Question 11:

a). (5 points) How many clusters have you determined? Please provide the Elbow and the Silhouette charts and state your arguments. The charts must be properly labeled.

Based on the Elbow and Silhouette values, four clusters are determined.

| N Clusters	Elbow Value	Silhouette Value:
2	3.1610	0.3939
3	3.1063	0.3973
4	2.7190	0.4123
5	2.9054	0.3857
6	2.8444	0.3654
7	2.8300	0.3839
8	2.8860	0.3731
9	2.9332	0.3759
10	2.8999	0.3719
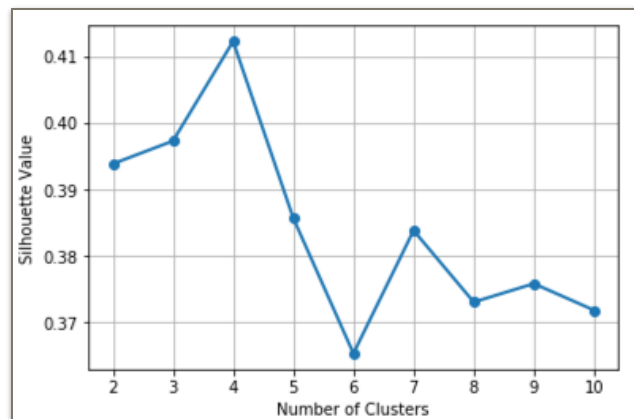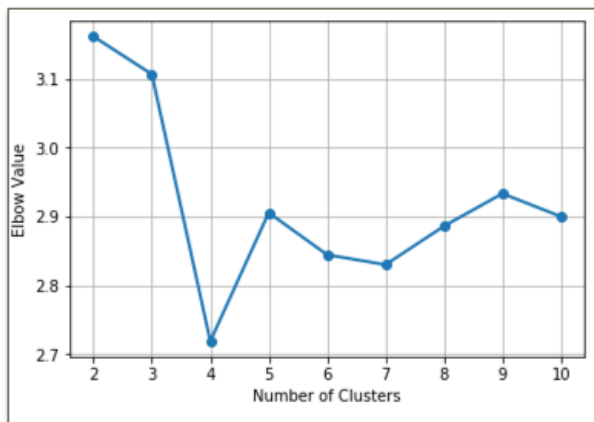
b). (5 points) Generate a scatterplot of LATITUDE (y-axis) versus LONGITUDE (x-axis) using the Cluster ID as the color response variable. You may need to adjust the marker size and set the aspect ratio to one in order to make the scatterplot more readable.
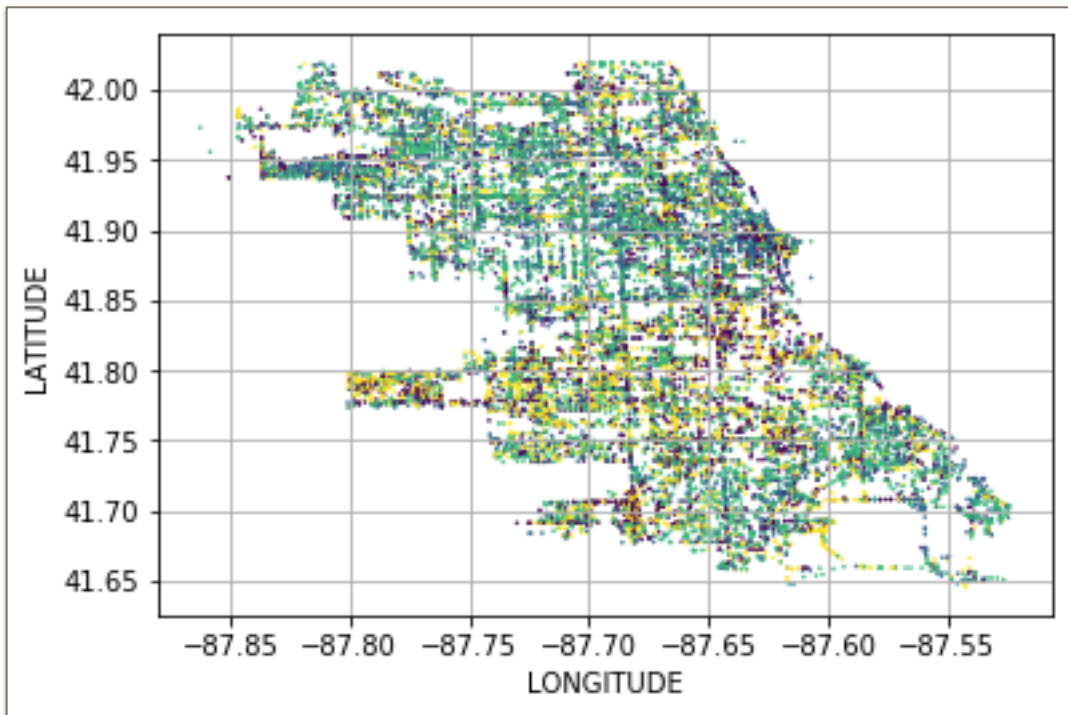
c). (5 points) How many leaves in your classification tree?   Please provide the properly labeled tree diagram.

