



FUNDAMENTALS OF MACHINE LEARNING

FINAL EXAM REPORT

JETANGI, LOKESH

Table of Contents:

S.No	Contents
1	Executive Summary
2	Introduction
3	Problem Statement
4	Analysis and Discussions
5	Conclusion

Executive Summary:

As the U.S. government plans to preserve fossil fuels, the goal of this project is to study many aspects of power generation in the U.S.A. and make recommendations regarding the kind of fuel that can be removed from the power generation process. In order to determine the fuel type on which the government spends the least, analysis has been done depending on how much is spent on each fuel. The government is willing to store this sort of fuel for usage in upcoming decades because it is thought to be the least consumed.

Data has been divided into three categories with one type of fuel in each group using a machine learning technique. On the basis of the average fuel price, the quantity of fuel received at power plants, and the chemical make-up of that fuel, additional analysis has been conducted. According to the findings, the United States government spends the greatest money on gas and the least on oil as fuel types. Despite being less expensive per MMBtu unit, coal is not commonly used, and this may be because it contains pollutants such as ash, mercury, and sulfur.

Introduction:

Information about monthly fuel contracts, purchases, and costs is included in the data. There are 30 variables in it, ranging from the `mine_ids` from which various suppliers will furnish fuel to different power plants to the mode of transportation to be employed to deliver these fuels. Data has been cleansed for effective analysis, and some variables with missing and redundant data have been eliminated. 2 percent of the data from 608,565 rows have been sampled using a set. Seed (1234), which aids in more accurate data interpretation. And data is divided into training and testing set 75% of the sampled data are used as the training set, while the remaining 25% are used as the test set. we got the output of 9000 training sets and 3000 test sets .

Problem Statement:

In the United States, In order to determine the type of fossil fuel on which they are spending the least money, the U.S. Power Generation Unit has employed a data analyst to analyze historical data of `monthly_fuel_contract`, `Purchases`, and `Costs` information. They are prepared to reduce spending on the fossil fuel type that they use the least frequently and stop using it for the purpose of producing electricity. This new strategy is being utilized, it is claimed, in order to protect fossil fuels, which are not now used very much.

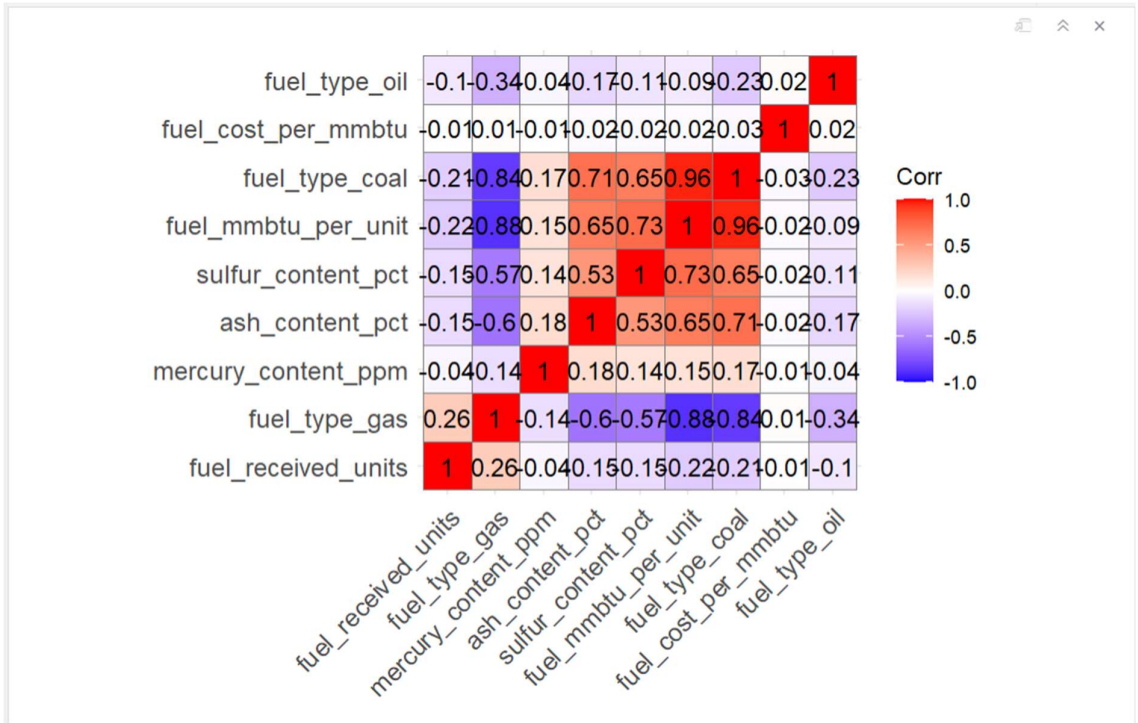
Analysis and Discussions:

