
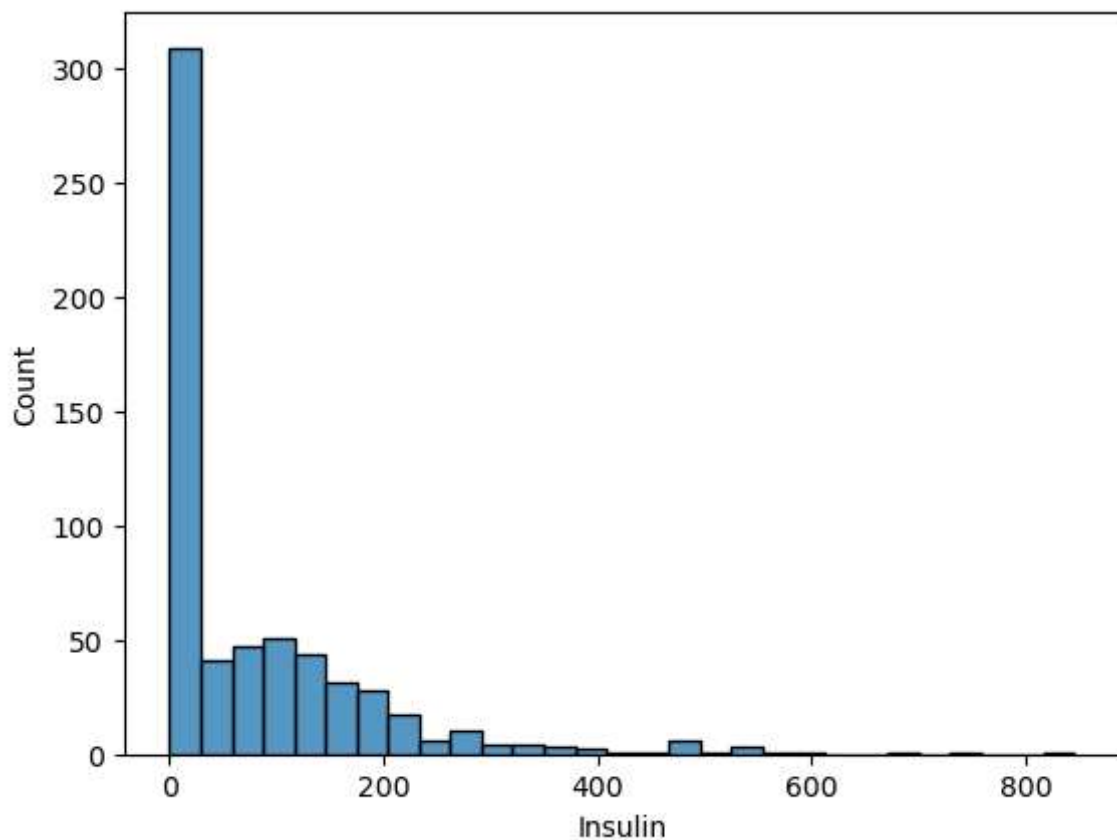


```
In [26]:  sns.histplot(data["Age"])
```

```
Out[26]:  <Axes: xlabel='Age', ylabel='Count'>
```

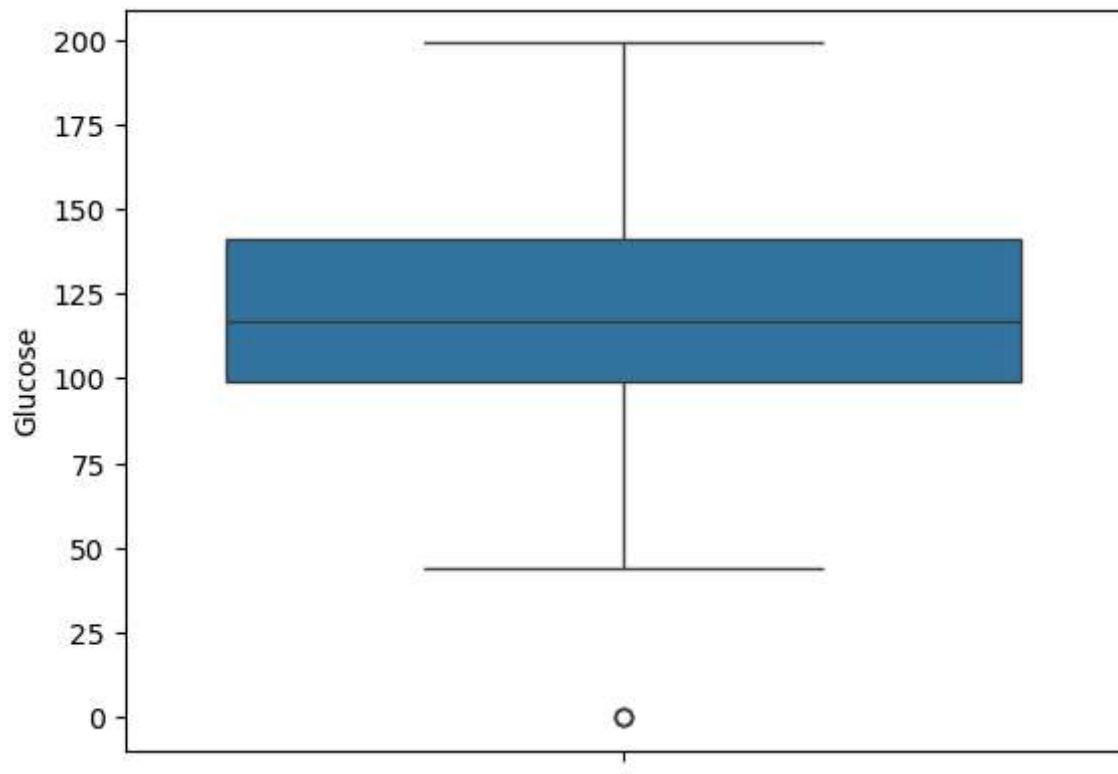In [ ]: # It can be seen that large number of people are from the age group of 20 to 30.

