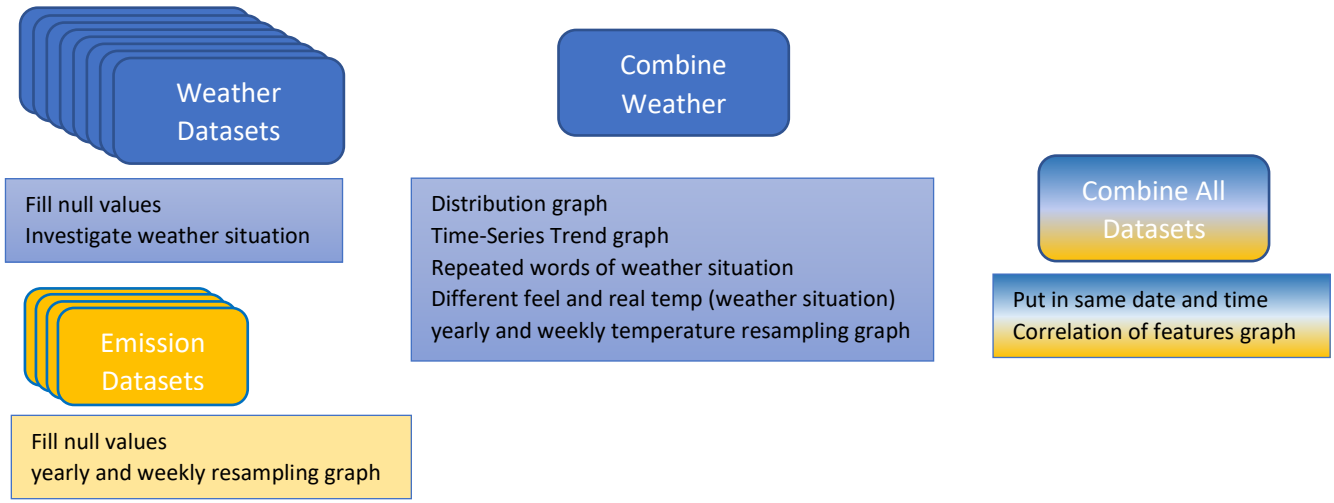
How emission of greenhouse gases can affect temperature

Abstract

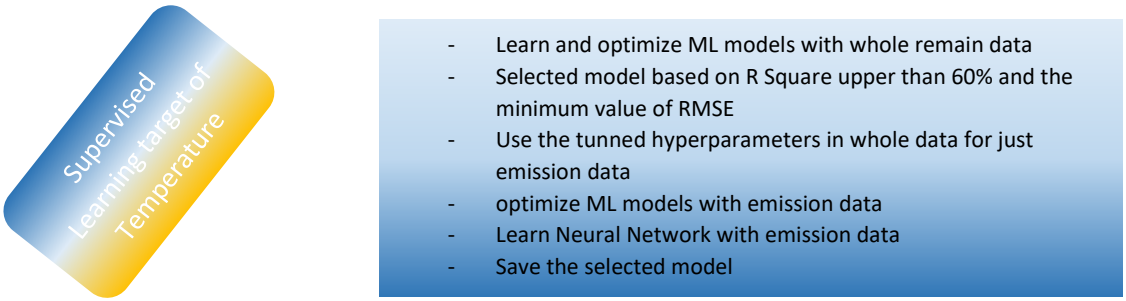
One of the most popular topics these days is global warming and what is human activity's role in that. In this article, I tried to look at the human role based on releasing greenhouse gases into the atmosphere and checking the changing temperature in the specific area in the heart of the European continent that is surrounded by three different countries and related cities, 'Lake Constance'. I use the data from the Visualcrossing website which has historical weather data and the NOAA website which has emission data of greenhouse gases.

I simply follow the below steps to

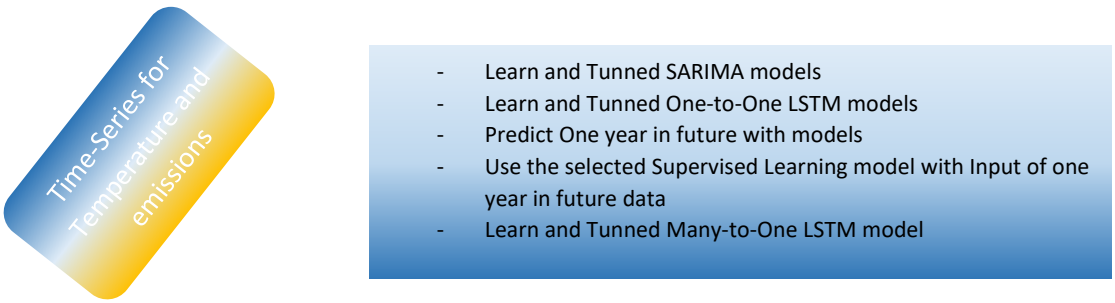
1- Understand the data I use in this research, (Exploratory Data Analysis)



