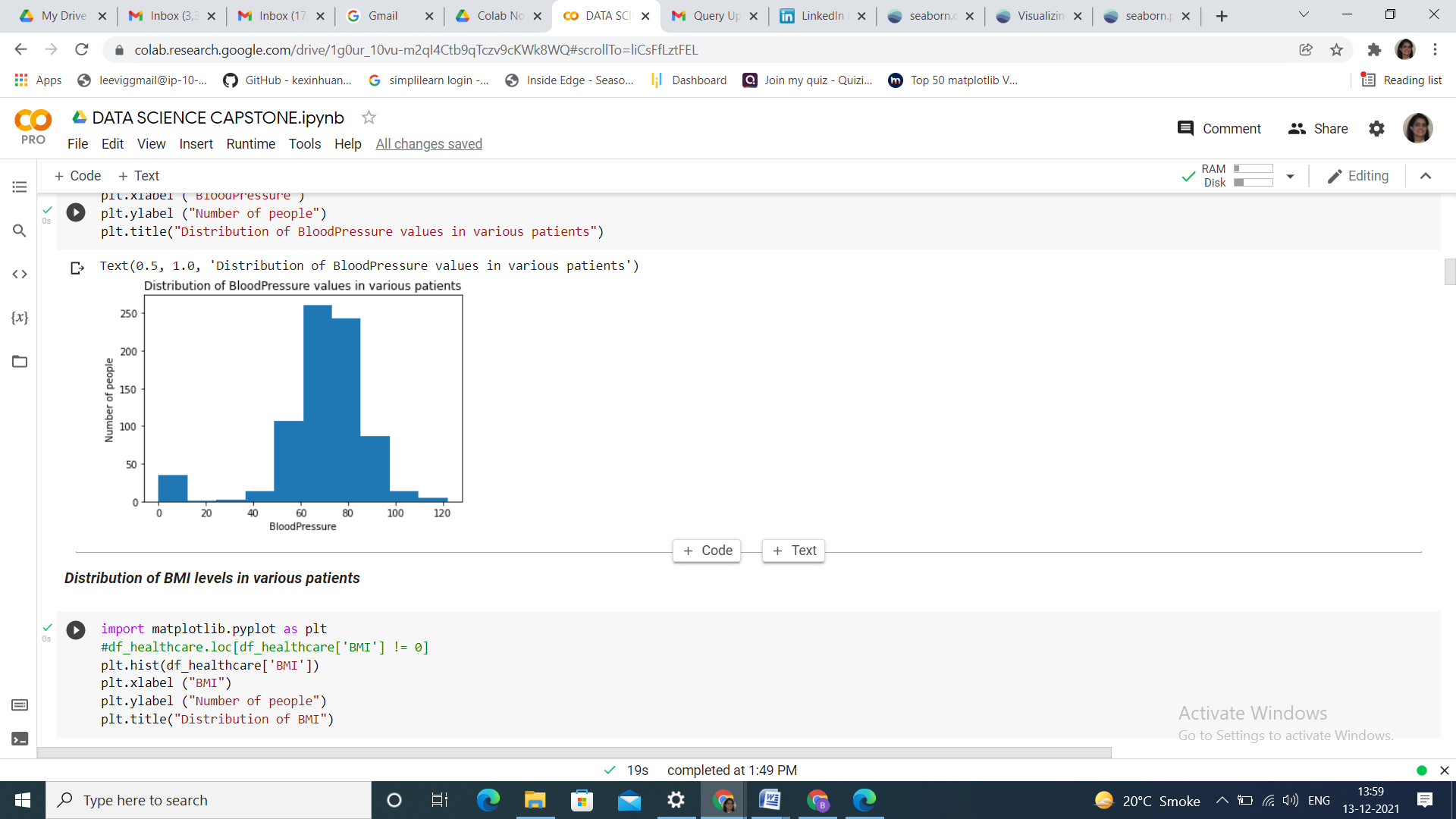
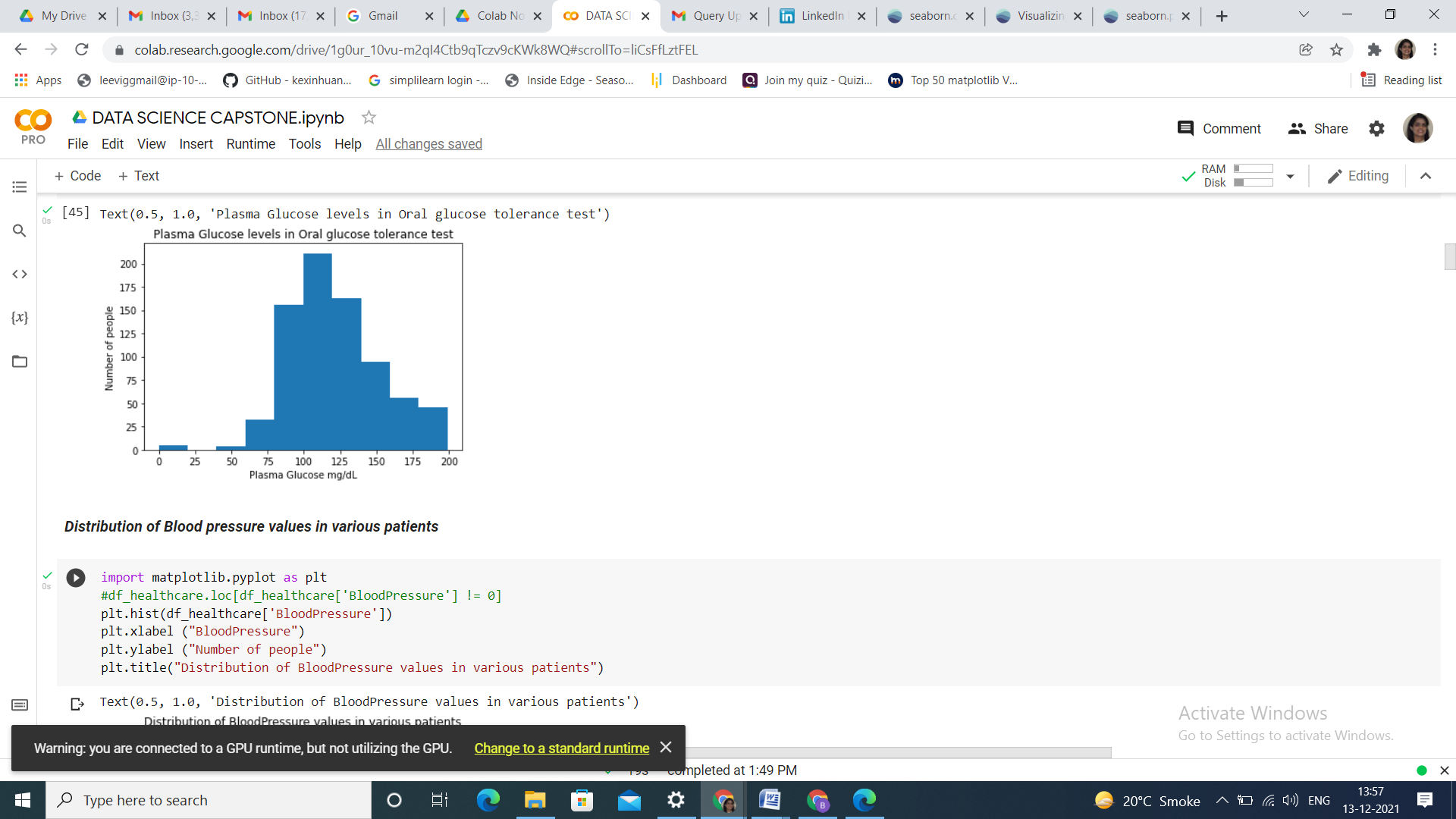
**Write up for the Healthcare dataset:**

