CS3121 – Data Analysis Lab Exercise

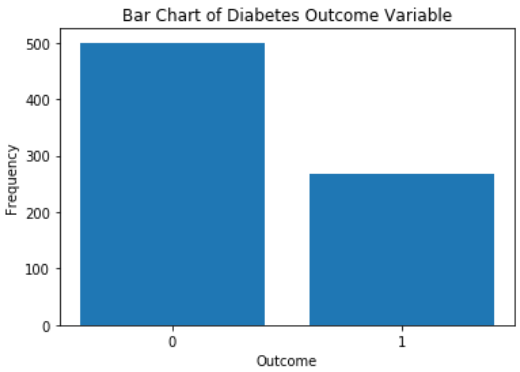
Name: K. O. S. Perera

Index no: 200458M

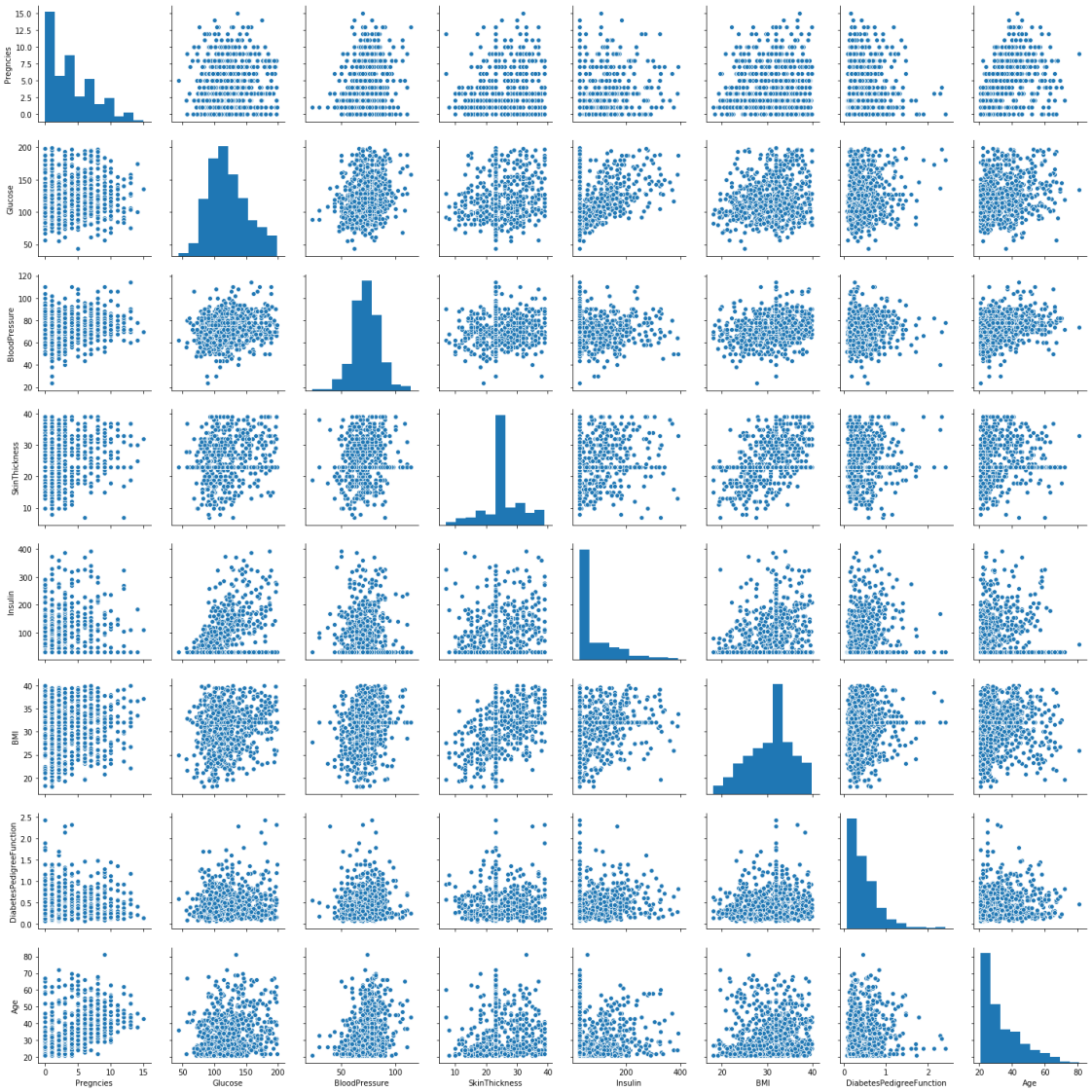
Five number summary

	Insulin	BMI	Age
min	32.000	18.200	21.000
25%	32.000	27.575	24.000
50%	32.000	32.000	29.000
75%	115.250	34.100	41.000
max	392.000	39.900	81.000

Bar chart for Outcome variable



Pairwise relationship scatterplot for predictor variables



Covariance with Outcome

Pregncies	0.343886
Glucose	7.177036
BloodPressure	0.986168
SkinThickness	0.484280
Insulin	4.106116
BMI	0.618582
DiabetesPedigreeFunction	0.027738
Age	1.337258
Outcome	0.227483

Correlation with Outcome

Pregncies	0.216399
Glucose	0.496169
BloodPressure	0.173258
SkinThickness	0.165973
Insulin	0.114810
BMI	0.270324
DiabetesPedigreeFunction	0.175637
Age	0.238505
Outcome	1.000000

Two most influential factors

Glucose, BMI

Glucose and BMI are the features that are highly correlated with the target variable (Outcome) which have the highest correlation values. The Glucose variable quantifies the patient's blood glucose level, a crucial sign of diabetes. An indication of insulin resistance, a prevalent aspect of type 2 diabetes, is a higher blood glucose level. Therefore, diabetes is more prone to occur in persons with high glucose levels. The BMI variable calculates the patient's body mass index, which is a measure of their level of body fat. The patient is overweight or obese, which is a risk factor for type 2 diabetes, according to a higher BMI. Insulin resistance brought on by too much body fat can raise the chance of developing diabetes. Therefore, based on the correlation coefficient values and the known factors, Glucose and BMI are the two most influential factors that increase the risk of diabetics.