Analyzing the diabetes datasheet using python

What is diabetes?

Diabetes is a chronic, metabolic disease characterized by elevated levels of blood glucose (or blood sugar), which leads over time to serious damage to the heart, blood vessels, eyes, kidneys and nerves.

- Type 1 diabetes (genetics & virus, where our immune system destroy the pancreas i.e.. Beta cell)
- Type 2 diabetes (pancreas produce less insulin)
- Gestational diabetes (due to pregnancy)
- Maturity onset diabetes of the young (MODY) (non-obese children, adolescents, and young adults)

Average blood glucose level is between 80Mg/dl to 12o Mg/dl

The average blood pressure for a healthy adult is less than 120/80 millimeters of mercury (mmHg)

Average BMI is 13 to 23

To import the csv file

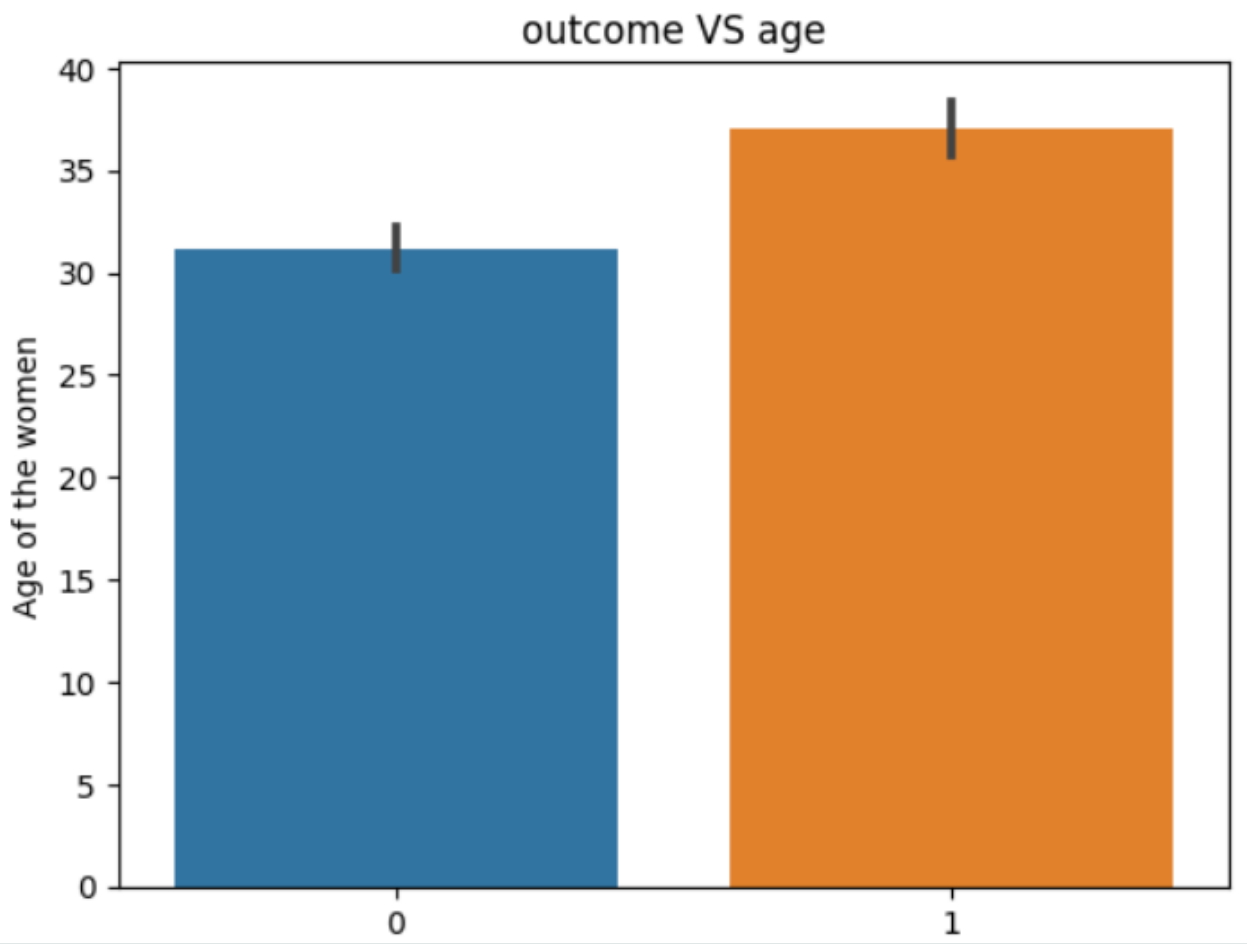
In [35]:
```
# similarly to print dataset in a structured manner

pd.DataFrame(df)
```

Out[35]:

	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
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1
...	...	...	...	...	...	...	...	...	...
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