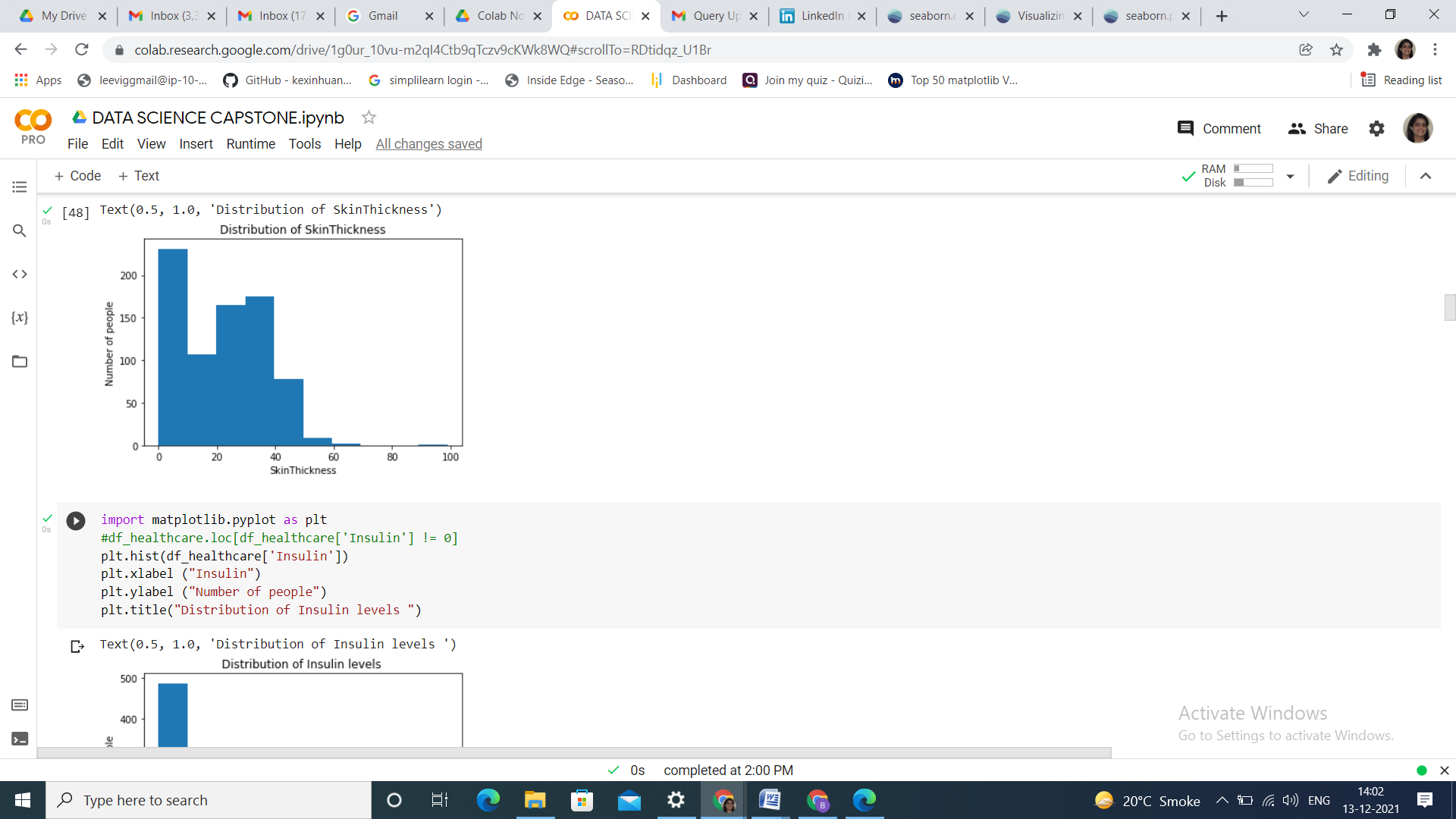
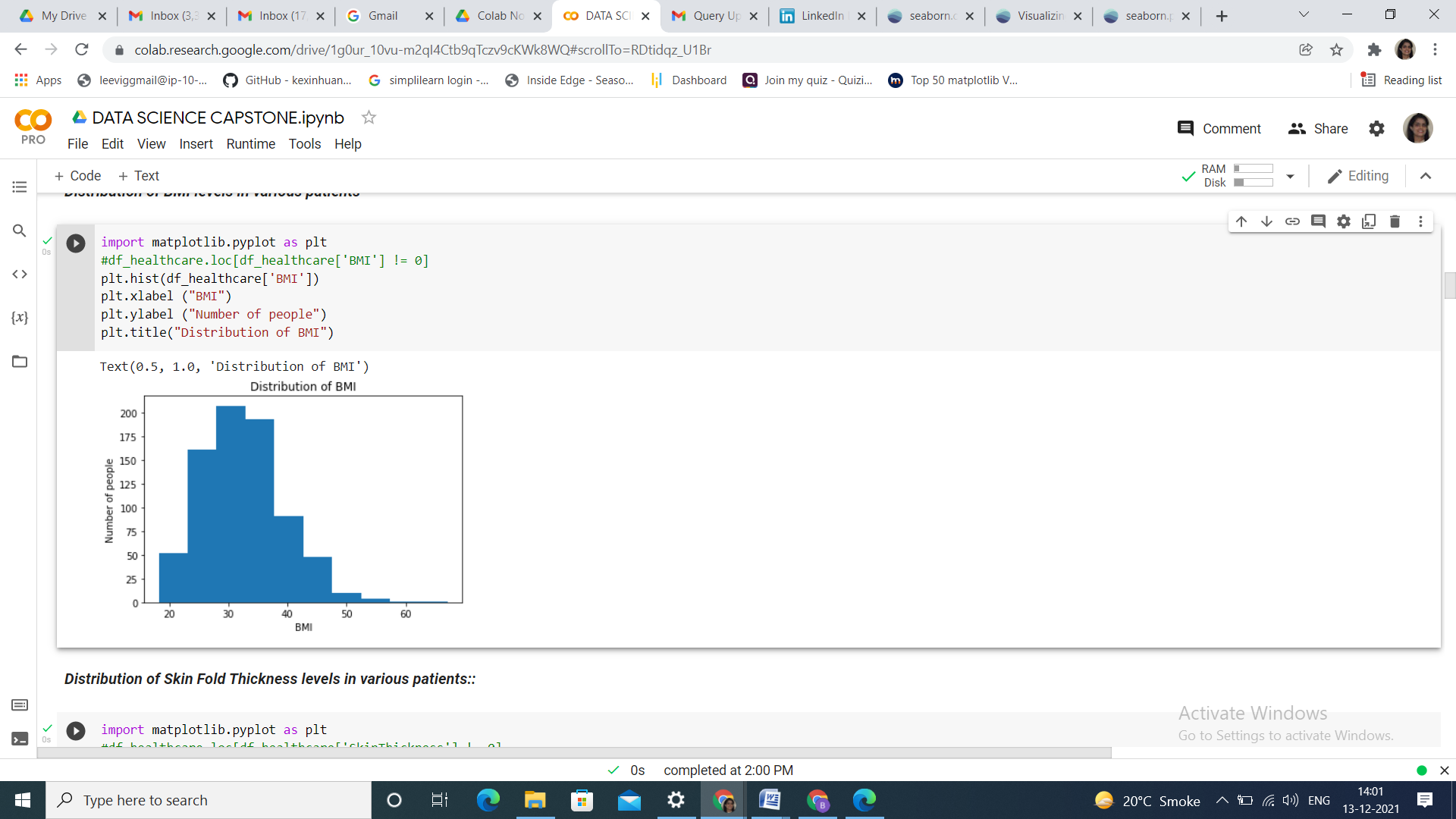
The dataset provided had zero values for age and other parameters of insulin, glucose,skin thickness,blood pressure and BMI.

