Carbon Emissions Forecasting

Milestone 1

Davis Lee

Problem Definition

1

Why does CO2 emission forecast matter?

Forecasting CO2 emissions can make an impact on decision-making in terms of emission reduction & choosing better methods of electricity production.

2

What is our goal?

Forecast the carbon emissions value for natural gas fuel type for the next 12 months and propose certain measures that can be adopted as policies to reduce these emissions

3

What are the questions that need answers?

1. How will the overall CO2 emission from energy production be under current policy?(under status quo)
2. What energy sources are the major contributors of CO2 emission?

4

How will data science help us?

Time series models in data science will allow us to accurately predict future CO2 emission by each energy source which will then provide insights that will help us to come up with data-driven policies and measures

Data Exploration

Original Dataset

	MSN	YYYYMM	Value	Description
0	CLEIEUS	197301	72.076	Coal Electric Power Sector CO2 Emissions
1	CLEIEUS	197302	64.442	Coal Electric Power Sector CO2 Emissions
2	CLEIEUS	197303	64.084	Coal Electric Power Sector CO2 Emissions
3	CLEIEUS	197304	60.842	Coal Electric Power Sector CO2 Emissions
4	CLEIEUS	197305	61.798	Coal Electric Power Sector CO2 Emissions
5	CLEIEUS	197306	66.538	Coal Electric Power Sector CO2 Emissions
6	CLEIEUS	197307	72.626	Coal Electric Power Sector CO2 Emissions
7	CLEIEUS	197308	75.181	Coal Electric Power Sector CO2 Emissions
8	CLEIEUS	197309	68.397	Coal Electric Power Sector CO2 Emissions
9	CLEIEUS	197310	67.668	Coal Electric Power Sector CO2 Emissions
10	CLEIEUS	197311	67.021	Coal Electric Power Sector CO2 Emissions
11	CLEIEUS	197312	71.118	Coal Electric Power Sector CO2 Emissions
12	CLEIEUS	197313	811.791	Coal Electric Power Sector CO2 Emissions
13	CLEIEUS	197401	70.55	Coal Electric Power Sector CO2 Emissions
14	CLEIEUS	197402	62.929	Coal Electric Power Sector CO2 Emissions

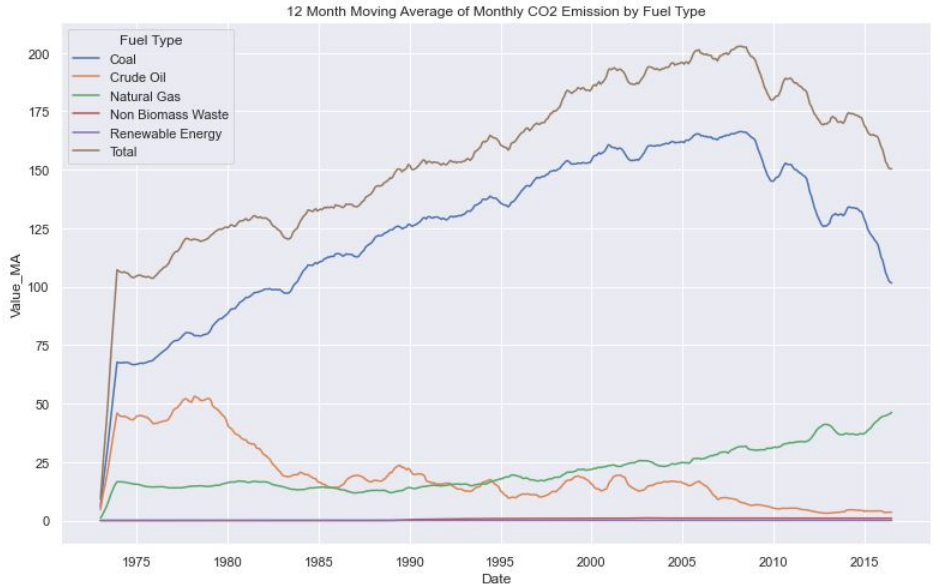
Transformed Dataset

	Date	Fuel Type	Value_num	Value_MA
0	1973-01-01	Coal	72.076	6.006333
1	1973-01-01	Crude Oil	54.739	4.561583
2	1973-01-01	Natural Gas	12.175	1.014583
3	1973-01-01	Non Biomass Waste	0.000	0.000000
4	1973-01-01	Renewable Energy	0.000	0.000000
5	1973-01-01	Total	111.621	9.301750
6	1973-02-01	Coal	64.442	11.376500
7	1973-02-01	Crude Oil	46.068	8.400583
8	1973-02-01	Natural Gas	11.708	1.990250
9	1973-02-01	Non Biomass Waste	0.000	0.000000
10	1973-02-01	Renewable Energy	0.000	0.000000
11	1973-02-01	Total	99.185	17.567167
12	1973-03-01	Coal	64.084	16.716833
13	1973-03-01	Crude Oil	42.068	11.906250
14	1973-03-01	Natural Gas	13.994	3.156417

Data Cleansing & Transformation

- "YYYYMM" column had 386 rows with month 13. These rows were removed reducing the total row count from 5093 to 4707. "Date" column was then created to leverage "datetime64" data type.
- Changed "Value" data type from "object" to "float64". The column had "Not Available" as value so there were replaced with 0.
- "Description" column is replaced by "Fuel Type" column and which was created by extracting fuel type information from "Description"
- Discarded "MSN" column as it was redundant
- Created a new column named "Value_MA" which is a 12 month moving average of the "Value" column. This was created to better observe trend without seasonality

Data Exploration



Key Observations

- CO_2 emissions from fuel sources other than natural gas and coal have become rather irrelevant since the mid 2000's.

- While CO_2 emission from natural gas has been steadily increasing, the overall CO_2 emission rate from all energy sources is decreasing since the mid 2000's.

  This is because CO_2 emission from its major contributor, coal, has been dropping more than the increased emission from natural gas. It does seem from the chart that natural gas is replacing coal as US energy source and it is contributing to the overall decrease of CO_2 emission.

- Therefore I would like to form the following hypothesis : **if we replace energy produced by coal to energy produced by natural gas, total CO_2 emission from energy production will be reduced.**

Proposed Approach

Objective

Forecast the carbon emissions value for **natural gas (NNEIEUS)** fuel type for the next 12 months and propose certain measures that can be adopted as policies to reduce these emissions.

Potential Techniques

- **AR,MA & ARMA Models**
- **ARIMA Model**

Above time-series model building techniques will be used to create natural gas forecasting model

Overall Solution Design

1. Split data into train and test data set
2. Decompose time series components into Trend, Seasonality, and Residual
3. Plot ACF/PACF on the Residual data to get p and q values for time series models
4. Fit the residual data into AR,MA,ARMA and ARIMA models and compare the results
5. Pick the best performing model and apply inverse transformation to test the model on test data set

Measure of Success

- **Model Accuracy(RMSE)**

Models will be trained using data excluding the last 2 years so the last 2 years data can be used as test dataset which we can use to measure accuracy of our model