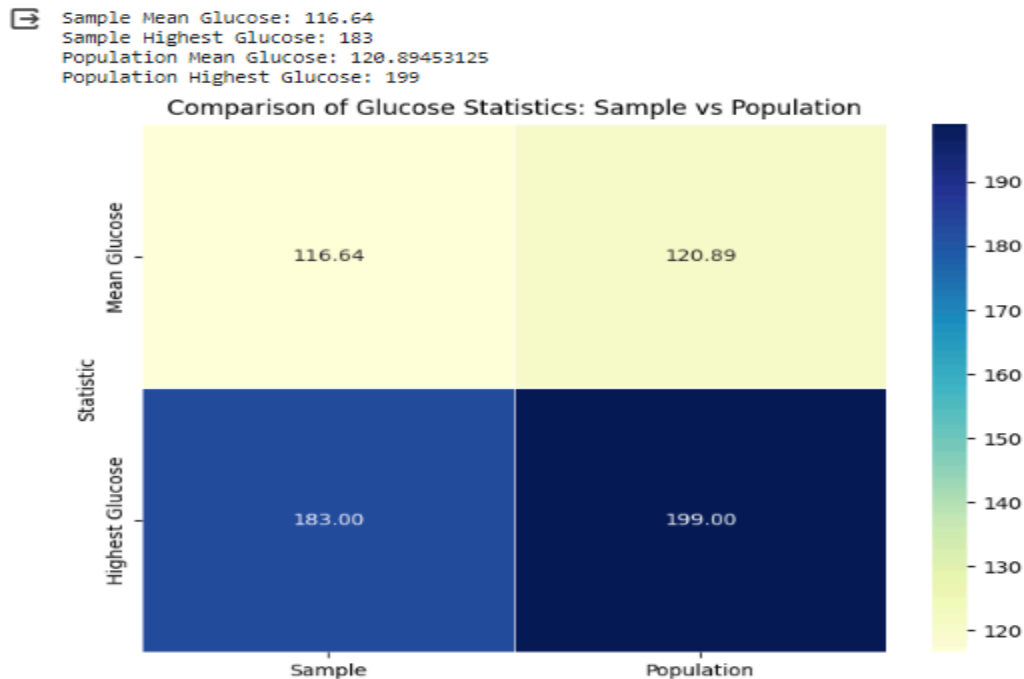# PDS RESULTS
ASSIGNMENT 4

**NAGINENI SAI LASYA**
16353784

2) The data file diabetes.csv contains data of 768 patients. In this data there are 8 attributes
(Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction,
and Age)
and 1 response variable (Outcome). The response variable, Outcome, has binary value (1
indicating the
outcome is diabetes and 0 means no diabetes). For this assignment purposes we will consider this
data
as a population. Use this data to perform the following:

a) set a seed (to ensure work reproducibility) and take a random sample of 25 observations and
find the mean Glucose and highest Glucose values of this sample and compare these statistics
with the population statistics of the same variable. You should use charts for this comparison.

Seed value : 30

```
Sample Mean Glucose: 116.64
Sample Highest Glucose: 183
Population Mean Glucose: 120.89453125
Population Highest Glucose: 199
```


Comparison of Glucose Statistics: Sample vs Population

b) Find the 98th percentile of BMI of your sample and the population and compare the results using charts

98th percentile of BMI for the sample 44.484
98th percentile of BMI for the population 47.52599999999996

