**POWER GENERATION IN THE USA**

The electrical board in the United States produces electricity using a wide range of energy sources and technology. They are attempting to determine which fuel produces electricity with the lowest cost model in a very inefficient manner. The power board will be able to locate the best sources and save money with the aid of this study. I would use the present data set and several machine learning R Studio algorithms to figure this out. The three most efficiently working types of energy to produce electricity are fossil fuels, which include coal, natural gas, and oil. Need to recommend which is best.

Introduction:

\*a more accurate analysis I'd take a different approach, but let's start with the data set I've chosen: the cost of fuel receipts data set, which has 608565 observations over 23 variables.

\*As required by the task, I removed most of the undesirable data from the data set during this process. \*The MICE method was used to remove the data in the following phase. The MICE method can assist in replacing the missing values in a data set from the other columns.

\* Partition the data as per the task 75% training data and remaining into test data.

\* I obtained K=5 clusters using the silhouette method to determine the ideal number of clusters for my investigation.

\* I would analyze the cluster that is strong and close to the decision boundaries by creating the cluster plot.

PROBLEM STATEMENT:

1.What Algorithm is used to form a cluster?

2.How should the value for the number of clusters be chosen?

3.Which Cluster show high heat content

ANALYSIS:

1.what Information is revealed by clustering

When utilizing the k Means silhouette approach and deleting all category variables, the optimal number of clusters is K=5,

2.How should the value for Number of clusters be chosen

In detail view by using Silhouette method to find optimal number of clusters K=5 below I explained each of every clusters

Cluster 1: It looks to be a higher fuel\_mmbtu\_per\_unit and high with respect to ash\_content\_pct and good with sulfur\_content\_pct (2.76) approximately.

Cluster 2: It represents least in sulfur\_content\_pct, ash\_content\_pct, and maintains above average value with fuel\_cost\_per\_mmbtu.

Cluster 3: is dominant in the fuel\_received\_units, very highly influenced with "fuel\_cost\_per\_mmbtu"

Cluster 4: is next in place with fuel\_mmbtu\_per\_unit

Cluster 5: might be either the fuel\_mmbtu\_per\_unit and fuel\_received\_units are optimum.

(3) Which Cluster show high heat content

cluster 1 and 5 have very high rate of heat produced per units, cluster 2,3 and 4 is less heat produced per units,

Cluster 3 have a very high fuel received with very less heat produced.

Increased heat production due to use of natural gas out three fuels Gas being uses is more than Oil, Coal.

Conclusion:

The cluster analysis of values is what determines which segment is the best. According to my understanding, I'd pick clusters 1 and 5. Depending on the total values, Clusters 1 and 5 refer to the Fuel mmbtu per unit and Fuel received units as the fuel's heat content. The fuel received units always depends on the uses of the fuel. For better understanding any Business will always try to Save Money.

For Example, several different types of fuels, including natural gas, coal, and oil, are used by the electrical board in the United States to generate power. In this report's analysis, Utilizing oil and coal as fuels to generate heat Natural gas has more uses. There are several reasons to use gas, which is the most efficient fuel for power generation. Even the electricity board is eager to employ more gas in production.