In [21]: sns.histplot(data["Insulin"])

Out[21]: <Axes: xlabel='Insulin', ylabel='Count'>
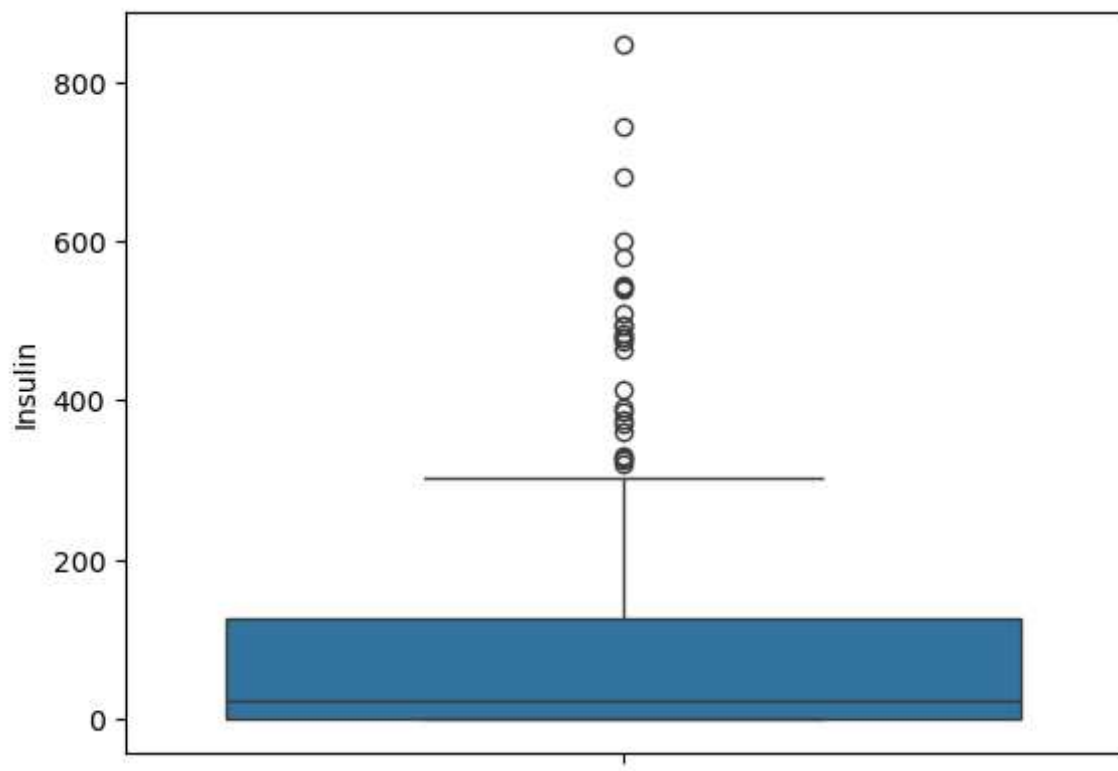

In [50]: sns.boxplot(data['Glucose'])

