Results:

Task 1:

Data:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

Taking random sample and calculating mean.

```
Glucose_mean_s=df1['Glucose'].mean()

print("Glucose mean sample : "+str(Glucose_mean_s))
```

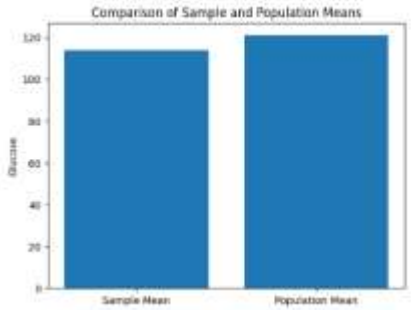
```
Glucose mean sample : 113.92
```

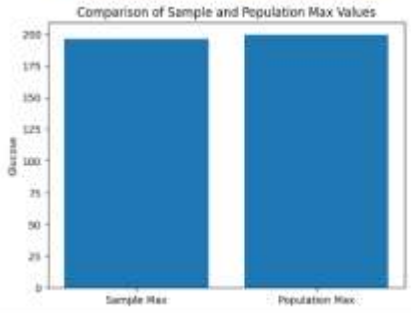
```
Glucose_max_s=df1['Glucose'].max()
print("Glucose max sample : "+str(Glucose_max_s))
```

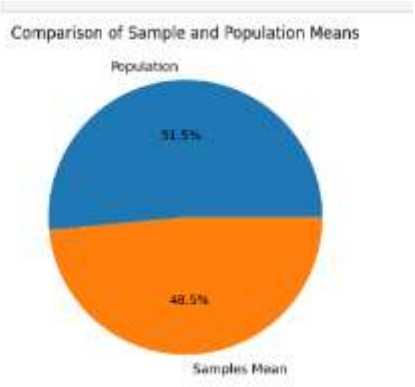
```
Glucose max sample : 196
```

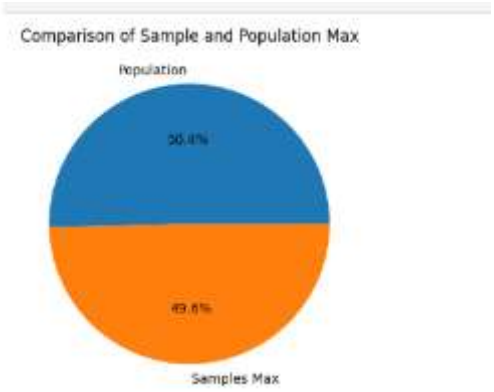
```
Glucose_mean_p=df['Glucose'].mean()

Glucose_max_p=df['Glucose'].max()
print("Glucose mean population : "+str(Glucose_mean_p))
print("Glucose max population : "+str(Glucose_max_p))
```

```
Glucose mean population : 120.89453125
Glucose max population : 199
```







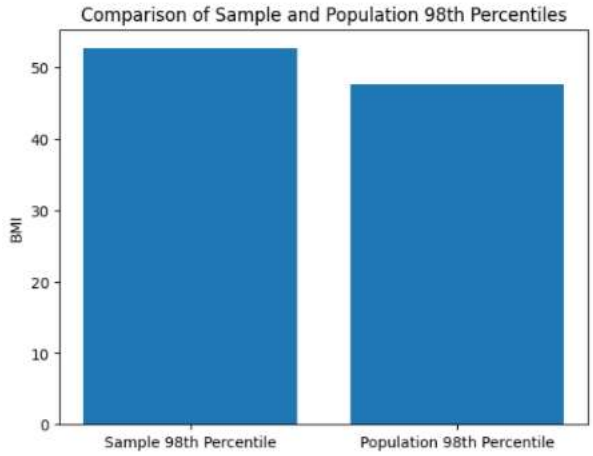


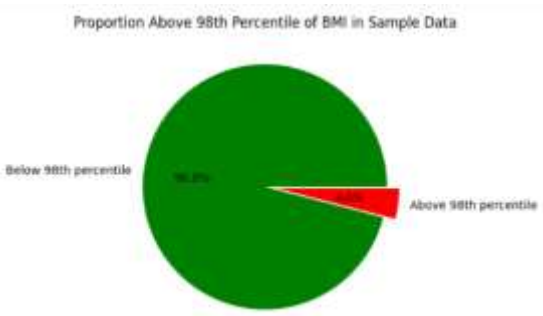
By the graphs the values of mean and max value of glucose in the sample and population are very closer.

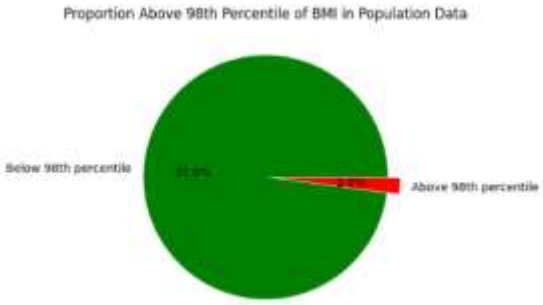
Task 2:

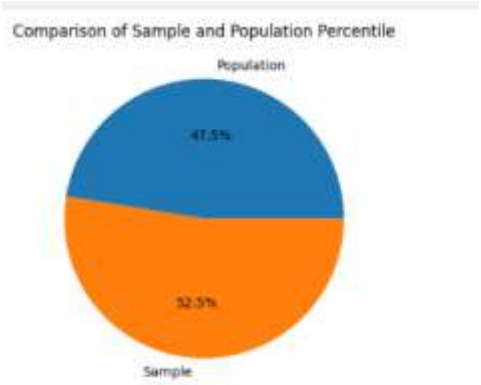
```
# Take a random sample of 25 observations
sample_data = df.sample(n=25)

# Calculate the 98th percentile of BMI for the sample and population
sample_percentile = np.percentile(sample_data["BMI"], 98)
pop_percentile = np.percentile(df["BMI"], 98)
print("Percentile value of the sample :" ,sample_percentile);
print("Percentile value of the population :" ,pop_percentile)
```

```
Percentile value of the sample : 52.63199999999999
Percentile value of the population : 47.52599999999996
```









In the total population we have 2% above the 98% percentile but in the taken sample of data we have 4%.