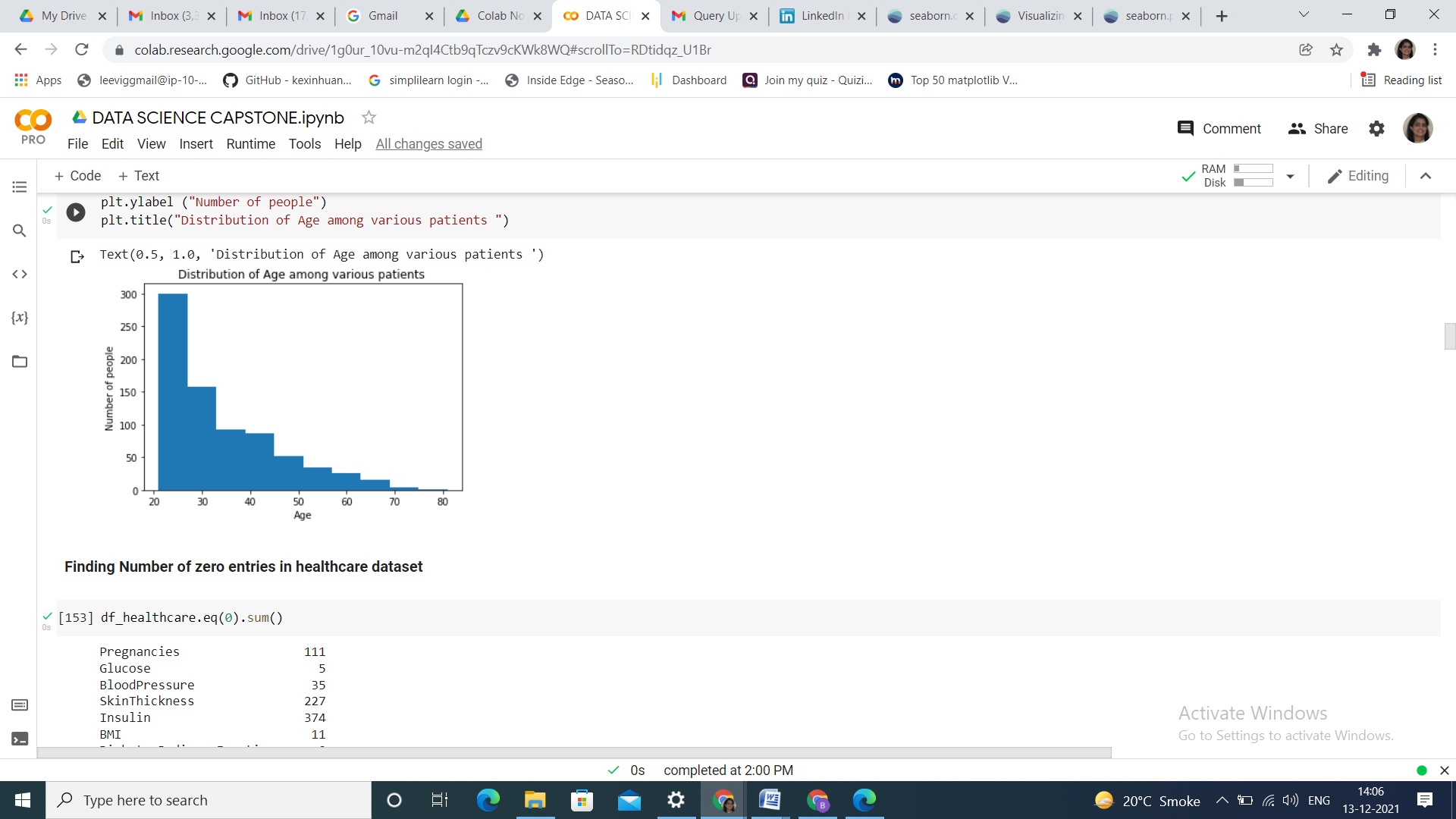
The zero values were adjusted with mean or median values.

