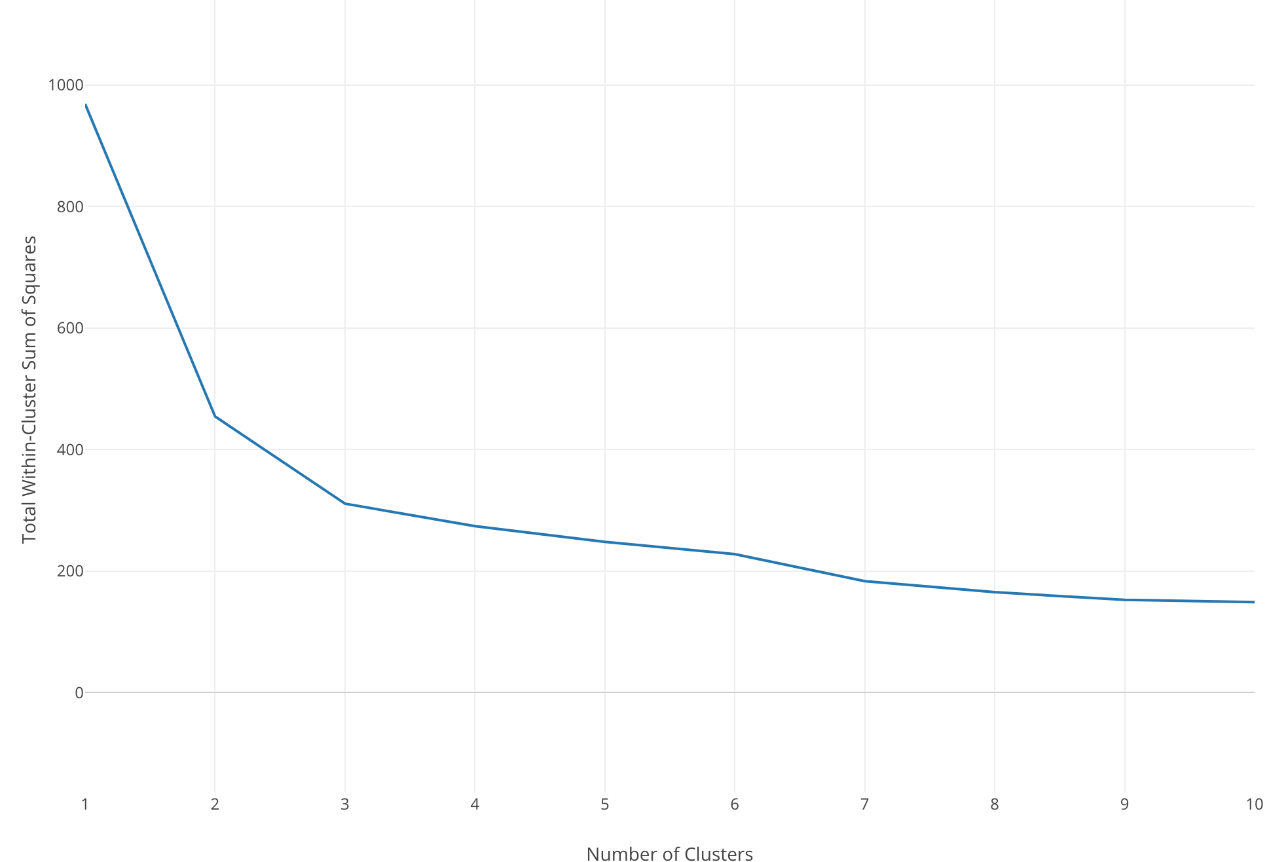
1. How to decide how many cluster to create

