PROJECT REPORT

Executive summary-

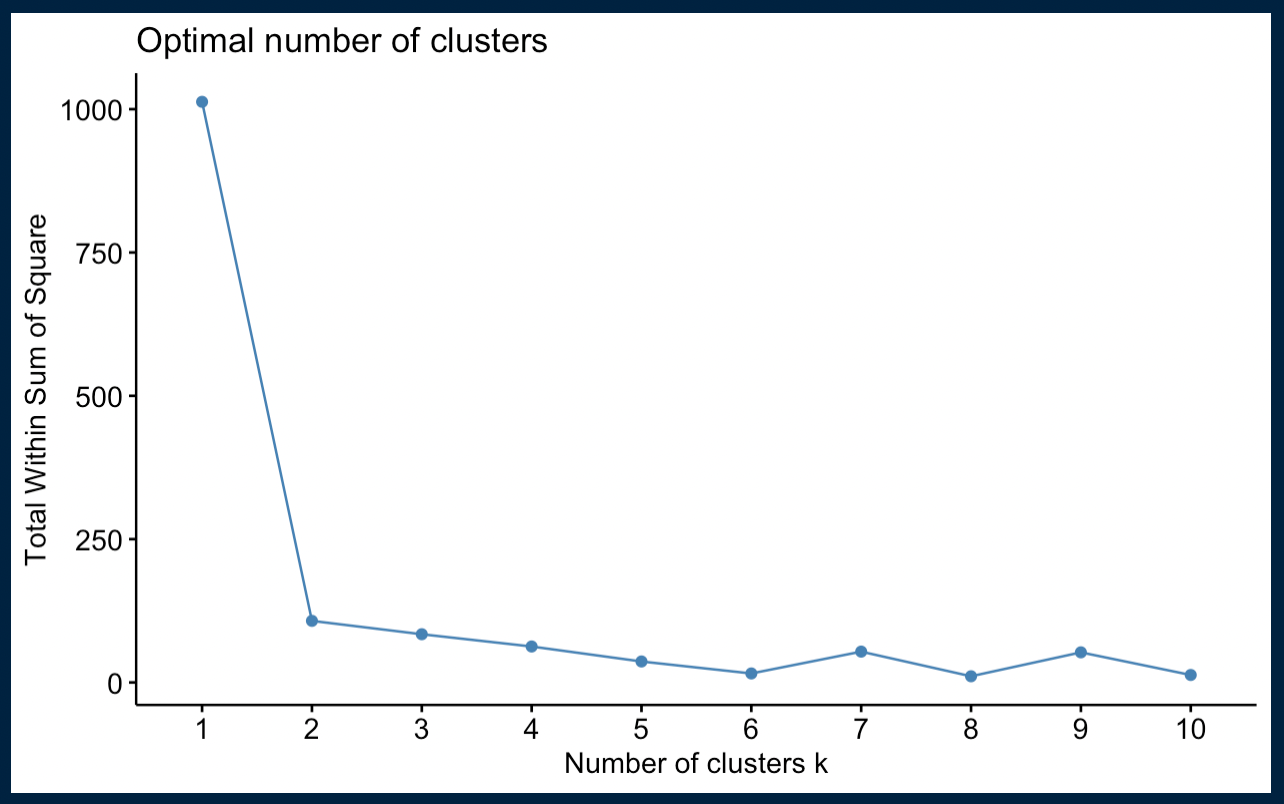
The main aim of the project was to find out which natural resource is demanded the most in the United States in order to generate electricity. Also, to see how cost impacts the viability of different types of generation. After applying K means clustering algorithm, 2 clusters are formed. The first cluster consists of coal and petroleum coke while the second cluster contains coal, natural gas, other gas and petroleum. As per the clusters formed, cluster 1 has the highest MMBtu which means that these resources can provide the automobile industry the highest energy providing natural resource. The quality which would be highly considered by the population in USA would definitely prefer on going for the best quality with the reasonable price i.e. coal. Therefore, coal would be the most in demand when compared to others.

