Objective:

To create a prediction model from global sustainable data, consisting of data related to different sources of electricity and other factors related to energy consumptions and access. The aim is to find which country has what percentage contribution in total CO2 emissions. For a country's percentage contribution, it can be calculated as the country's CO2 emissions in kt divided by the total global CO2 emissions.

Secondary aim is to find the features which have the most impact on global CO2 emissions.

Data Used:

Data used in this project can be downloaded from Kaggle: Global Data on Sustainable Energy (2000-2020) (Provided by Ansh Tanwar.)

NOTE: Please refer to the kaggle link for key features.

Methodology:

The project is divided into 3 stages:

Stage 1: Consists of primary data understanding and data cleaning.

Stage 2: Consists of exploratory data analysis and data transformation.

Stage 3: Consists of building data prediction models and their testing.

In Stage 3 data is divided into three parts corresponding to the group of countries on the basis of their emission behavior. Pareto analysis is used as a reference for deciding the countries in the group. Also, three models corresponding to the groups are tested both on the test data and the data of 2010.

Below are the screenshots of pareto analysis and the difference between actual and predicted values of CO2 emission percentage contribution of 2010 data.

Figure 1: Pareto Analysis

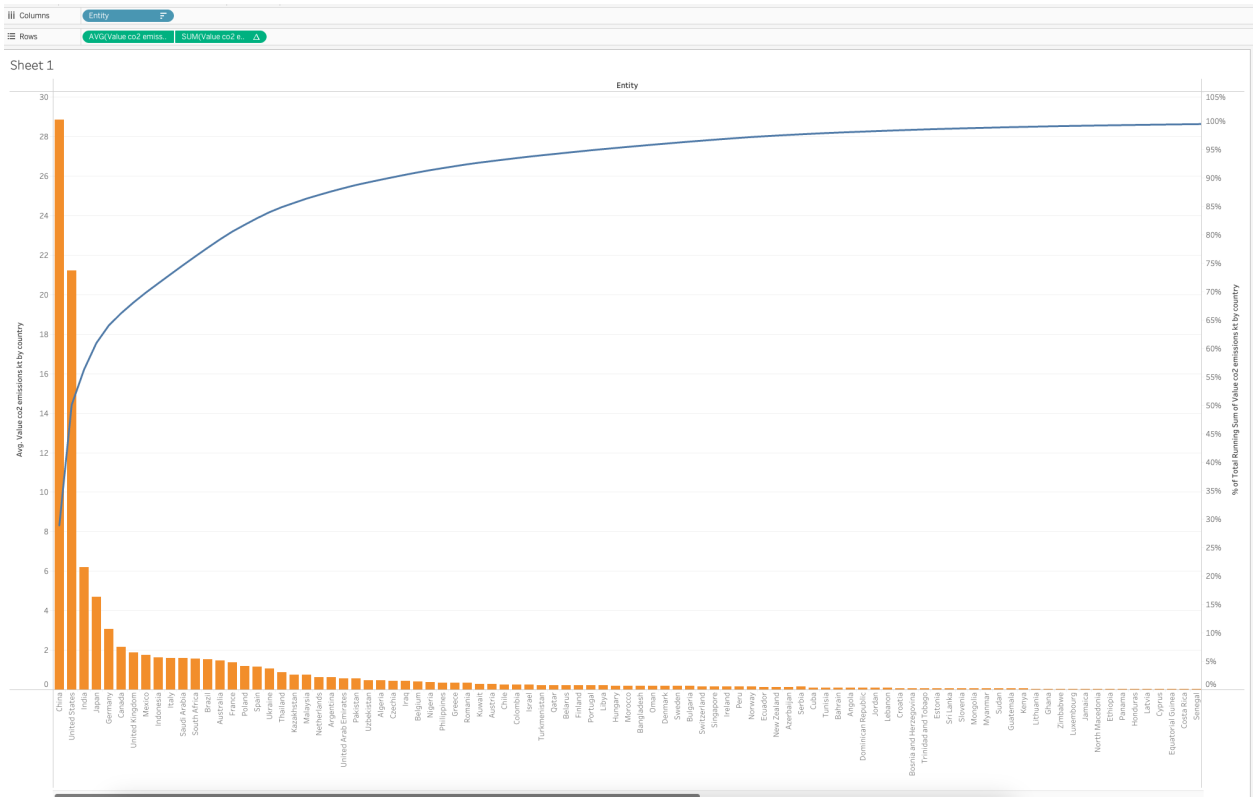
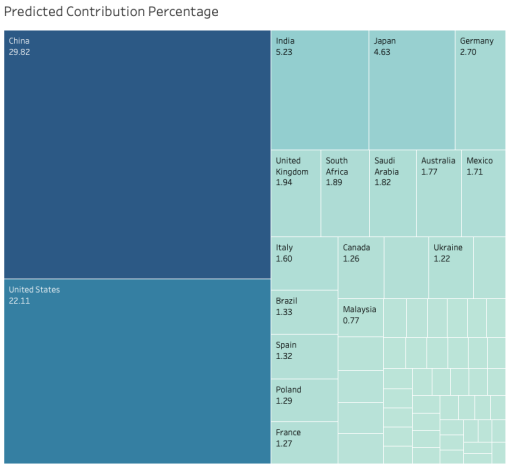


Figure 2 : Predicted vs Actual Values. (2010 Data)



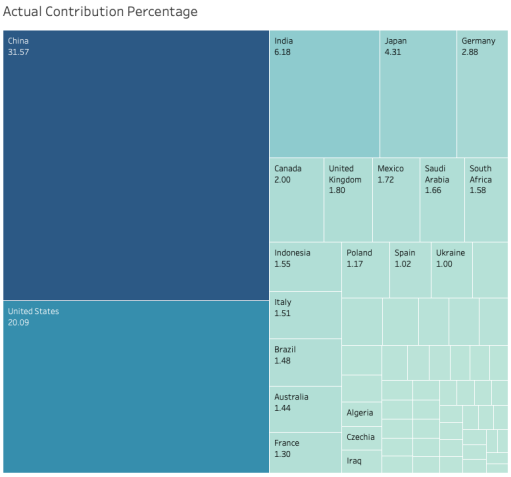


Figure 3: Predicted Vs Actual values for some selected countries. (2010 Data)