768 rows × 9 columns

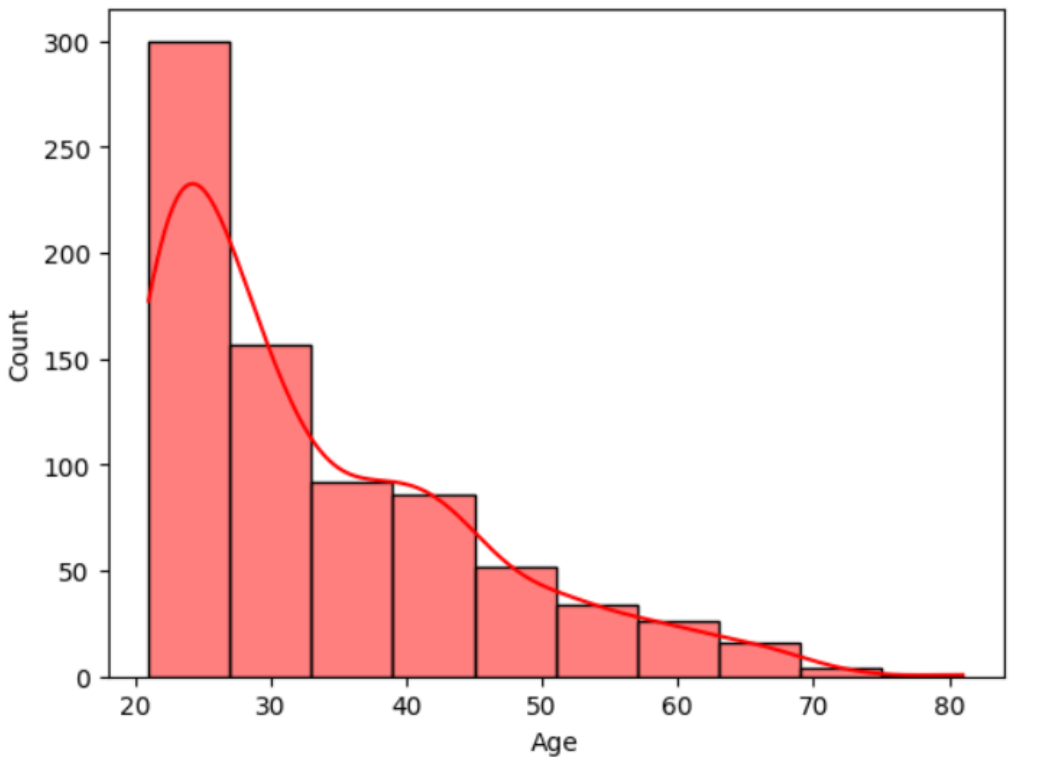
In [59]:
```
# plotting bar graph between age and outcome
# here data = data, this refers to complete datas in the datasheet otherwise DataFrame

sns.barplot(x="Outcome",y="Age",data= data)
plt.xlabel("outcome")
plt.ylabel("Age of the women")
plt.title("outcome VS age")
plt.show()
```