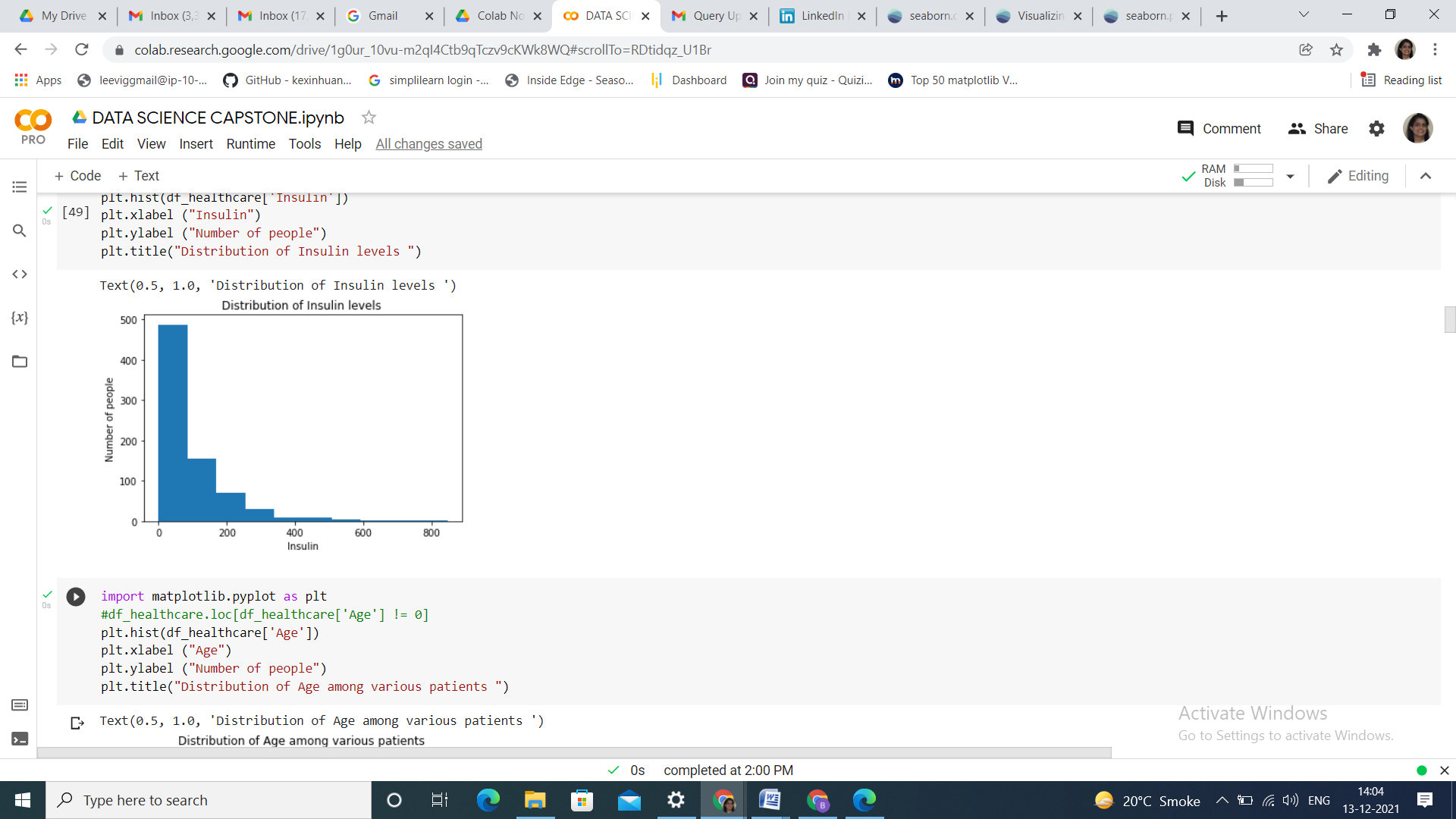
**SCREENSHOTS**

**Initial Exploration of Healthcare Dataset**

