Hertfordshire University

Applied Data Science

Assignment 01

Wijekoon Mudiyanselage Bhagya Wijekoon
22031471

Topic: Find the relationship between Electricity generation emission and air pollution for past 3 decades.

Considering industry revolution, Electricity is major requirement of mankind. I have analyzed the data of Electricity generated sources from 1970 – 2021. Following line plot (Figure 1) is showing each power sources used to generate to electricity and how much power (GWh) generated for the time series.
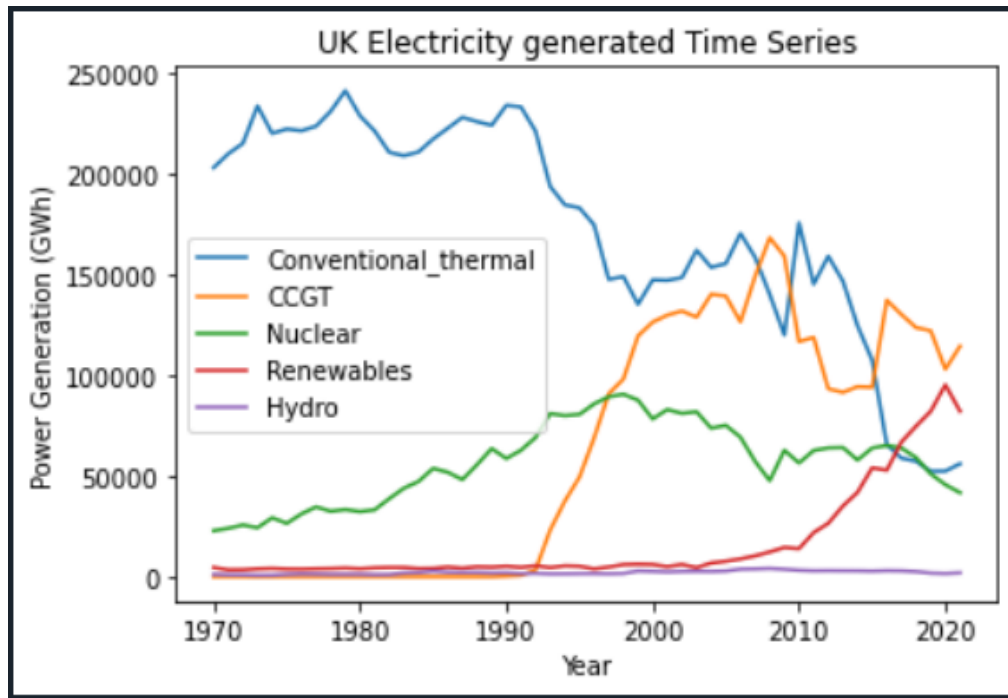

Figure 1: UK Electricity generated Time Series 1970-2021

Data source [Digest of UK Energy Statistics (DUKES): electricity - GOV.UK (www.gov.uk)](#)
Excel File :
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1094474/DUKES_5.1.3.xlsx

The majority of the UK's electricity is produced from fossil fuels, such as coal and natural gas. The UK has made efforts to reduce the amount of pollution generated by electricity production through the use of cleaner energy sources, such as wind, solar, and nuclear power, as well as implementing emissions controls on fossil fuel power plants.

List of power sources used to generate electricity.

- Conventional thermal power – Convert heat energy into electrical energy using steam turbine.
- Combined cycle gas turbine power (CCGT) – Convert heat energy to mechanical energy with help of combined gas. Improved efficiency and low fuel cost to generate electricity.
- Nuclear power – Convert nuclear reactor thermal energy to electrical energy.
- Renewables power - These energy sources include wind power, solar power, bioenergy.
- Hydro power - Uses the natural flow of moving water to generate electricity

From early 1970s we can see conventional thermal energy sources had been used for power generation. Hydro power one of the oldest renewable power was there from 1970 but not much contribution by power generation. It was continued until 2010 and then supply was reduced.

Nuclear power generation also had significant improvement for past 3 decades. CCGT power helped to reduce the requirement of conventional thermal energy. But start with renewable energy other power source demand has significantly dropped. With year 2000 conventional power generation dropped by 80% and CCGT, Nuclear power generation also running at average level. We can observe that renewable energy has done significant contribution to electricity generation.