As a set seed, a 4-digit random number, i.e., 1234, was used to randomly sample around 2% of the total data to ensure that the data remained unique. A training set of 75% of the 2% data and a test set of 25% of the data were each taken. Since Z-score compares observations across disparate variables and spots outliers, it was chosen as the data-normalizing strategy. Silhouette and WSS analysis were both employed in the Clustering algorithm, which is denoted by the letter K. due to the fact that it scales to enormous data sets and is relatively straightforward to implement. As well as elliptical clusters, it generalizes to clusters of various sizes and shapes.

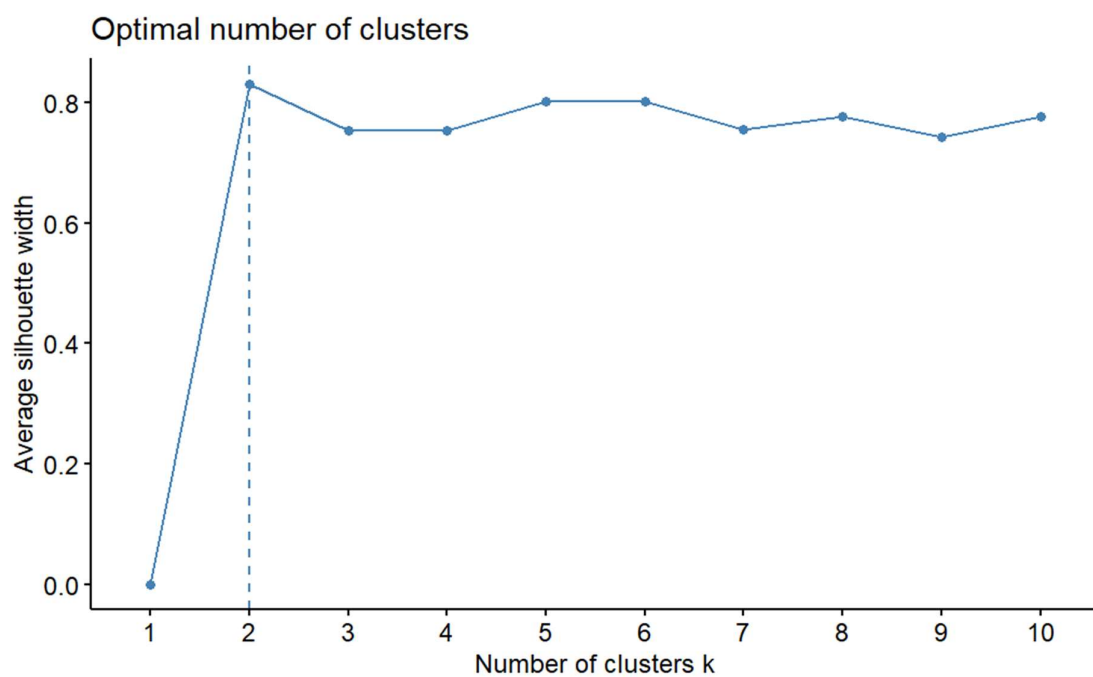
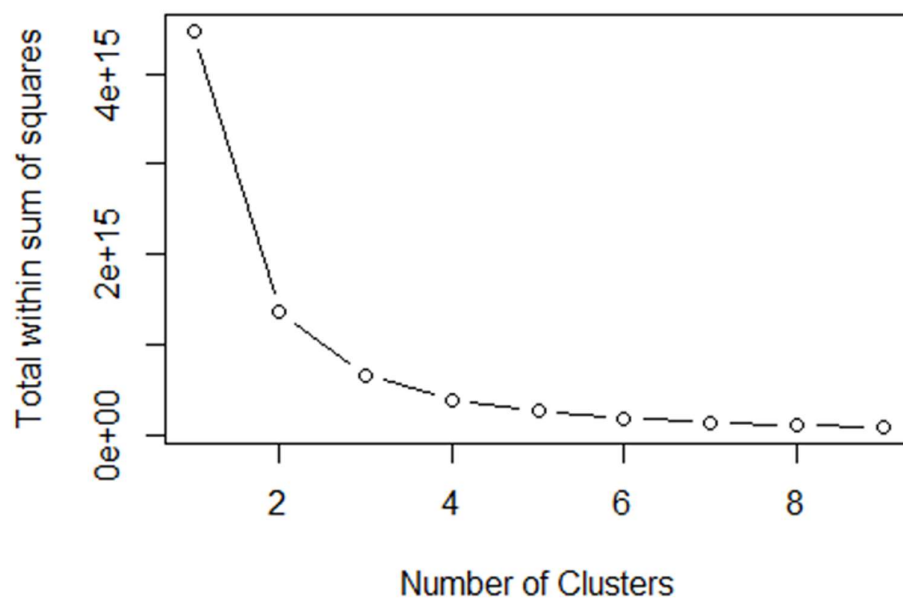
The U.S.A power generating spends the most on gas because it is the fuel that is used in the most units at the lowest cost per unit. They spend the least money on oil as a fuel source. According to the U.S. Energy Information Administration, oil is used to produce power as well as to propel vehicles and in the petrochemical industry to create hundreds of different intermediate and end-use products, including polymers and solvents. Because oil can be used for a variety of purposes, it is not advised to eliminate it from power generation.

