Name : Gandepalli Sasi Kiran

Student ID: 16336380

PDS Assignment 2

Report: Diabetes Data

Our current code was developed in R, and it cleans, analyzes, and visualizes the diabetes dataset. The dataset is first read in, missing values are checked for, and zeros in some columns are replaced with the median values. The cleansed data is then saved in a new file after being filtered to remove certain excessive values.

It then takes a random sample of 25 rows from the clean data, determines the sample's mean and maximum glucose levels, and records the results in a file. Additionally, it computes the same values for the full population and stores the results in a different file.

Following that, a bar chart is used to visually compare the mean and maximum glucose readings between the sample and the population.

The code then determines the BMI value at the 98th percentile for both the sample and the population, saving the findings in a file. Additionally, a scatter plot is used to compare the sample and population's BMI values at the 98th percentile. A file with the chart's picture has been saved.

The code then conducts bootstrapping on the cleansed data's Blood Pressure column. The software creates 500 bootstrap samples with 150 values each, computes each sample's means, SDs, and 98th percentiles, and then stores the data in a file. The same data are also computed for the full population and saved in a different file.