**CURNEU MEDTECH SOLUTIONS PVT. LTD.**

**Problem Statement 2:**

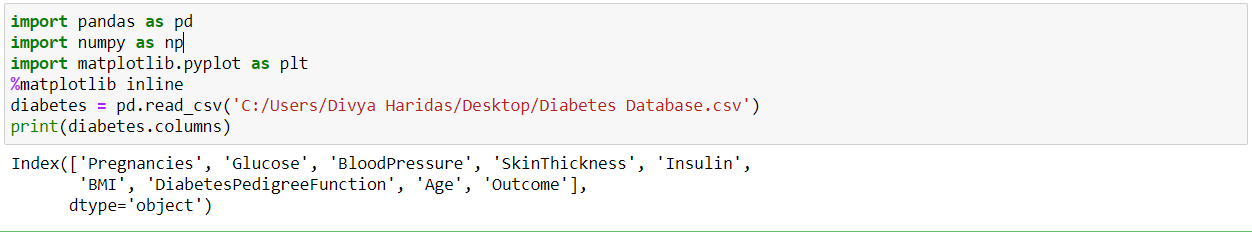
Predict which patient has diabetes from Diabetes Database.csv and try to understand the dataset attributes and try to figure out type ML model suits and build from scratch.

DATASET DESCRIPTION:

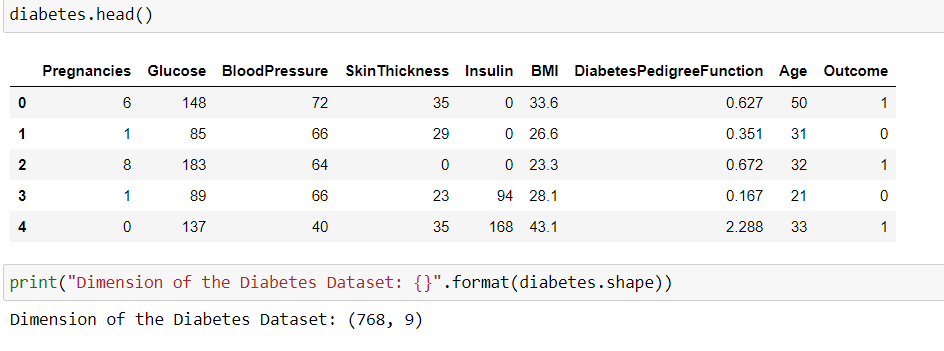
The dataset contains various independent attributes such as pregnancies, glucose, blood pressure, skin thickness, insulin, BMI, DiabetesPedigreeFunction, age which contributes to the result whether a person has diabetes or not given as Outcome. 0 indicates that the person has no diabetes and 1 indicates that the person has diabetes from which we can tell that the dataset is a supervised binary data.

EXPLORATORY DATA ANALYSIS:

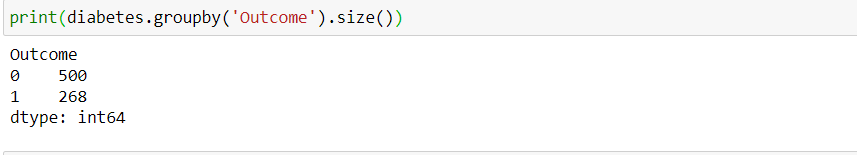
Before moving into the analysis of the data some EDA process so that we come to know some more information about the data and if the dataset contains any missing values.



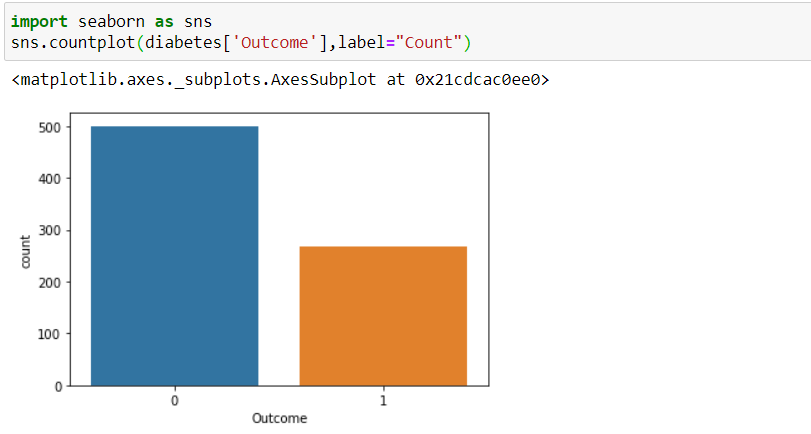
The necessary libraries has been imported and the columns has been printed



