Formula used to calculate Linear Discriminant ratio:

Where m, n are cluster numbers

This formula was used to calculate the ratio for each of the variables involved in the train set. These values are tabulated on the Observations Sheet in the Excel file.

Based on the values of the ratio, we picked out the most dominating variables between two clusters. Further, we tried to visually analyze and figure out similarities.

The observations noted after combining all 6 possible

Cluster 1: Very High Na, K, Fiber, Sugar Cluster.

Cluster 2: Low Sodium, Calories, Fat Cluster.

Cluster 3: High Fat and average values

Cluster 4: High Na, Carbs and Low Fiber, Sugar, K