

INT 353 CA 1

NAME : MADDIRALA PRAVEEN REDDY

REG. NUMBER : 12014718

SECTION : K20CH

ROLL NUMBER : K20CHB62

DATASET: CO2 EMISSIONS DATASET

**About Dataset:**

The dataset which I have chosen is that of CO2 emissions with respect to various energy sources in all countries around the world from the year 1980 to the year 2019.**CO2 Emission** refers to the Carbon Dioxide emitted throughout the world. In this dataset it is broadly categorizing all emitters into one total amount for each energy type. CO2 emissions from the burning of fossil fuels are the primary cause of global warming which happens to be one of the biggest threats facing humanity. Although there are plenty of other emissions that are emitted on this earth, including Methane, nitrous oxide, and CFCs, none compare to the emission of CO2, and we as humans are mostly to blame for this. For this analysis I will be focusing on CO2 Emissions and its effect on the world we live in as well as some key factors and stats that may play a role in the emission of CO2 globally.

**Reason for Choosing This Dataset:**

The reason behind choosing this dataset is that in this dataset the information is well sorted out and it contains some important factors to precisely measure the Carbon Footprint. In this 21st Century, the imminent threat to all the species on Earth is primarily climate change which is in fact an outcome of over exploitation of resources.

To reduce the climate change we should find the carbon dioxide emissions in different parts of the world and also perform an analysis on how different energy resources are causing carbon dioxide emissions, in doing so we can quite accurately find the relationship between the energy resources and the CO2 emissions. This will help us in focusing on the problem where it is severe.

**Dataset Structure:**

The data set consists of 10 columns which are as follows:

* Country - Country in question
* Energy\_type - Type of energy source
* Year - Year the data was recorded
* Energy\_consumption - Amount of Consumption for the specific energy source, measured (quad Btu)
* Energy\_production - Amount of Production for the specific energy source, measured (quad Btu)
* GDP - Countries GDP at purchasing power parities, measured (Billion 2015$ PPP)
* Population - Population of specific Country, measured (Mperson)
* Energy\_intensity\_per\_capita - Energy intensity is a measure of the energy inefficiency of an economy. It is calculated as units of energy per unit of capita (capita = individual person), measured (MMBtu/person)
* Energy\_intensity\_by\_GDP- Energy intensity is a measure of the energy inefficiency of an economy. It is calculated as units of energy per unit of GDP, measred (1000 Btu/2015$ GDP PPP)
* CO2\_emission - The amount of C02 emitted, measured (MMtonnes CO2)

**DOMAIN:**

This data set deals with the domain of Environmental studies emphasizing mainly on the consequences of climate change.Environmental studies include various concepts that affect us in daily lives. This data set mainly focuses on the concept of carbon dioxide emissions with respect to the energy Consumption around the world.

**Tasks To Perform On The Dataset:**

* 1. First and foremost, The data set should be cleaned by removing unnecessary values or columns.
  2. I need to find if there are any null Values in the Dataset And try to replace those null values.
  3. I need to find outliers, if there are any in the Dataset.
  4. After cleaning the data set, I need to find if there is any relationship between any of the columns and interpret the relationship.

**Reference:**

This dataset is extracted from Kaggle.