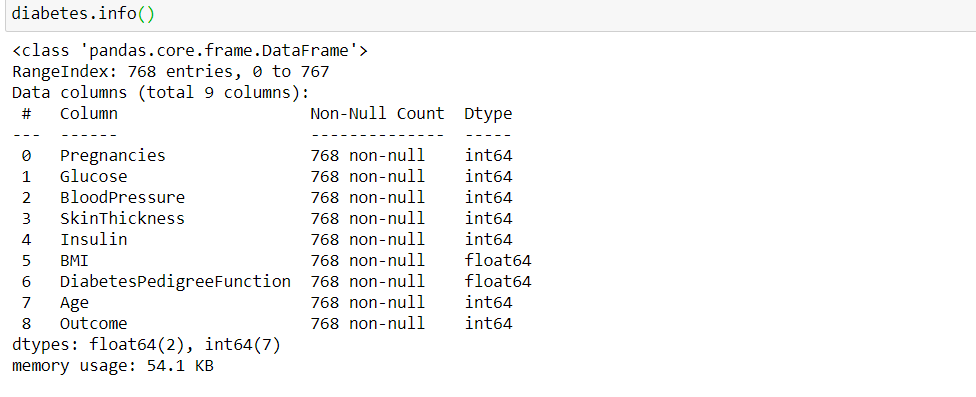
The dimension of the dataset has found to be as (768,9) i.e. 768 rows and 9 columns



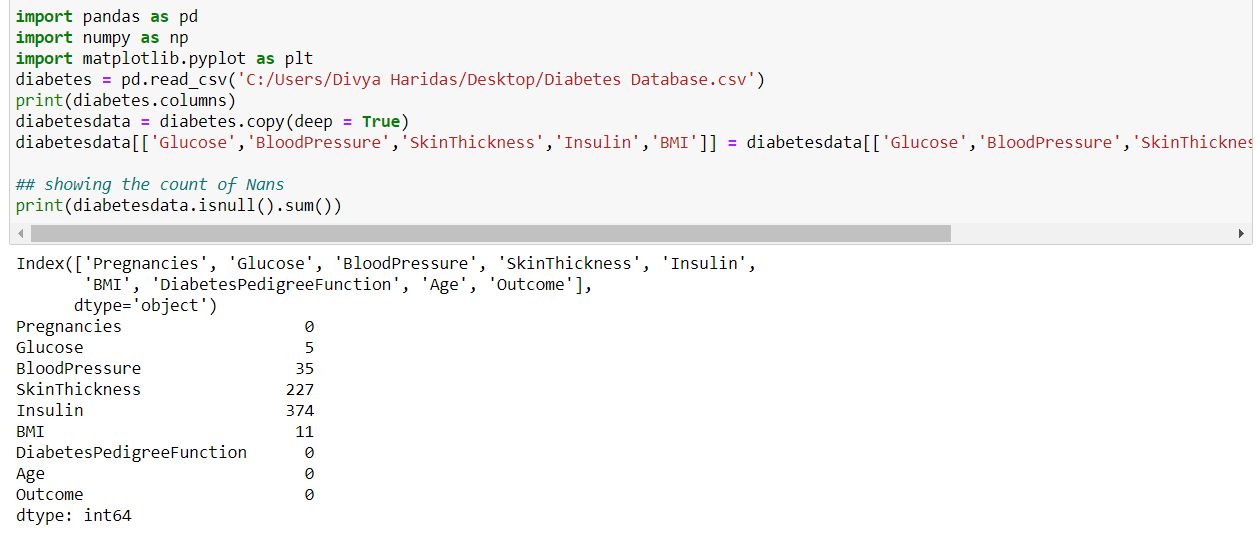
The class ‘Outcome’ is the feature that has to predicted. From the above command the count of persons with and without diabetes has been found. There are 500 persons without diabetes and 268 persons with diabetes.