Comparison of 98th Percentile of BMI
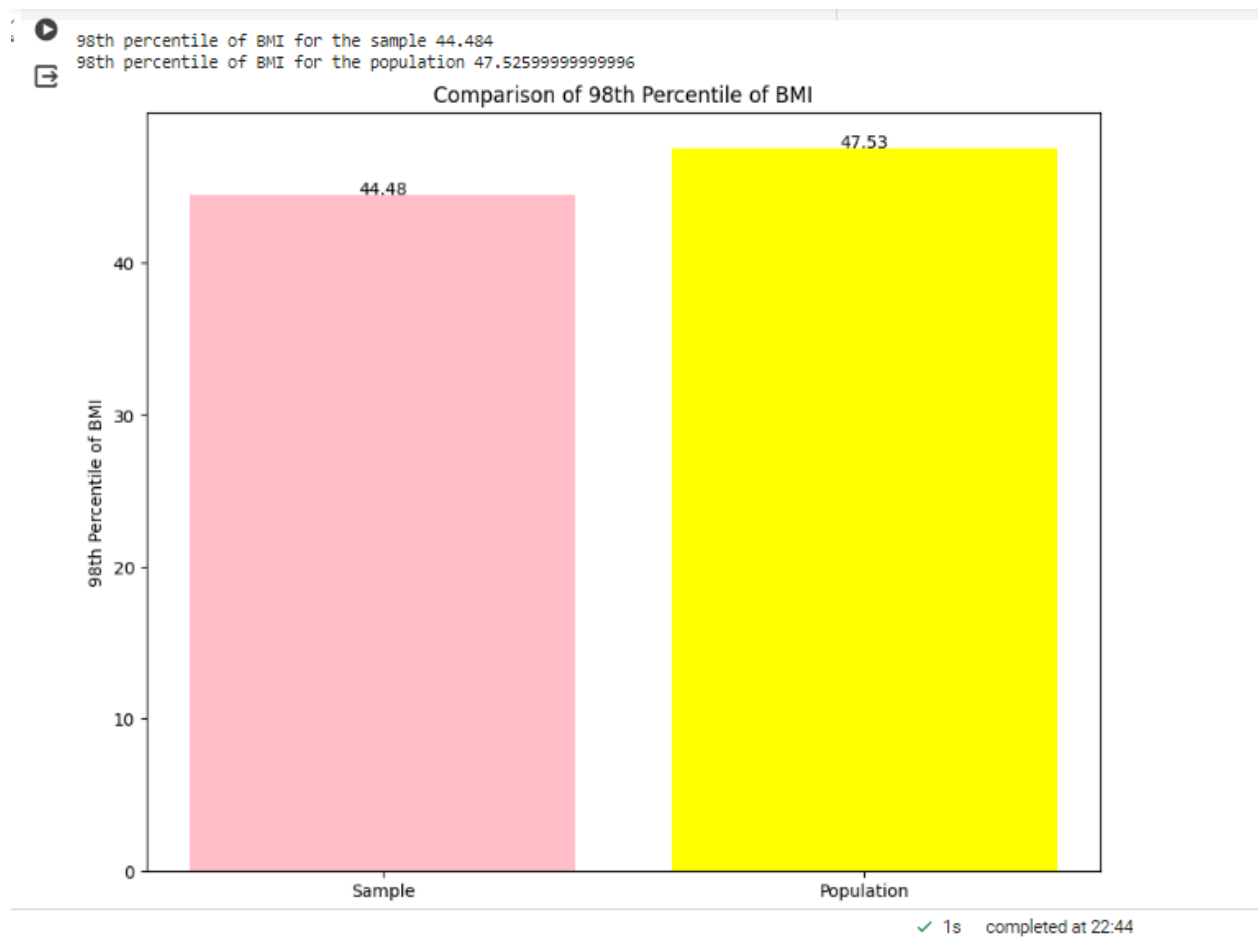
c) Using bootstrap (replace= True), create 500 samples (of 150 observation each) from the population and find the average mean, standard deviation and percentile for BloodPressure and compare this with these statistics from the population for the same variable. Again, you should create charts for this comparison. Report on your findings

