Clustering and Fitting

**Name:** Manikanta Sasidhar Reddy **Student ID:**

**GitHub Link:**

**Introduction:**

In this report we going to see a comprehensive approach of understanding Clustering and fitting. This report examines the carbon dioxide (CO2) emissions that result from the burning of liquid and solid fuels, which are significant factors in the overall levels of greenhouse gases worldwide. By employing meticulous visual analytics, we examine the distributions, correlations, and clustering of emissions in order to extract valuable insights about emission trends. Moreover, a logistic model provides predictions of future patterns, providing a strategic perspective for the establishment of environmental policies and initiatives to address climate change.

**Data Description:**

The analysis we conducted is based on three CSV datasets that provide detailed information on CO2 emissions. One dataset focuses on emissions resulting from the use of liquid fuel, another dataset focuses on emissions resulting from the consumption of solid fuel, and the third dataset captures the total CO2 emissions from all sources. These datasets offer a detailed examination of the many types of fuels that contribute to emissions and enable a thorough evaluation of their environmental effects.   
  
The data underwent thorough cleaning and preprocessing, which included excluding non-essential metadata rows, addressing missing values, and consolidating datasets by country to achieve a coherent perspective. We prioritised the most recent full year of data, standardised the variables to ensure consistency in analysis, and utilised logarithmic transformations as needed to normalise the distribution. These measures were taken to ensure that our methodologies and findings are based on dependable data.

**Data Analysis:**

The dataset's structure, top and bottom rows, size, and form are examined first. It summarises 2015 CO2 emissions from liquid fuel, solid fuel, and total emissions across countries.

Data tails and symmetry are assessed by computing numerical column kurtosis and skewness. Kurtosis measures distribution tail heaviness, which affects outliers. Skewness quantifies data asymmetry, exhibiting biases towards higher or lower values.

|  |  |
| --- | --- |
| Fig: Box Plot  The significant scatter of data points indicates a high degree of variability in the correlation between the two variables. Several data points are located significantly distant from the central cluster, suggesting that certain countries exhibit an imbalanced amount of emissions originating from liquid fuel as opposed to solid fuel | Fig: Scatter Plot  The elongated whiskers and the presence of data points above the whiskers signify the existence of outliers, which are countries exhibiting notably elevated levels of CO2 emissions. The median values, represented by the lines inside the boxes, differ significantly among the three categories, with total CO2 emissions having the highest median value. |
| Fig: Correlation Heatmap  The heatmap illustrates a strong correlation between the emissions of CO2 from total and liquid fuel, suggesting that liquid fuel has a substantial effect on overall emissions, whilst the effects of solid fuel is very minor. | Fig: Histogram  The histogram depicting CO2 emissions from solid fuel indicates a distribution that is skewed, with the majority of countries exhibiting low emissions and a tiny subset of countries displaying exceptionally high emissions. This reflects a notable level of variability among countries. |