When we plot pie chart of Year 2020 Electricity generation (Figure 2) we can see that 32% renewable energy resources used for UK electricity generation. If renewable energy generation increased it is able to shutdown most of the thermal power plants which are using Gas and coal. I have found news article in the internet which confirm 2020 highest renewable energy source recorded so far, which is nearly close. Here is the link to news article : https://www.pv-magazine.com/2021/07/29/uk-power-generation-in-2020-renewables-up-to-43/#:~:text=Power%20production%20from%20renewable%20sources%20again%20provided%20record,solar%2C%20and%20biogas%2C%20up%20from%2037%25%20in%202019
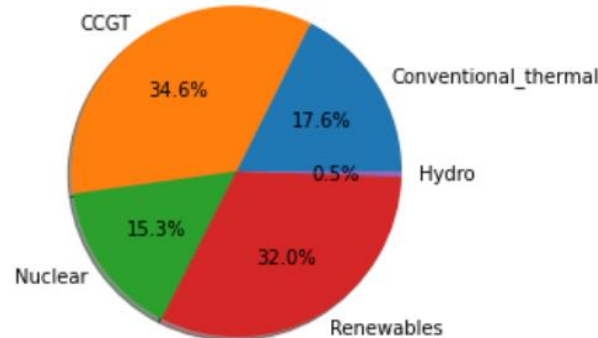

UK Electricty generating sources for the year 2020

Figure 2:UK Year 2020 Electricity generating sources.

Data source Digest of UK Energy Statistics (DUKES): electricity - GOV.UK (www.gov.uk)
Excel File :
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1094474/DUKES_5.1.3.xlsx

There are many environment effects happened around the world. One major impact is air pollution. It can be measured with Air quality monitor and units are calculated as Air Quality Index. Air pollution from electricity generation remains a concern, particularly in urban areas with high levels of electricity demand. The UK government has set targets to reduce air pollution and improve air quality, including

reducing emissions from electricity production. I have gathered air quality statistics and plot against time series. Airborne particles are called aerosols. Particulate matter (PM) includes microscopic matter suspended in air or water. PM10 includes particles less than 10 μm in diameter, PM2. 5 those less than 2.5 μm. The toxicity of suspended particles is mainly due to particles with a diameter of less than 10μm.
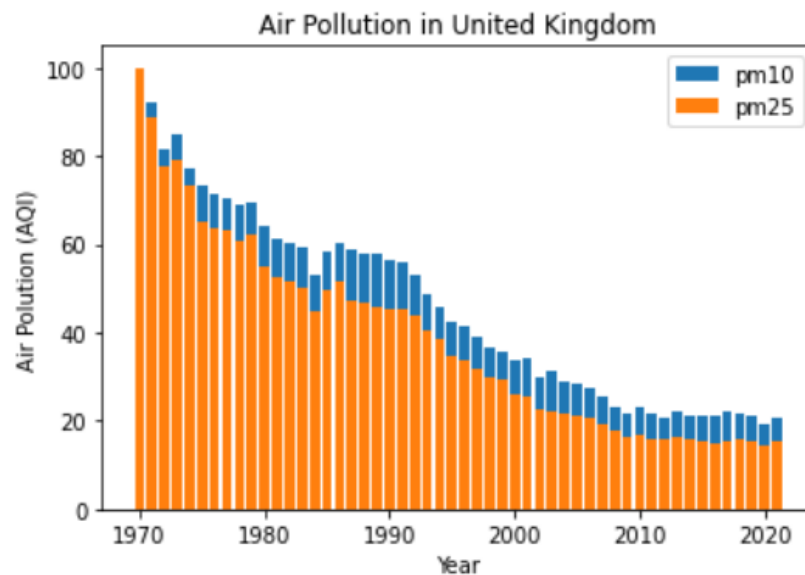

Figure 3: Air Pollution in UK

Data Source https://www.gov.uk/government/statistics/air-quality-statistics

CSV File:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1135477/Figure01.csv

With the above figure we can clearly observer Air pollution indicators PM10 and PM2 5 has been dropping with the time. Air quality indicators are showing nearly 25% for both PM10 and PM2 5.

Conclusion

In the analysis we can observe that air pollution and energy generation emission have direct correlation. But in order to verify we need to gather more data relevant to polluted air emission from vehicles, factories, forest fire and other varies human related activities.

Python Program code link

https://github.com/Bhagyaw/ads_01/blob/bc00d0854ab08fca74da17a9a2ac263f9ac27a83/ADS-AS01-Final.py

Program code

```python
# -*- coding: utf-8 -*-
"""
Created on Sat Mar  4 08:16:00 2023

@author: bhagy
"""


#Import required modules
import pandas as pd
import matplotlib.pyplot as plt


def loaddfxy(df, headers):
    """ Function to load data to Pandas dataframe with selected columns.
        Arguments:
        A dataframe with a index "head" and other columns to be taken selected headers.
        A list containing the selected headers of the columns to create datadrame.
        For loop is used to fetch and assign data to dataframe(data_xy)
        After data is inserted into dataframe it will be return from the function.
    """
    #For loop to load required headers to DataFrame
    for head in headers:
        data_xy[head] = pd.DataFrame(data[head], columns=[head])


    return  # Function return the DataFrame data_xy


def loaddfjk(df, headers):
    """ Function to load data to Pandas dataframe with selected columns.
        Arguments:
```

A dataframe with a index "head" and other columns to be taken selected headers.

