Clustering and Forecasting Population Density & Clean Fuel

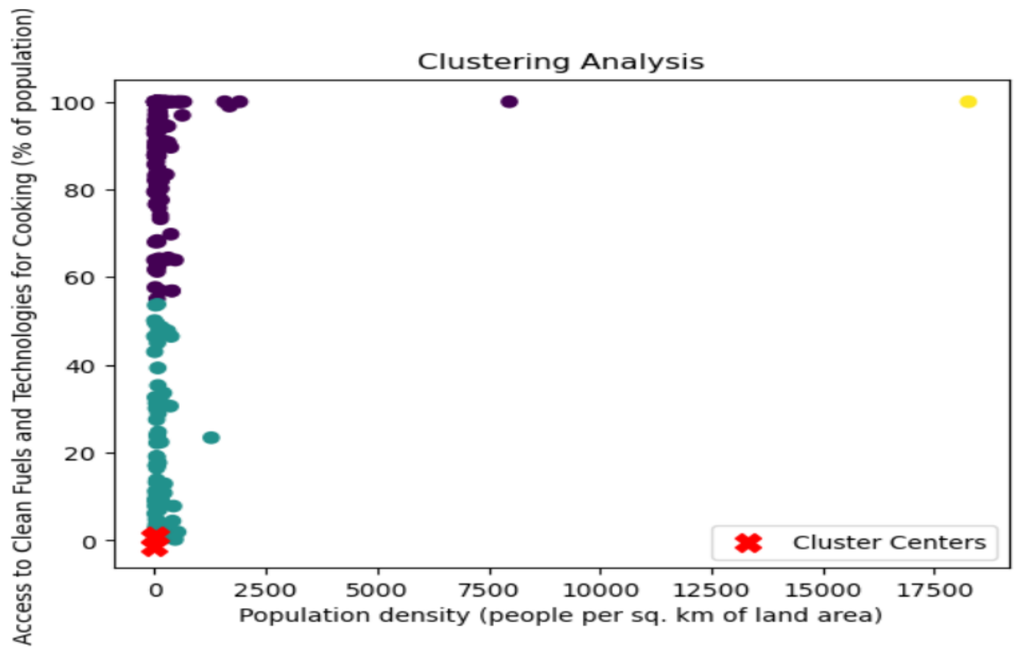
Introduction:

The analysis involved world bank data concerning Population density data and Clean fuels data for all the countries in the world. The objective of this study is to investigate patterns and trends in this dataset through clustering as well as fitting the data to a linear model to help predict its future values.

Data Clustering:

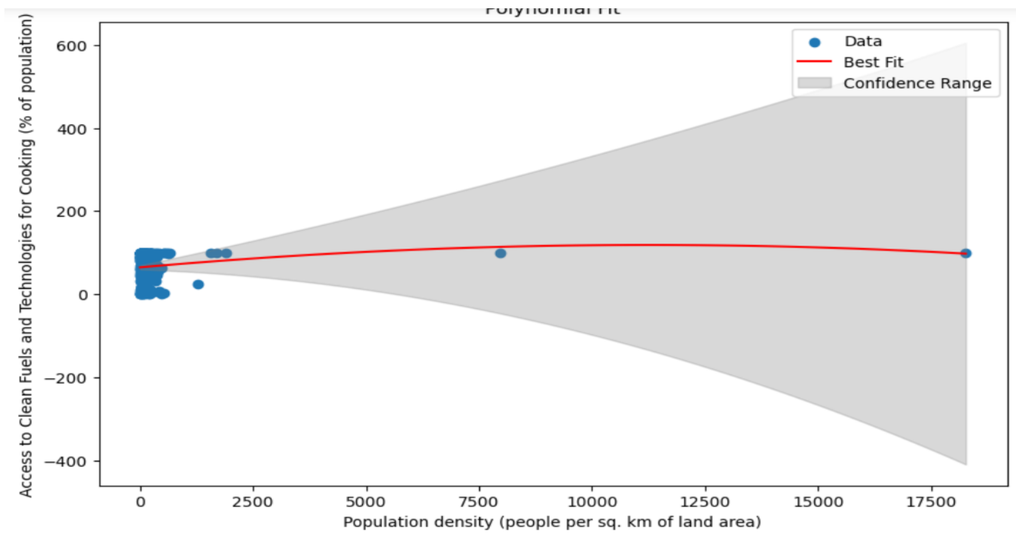
Data clustering deals with fitting the data set into different clusters each with various characteristics and it has the following steps:

I. Determining the optimal number of clusters using K means



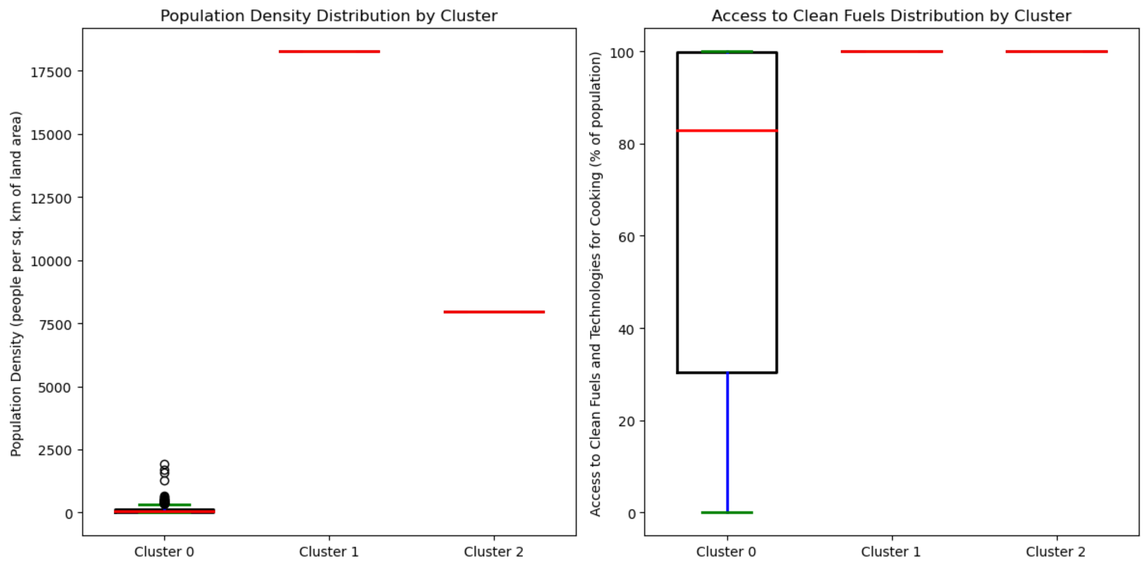
- K – means is a clever method that helps recognize the number of clusters available in a data set.
- From the above example, it is derived that there are three data clusters which are going to be used for this analysis.

II. Plotting the clusters and their cluster centers



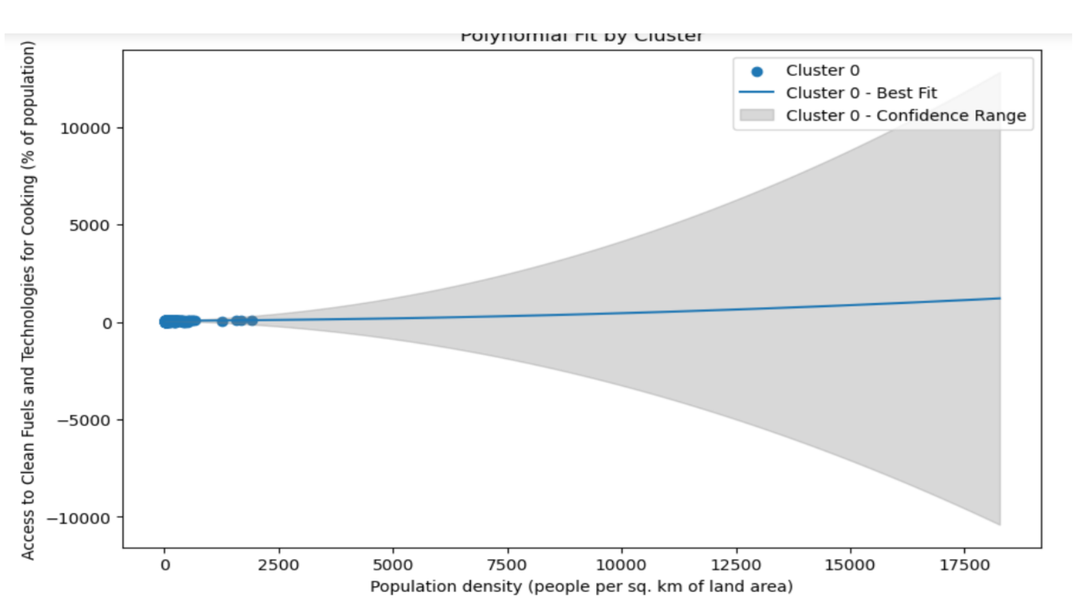
- Based on the provided variables, "Population density (people per sq. km of land area)" and "Access to Clean Fuels and Technologies for Cooking (% of population)," we can explore the relationship between population density and access to clean fuels.
- The plot illustrates the scatter plot of the data points, the best-fitting polynomial function (red line), and the confidence range (gray shaded area).

Cross Comparison Between Different Clusters:



- The Python code classifies both the density distribution by cluster and access to fuel distribution into three distinct clusters both with different characteristics.
- The clusters for population density for people per square kilometer, the first cluster has an average of 0 population, 20,000 population, and the third has a population of 8,000 people.
- For clusters containing access to clean fuel, there are three clusters, the first with 80 percent of the people, the second with about 95 percent, and the third with 100 percent.

Data Modelling:



- The polynomial fit suggests that there is a non-linear relationship between population density and access to clean fuels. As population density increases, the percentage of the population with access to clean fuels initially rises rapidly and then levels off. This pattern suggests that as population density increases, there is an initial effort to provide clean fuels and technologies for cooking to a larger portion of the population. However, beyond a certain threshold, the rate of improvement in clean fuel access slows down.
- The confidence range represents the uncertainty in the predictions. It provides a range within which future values are likely to fall given the fitted model and the associated uncertainties.
- Based on the polynomial fit, we can make predictions for future values of access to clean fuels based on different levels of population density. However, it's important to note that as we move further away from the observed data, the uncertainty in the predictions increases.

Conclusion:

This analysis highlights the importance of considering population density when addressing access to clean fuels and technologies for cooking. It suggests that policies and interventions should take into account the specific challenges associated with population density and tailor solutions accordingly. Additionally, the confidence range serves as a reminder of the uncertainty inherent in making predictions and underscores the need for ongoing monitoring and evaluation to refine our understanding of the relationship between population density and access to clean fuels.