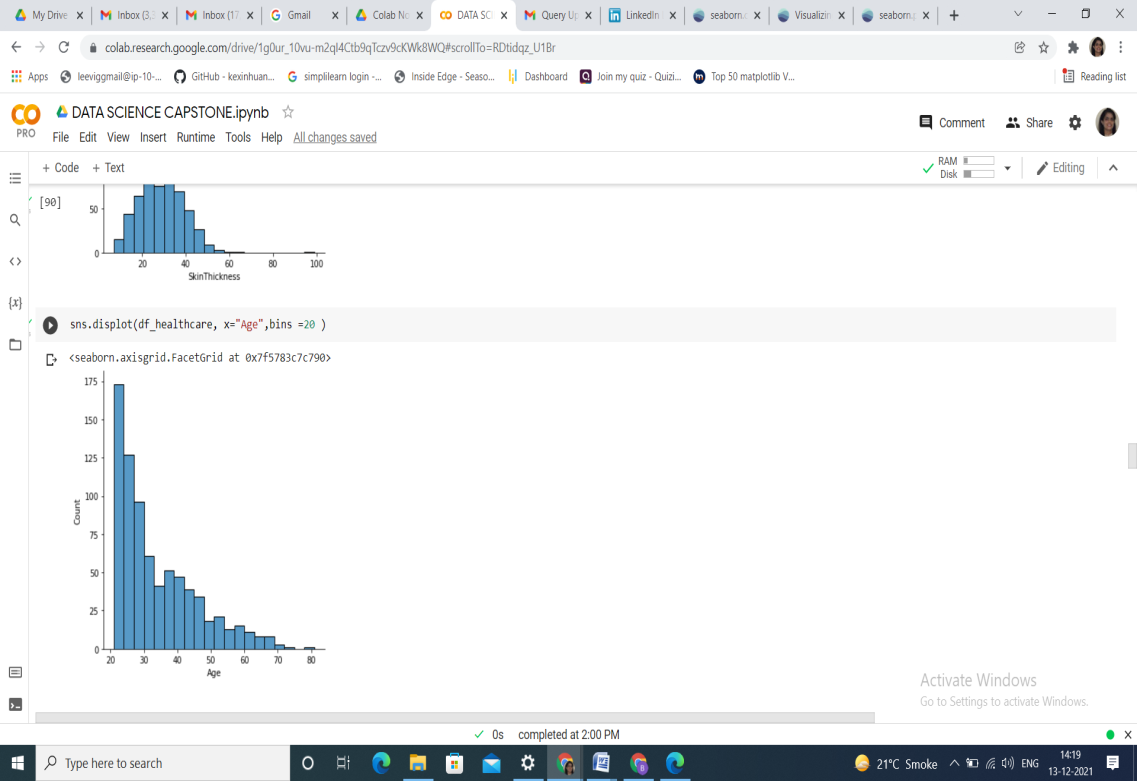
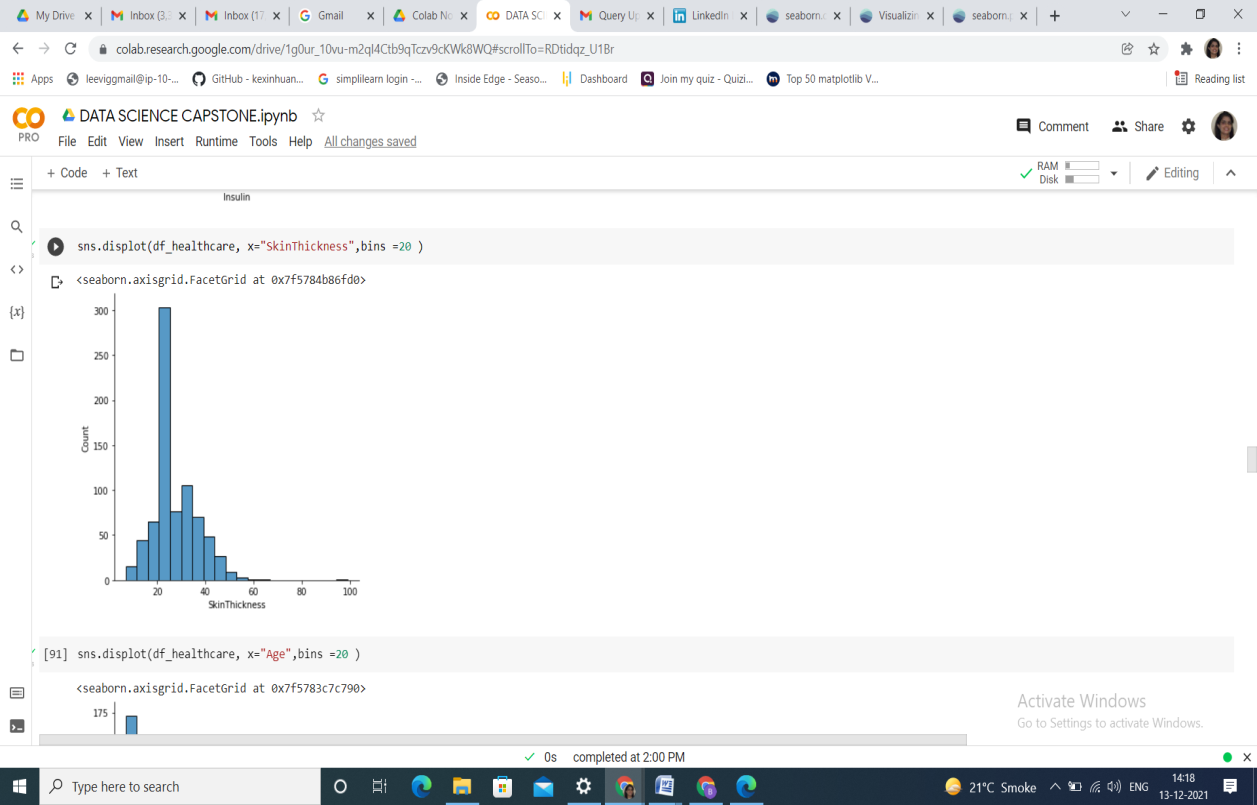
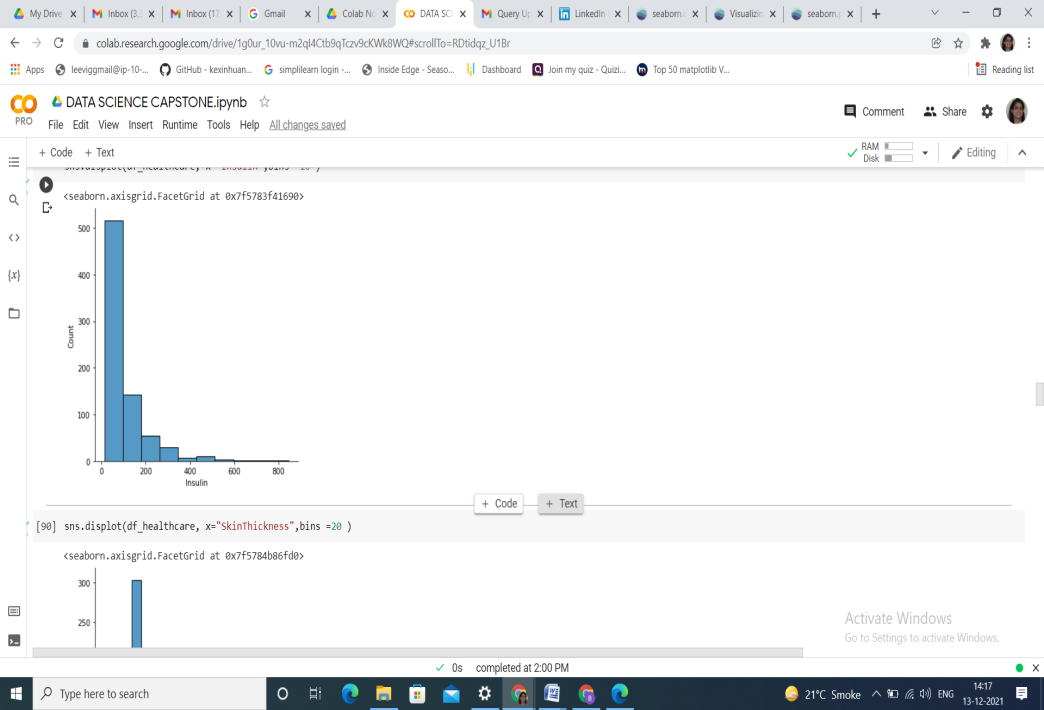