2- Find the model of how the emission of greenhouse gases can affect the temperature,

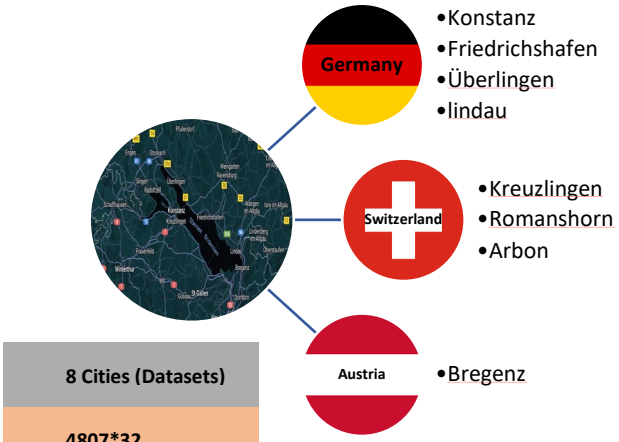


3- Investigate on what is the cause of the release of greenhouse in the same trend as temperature.

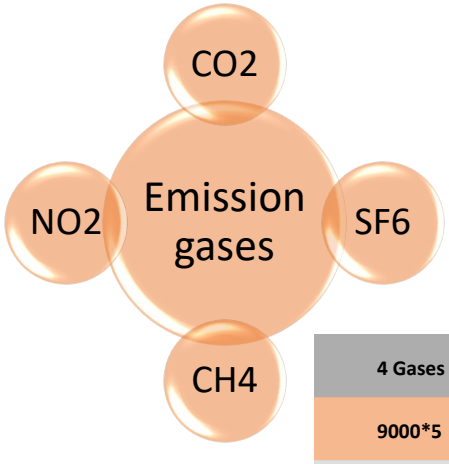


In this document, I focus on the Exploratory Data Analysis.

Exploratory Data Analysis



8 Cities (Datasets)
4807*32
3 Date and Time
3 Categorical
26 Numerical



4 Gases (Datasets)
9000*5
All Numerical

Fill Null values

Numeric data manipulation

Fill null values in time-series has different technics and it is not like fill null values in another dataset (for other datasets, we will find the distribution of feature and after that based on that we make decision).

In time-series we follow one of the below technics,

- back fill/forward fill,
- interpolation (has different methods: linear, polynomial, spline, nearest, quadratic)
- K-Nearest Neighbor Imputation (KNNImputer)
- Multivariate Imputation by Chained Equations (MICE)

Information Box

In this research, I use the measuring MSE to find the best method for filling the null values of my datasets on each feature separately. I follow the below steps to get the method,

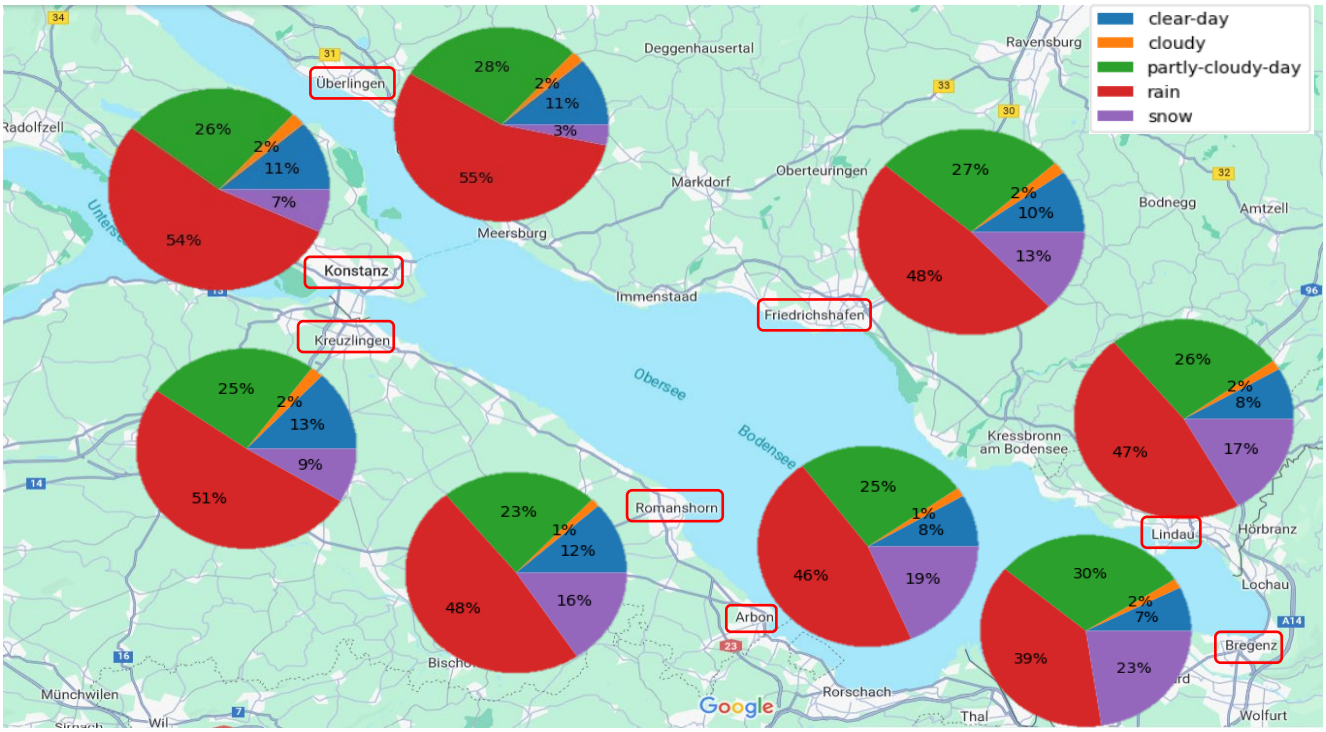
- Take 10% of data that is not null, save it in different data frames, and make them null
- Use each of the above technic and measure MSE for that 10% I saved before as validation
- Use the method in the main data and fill in null values

