1. Diabetes Dataset (Problem Statement 2):

TASKS:

1. Data Cleaning

2. Data Visualizing

3. Model Building

4. Model Testing.

* Data Cleaning: Here 1st we have started with the exploration of the given dataset and analyze it. Since we need to clean data this understanding of the data is important. Factors that are considered are: Duplicate Observations, Bad Labeling of data, Missing or Null values, Unexpected Outliers.
* Data Visualizing: In this data, “Outcome” column is what we are going to analyze and predict since that column says whether the patient is diabetic or not. Visualization helps to understand data. Here the visualization libraries used are Matplotlib, Seaborn which are used for statistical graphing.
* Model Building: Model Building is more important for the functioning of the algorithm and it should be more affective.