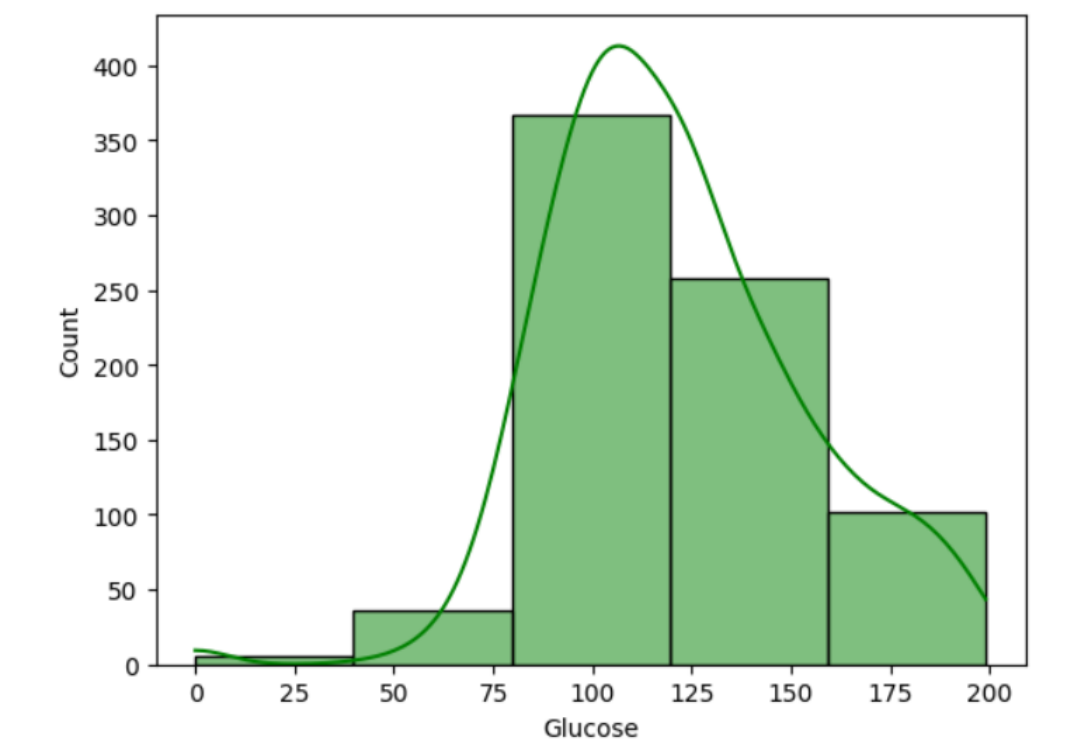


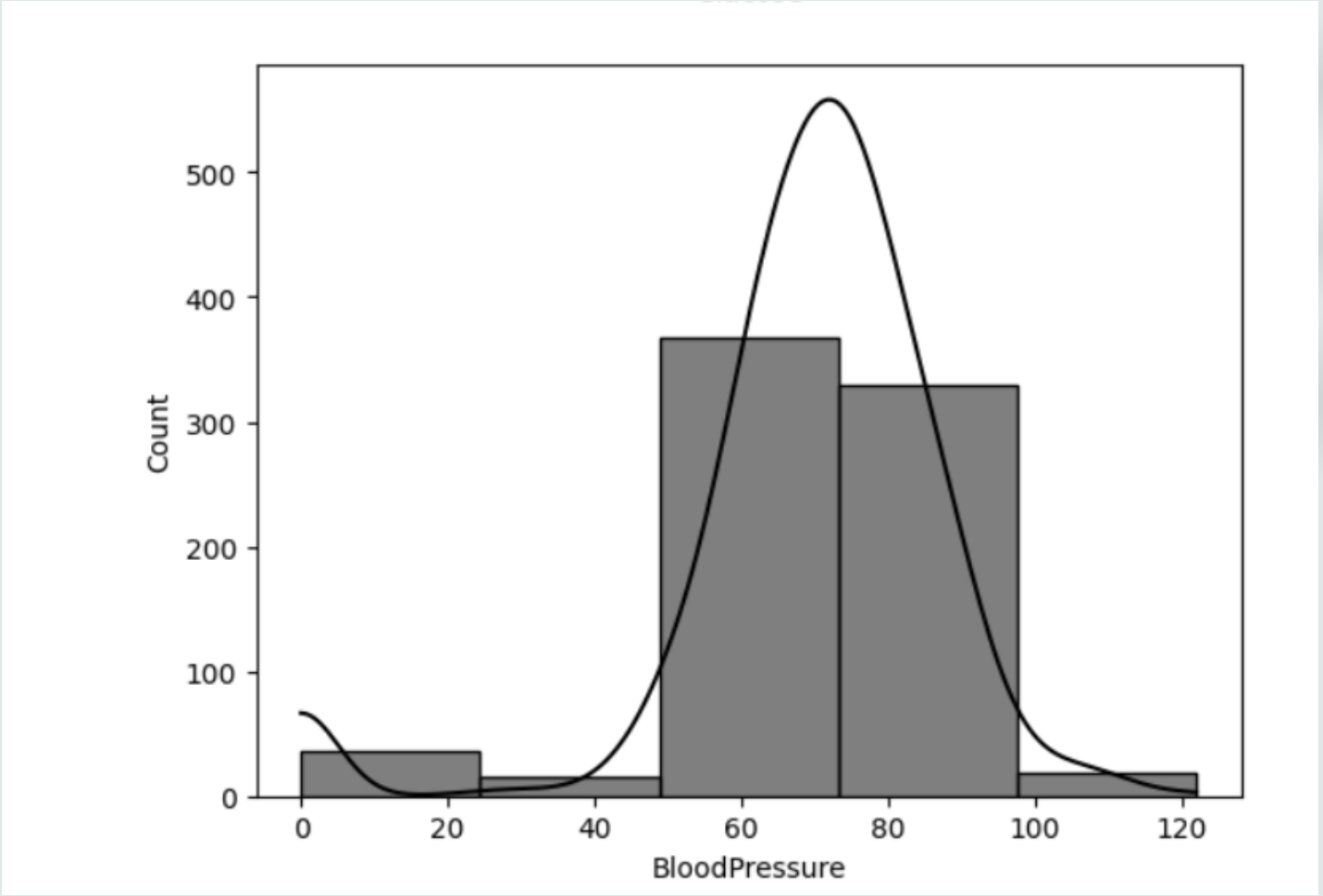
Insights from barplot

- The plot also shows a trendline, which is a line that best fits the data. The trendline shows that the average risk of diabetes increases by about 0.5% for each additional year of age.

