

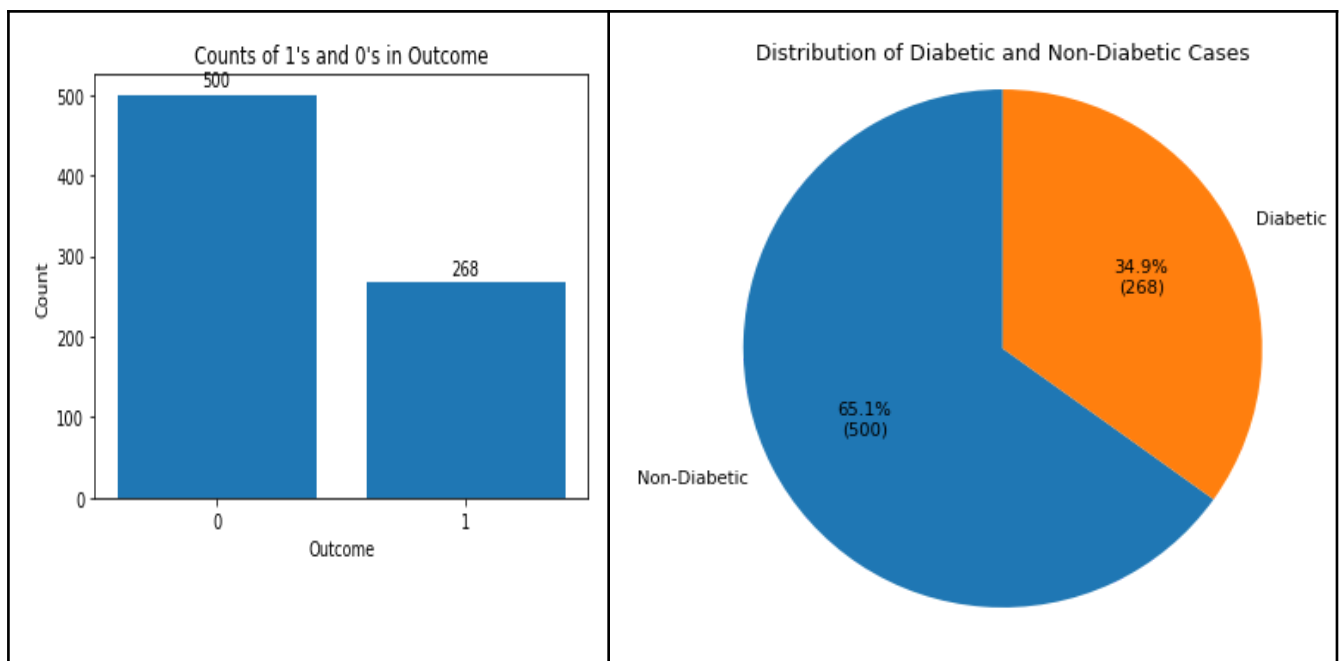
EDA ON DIABETES DATASET:- (9 * 768)

1. Data distribution (Histogram & Pie chart)
2. Correlation(Heatmap and correlation matrix)
3. Significant data (Normal distribution curve)
4. Outliers identification (u-3 SIGMA ____ u +3 SIGMA)

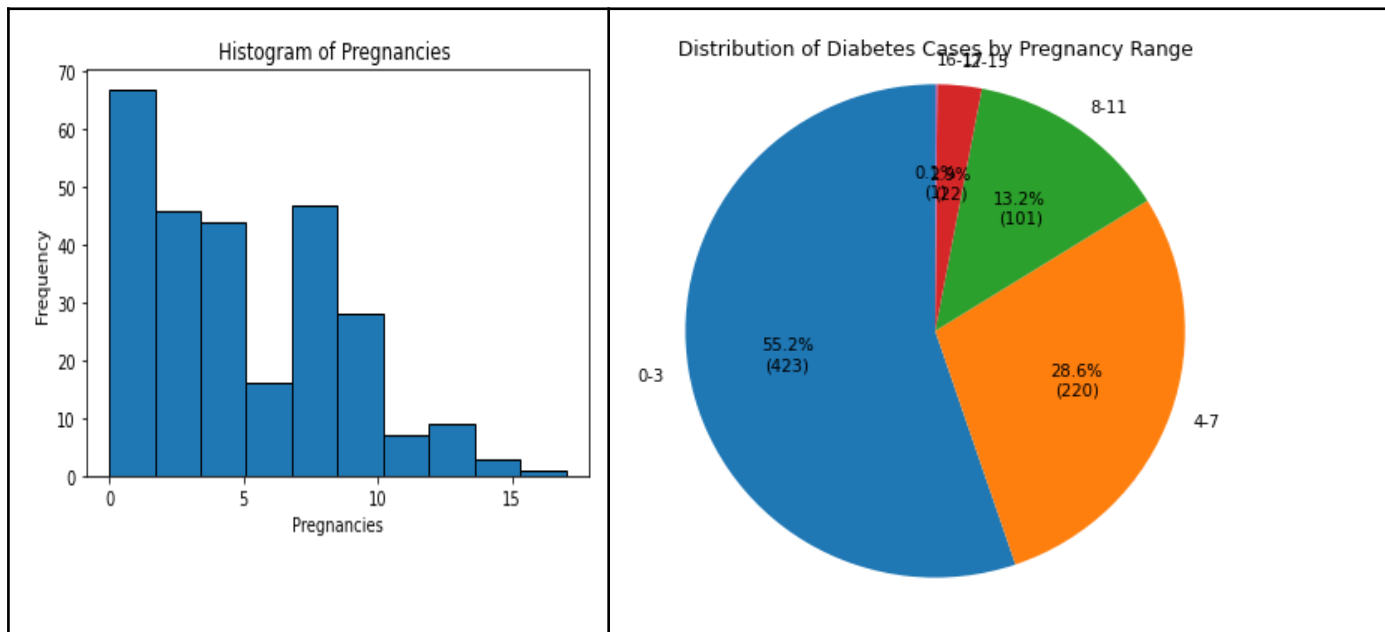
Pregnancies	int64
Glucose	int64
BloodPressure	int64
SkinThickness	int64
Insulin	int64
BMI	float64
DiabetesPedigreeFunction	float64
Age	int64
Outcome	int64