Mean blood pressure 69.10546875
Standrad Deviation blood pressure 19.355807170644777
Percentile blood pressure 80.0
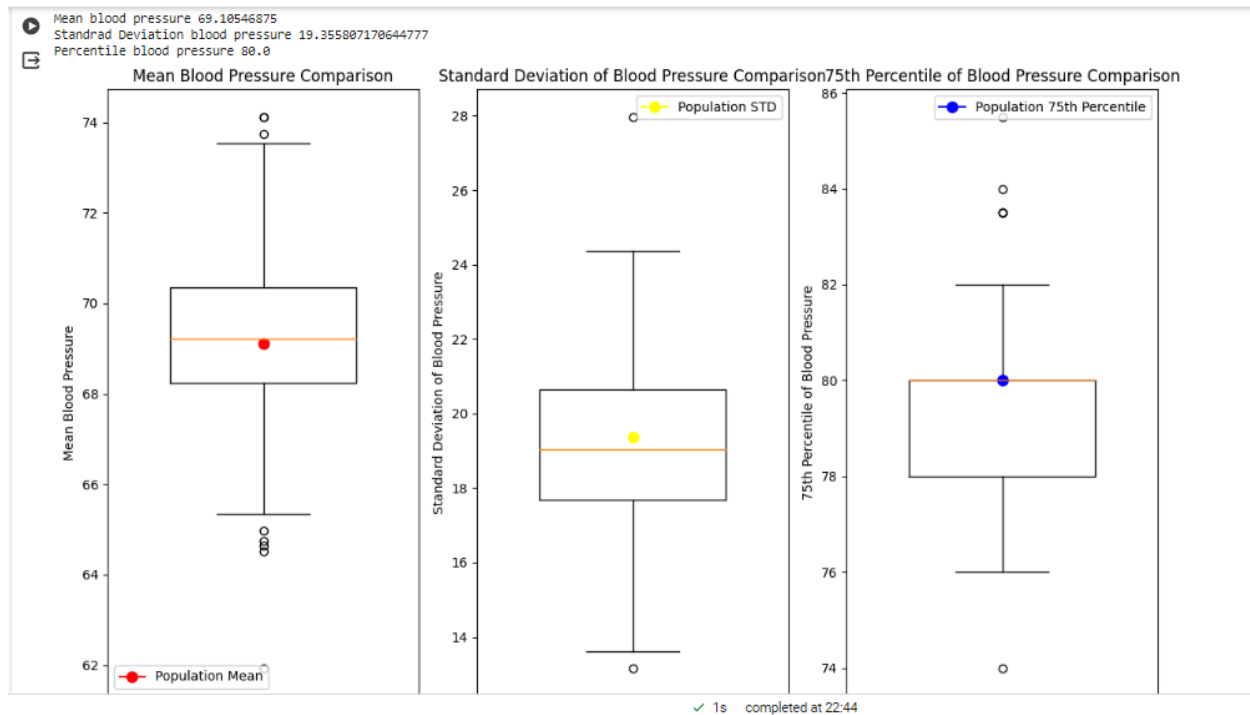
Mean Blood Pressure:

The population's mean blood pressure is around 69.1 mm Hg.
The bootstrap sampling approach yields a variety of estimates for the mean blood pressure, as seen by the variation in the mean blood pressure of the samples around the population mean.

The bootstrap samples' mean blood pressures tend to cluster around the population mean, suggesting that the sampling strategy captured the population's central tendency.

Standard Deviation of Blood Pressure:

The population's blood pressure standard deviation is around 19.3 mm Hg.
In the bootstrap samples, the blood pressure standard deviation also fluctuates within the population standard deviation.

With certain samples having more or less variability than the population, the bootstrap samples' standard deviation value spread represents the variability in the population.

Blood pressure :

Approximately 80.0 mm Hg is the population's 75th percentile blood pressure.
In the bootstrap samples, the blood pressure percentile likewise fluctuates around the population percentile.

All things considered, the bootstrap sampling technique offers a way to calculate the degree of variability and uncertainty in the population's blood pressure data. The evaluation of the sample's representativeness and the comprehension of the variations in the estimates derived from various samples are facilitated by the comparison with population statistics.