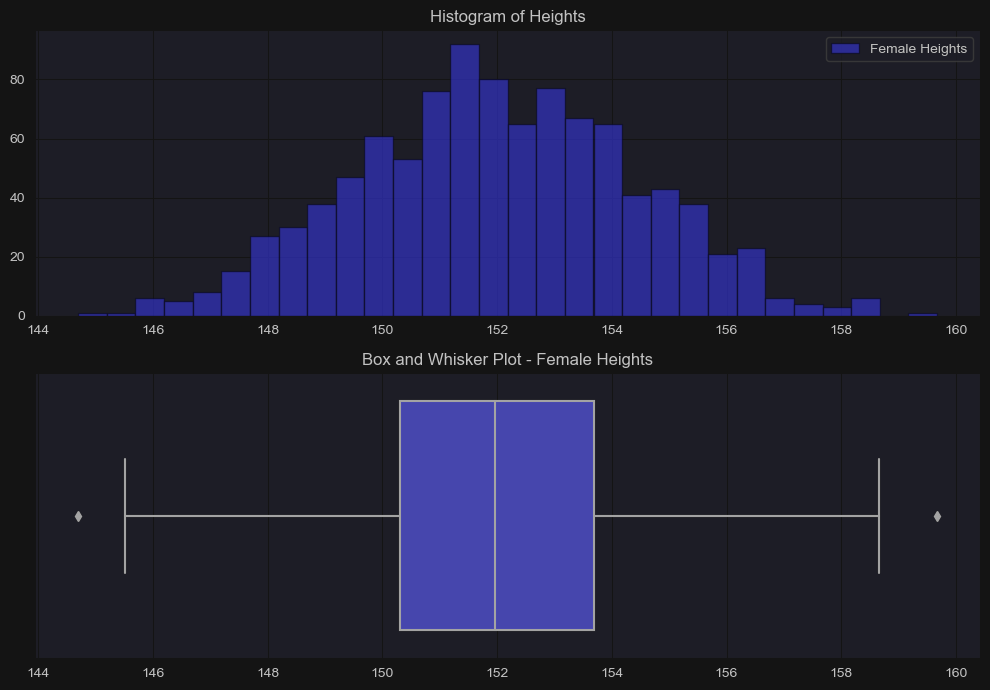
Name: Shreeyash S. Dongarkar

PRN: 22510025

Machine Learning Lab Assignment 3

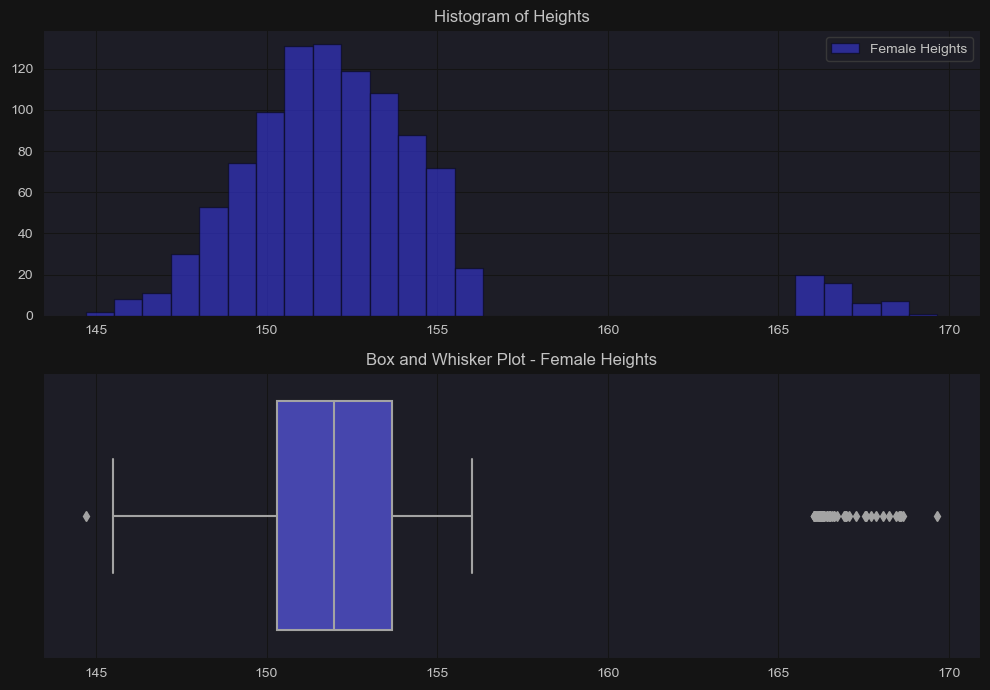
After generation of dataset:

* Values of mean and standard deviation were close to that of the actual mean and standard deviation



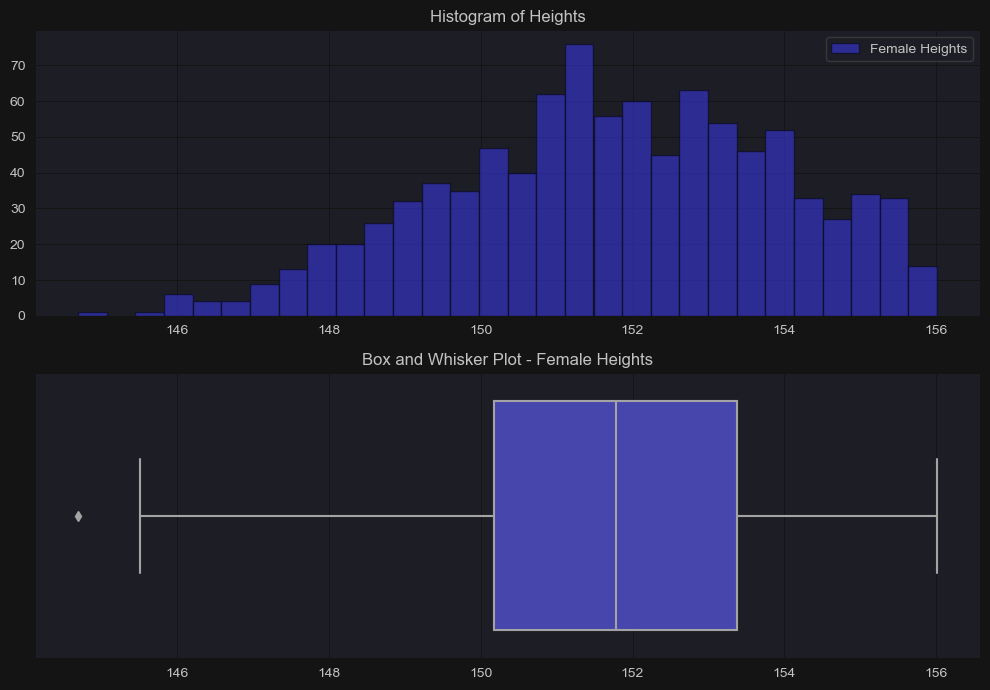
After introduction of outliers in dataset:

* There was increase in mean and standard deviation due to introduction of outliers impacting overall mean and distribution



Using z-score to decide cutoff for removing outliers:

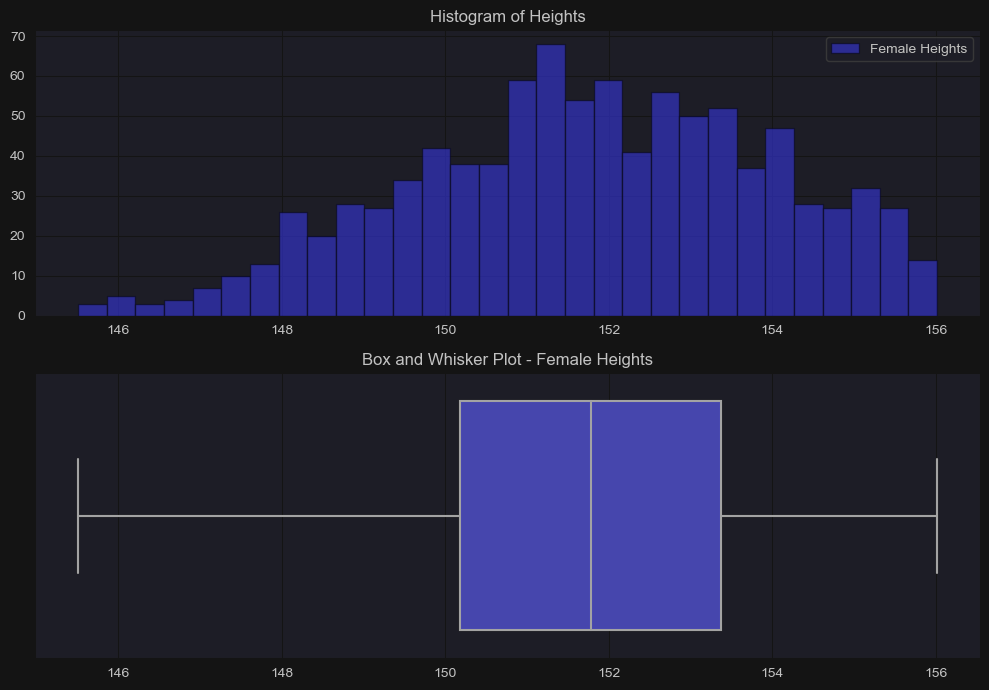
* Values having z-score greater than certain threshold are very extreme values and hence considered as outliers
* Here that cutoff is threshold is considered as 2



Most of the outliers are removed but the distribution is slightly skewed

Using IQR to decide cutoff to remove outliers:

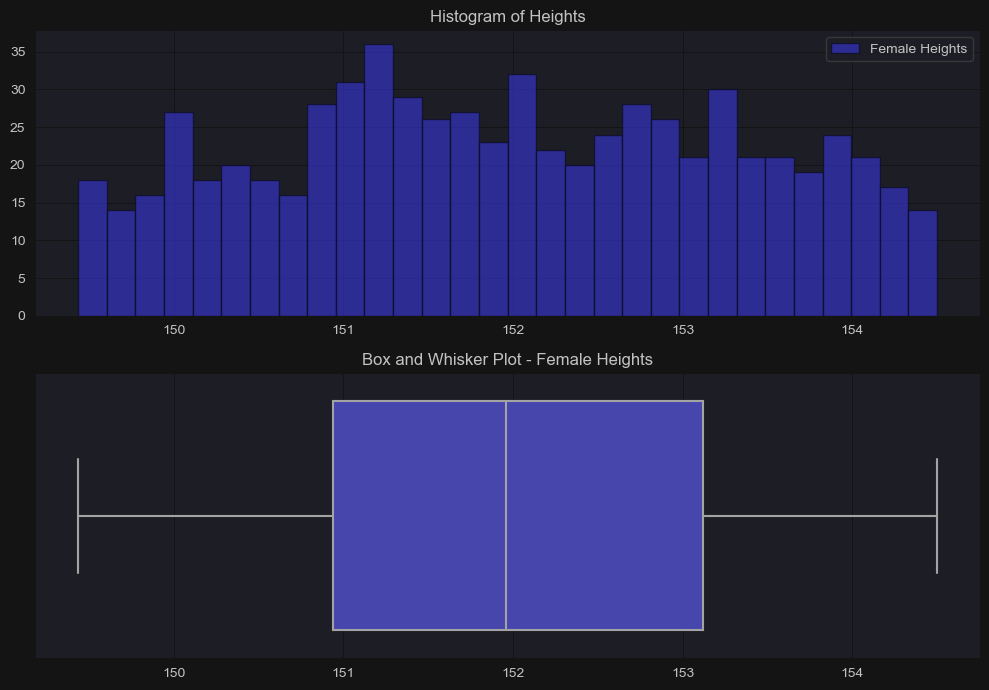
* Can be used to detect outliers by identifying the values that fall significantly outside the central range of data