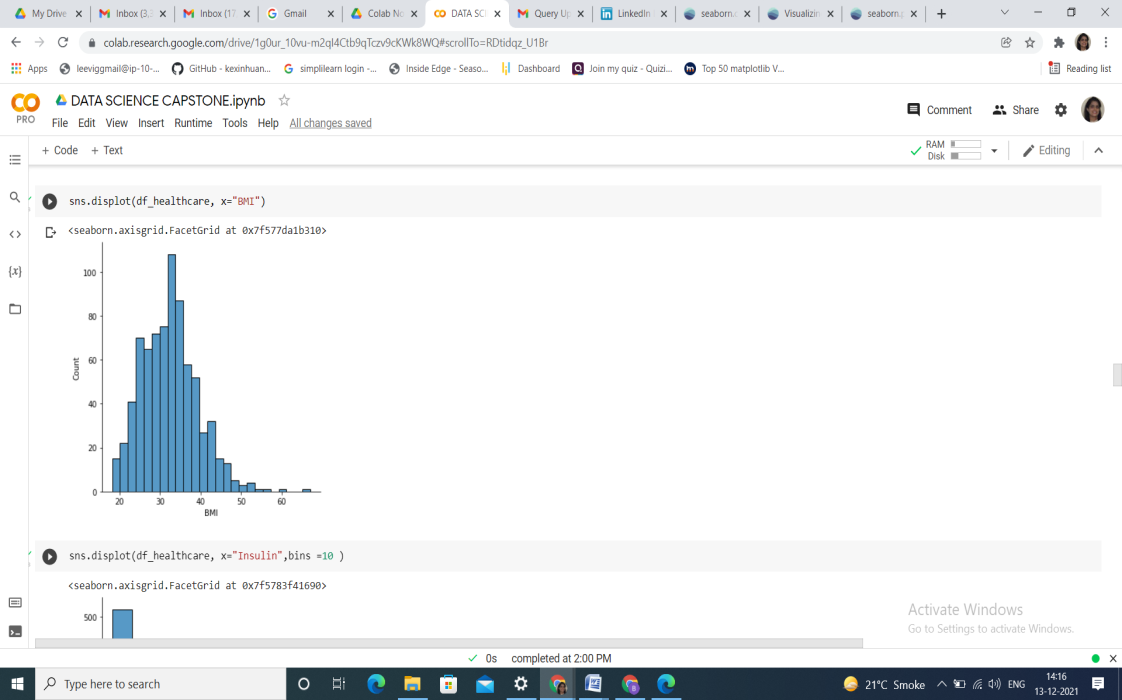
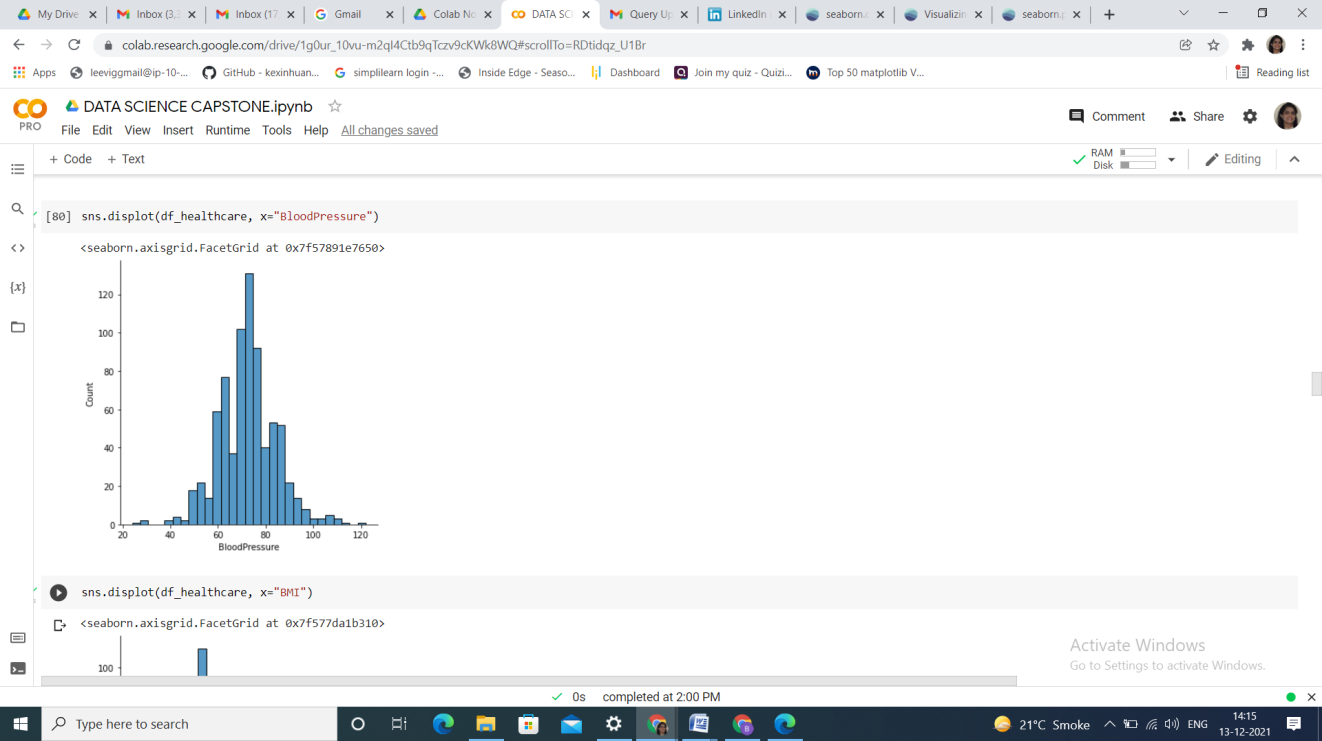
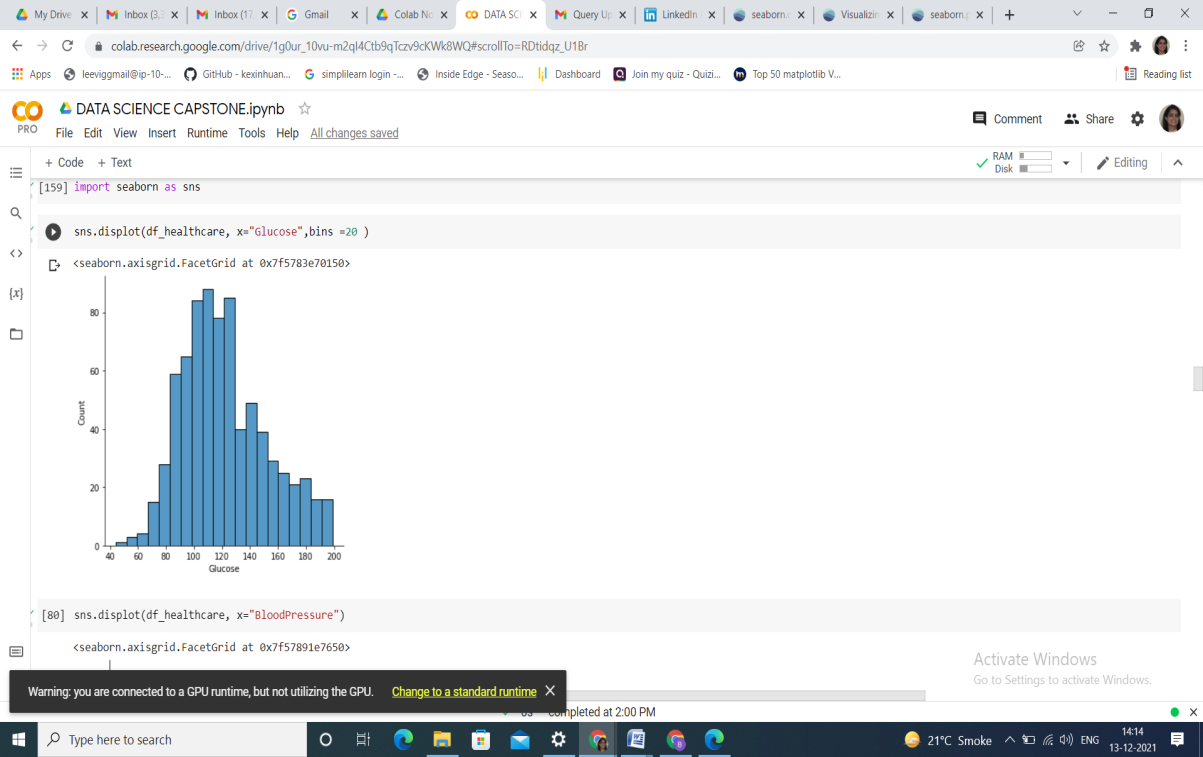