There are four leaves in my classification tree.
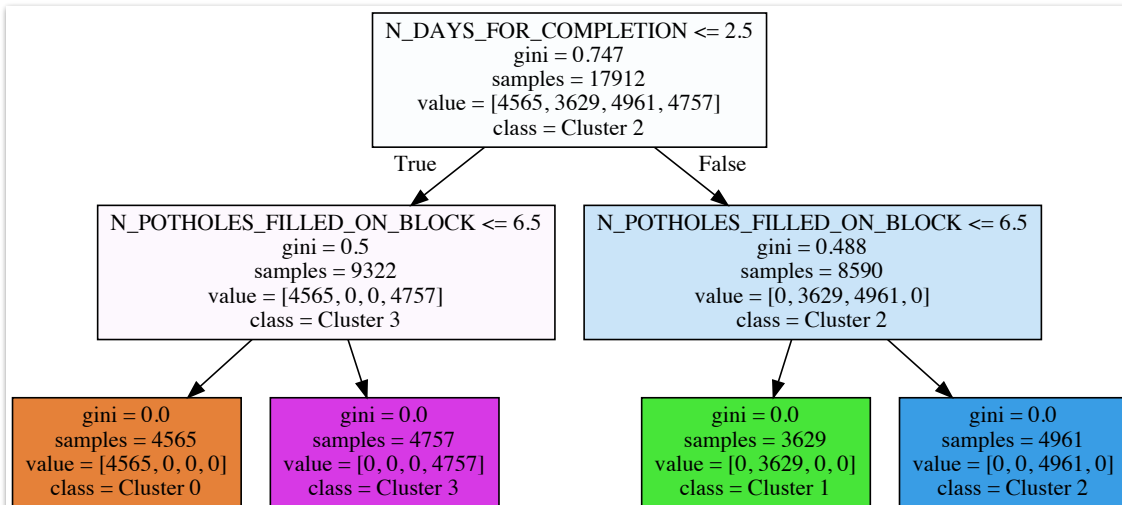
d). (5 points) What is the Root Average Squared Error of your classification tree? Please give your answer up to four decimal places.

The Root Average Squared Error of the classification tree is 0.0000.

e). (5 points) Based on your classification tree, please describe the profiles of the clusters which are at least 99% correctly classified by the classification tree.

4 Clusters: identified as 0, 1, 2 and 3
100% Accuracy and 0 Root Average Squared Error.

Cluster	Samples	N_DAYS_FOR_COMPLETION	N_POTHOLES_FILLED_ON_BLOCK	gini
0	4565	<= 2.5	<= 6.5	0.0
1	3629	> 2.5	<= 6.5	0.0
2	4961	> 2.5	> 6.5	0.0
3	4757	<= 2.5	> 6.5	0.0

Cluster 0 has 4565 observations
100% of these observations are N_DAYS_FOR_COMPLETION <= 2.5 and N_POTHOLES_FILLED_ON_BLOCK <= 6.5
We, therefore, can conclude that Cluster 0 is characterized by relatively less days for completion and less potholes filled on block.

Cluster 1 has 3629 observations
100% of these observations are N_DAYS_FOR_COMPLETION > 2.5 and N_POTHOLES_FILLED_ON_BLOCK <= 6.5
We, therefore, can conclude that Cluster 1 is characterized by relatively more days for completion and less potholes filled on block.

Cluster 2 has 4961 observations
100% of these observations are N_DAYS_FOR_COMPLETION > 2.5 and N_POTHOLES_FILLED_ON_BLOCK > 6.5
We, therefore, can conclude that Cluster 2 is characterized by relatively more days for completion and more potholes filled on block.

Cluster 3 has 4757 observations
100% of these observations are N_DAYS_FOR_COMPLETION <= 2.5 and N_POTHOLES_FILLED_ON_BLOCK > 6.5
We, therefore, can conclude that Cluster 3 is characterized by relatively less days for completion and more potholes filled on block.