Qn1) Objective: Prepare a classification prediction model using Boosting techniques to find Diabetes among patients.

Conclusion:

1. Loading Data & checking data shape:

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

%matplotlib inline

import seaborn as sns

import scipy.stats as stats

import pylab

from sklearn.preprocessing import StandardScaler

from sklearn.preprocessing import LabelEncoder

from sklearn.feature\_selection import SelectKBest

from sklearn.feature\_selection import f\_regression

from sklearn.ensemble import ExtraTreesRegressor

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import r2\_score

from sklearn.metrics import mean\_squared\_error

from math import sqrt

from sklearn.linear\_model import Ridge

from sklearn.model\_selection import GridSearchCV

from sklearn.linear\_model import Lasso

1. EDA:
2. Segregating data into numerical & categorical
3. Distribution Plot
4. Prob plot
5. Correlation heat map
6. Pair Plot
7. State wise spending & profit
8. Feature Engineering:
   1. Feature Scaling
   2. Label Encoding
   3. Segregating data into feature & target column
   4. Feature Importance using ExtraTreeClassifier & Univariate Selection
9. Modelling:
10. Linear Regression:

* Accuracy: 0.93
* Rmse: 0.22

1. Ridge regression with hyper parameter tuning:

* Accuracy: 0.93
* Rmse:0.22

1. Lasso Regression with hyper parameter tuning:

* Accuracy:0.94
* Rmse:0.21