Out[50]:   <Axes: ylabel='Glucose'>

In [ ]:   #There is only one outlier value in glucose datapoints

In [52]:   sns.boxplot(data['Insulin'])
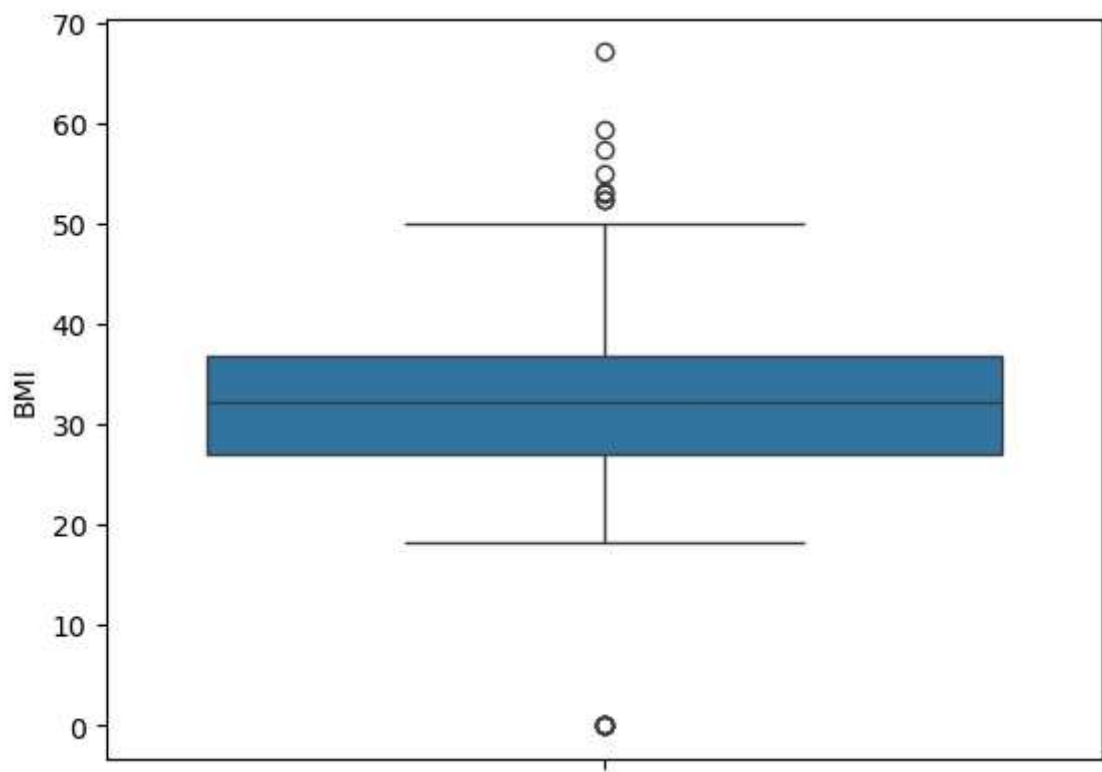
Out[52]:   <Axes: ylabel='Insulin'>

In [ ]:   # Compared to glucose datapoints, insulin datapoints has more outliers above the
          # it can be seen that a large majority of people do not use insulin. Since the m
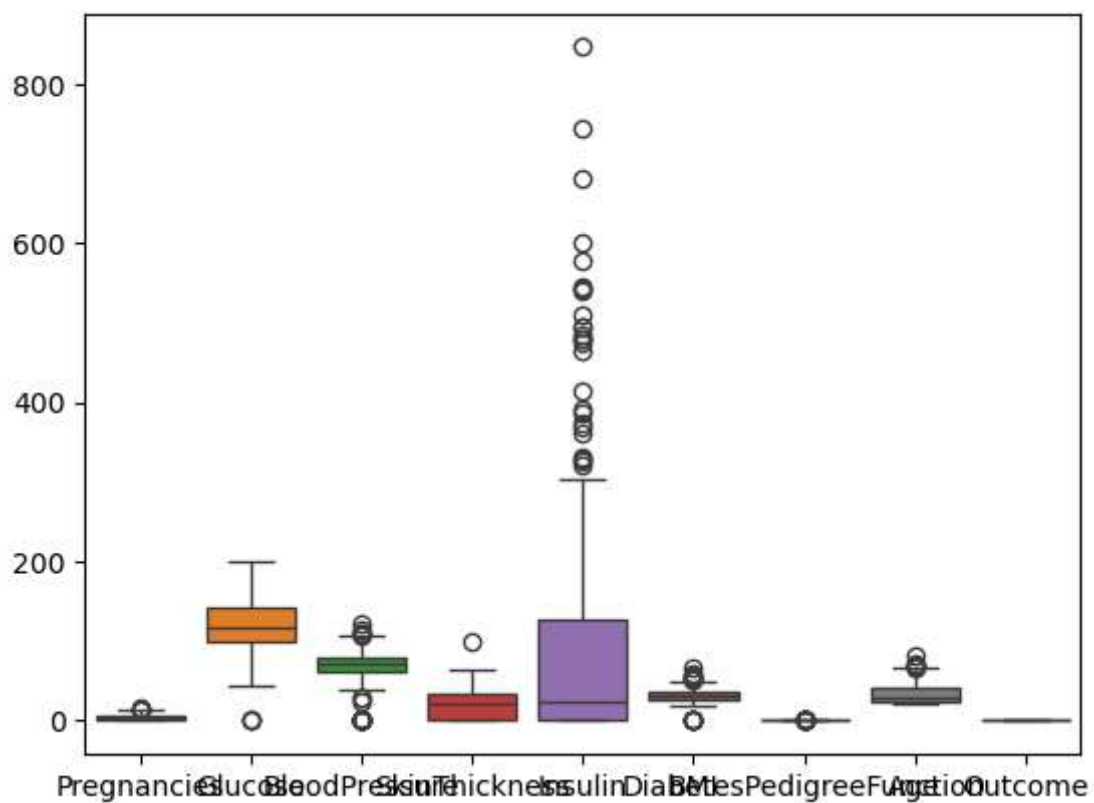
In [54]: `sns.boxplot(data["BMI"])`