DATA DISTRIBUTION USING HISTOGRAM AND PIE CHART

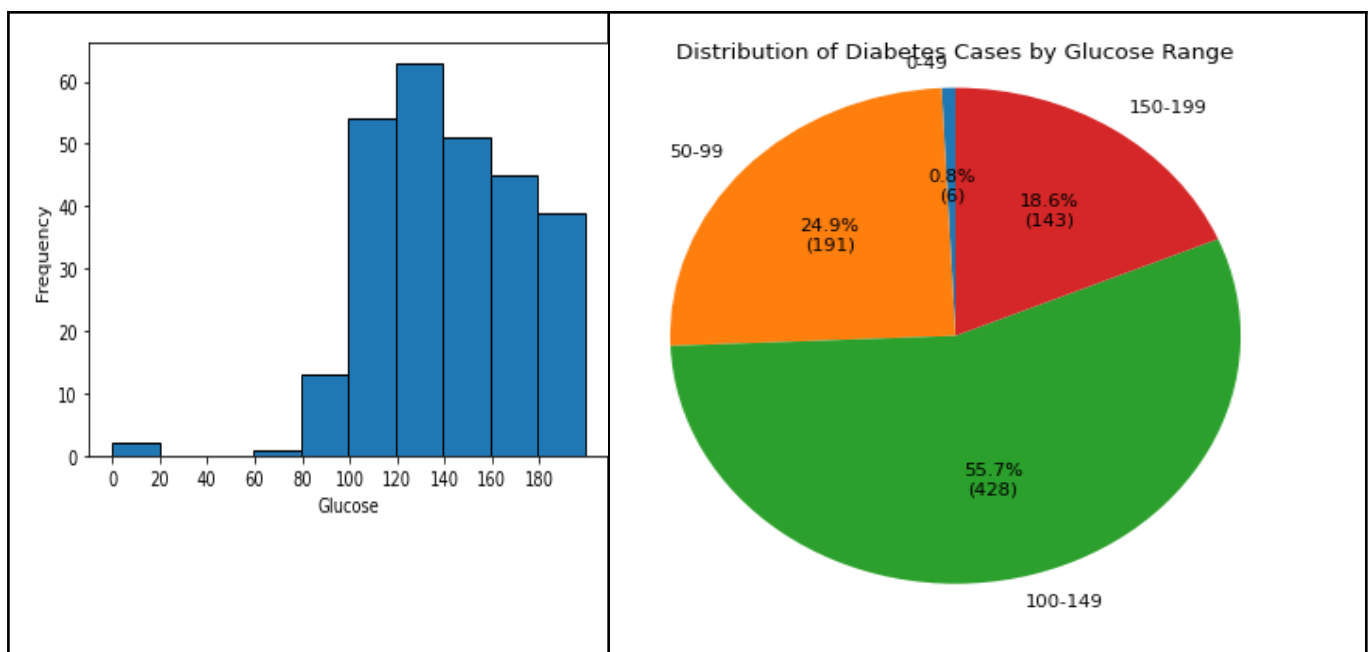
Outcome:



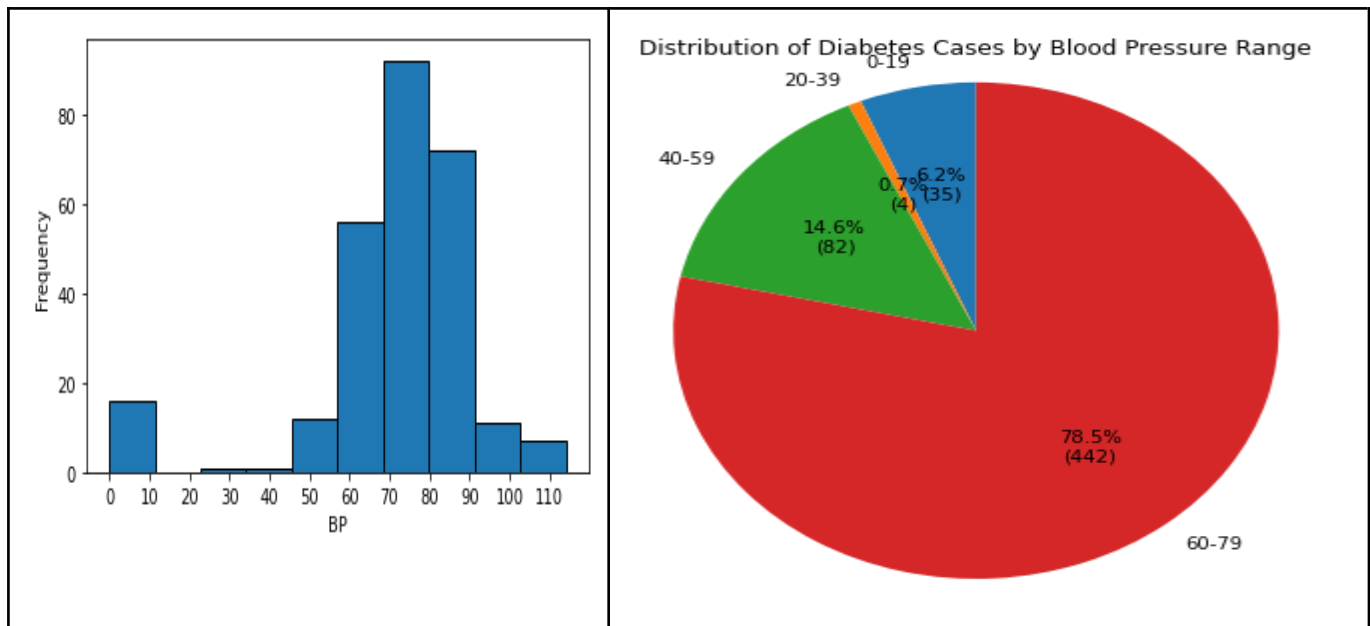
1)Pregnancies:-



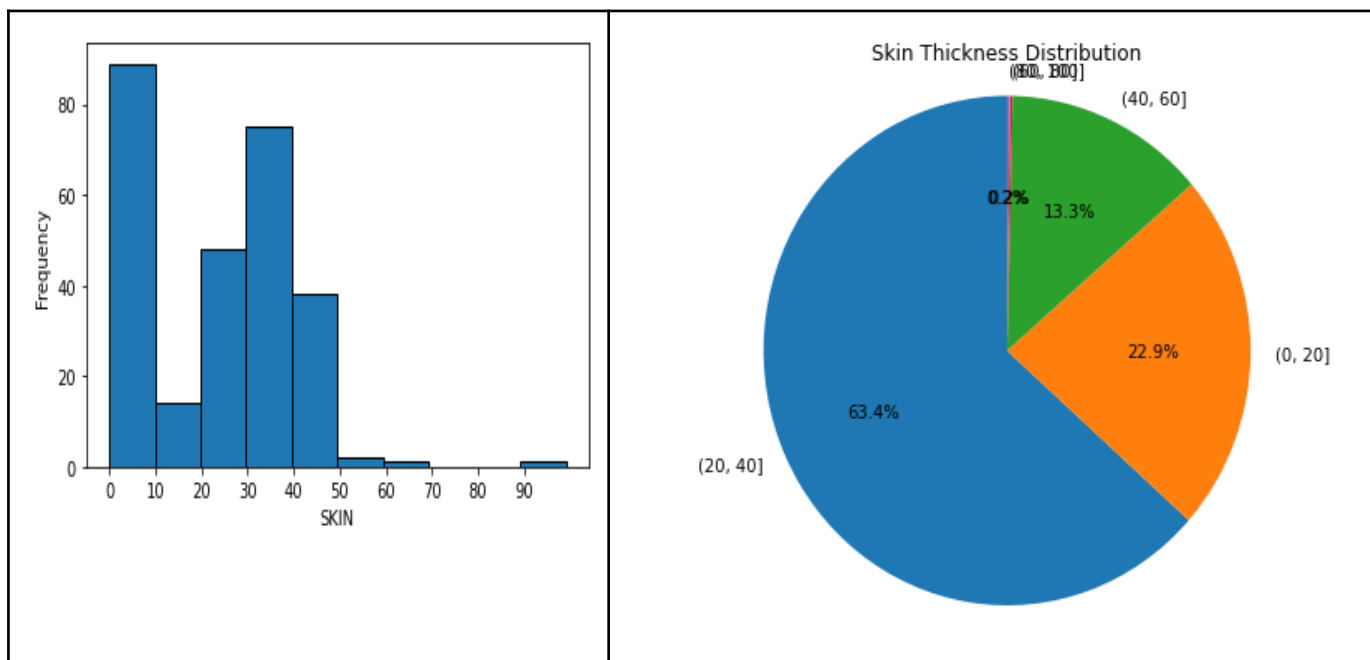
2)Glucose:-



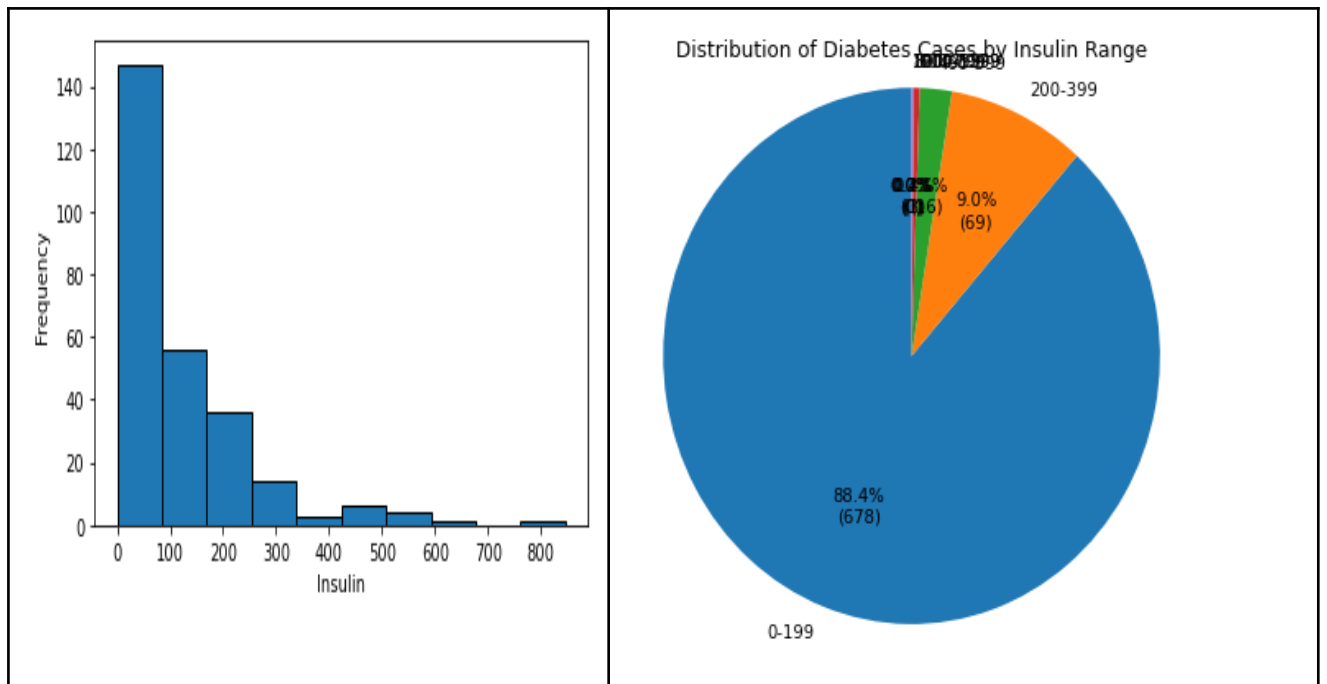
3)Blood pressure:-