Using histplot

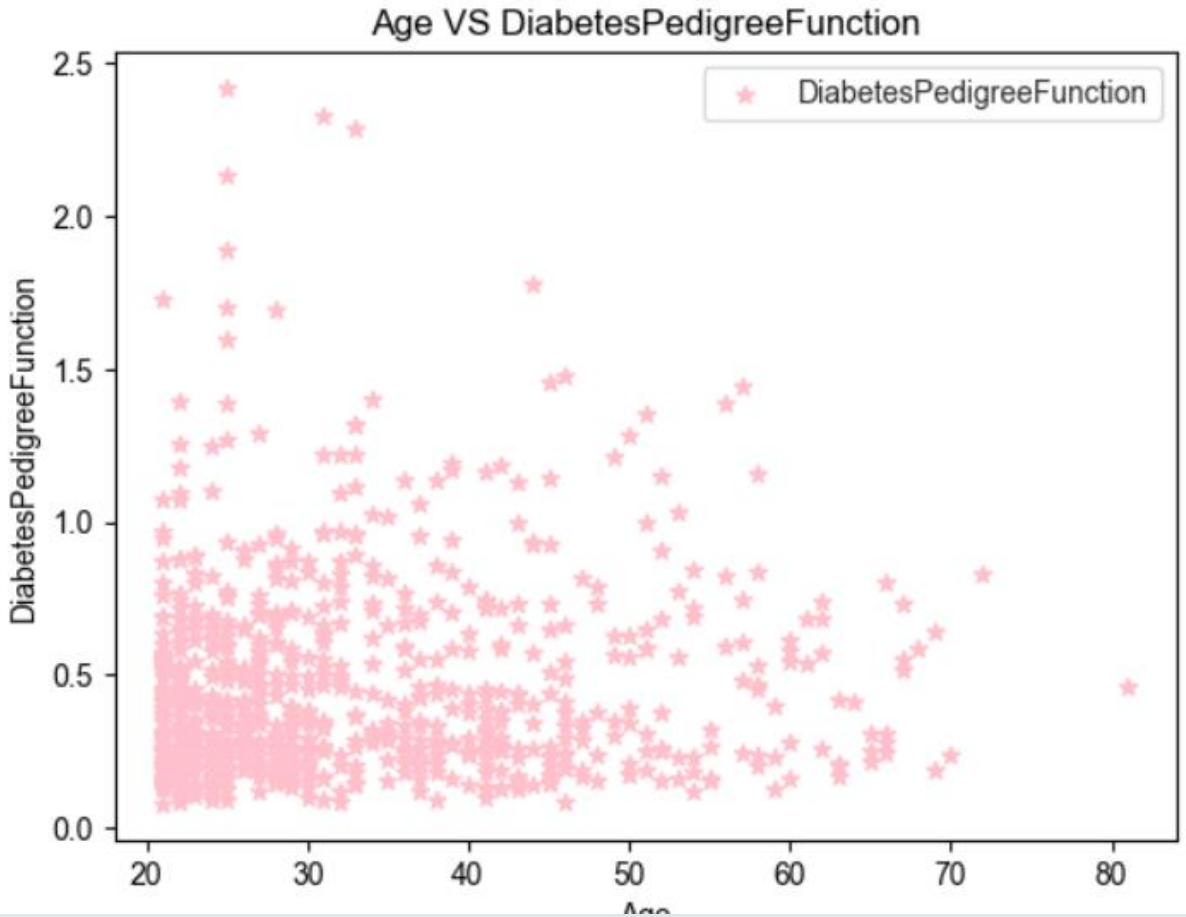





Count
500
400
300
200
100
0
0
20
40
60
80
100
120
BloodPressure

In [201]:
```
plt.scatter(x="Age",y="DiabetesPedigreeFunction",c= "pink",marker='*', data =data)
plt.xlabel("Age")
plt.ylabel("DiabetesPedigreeFunction")
plt.title("Age VS DiabetesPedigreeFunction")
plt.legend()
```