Out[54]: `<Axes: ylabel='BMI'>`

In [56]: `# BMI Boxplot have a fairly distributed plot with a few outliers on either of th`

In [87]: `sns.boxplot(data)`
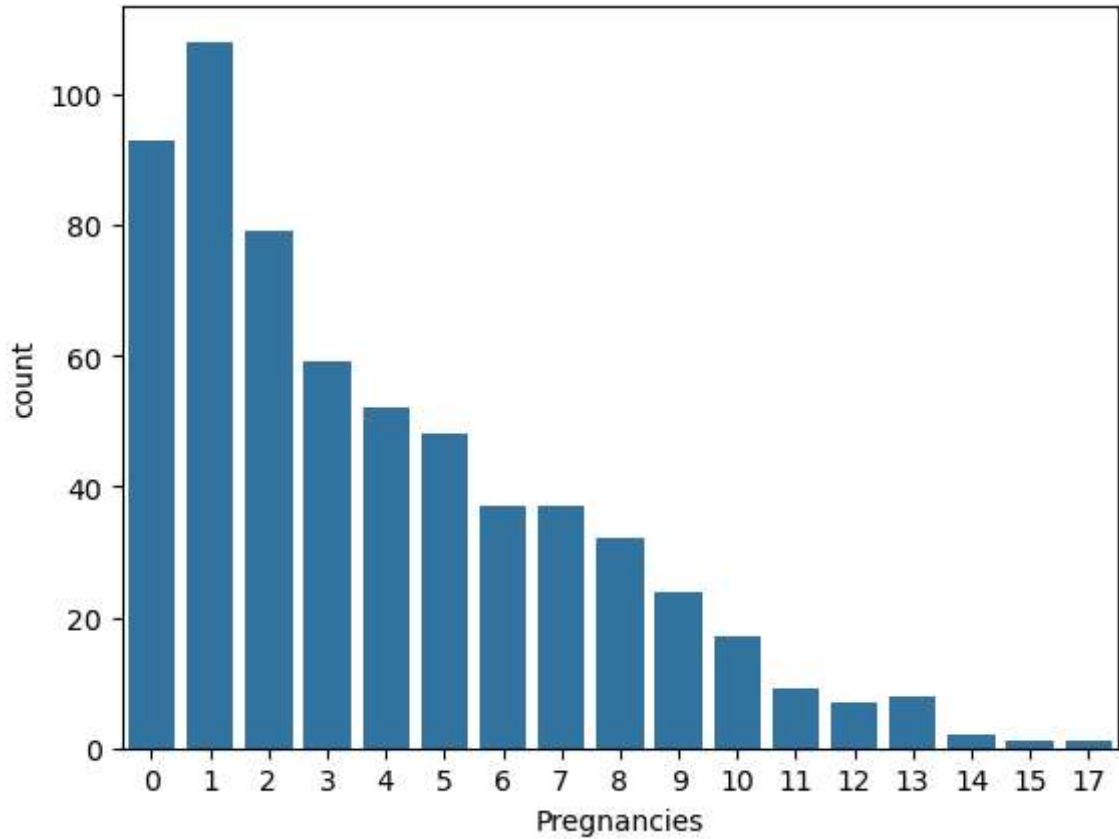
Out[87]: `<Axes: >`

In [64]:   *#From this boxplot, we can infer that the insulin datapoints have the highest*