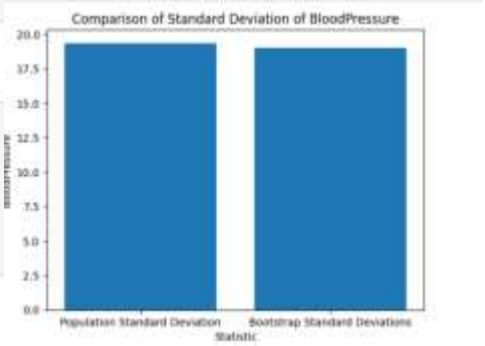
Task 3:

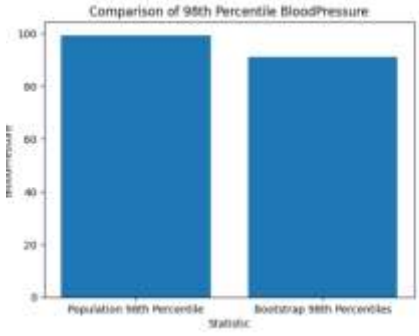
```
# Generate the bootstrap samples and calculate summary statistics for BloodPres
bootstrap_means = []
bootstrap_stds = []
bootstrap_percentiles = []
for i in range(num_samples):
    sample = bootstrap_sample(df, sample_size)
    bootstrap_means.append(sample["BloodPressure"].mean())
    bootstrap_stds.append(sample["BloodPressure"].std())
    bootstrap_percentiles.append(np.percentile(sample["BloodPressure"], 95))

print("Bootstrap Sample Mean :",np.mean(bootstrap_means))
print("Bootstrap Sample Standard Deviation :",np.mean(bootstrap_stds))
print("Bootstrap Sample Percentile :",np.mean(bootstrap_percentiles))
```

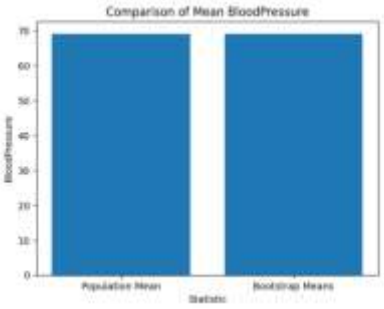
```
Bootstrap Sample Mean : 69.14689333333332
Bootstrap Sample Standard Deviation : 19.012701352276366
Bootstrap Sample Percentile : 91.08160999999998
```

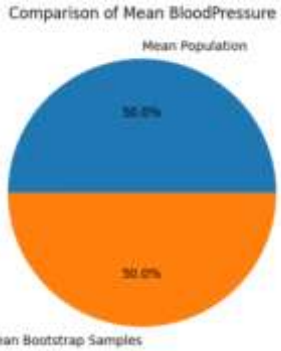
```
# Calculate summary statistics for BloodPressure in the population data
pop_mean = df["BloodPressure"].mean()
pop_std = df["BloodPressure"].std()
pop_percentile = np.percentile(df["BloodPressure"], 98)
print("Population Mean :",pop_mean)
print("Population Standard Deviation :",pop_std)
print("Population Percentile :",pop_percentile)
```

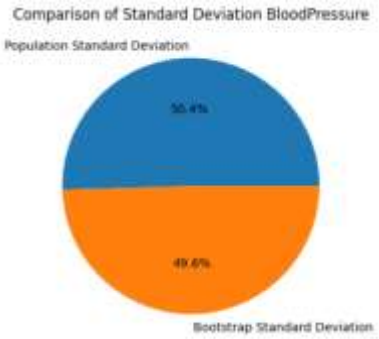
```
Population Mean : 69.10546875
Population Standard Deviation : 19.355807170644777
Population Percentile : 99.31999999999994
```

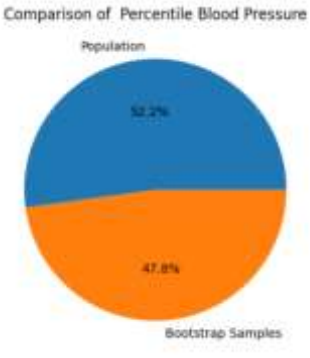












The population mean is 69. 10546875.The population standard deviation is 19.355807170644777, which indicates the spread of the variable of interest in the population, we might infer that the variable has a wide range of values in the population. The population percentile at 99.31999999999994 indicates that there may be some extreme values or outliers in the population.

The bootstrap sample mean is 69.14689333333332, which is very close to the population mean. This suggests that the sample is a good representation of the population. The bootstrap sample standard

deviation is 19.012701352276366, which is slightly lower than the population standard deviation. This suggests that the spread of the variable in the sample may be slightly smaller than in the population. The bootstrap sample percentile at 91.08169999999998 is lower than the population percentile at 99.31999999999994. This suggests that the extreme values or outliers in the population may not be present in the sample, or that the sample has a different distribution than the population.