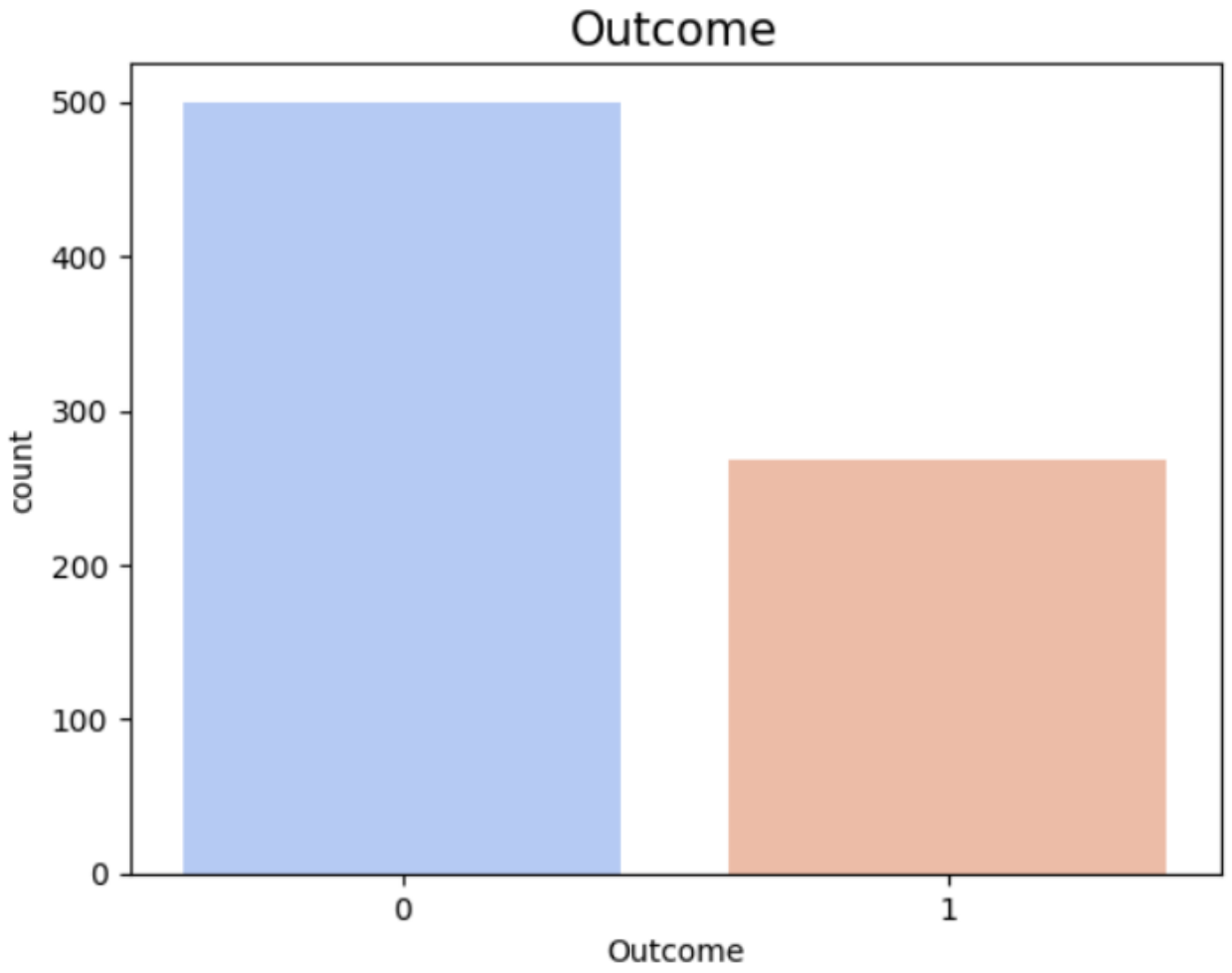
Out[201]: <matplotlib.legend.Legend at 0x147fa459e50>