In [109…
```python
import matplotlib.pyplot as plt
sns.countplot(x=data["Pregnancies"])
plt.figure(figsize=(50, 50))
```
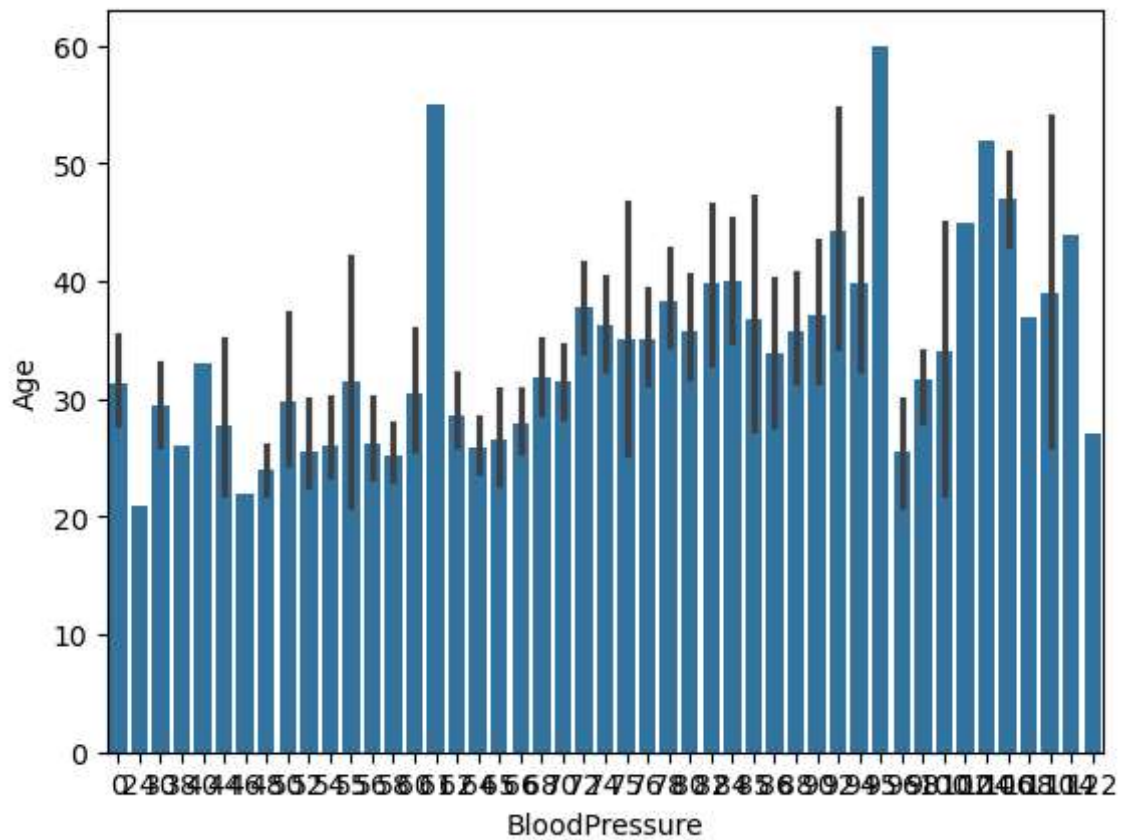
Out[109…   <Figure size 5000x5000 with 0 Axes>