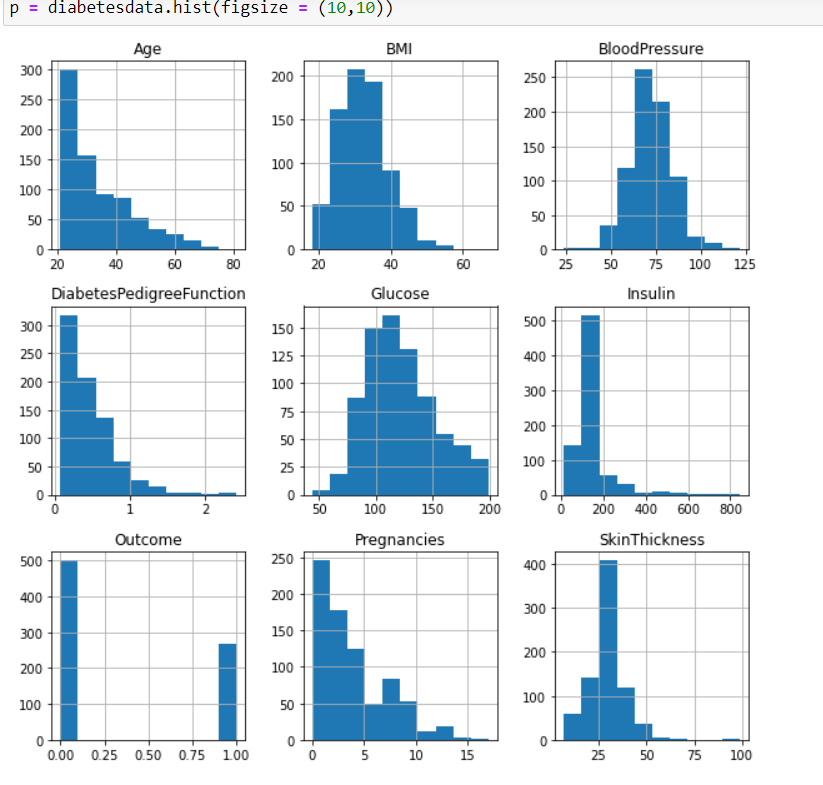
The above figure shows the ‘count’ and the ‘outcome’ of persons with and without diabetes. We can clearly see that the persons without diabetes is more than that of persons with diabetes.

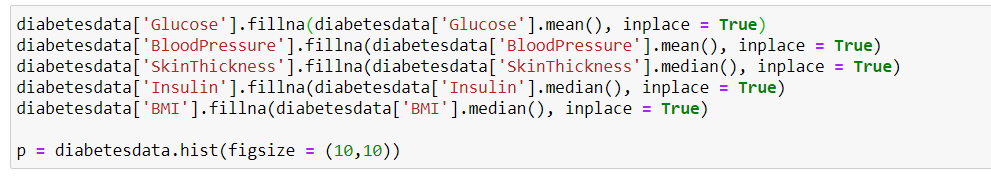


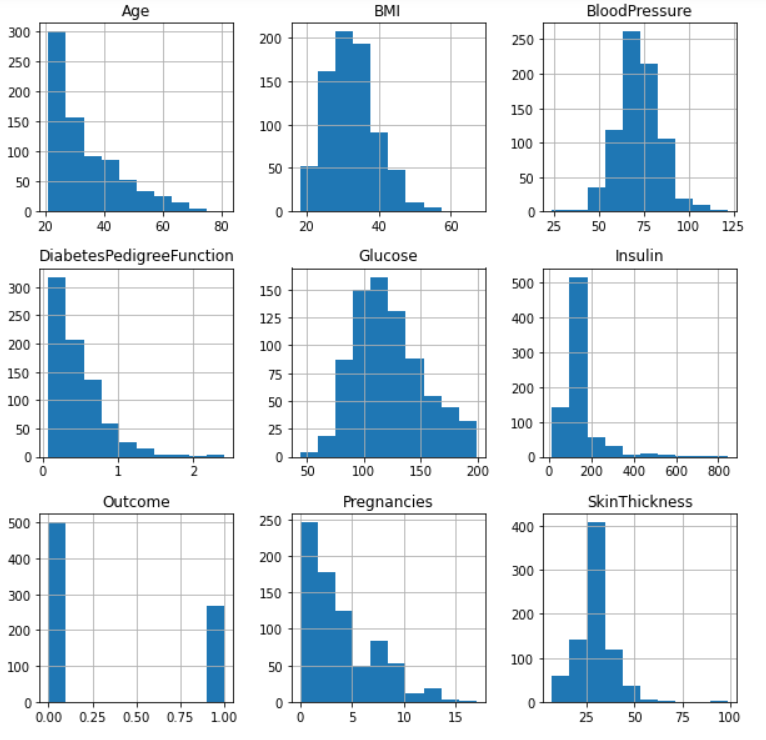
The info command gives a description of number of columns, rows and if the data contains any null values. Here there are no null values present. There are two features with data type float and 7 features with datatype int.