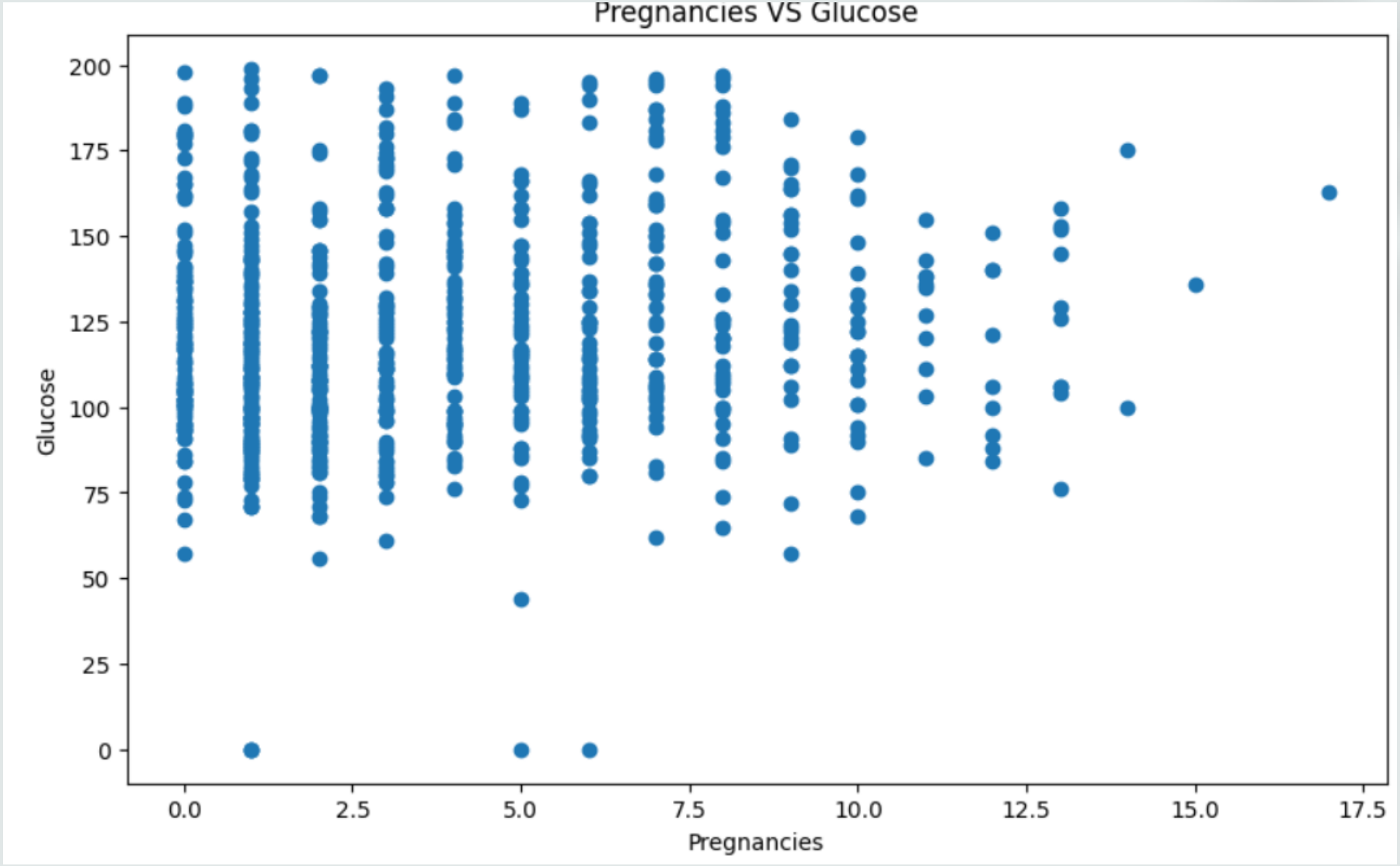


Insights from scatterplot

- The scatter plot shows a positive correlation between the two variables, which means that as age increases, the incidence of diabetes also tends to increase.
- The trendline shows that the average diabetes incidence increases by about 0.1% for each additional year of age

In [82]:
```
# no.of.women who is diabetic

sns.countplot(x="Outcome",palette="coolwarm",data= data)
plt.title("Outcome",fontsize= 15)
plt.show()
```

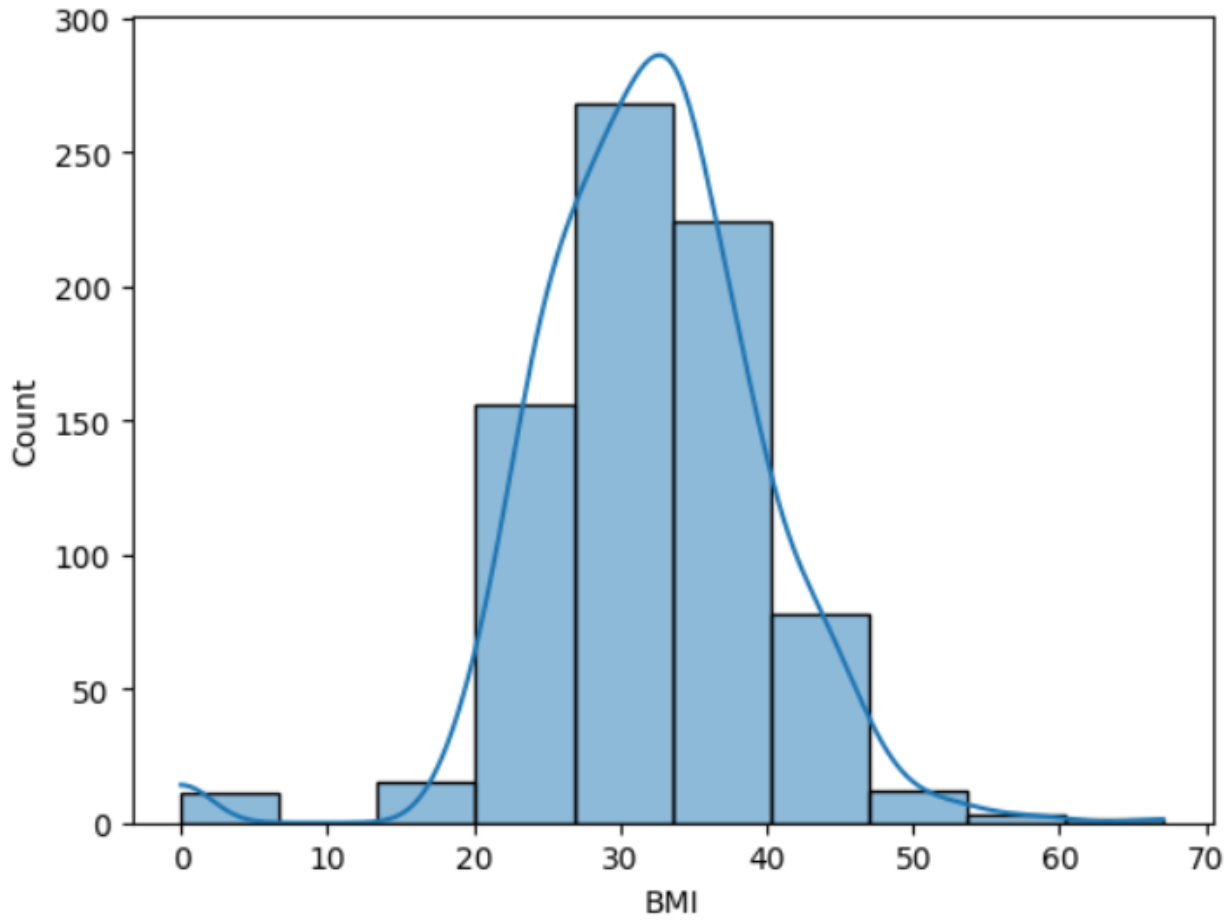



Pregnancies VS Glucose
Glucose
200
175
150
125
100
75
50
25
0
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Pregnancies