<Figure size 5000x5000 with 0 Axes>

In [111…   *#most number of people have one pregnancy.*

In [115…
```python
sns.barplot(x=data['BloodPressure'], y=data['Age'])
```
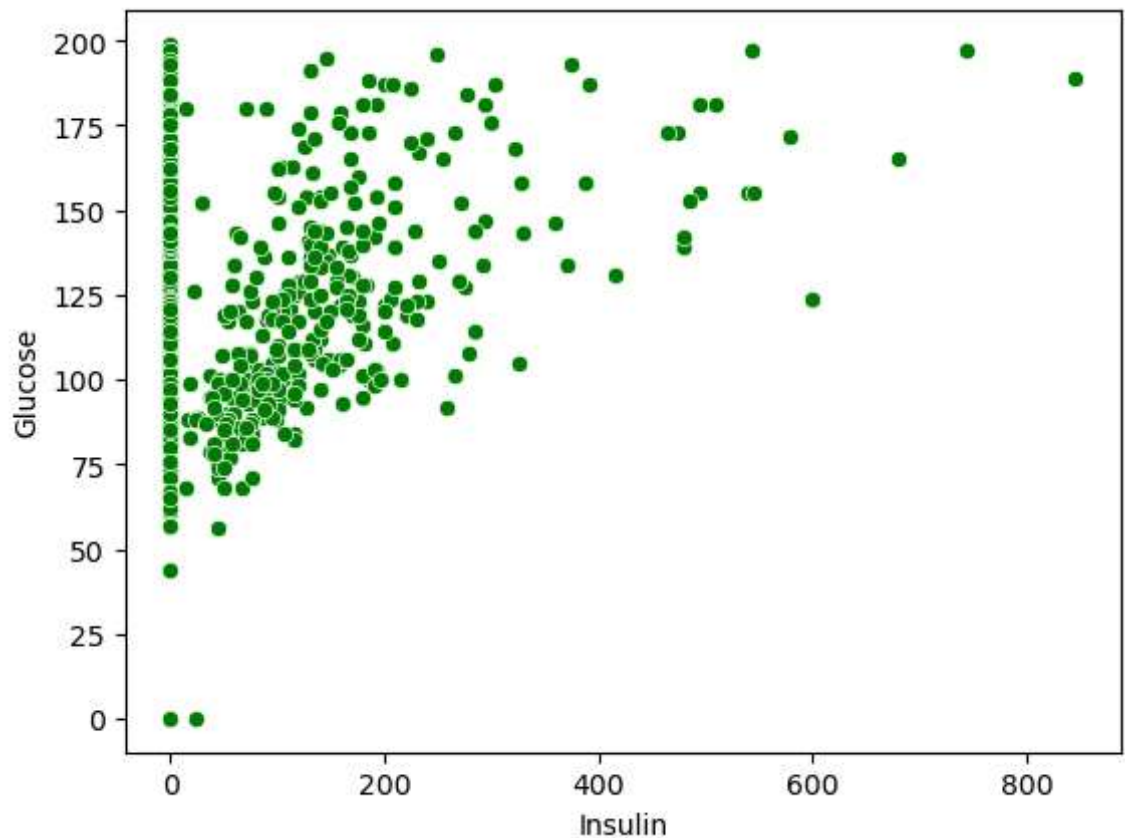
Out[115…   <Axes: xlabel='BloodPressure', ylabel='Age'>

```
In [117…   #While comparing boxplot and blood pressure, it can be seen there is a major dif

In [146…   sns.scatterplot(x=data["Insulin"], color="green",
                           y= data["Glucose"])

Out[146…   <Axes: xlabel='Insulin', ylabel='Glucose'>
```

In [162…
```python
plt.figure(figsize=(8,6))
plt.subplot(2,1,1)
plt.scatter(data["Insulin"], data["Glucose"], c="red", alpha=0.3)
plt.ylabel("Glucose")
plt.xlabel("Insulin")
plt.subplot(2,1,2)
plt.scatter(data["BloodPressure"], data["Glucose"], c="blue", alpha=0.1)
plt.xlabel("Blood Pressure")
plt.ylabel("Glucose")
plt.show()

#Glucose and Insulin scatter plot is heavily distributed in the 0 to 200 range.
#taken also increases. Glucose levels of those who take insulin within the range
#be seen between gluose and insulin, because as the glucose levels increase, amo
```
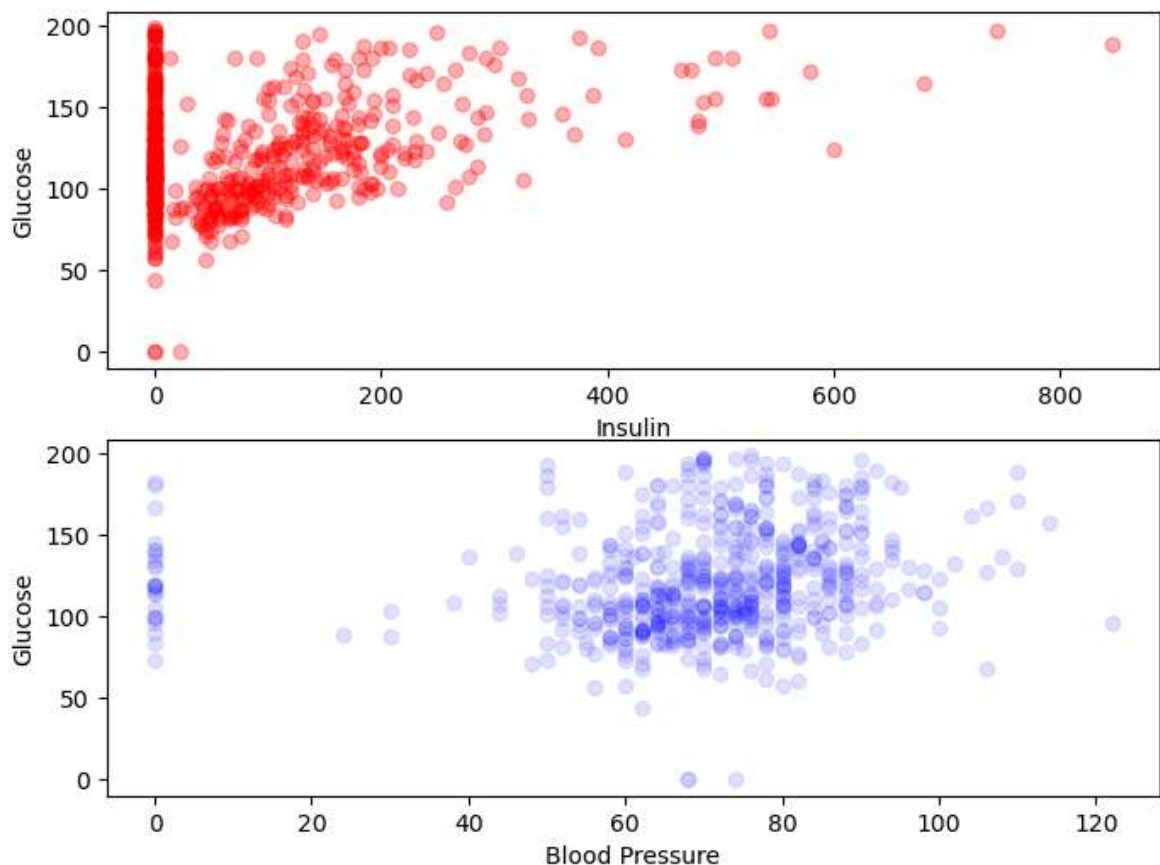