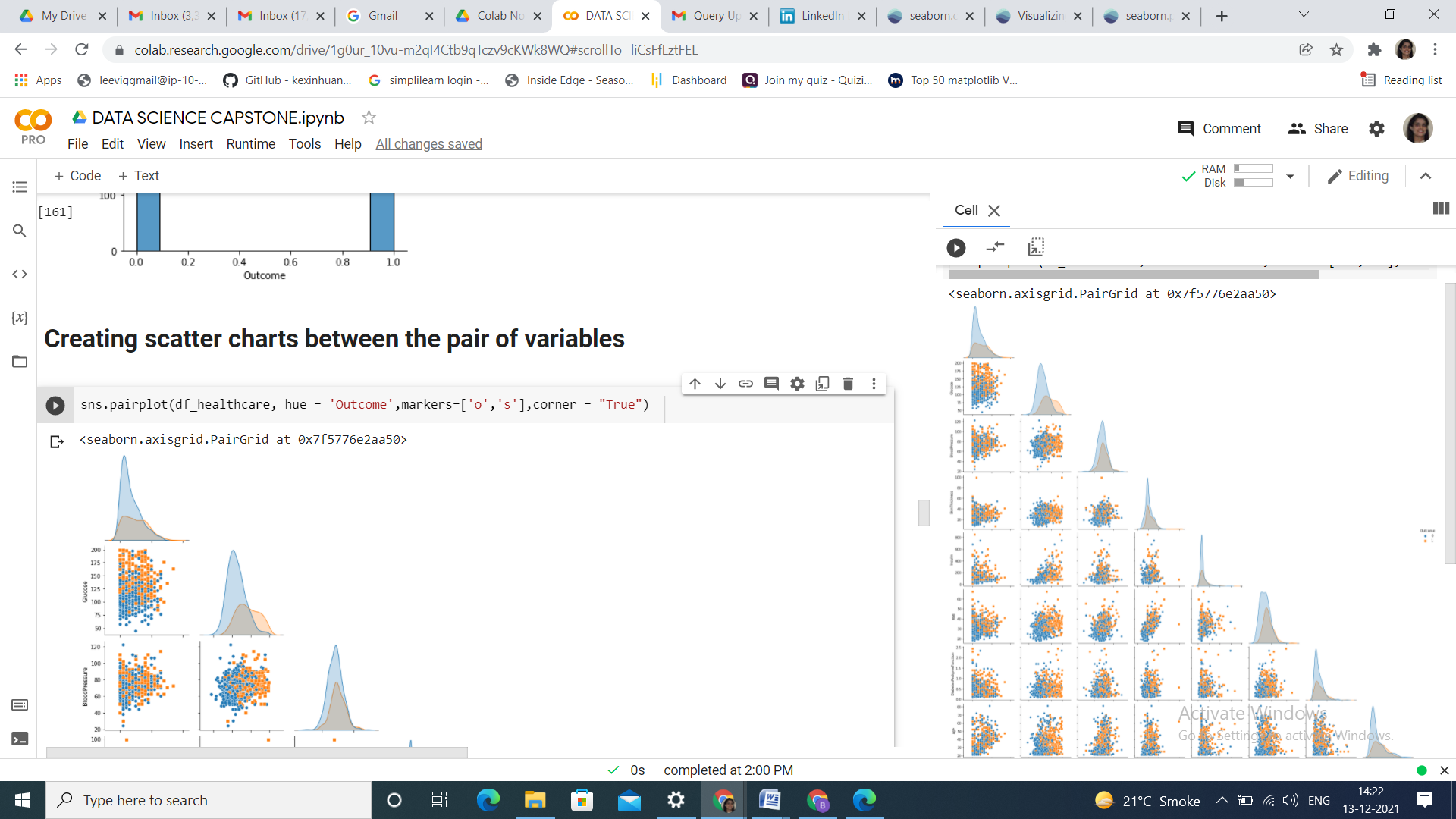




**Frequency Plot for the different variables**

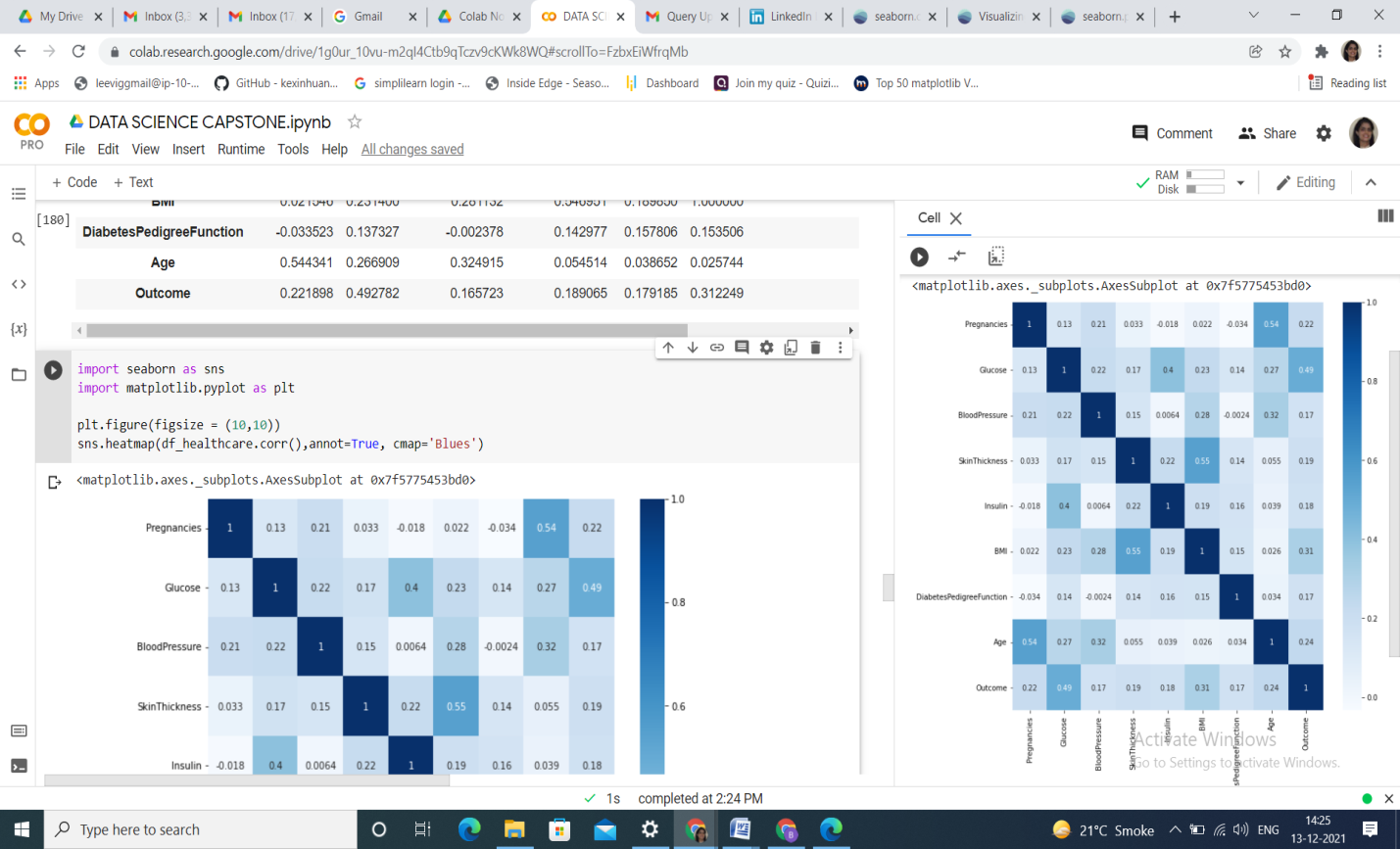


**Scatter plots:**



Correlation between the different variables were obtained and heat map was constructed

**Heat map: correlation analysis**



Significant correlation values is seen between the following features(in decreasing order of correlation)

1. Skin Thickness and BMI : 0.55
2. Glucose & Insulin : 0.4
3. Age & Outcome : 0.34
4. Glucose & Blood Pressure : 0.22
5. Glucose & BMI : 0.21
6. Insulin & BMI : 0.19
7. Glucose & skin thickness : 0.17

KNN modeling was done with the following results:

ROC\_AUC score from KNN model is 0.7181818181818181

Confusion Matrix:

[[81 18]

[21 34]]

Logistic regression was done with the following results:

ROC\_AUC score from logistic regression model is 0.7464646464646464

[[83 16]

[19 36]]

Decision Tree modeling was also done with the following results:

ROC\_AUC score from decision tree model is 0.7555555555555555;

Confusion Matrix:

[[83 16]

[18 37]]

Naive Bayes modeling was also done with the following results:

ROC\_AUC score for naive bayes is 0.7393939393939394

[[78 21]

[17 38]]

Support Vector Machines modeling was also done with the following results:

ROC\_AUC score from svm model is 0.7212121212121211

[[87 12]

[24 31]]

**Pairwise scatter plots:**

