**PROJECT REPORT**

**EXECUTIVE SUMMARY**

The ABC Corporation is in the midst of a financial crisis, thus in order to halt the downfall of business, they require a conveniently available and reasonably priced natural resource that aids in power generation.

To draw conclusions based on the problem that has been addressed, we have considered a total of four numerical variables namely fuel type code pudl, fuel received units, fuel mmbtu per unit, and fuel cost per mmbtu. K-means clustering has been used to identify the number of clusters i.e. k value =3. This helps to enhance the interpretation of the facts. The clusters generated and data interpreted leads to the conclusion that gas would be the ideal fuel for the company because it is easily accessible and reasonably priced compared to other fuels. Although its mmbtu value, or the amount of energy it can produce, is substantially lower than that of coal, this drawback has been overcome by the fact that gas is widely available in the U.S. If gas is used in large number to generate power then it can provide the same amount of energy as coal. As a result, the organization would surely think about using gas under the given circumstances and enhance the business

**INRODUCTION**

In order to do our research , we have used a dataset that contains details about monthly fuel contracts, purchases, and expenses that is reported in EIA-923 Schedule 2, Part A. There are 20 variables and 608,565 rows in this table, although a few of the variables have significant missing data. In the four variables considered, fuel cost per mmbtu has certain missing values which has been removed using

imputation. From the entire dataset only 2% has been taken into consideration and out of which 9000 are train sets and 3000 are test sets.

**PROBLEM STATEMENT**

The corporation is in financial crisis, thus its goal is to determine which natural resource would be deemed the greatest for generating electricity under the condition that it should be more readily available and less expensive.

**TECHNIQUE**

The data has been normalized using the min-max normalization as for every feature has a minimum value that is converted to a 0 and a maximum value that is converted to a 1. This makes it easier to interpret and derive conclusions.

In the next step k-means clustering has been used through which we have obtained the k-value= 3. We have used both the silhouette method and WSS method in our Analysis. K-means is a better option to this type of data as it deals well with large datasets and generalizes to clusters of different shapes and sizes.

**ANALYSIS AND FINDINGS**

Chart

Description automatically generated

**Graphical user interface, application

Description automatically generatedBackground pattern

Description automatically generated with medium confidence**

The three clusters formed consists of the following information :

Cluster1 = Coal(45), Gas(4598) and Oil(776).

Cluster2 = Coal. (3275)

Cluster3 = Gas. (436)

Fuel expenditures and mmbtu are roughly the same for Clusters 1 and 3. When compared to clusters 1 and 3, Cluster 2 (Coal) is comparatively less expensive and produces significantly more mmbtu. It is also observed that Cluster 3 has the most fuel received units.

**CONCLUSIONS**

As the cost and availability of the fuel are the key issues in this situation rather than the element affecting power generation, gas is considered to be the ideal fuel for the organization. Despite the fact that coal has a greater mmbtu, gas can nevertheless provide more energy than coal as the units of fuel received is several times higher than coal. There will not be any further investment by the company in this, unlike for the import or export of fuels, due to the ample supply of gas within the country itself. This has surmounted the fact that coal can provide higher levels of energy. As per the analysis it is evident the Cluster 3 is the best among the three.

**APPENDIX**

References: - <https://data.catalyst.coop/pudl/fuel_receipts_costs_eia923>