In [ ]:
```python
#Blood pressure v/c Glucose scatter plot :
#There are many outliers. as blood pressure increases, glucose levels also incre
#side. there are two values within the range 60 to 80 bp that have very low gluc
```
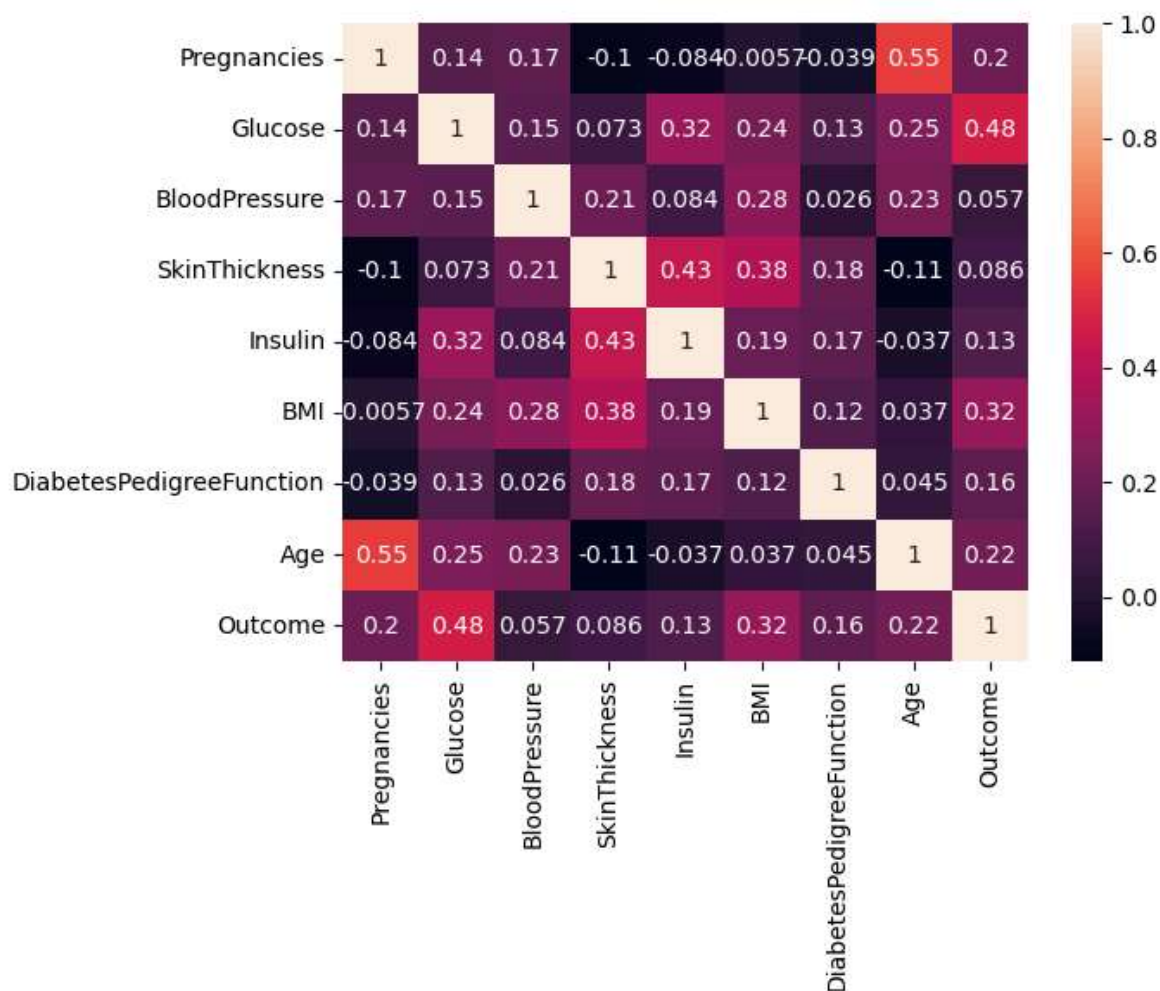
In [148…
```python
sns.heatmap(data.corr(), annot=True)
```