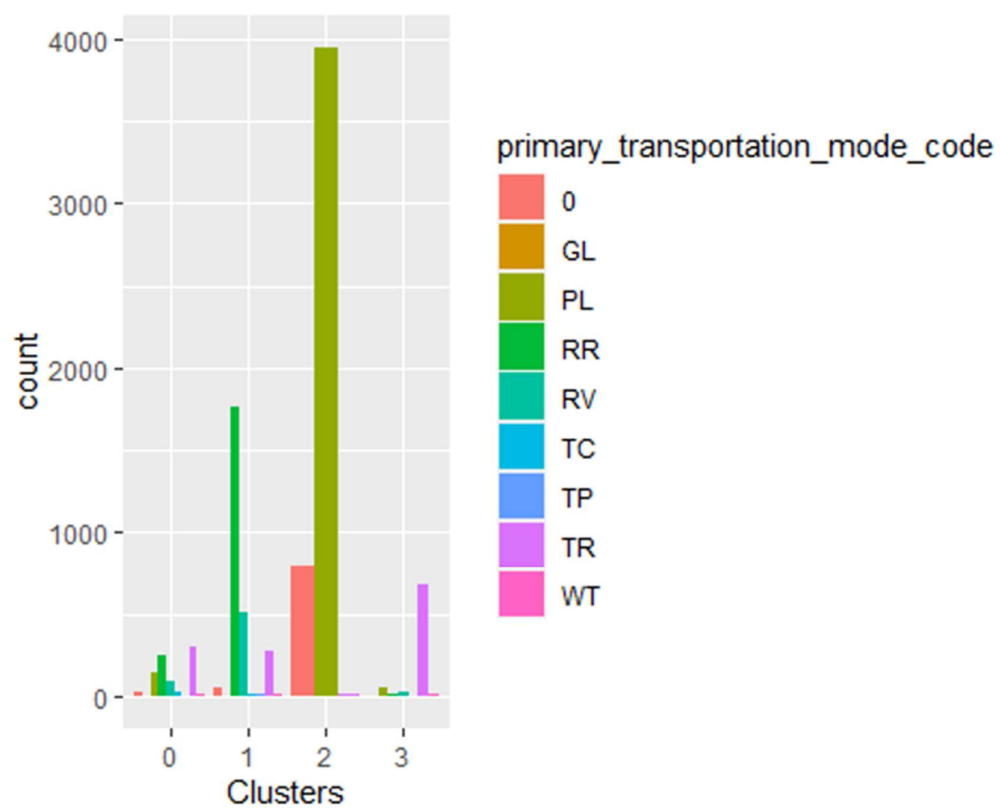
The cost of coal is higher than that of oil, and it has larger percentages of ash, mercury, and sulfur impurities according to its chemical makeup. When power is produced in power plants using this sort of fuel with impurities, additional costs must be met because the more severe the contaminants, the higher the cost would be. The government ought to take into account these extra expenses. According to the International Energy Agency's report, all unabated coal production must stop by the year 2040 if fossil resources are to be preserved for future generations. Through 2030, coal-fired power plant emissions must be reduced on average by about 8% annually to be on pace with the Net Zero by 2050 Scenario.