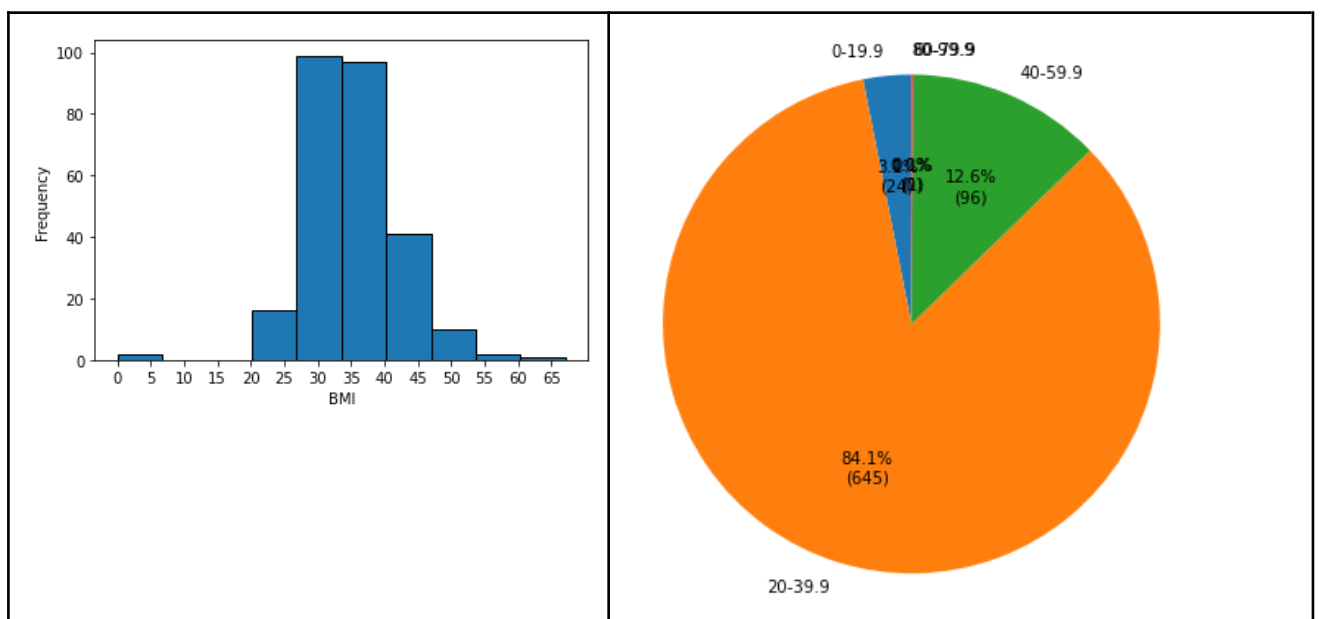
4)Skin Thickness:-



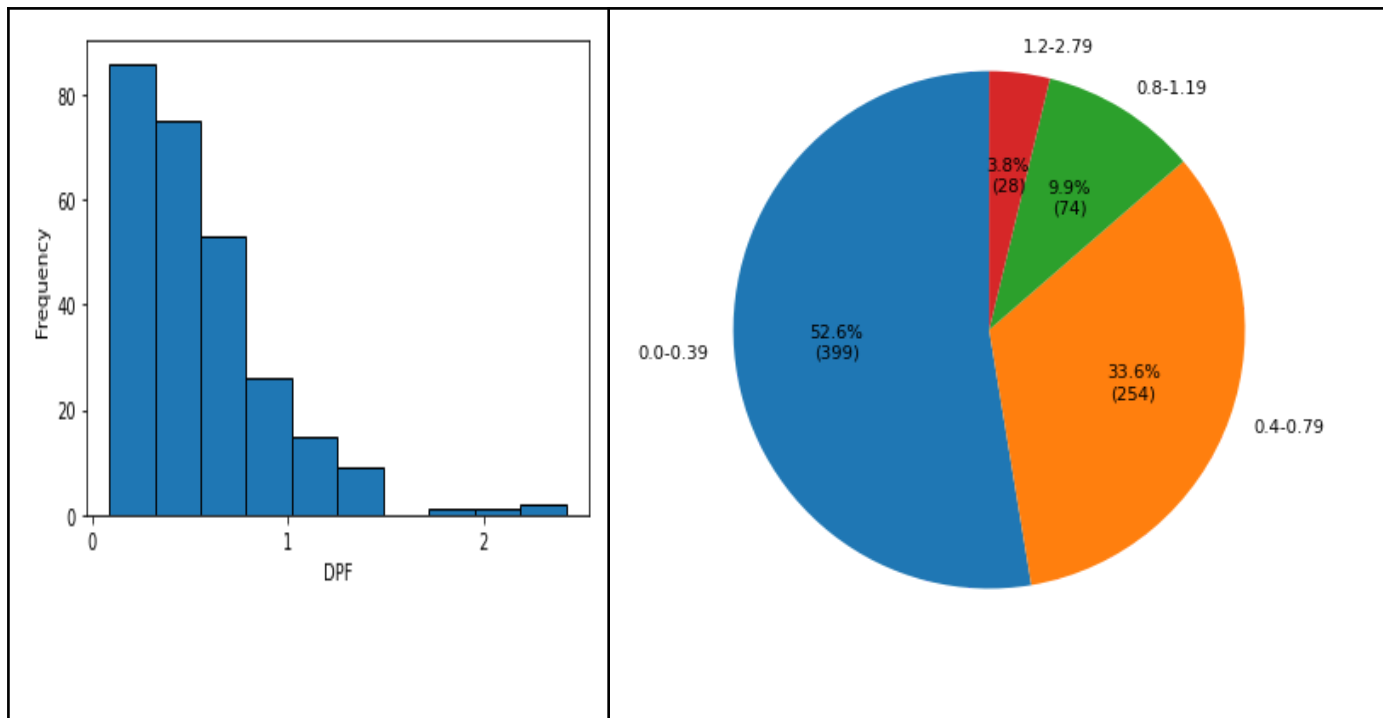
5)Insulin:-



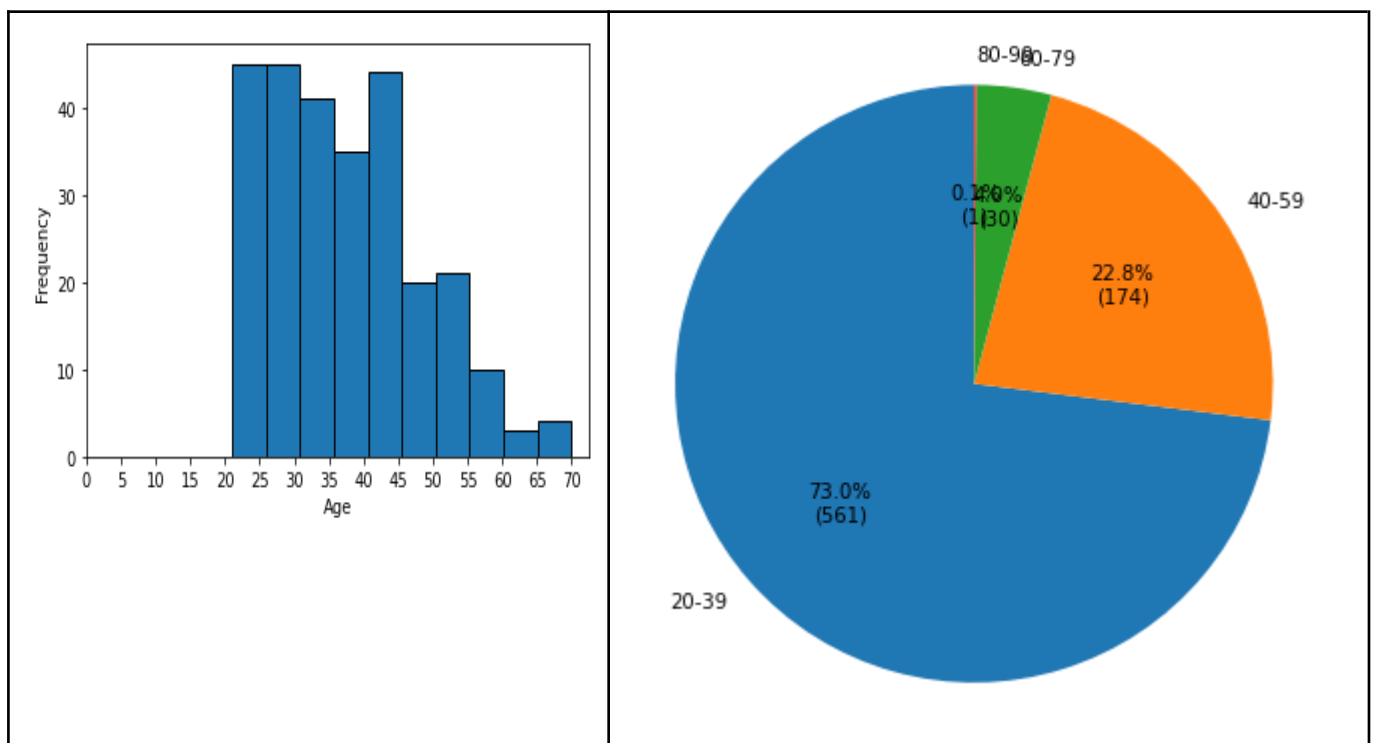
6)BMI



7)DiabetesPedigreeFunction