Insights from scatterplot

- The scatter plot also shows a trendline, which is a line that best fits the data. The trendline shows that the average glucose level increases by about 10 mg/dL for each additional pregnancy.
- The trendline shows that the average glucose level increases by about 10 mg/dL for each additional pregnancy.

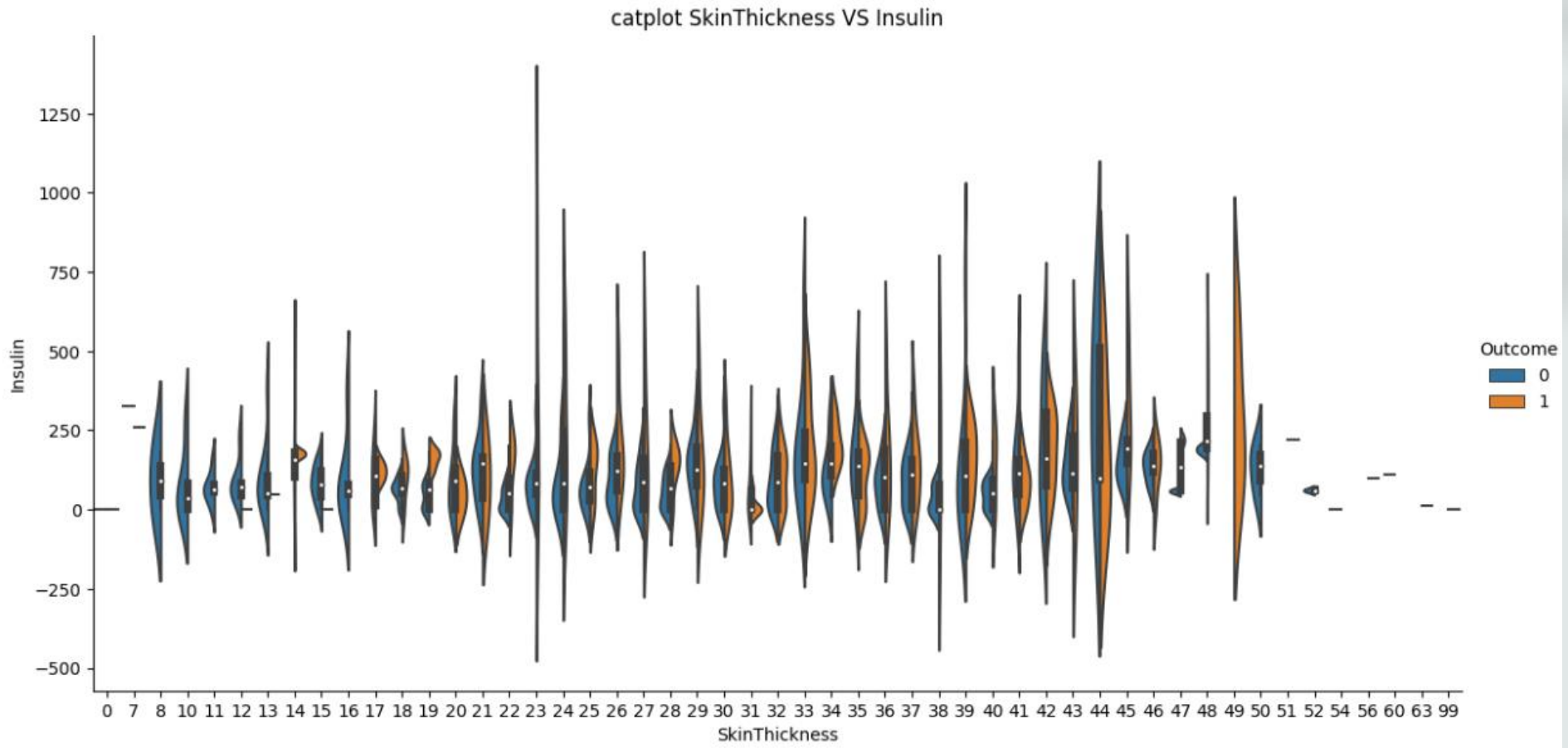
In [119]:
```
sns.histplot(x="BMI",bins=10,data=data,kde= True)
plt.show()
```