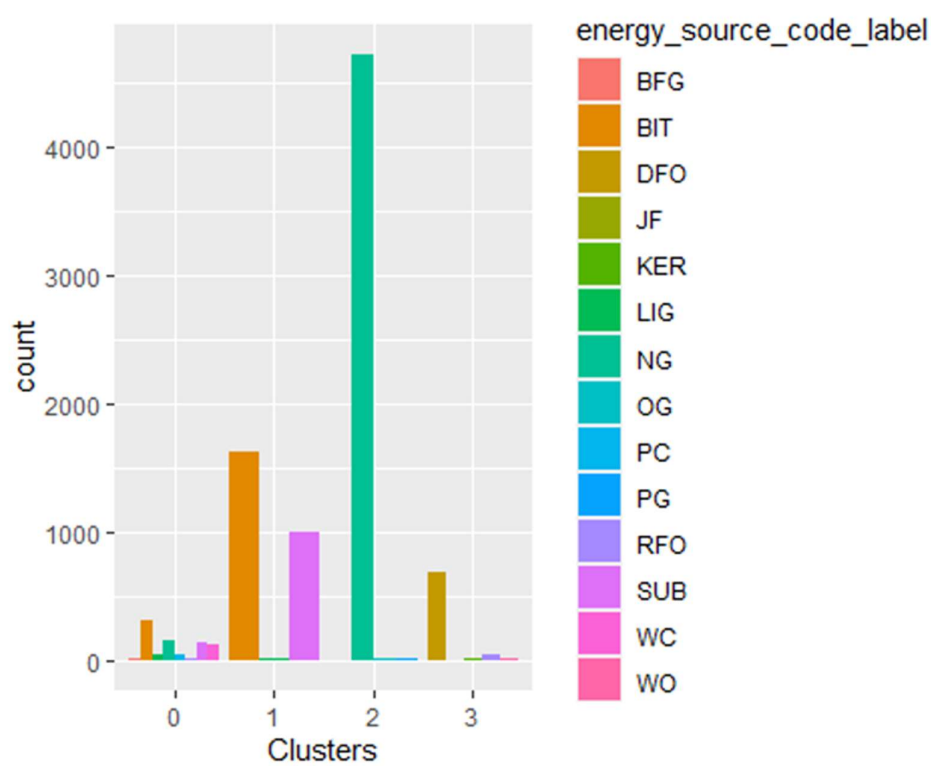
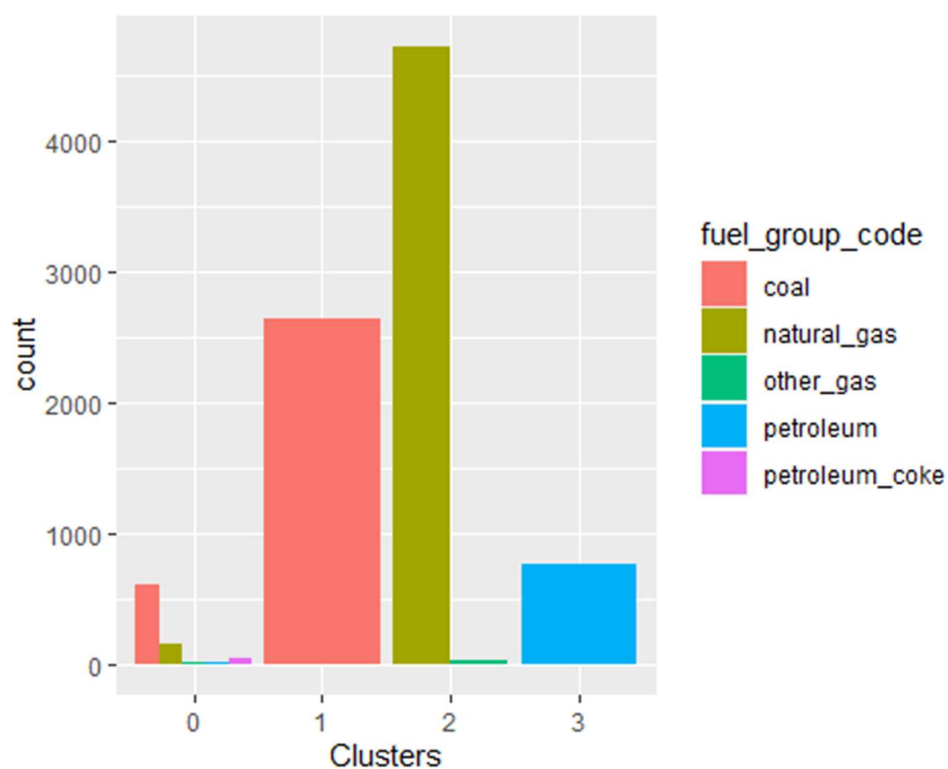
Therefore, it is important to protect coal for future generations. In order to protect fossil fuels, I would want to suggest that the government use more effective technology like carbon capture, utilization, and storage.

