The World Development Indicators from the World Bank contains about 800 annual indicators of economic development from over 190 countries across 5 decades. These indicators are spread across economy, energy consumption, birth/death rate, literacy rate, pollutants, trade, logistics and several others.

Machine Learning Algorithms used - k-means clustering, OLS regression, Panel Data Regression(Fixed Effects Model)

Pre-processing of Data:

1. Missing value treatment - MICE function(CART method) to impute missing values

2. Bivariate Analysis - Variables which were highly correlated were removed from model to avoid multicollinearity

Method:

1. k-means clustering to classify countries as developed and developing

2. Panel Data Regression to study impact of renewable sources on CO2 emissions

3. Exploratory analysis to determine CO2 emissions by type of fuel and type of sector

4. AIC regression to identify most important factors affecting CO2 emissions across 5 different decades

5. OLS regression to identify most important factors affecting CO2 emissions in developed and developing countries