A list containing the selected headers of the columns to create datadrame.

For loop is used to fetch and assign data to dataframe(data_xy)

After data is inserted into dataframe it will be return from the function.

"""

#For loop to load required headers to DataFrame

for head in headers:

    data_jk[head] = pd.DataFrame(data_air[head], columns=[head])

return # Function return the DataFrame data_xy

def lineplot_xy(df, headers):

   """ Function to create a lineplot.

    Arguments:

    A dataframe with a column "Year" and other columns to be taken as headers.

    A list containing the headers of the columns to plot.

    For loop will be used to plot and matplotlib.pyplot used.

   """

   #Start to plot

   plt.figure()

   #For loop to load required data using headers.

   for head in headers:

      plt.plot(df[['Year']], df[head], label=head)

   #labelling x axis and y axis

   plt.xlabel('Year')

   plt.ylabel('Power Generation (GWh)')

   #Add legend to explain the each lineplot

```python
    plt.legend()

    #Add Title to the graph

    plt.title("UK Electricity generated Time Series")

    # save the plot output as png

    plt.savefig("linplot-UK Electricity generated Time Series.png")

    #Display the plot

    plt.show()


    return  #Function return the lineplot and save the image


def barplot_jk(df, headers):
    """ Function to create a barplot.

        Arguments:

        A dataframe with a column "Year" and other columns to be taken as headers.

        A list containing the headers of the columns to plot.

        For loop will be used to plot and matplotlib.pyplot used.
    """

    #Start to plot

    plt.figure()


    plt.bar(df["year"], df["pm10"], label="pm10")

    plt.bar(df["year"], df["pm25"], label="pm25")


    #labelling x axis and y axis

    plt.xlabel('Year')

    plt.ylabel('Air Polution (AQI)')


    #Add legend to explain the each lineplot
```

```python
    plt.legend()

    #Add Title to the graph

    plt.title("Air Pollution in United Kingdom")

    # save the plot output as png

    plt.savefig("barplot-UK Air Pollution.png")

    #Display the plot

    plt.show()


    return  #Function return the lineplot and save the image


def pieplot(df, headers, RYear):
    """ Function to create a pieplot.

        Arguments:

        A signle dataframe with selected column and other columns to be taken as headers.

        The values are only containing for the selected headers of the columns to plot.

        Static passed data as argument will be used to plot the pie chart using matplotlib.pyplot.
    """

    #Start to plot

    plt.figure()


    plt.pie(df, labels=headers,  shadow = True, autopct='%1.1f%%')


    #Add Title to the graph

    plt.title("UK Electricty generating sources for the year {}".format(RYear))


    # save as png

    plt.savefig("pieplot-Electricty generating sources for the year {}.png".format(RYear))

    #Display the plot

    plt.show()
```

```python
    return  #unction return the pieplot and save the image



#Start of main program

#Reading data from the source file. Here source file type is xlsx.

#Therefore pandas reading funtion must be pd.read_excel

data = pd.read_excel("power_sources.xlsx")


#Define required headers to fetch data from required columns.

headers = ["Conventional_thermal", "CCGT", "Nuclear", "Renewables", "Hydro"]


#Initialize pandas dataframe variable to load data. Here Year is used  for indexing.

data_xy = pd.DataFrame(data["Year"], columns=["Year"])


#Pass the data fetch from Convert data to pandas dataframe

loaddfxy(data, headers)


#Plot1


#Calling lineplot with a list of the columns to be plotted.

lineplot_xy(data_xy, headers)


#Plot2


#Calling pieplot to display Energy power source contribution for year 2021

RequestYear = data_xy.iloc[50, 0]

RequestData = data_xy.iloc[50, 1:6]

pieplot(RequestData, headers, RequestYear)
```

```
#Plot3


#Reading data from the source file. Here source file type is csv.

#Therefore pandas reading funtion must be pd.read_csv

data_air = pd.read_csv("air_pollution.csv")



#Define required headers to fetch data from required columns.

headersjk = ["nh3", "nox", "voc", "pm10", "pm25", "so2"]



#Convert data to pandas dataframe

data_jk = pd.DataFrame(data_air["year"], columns=["year"])

loaddfjk(data_air, headersjk)


#Calling barplot to display Air Pollution for past 3 decades.

barplot_jk(data_jk, headersjk)
```