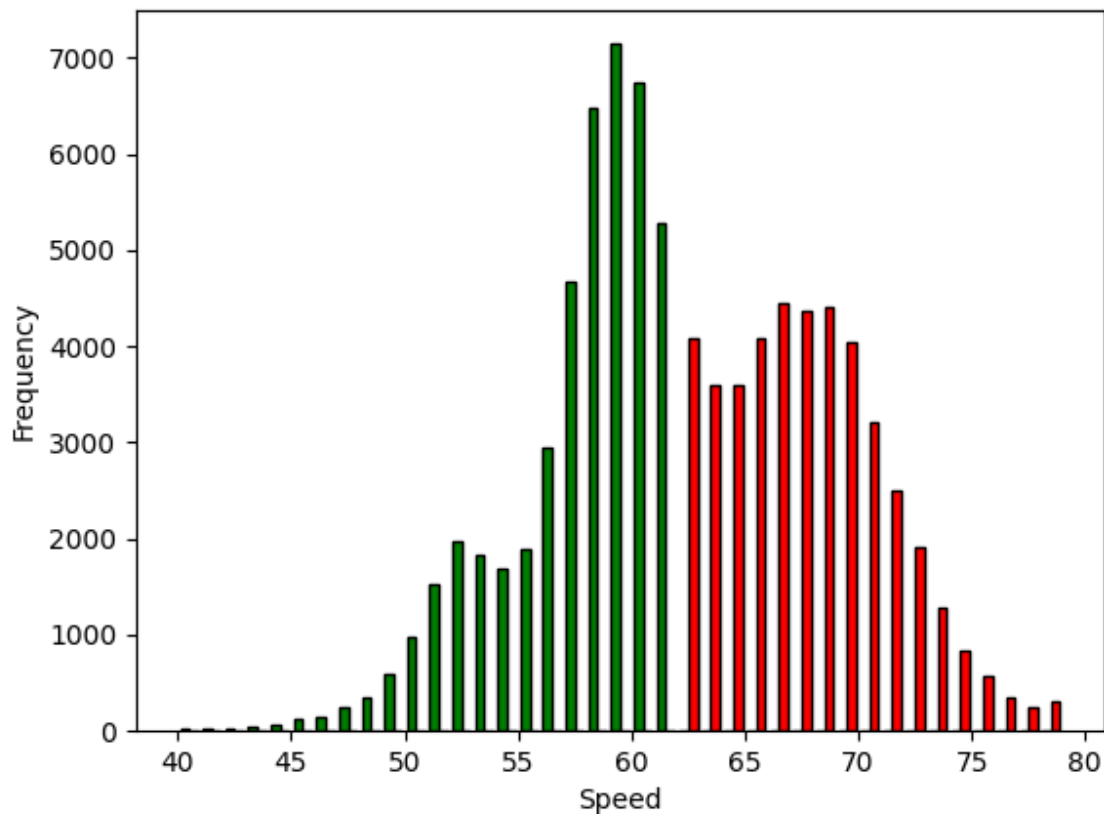


1.

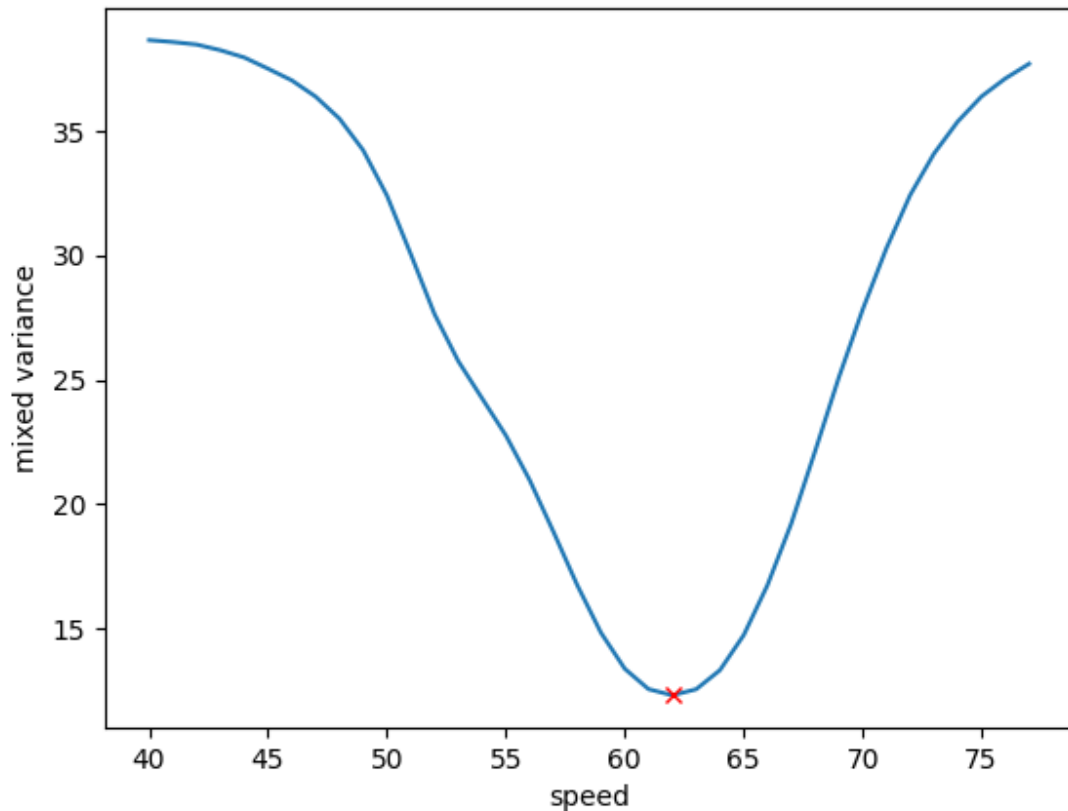
- a) I do not see any ethical issues in designing such a system.
- b) If we just send speeding tickets to all the drivers who are above the speed limit without any other consideration, this could cause issues. E.g By this design a speeding ambulance or other emergency vehicles might be flagged as reckless drivers. So, yes I would change my considerations.

