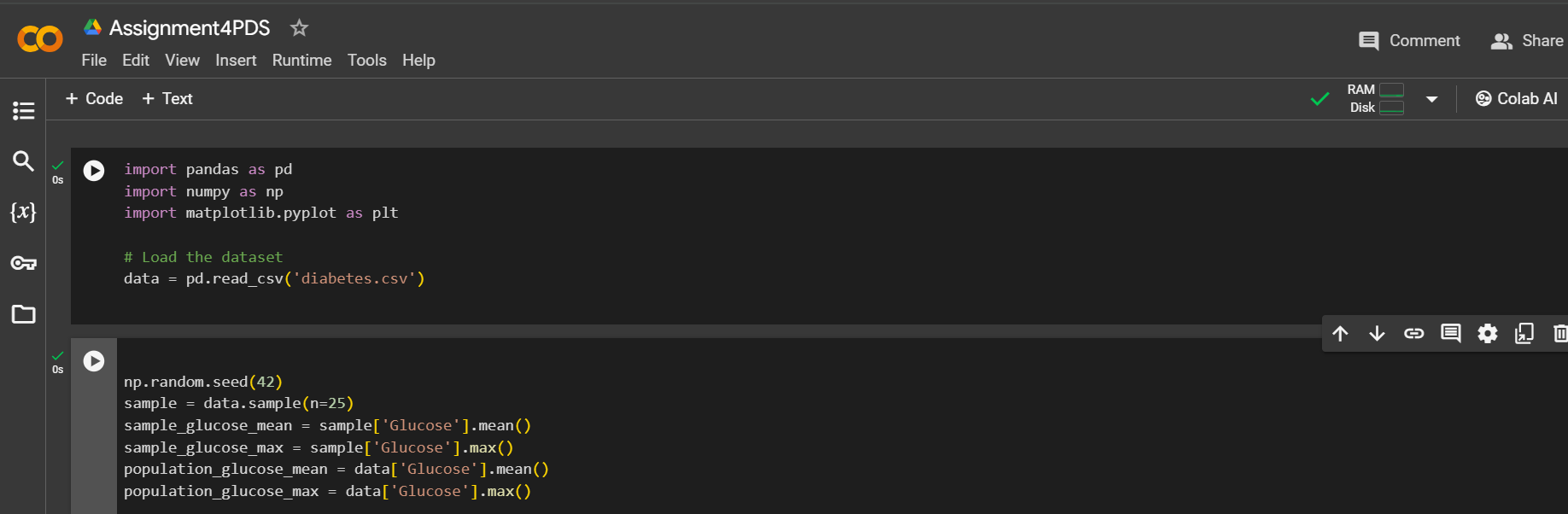
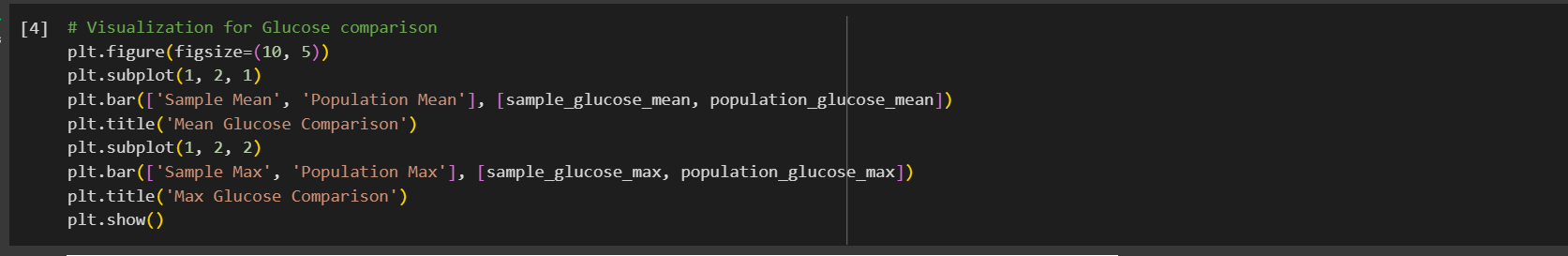
**Assignment 4**

2) The data file diabetes.csv contains data of 768 patients. In this data there are 8 attributes (Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, Diabetes Pedigree Function, 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.(5 points)