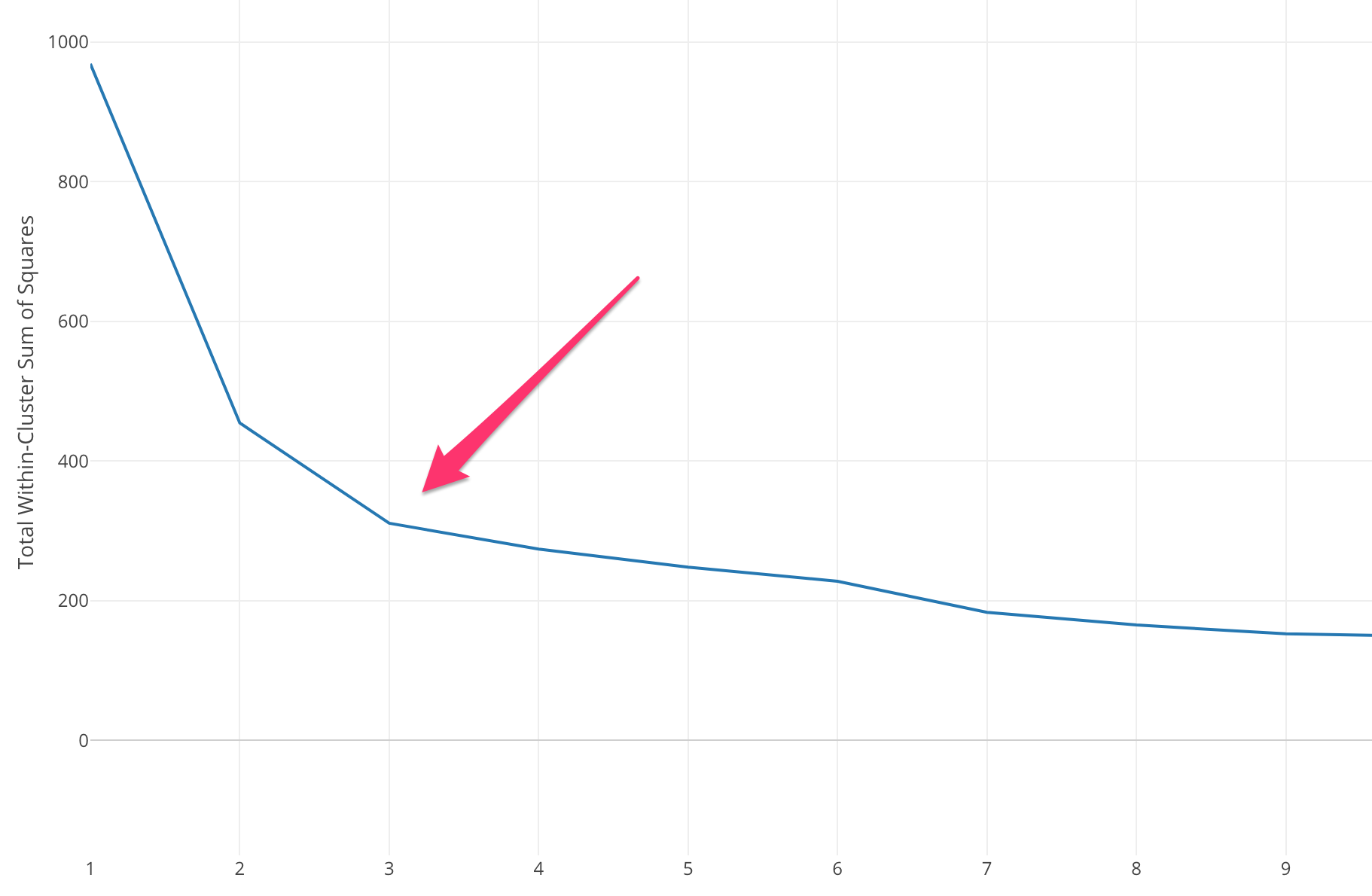
We iteratively build the K-Means Clustering models as we increase the number of the clusters starting from 1 to, let’s say, 10. Then we can calculate the distance between all the members (in our example they are the counties) that belong to each cluster and the center of each cluster every time we build a new model.

This will give us a series of models with different distance values. Finally, we draw a line that connects all the distances of the clusters like the below.



The Y-Axis is the distance and the X-Axis is the number of the clusters.

And when you see a huge drop that means increasing the number of the clusters for one at that point dramatically helps to optimize the model in terms of how good it separates the members. The distance keeps decreasing as you increase the number of the clusters, but usually, there is a point where the decline settles.



2, How to create K-means

(a). Question 1

Select the feature of VEHICLE COLOR and Violation Location to train the model.