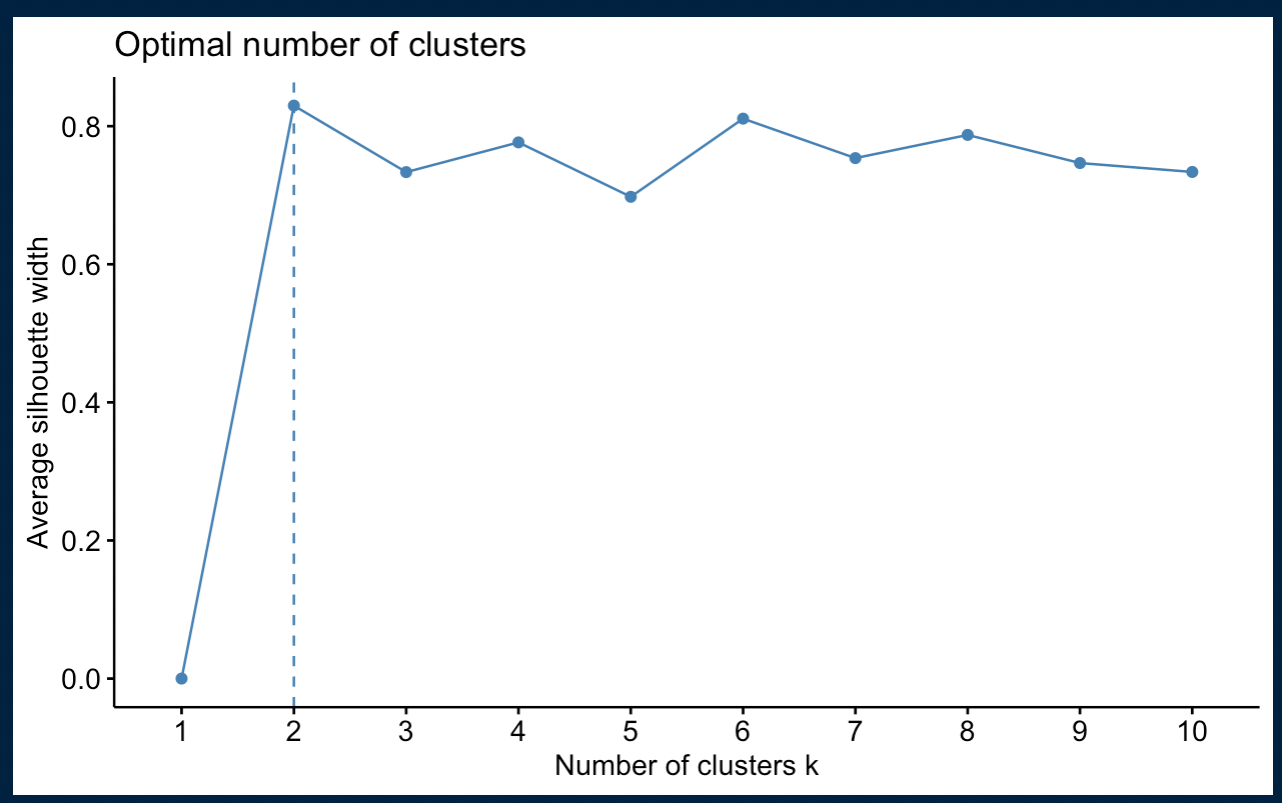
Problem-

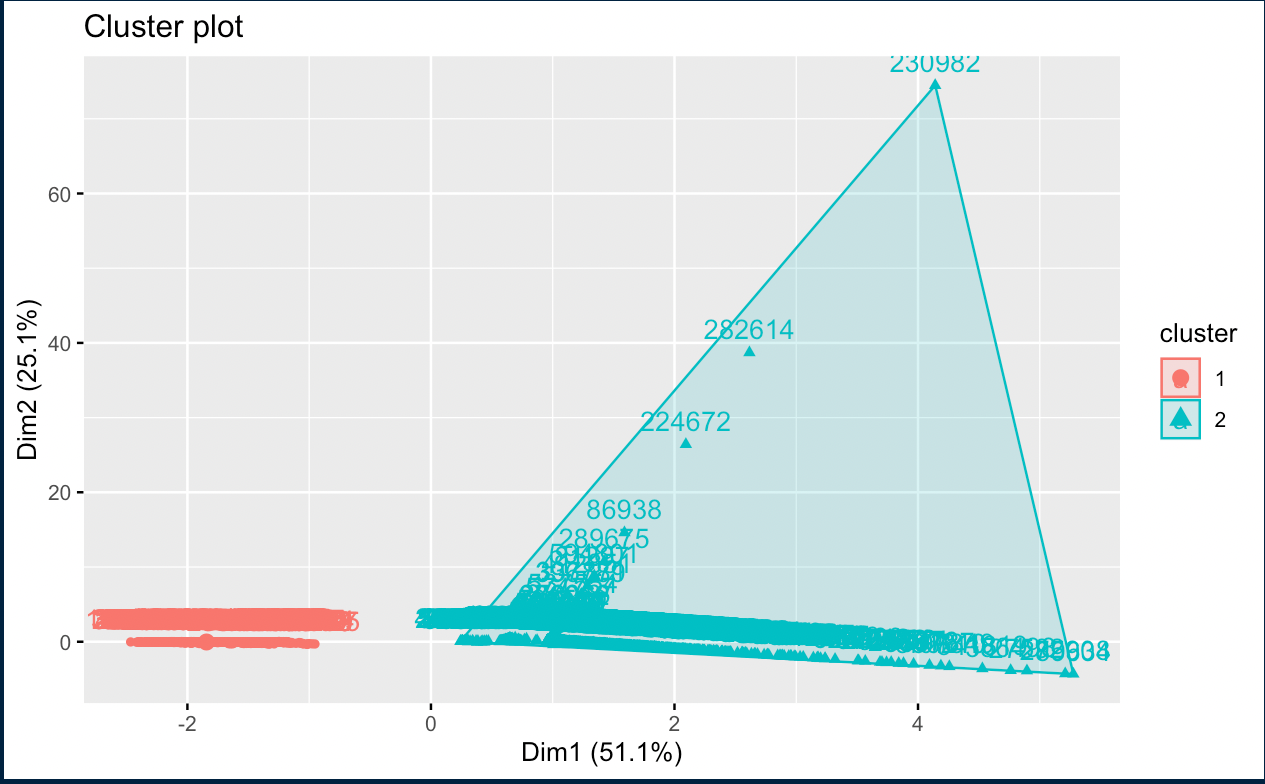
In the United States, a range of resources are used to produce electricity. Natural gas, coal, and nuclear energy are the three most popular. About 4.11 trillion kilowatt-hours (kWh) of energy were produced by utility-scale electricity producing facilities in the US in 2021, which is approximately 4,108 billion kWh. Coal, natural gas, petroleum, and other gases made up about 61% of the fuel used to generate this power. The focus of the project was to find out which amongst these natural resources is demanded the most in the USA in order to generate power. Also, to see how cost and availability impact it.

Technique-

For the data to stay unique, about 2% of the whole data was randomly sampled using a random 4-digit number i.e. 1234 as set seed. Of the 2% data, 75% of it was taken as training set and the rest 25% was taken as test set. The data normalization technique that was used was Z-score as it compares observations between dissimilar variables and identifies outliers. K means Clustering algorithm was performed, both Silhouette and Wss analysis was used. As it is relatively simple to implement and scales to large data sets. It also, generalizes to clusters of different shapes and sizes, such as elliptical clusters.