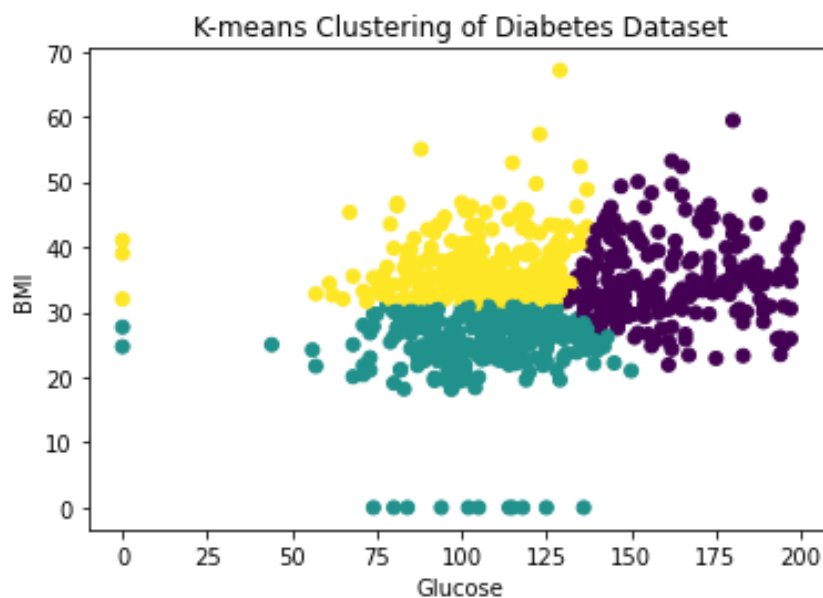
8) Age



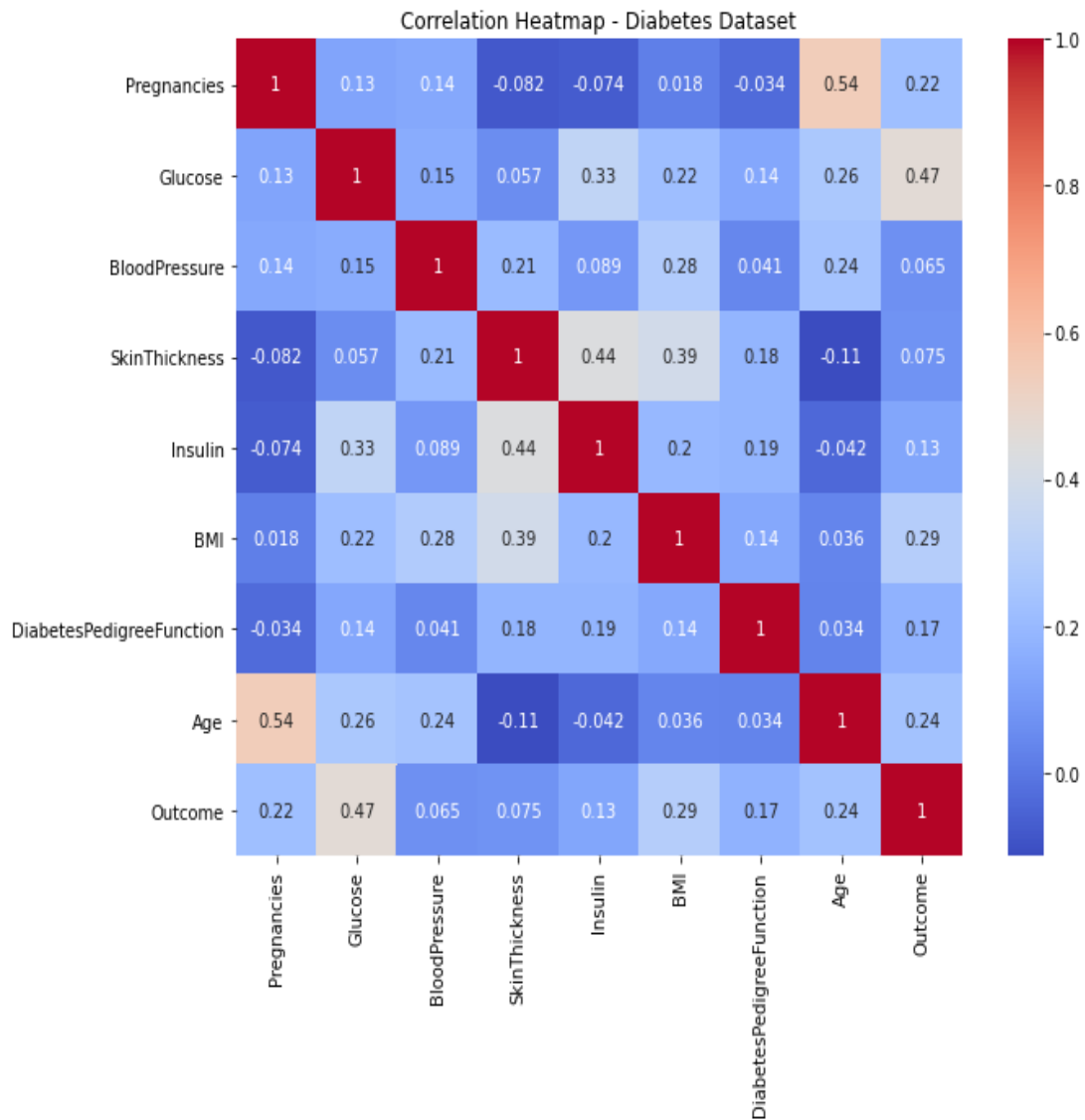
2)CORRELATION:

CORRELATION MATRIX: WITH OUTCOME COLUMN

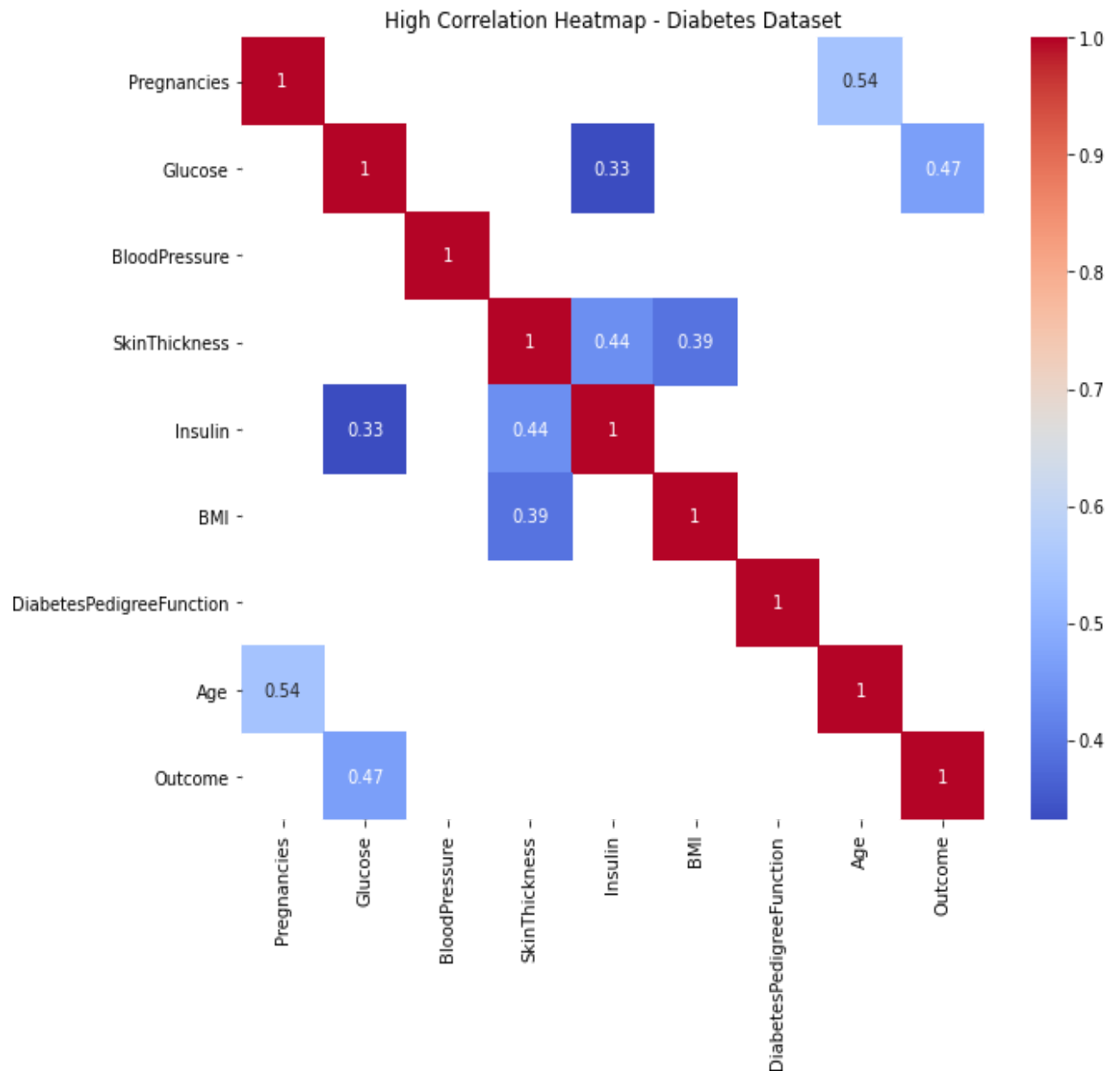
Glucose	0.466581
BMI	0.292695
Age	0.238356
Pregnancies	0.221898
DiabetesPedigreeFunction	0.173844
Insulin	0.130548
SkinThickness	0.074752
BloodPressure	0.065068