Out[148…   <Axes: >

```
In [164…   #Highest positive correlation is between pregnancies and age(0.55) and lowest be

In [ ]:

In [173…   sns.pairplot(data)

Out[173…   <seaborn.axisgrid.PairGrid at 0x167dce25f10>
```
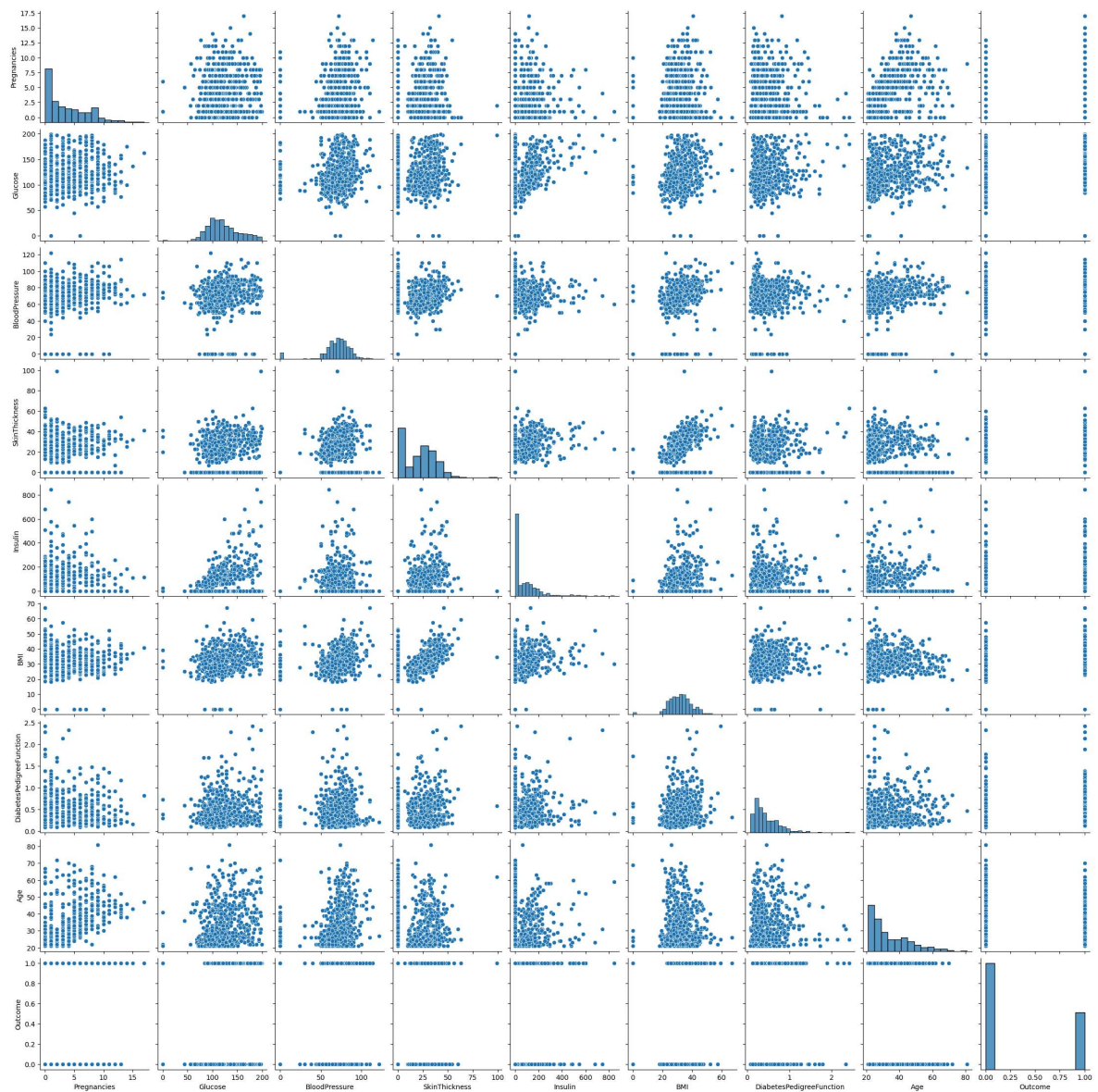
In [ ]: #from the pair plot it can eb seen that all individual data values have similar