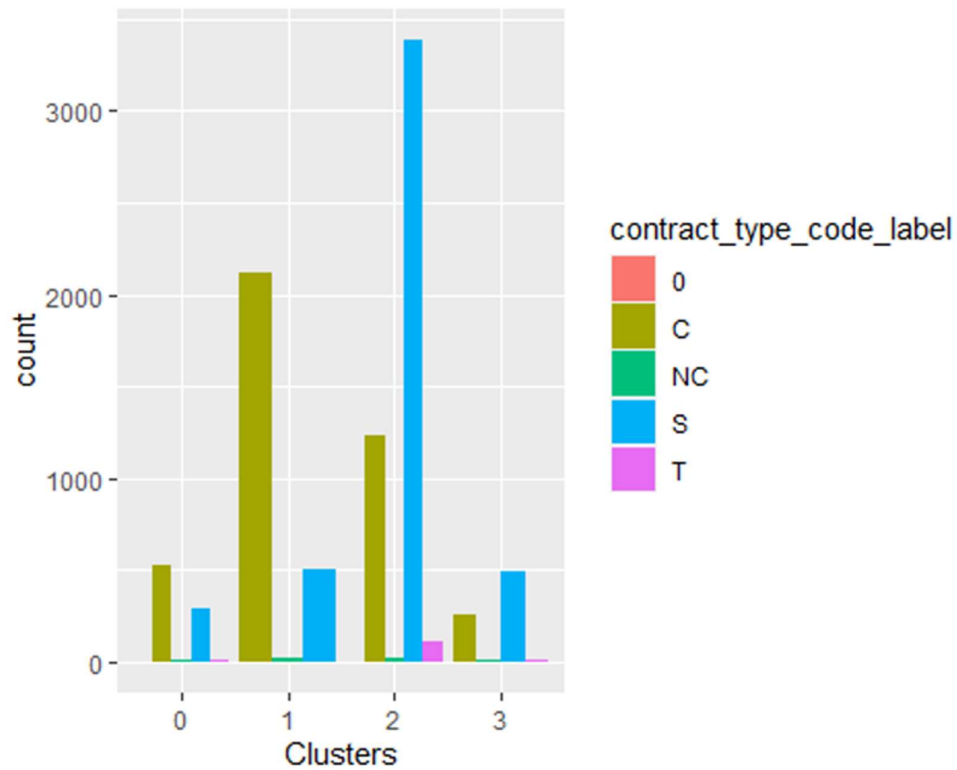


optimal number of clusters









Conclusion:

According to the clusters created, The fuel that must be excluded from power production is coal since there is an urgent need to reduce coal consumption and protect it for future generations. Additionally, compared to Gas and Oil, Coal expenditures are much too costly. As a result, rather than excluding the gasoline on which they are spending the least, I advise taking into account the fuel on which they are spending the most, especially when that fuel isn't even the best kind.

As a result, the government can employ Oil, which is available in its purest form and can be used for a variety of reasons, instead of spending a lot of money on coal.