HEAT MAP (-0.7 TO 0.7 -> HIGHEST), (-0.3 TO 0.3 -> AVERAGE)



HIGHEST CORRELATION GRAPH(DISPLAYING ONLY -0.3 TO 0.3)



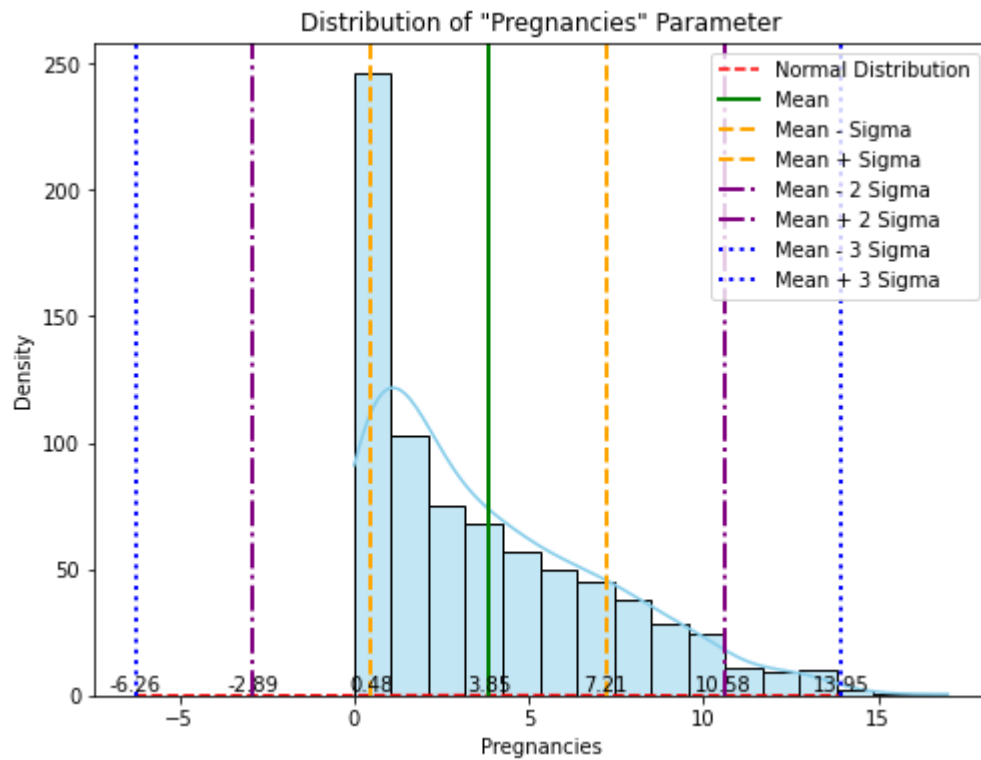
Age & Pregnancies :- 0.54

Age and Glucose:- 0.47

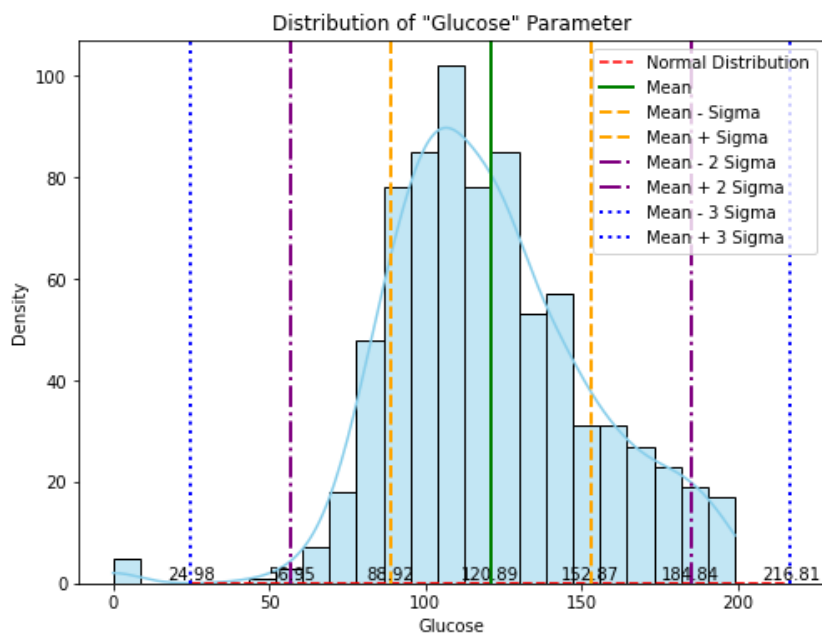
Insulin & skin Thickness 0.44

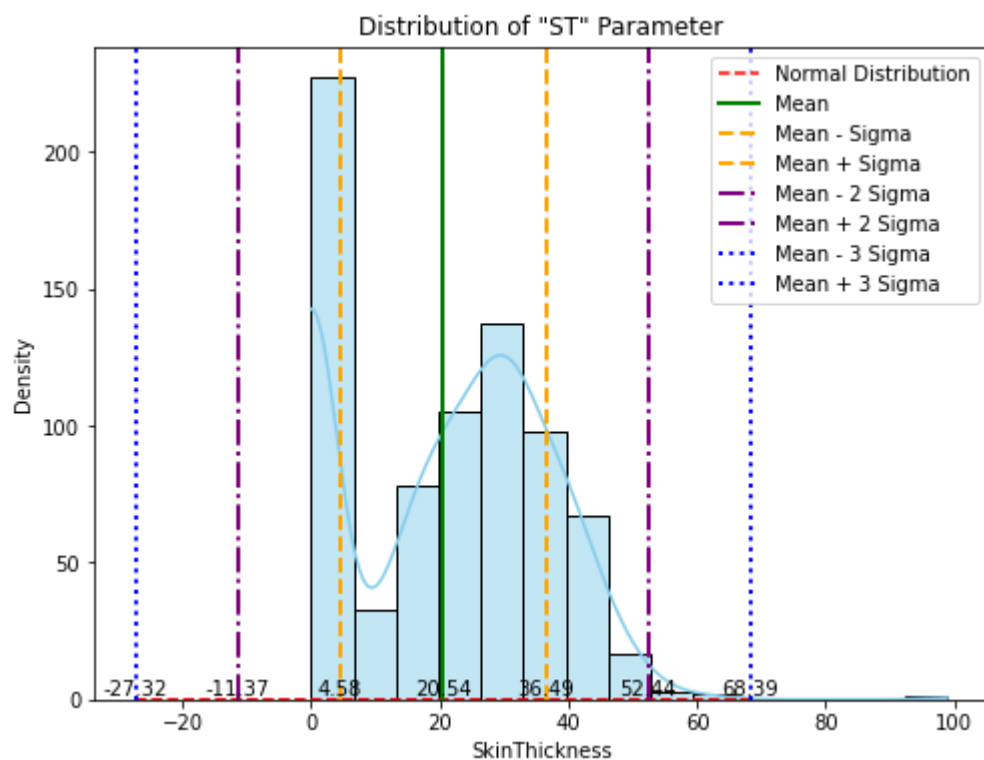
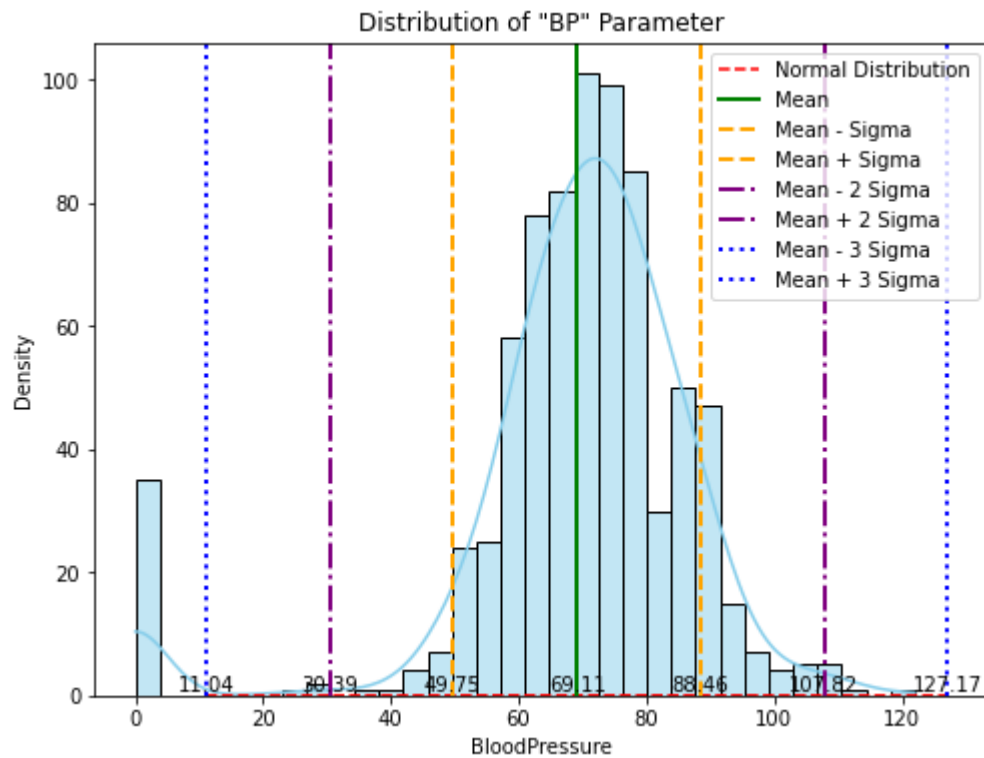
Bmi & skin Thickness :- 0.39

3)NORMAL DISTRIBUTION CURVE TO IDENTIFY OUTLIERS

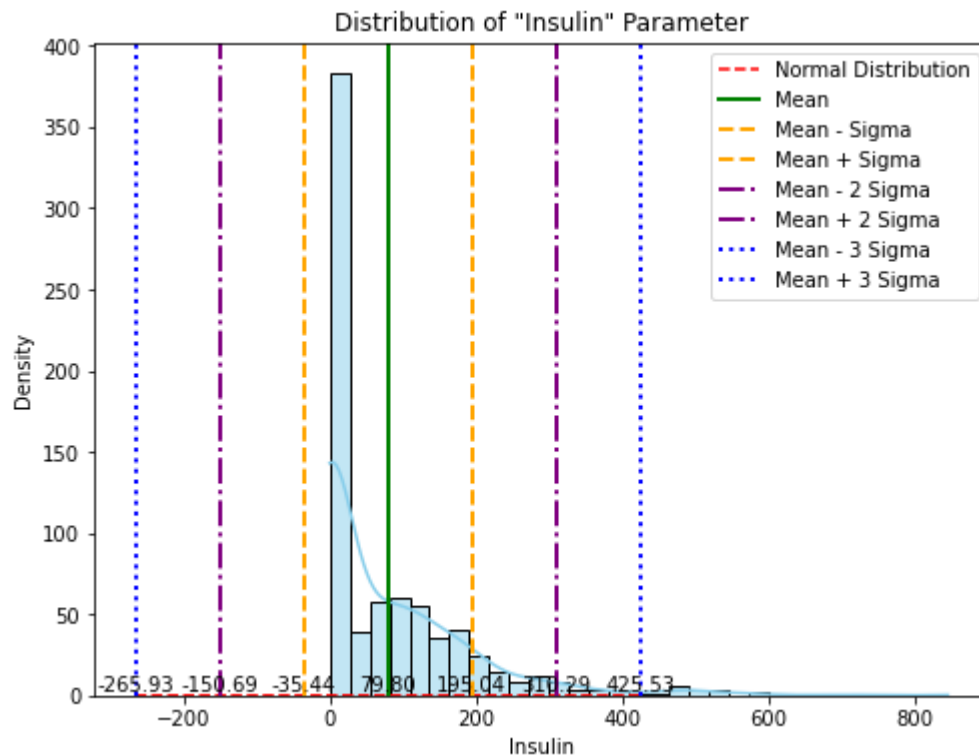


Outlier found for Pregnancies (Row 89): 15
Outlier found for Pregnancies (Row 160): 17
Outlier found for Pregnancies (Row 299): 14
Outlier found for Pregnancies (Row 456): 14

