**Exploratory Data Analysis:**

* The BMI feature have outliers.
* Outliers are treated by using IQR technique.
* The age is distributed between 0 to 80. So, this data set contains all age group’s data.
* As the number of “Other” category gender people are less in count we removed that class.
* More than 30% data have no info about smoking history.
* There is a significant class imbalance in the dataset. Class 0 (indicating the absence of diabetes) has a much larger number of instances (83559) compared to Class1 (indicating the presence of diabetes) which has smaller number of instances (7215).
* In this data set higher indicators of getting diabetes are “bmi”, “HbA1c\_level”, “blood\_glucose\_level” features.
* The people who have age below 20 also have the diabetes.