**Analysis of Heat Content in US Power Generation**

**Fall-2022 FML Final Project Sony Kiran**

* **Executive Summary**

By cleaning, standardizing, and cross-linking utility data from various sources in a single database, the Public Utility Data Preparation (PUDL) takes information that is already publicly available and makes it publicly usable. We want to study how effectively we can comprehend and suggest solutions to increase power generation in the US by performing some analysis and performing some segmentation on the data. The model was created and verified using a massive dataset created by translating massive quantities of raw data provided by PUDL. The dataset contains all the information gathered and was used to train, test, and evaluate the system. According to the analysis, the three most utilized fuel categories are coal, gas, and oil. After the analysis, Gas is suggested as a fuel type for power generation after a comprehensive investigation using K-Means and it concluded that Average fuel cost per mmBTU of heat content increase, Heat content of the fuel in millions of Btus per physical unit will getting increase and vice versa. So, the better the fuel cost, the better the heat content.

* **Problem**

To start a new company in the Power Generation Sector, the company needs to know which fuel gives the most power and an eco-friendly. Analyzing the data from the previous company will give the required information.

* **Technique**

The dataset from the PUDL has large number of variables. Since K-means handles the large number of variables in the best way and generalizes the clusters of different shapes and sizes, it can be considered as our algorithm. Consequently, a model would be created that accepts a data sample as input and returns the cluster to which the new data point belongs based on the training that the model received. The optimal number of clusters from the model is 3. Within the sum of the squares by cluster, one cluster has the minimal. So that cluster is the efficient cluster because all the points are clubbed together with minimal distance

* **Conclusions**

The better the fuel cost, the better the heat content. Ash content level is zero in Gas and it’s good for the eco-friendly. Trucks and railways are best for the early stage of production since building a pipeline will take some time. From the Analysis, gas is the best fuel among the three which doesn’t cause any harm to environment and gives the good profit to the company

* **Reference**

<https://catalyst.coop/pudl/>

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

https://datasette.io/