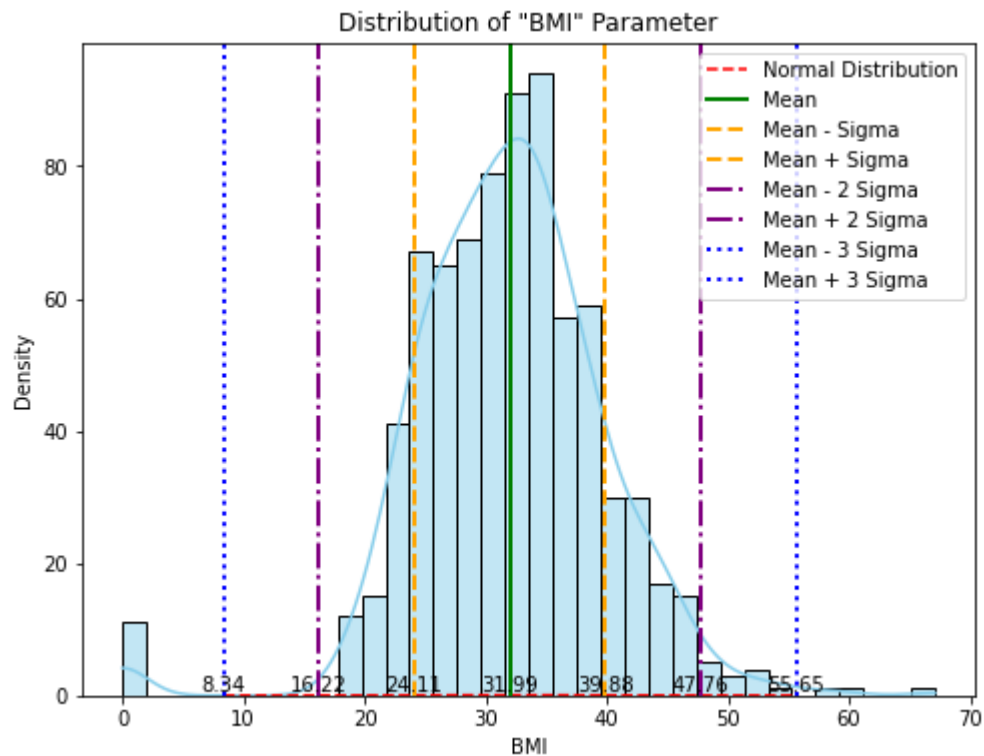




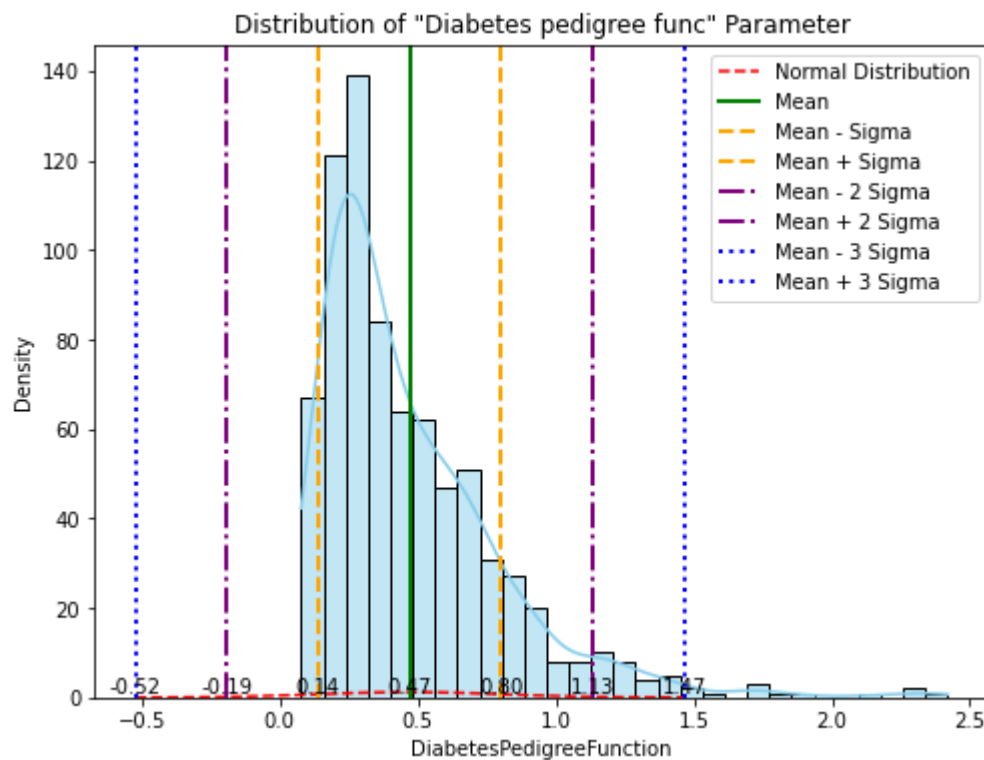
Outlier found for SkinThickness (Row 580): 99



Outlier found for Insulin (Row 9): 543
Outlier found for Insulin (Row 14): 846
Outlier found for Insulin (Row 112): 495
Outlier found for Insulin (Row 154): 485
Outlier found for Insulin (Row 187): 495
Outlier found for Insulin (Row 221): 478
Outlier found for Insulin (Row 229): 744
Outlier found for Insulin (Row 248): 680
Outlier found for Insulin (Row 287): 545
Outlier found for Insulin (Row 371): 465
Outlier found for Insulin (Row 410): 579
Outlier found for Insulin (Row 416): 474
Outlier found for Insulin (Row 487): 480
Outlier found for Insulin (Row 585): 600
Outlier found for Insulin (Row 646): 440
Outlier found for Insulin (Row 656): 540
Outlier found for Insulin (Row 696): 480
Outlier found for Insulin (Row 754): 510

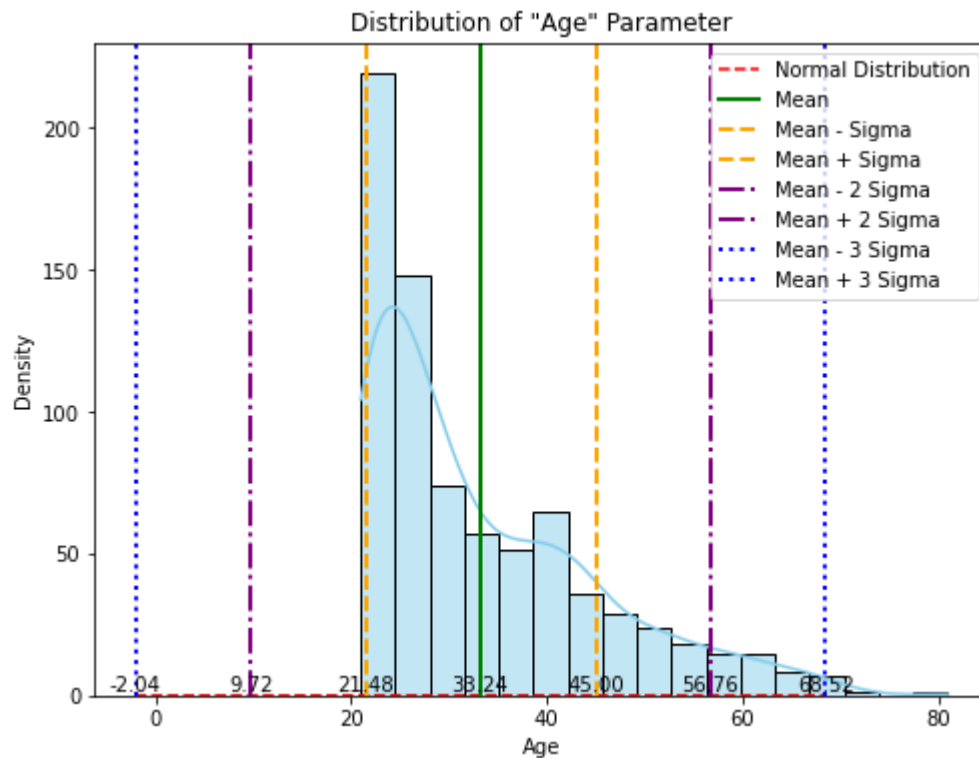


Outlier found for BMI (Row 178): 67.1
Outlier found for BMI (Row 446): 59.4
Outlier found for BMI (Row 674): 57.3



Outlier found for DiabetesPedigreeFunction (Row 5): 2.288
Outlier found for DiabetesPedigreeFunction (Row 46): 1.893

Outlier found for DiabetesPedigreeFunction (Row 59): 1.781
Outlier found for DiabetesPedigreeFunction (Row 229): 2.329
Outlier found for DiabetesPedigreeFunction (Row 331): 1.476
Outlier found for DiabetesPedigreeFunction (Row 371): 2.137
Outlier found for DiabetesPedigreeFunction (Row 372): 1.731
Outlier found for DiabetesPedigreeFunction (Row 396): 1.6
Outlier found for DiabetesPedigreeFunction (Row 446): 2.42
Outlier found for DiabetesPedigreeFunction (Row 594): 1.699
Outlier found for DiabetesPedigreeFunction (Row 622): 1.698



Outlier found for Age (Row 124): 69
Outlier found for Age (Row 454): 72
Outlier found for Age (Row 460): 81
Outlier found for Age (Row 667): 70
Outlier found for Age (Row 685): 69