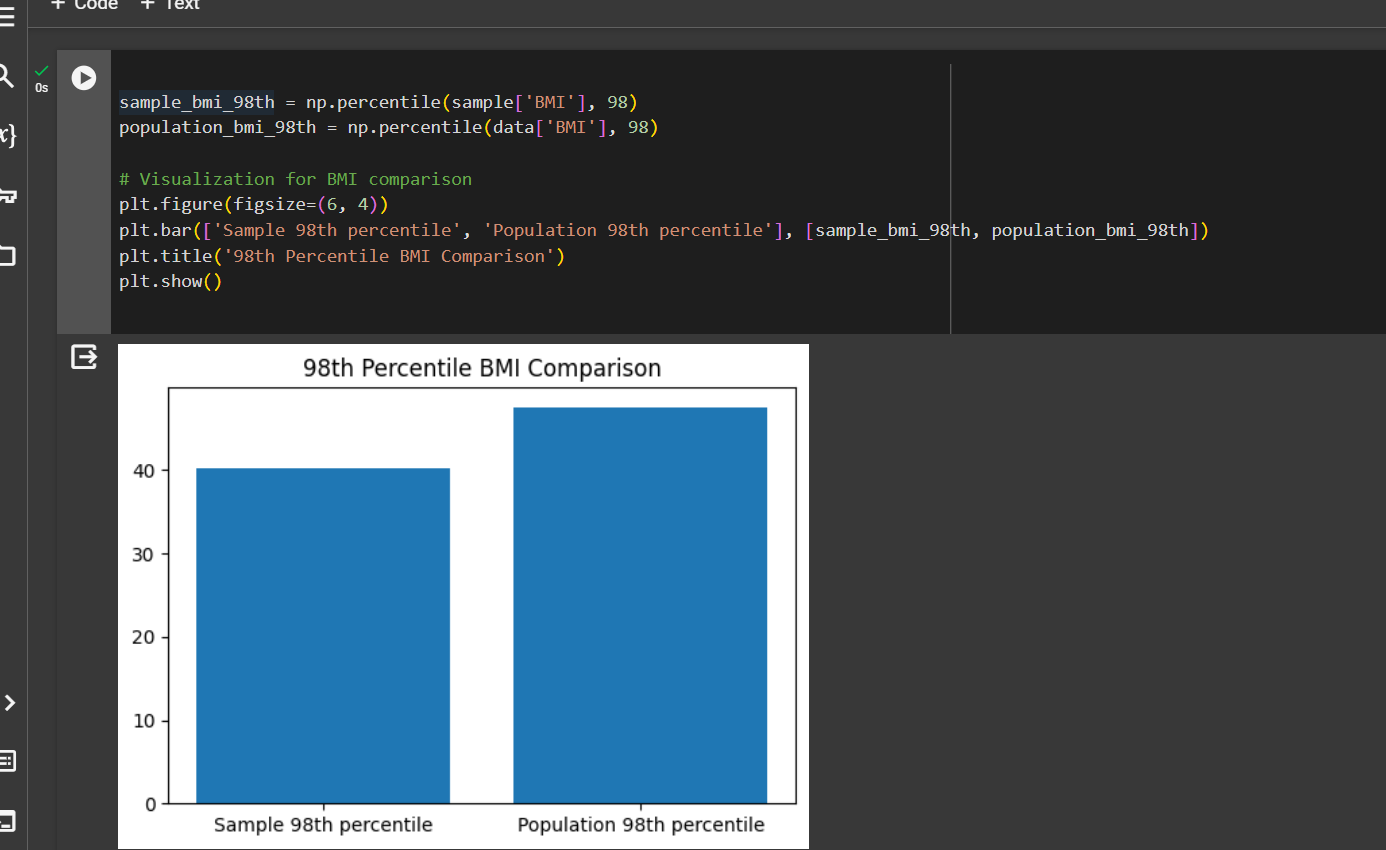
By comparing the mean and highest glucose levels between the sample and the population, the sample's mean glucose falls slightly below the population mean, indicating a tendency toward lower glucose levels in the sample. However, the maximum glucose value in the sample is similar to that of the population, suggesting that extreme values are consistent across both groups.

A screenshot of a graph

Description automatically generated

b) Find the 98th percentile of BMI of your sample and the population and compare the results using charts. (5 points)

By examining the 98th percentile of BMI, I found that both the sample and population exhibit comparable high BMI values. This indicates a consistency in extreme BMI values between the sample and the broader population, suggesting that individuals with high BMI are represented similarly in both groups.



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 Blood Pressure 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. (10 points)

A screen shot of a computer program

Description automatically generated

A screen shot of a computer code

Description automatically generated

A screenshot of a graph

Description automatically generated

By utilizing bootstrap sampling to compare blood pressure statistics between samples and the population, I can observe that the mean, standard deviation, and 95th percentile of blood pressure in the bootstrap samples align closely with those of the population. This suggests that the characteristics of blood pressure within the sample are representative of the broader population, reinforcing the reliability of the sample for inference.