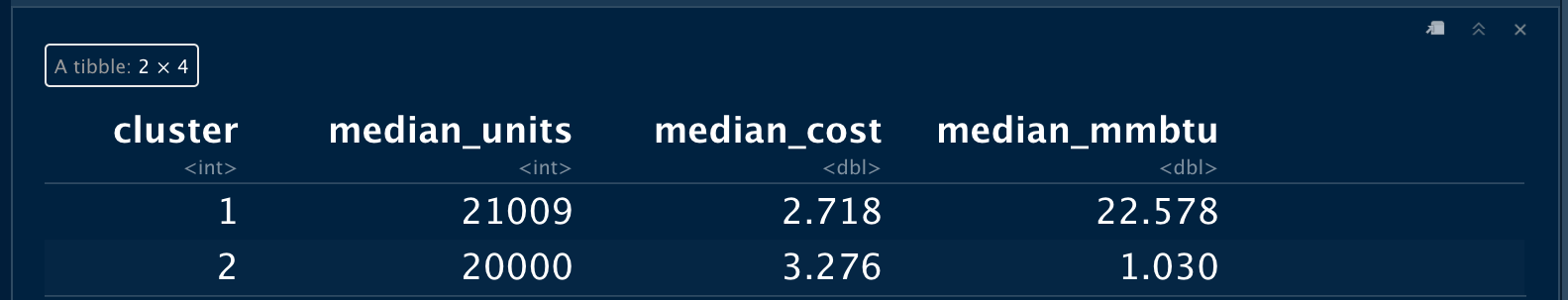


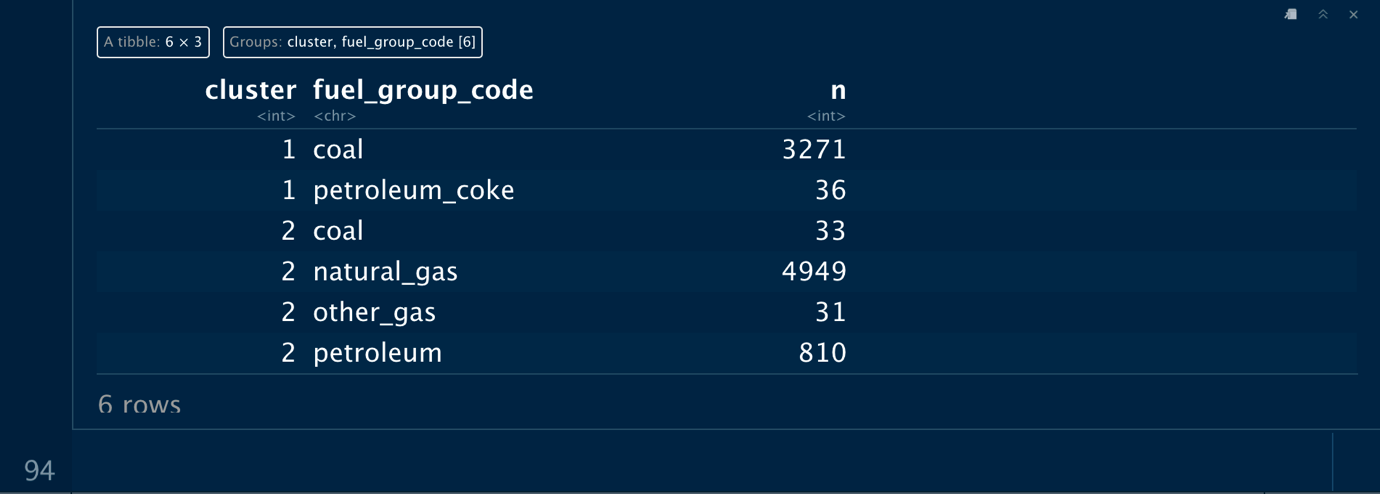




Conclusion-

* As per the clusters formed, cluster 1 has the highest MMBtu which means that these resources can provide the automobile industry the highest energy providing natural resource.
* The quality which would be highly considered by the population in USA would prefer on going for the best quality with the reasonable price i.e., coal.
* Coal is still the second-largest primary energy source in the world (by consumption), and the U.S. has the richest base of coal reserves due to its richness and accessibility to markets.
* No other nation in the world has bigger coal reserves than the United States. In actuality, the United States is home to one-fourth of all known coal reserves in the globe. In comparison to the rest of the world's oil reserves, the United States possesses more coal that can be mined. It is utilized in power plants to generate more than half of the nation's electricity.
* Therefore, coal would be the most in demand when compared to others.





References-

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

---> <https://www.eia.gov/energyexplained/coal/use-of-coal.php>