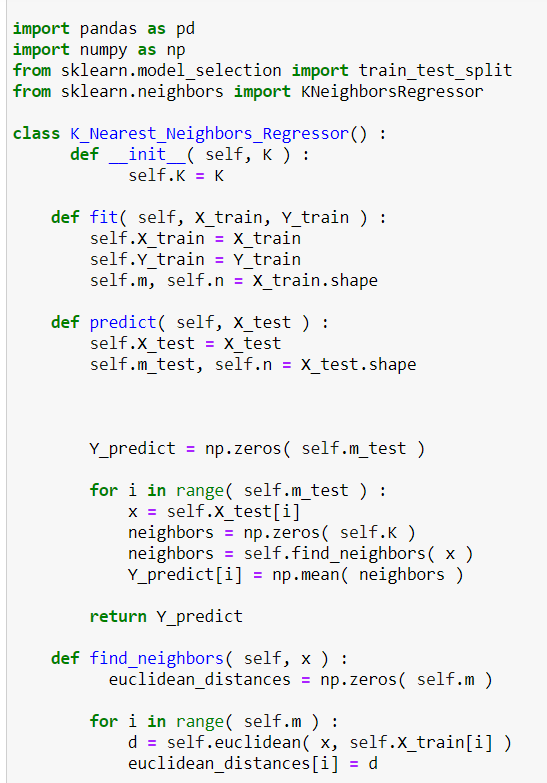
The dataset is checked for any null values. It can be seen that five attributes contains null values.

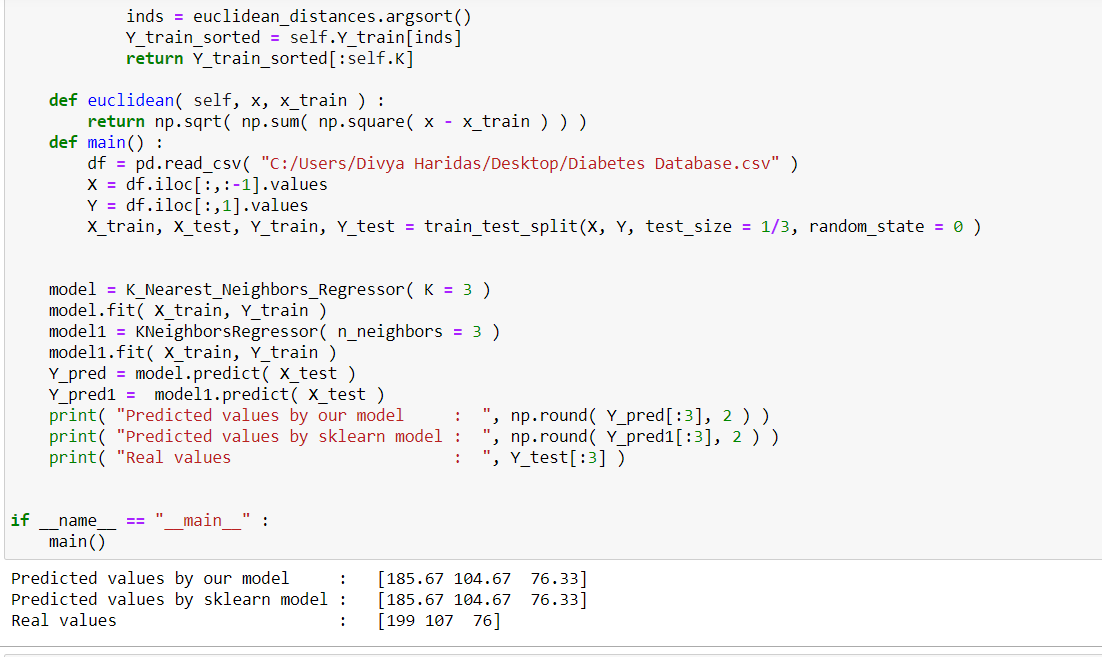




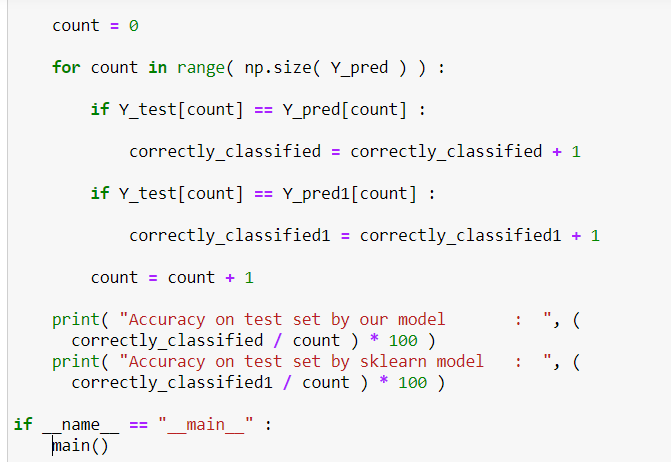


A histplot is done to have a view of null values and later the null values are filled with mean values and the histplot of that is also displayed.











The KNN model has been fit for prediction and to find the accuracy with K=12. The accuracy rate comes out to be 75%.