2.

- a) We should use the speed threshold given by the otsu's method, i.e. 62.
- b) Otsu's method returns the minimum mixed variance which is 62.
- c) If this situation occurs, another measure should be used to break the tie. The program handles it by choosing the first value as the condition when a tie occurs is handled by the if condition where we check if the minimum is greater than the current.
- d)



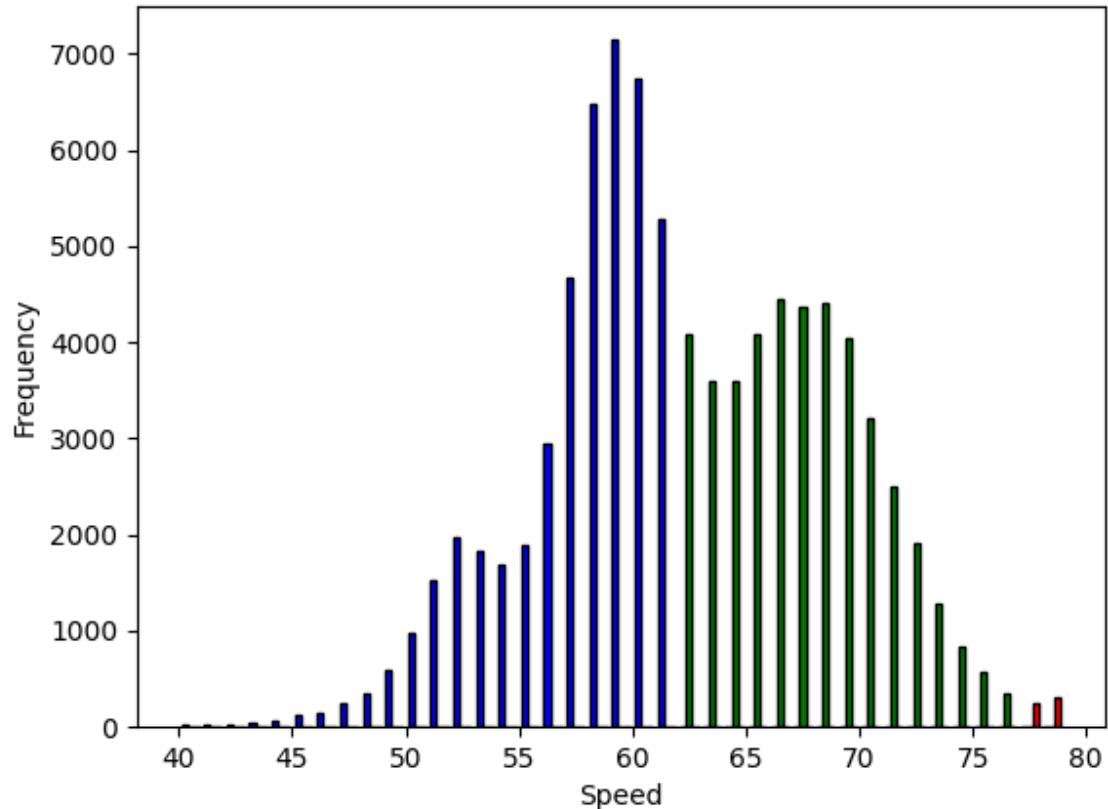
3.



4. Yes, this causes the threshold to change as demonstrated by the program execution console output. The new threshold comes out to be 61.
5. We have to partition the given data in 3 clusters. That means vanilla Otsu's method which returns the best threshold can not be directly used. To make this partition we need a way to get two threshold values so that we can split the data three ways.

I will choose a group which has the largest variance from the 1st application of the Otsu's and apply Otsu's method on that cluster. This is because we want to minimize variance and minimizing the one which is more in value makes sense.  
Speeds are - [40, 62), [62, 77) and 77 and higher.

After doing this we get the following histogram.



We can also choose the group with the largest speed range. This will divide the dataset more evenly, but the variance of the last cluster will be higher.

6.

a) Data cleaning is necessary to perform analysis on it within the boundaries of the given parameters. I had to exclude values of speed which were above or below the given limits of 40-80.

b) Otsu's method can be used when we need to cluster one dimensional datasets. It is a simple method which gives us a threshold value which minimizes the mixed variance that can then be used to partition the dataset as under or over the threshold value.

c) We can add a regularization term to the cost function used in Otsu's method that can

help in breaking ties. Regularization term needs to be appropriately scaled so that it will not dominate the cost function.

d) Python has libraries which provide the developer with various useful methods to perform analysis on data and then create visualizations. These visualizations can be customized as needed.