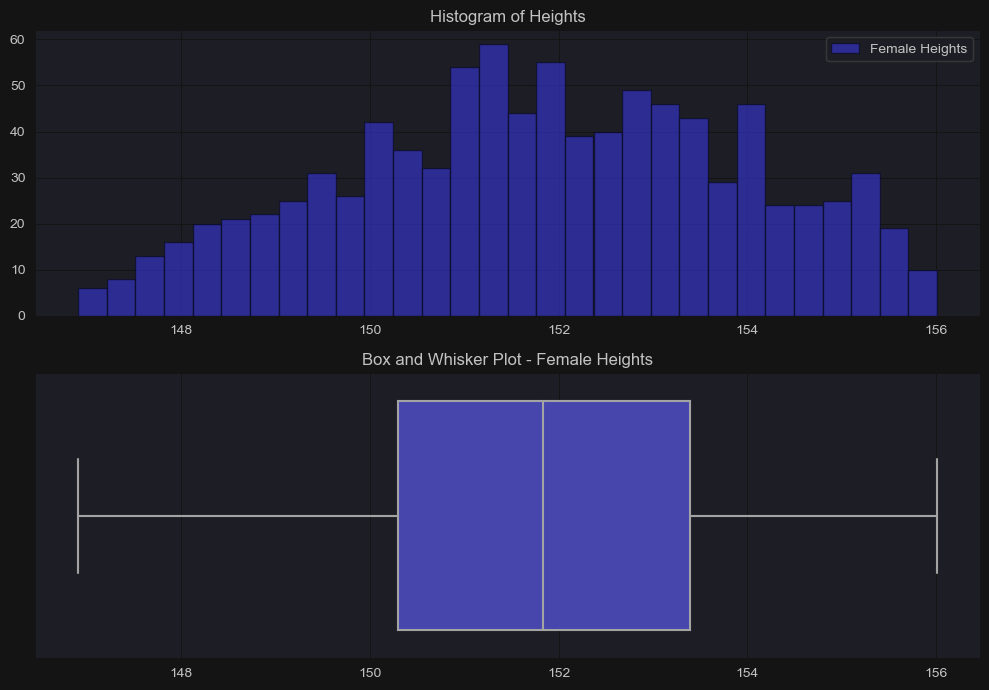


Here also most of the outliers are removed but the distribution is slightly skewed

Using MAD to decide cutoff for removing outlier:

* Here multipliers to MAD value can be used to define what qualifies as an outlier



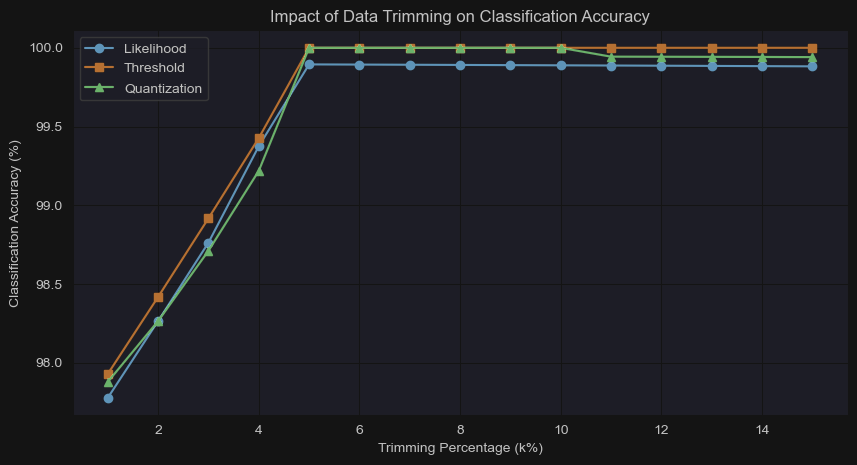


Here 3 is suitable as multiplier if lesser value is selected it identifies more values as outliers thus leading to more trimming of data and converting distribution to uniform whereas for 3 it is almost normal with no skewness

Accuracy Computation after removing outliers:

* Here z-score works well as number of datapoints is not that much large.
* Also, accuracy is increased in other to cases

Data Trimming:



Here as we go on trimming greater percentages of data as most of the outliers are removed accuracy goes on increasing