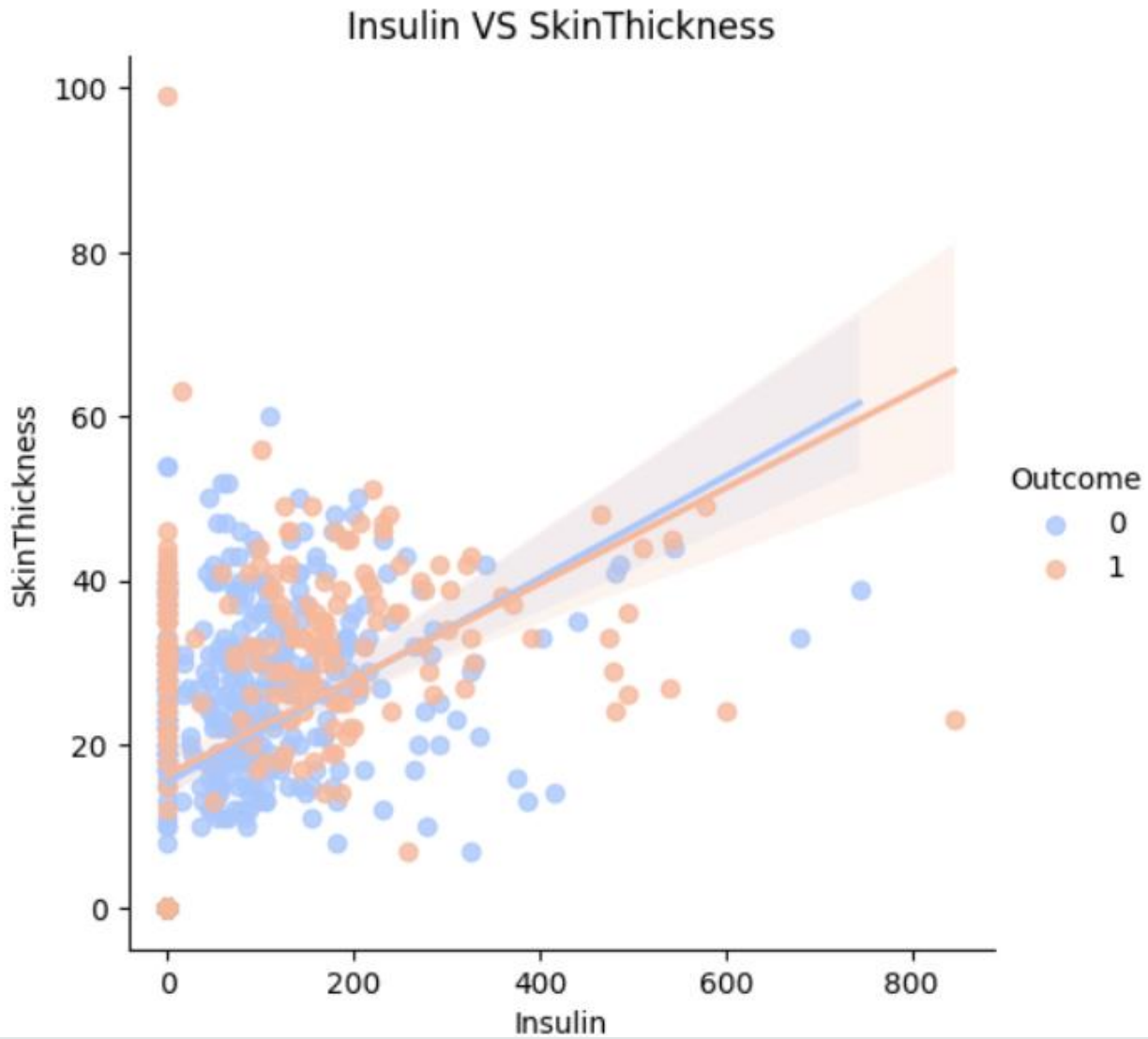
Insights from histplot

- There are more people with a BMI in the normal range (18.5 to 24.9) than there are people with a BMI in the overweight (25 to 29.9) or obese (30 or higher) ranges.

In [192]:
```
# catplot SkinThickness VS Insulin
sns.catplot(x="SkinThickness", y="Insulin", data=data, hue="Outcome", kind="violin", split=True, height=6, aspect=2)
plt.title('catplot SkinThickness VS Insulin')
plt.show()
```



In [200]:
```
sns.lmplot(x="Insulin",y="SkinThickness",data=data,hue= "Outcome",palette="coolwarm")
plt.title("Insulin VS SkinThickness" )
plt.show()
```



Insights from regression and catplot plot

insulin levels increase, skin thickness also increases. The correlation coefficient (R-squared) is 0.75, which means that this trendline explains 75% of the variation in the data.

positive correlation between insulin and skin thickness

The violin plot shows that people with diabetes tend to have higher skin thickness than people without diabetes. The median skin thickness for people with diabetes is 3.1 mm, while the median skin thickness for people without diabetes is 2.5 mm.

Thank you