Most of my series are very good at filling with the interpolation linear method, but the interpolation method can not fill the first and end of series null values, so based on the data I want to use in the next steps I fill them with back/forward fill.

Object data manipulation

According to the definition of object features with null values, there is no need to fill them and after more investigation, I drop the mentioned feature.

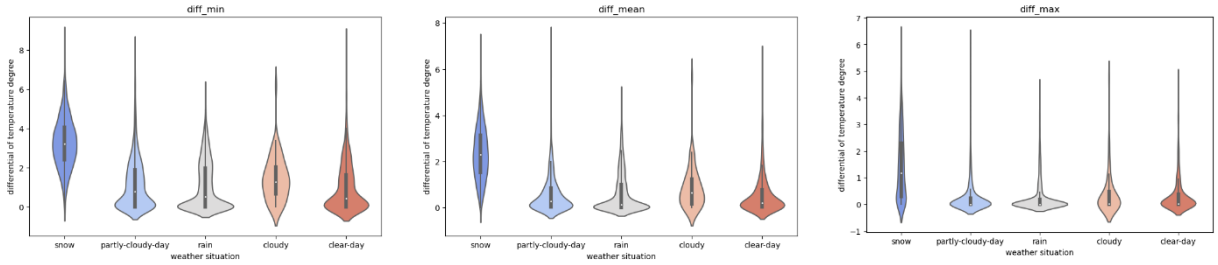
Investigate weather situation



Repeated words about the weather situation

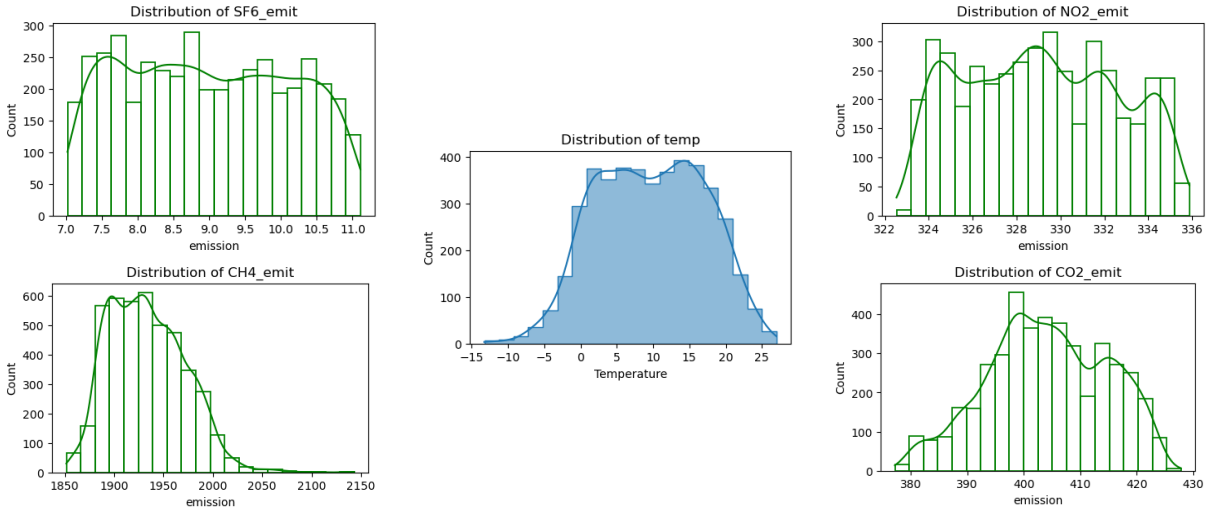
There is rarely clear condition weather exists in this area and most of the year they face clouds and cold weather

Different feel and real temp (weather situation)



- When the weather is clear, rainy, and partially cloudy there is less difference between real temperature and feeling temperature
- Snow and cloudy weather are tricky and the difference between feeling the temperature and a real one is a sense

Distribution graph

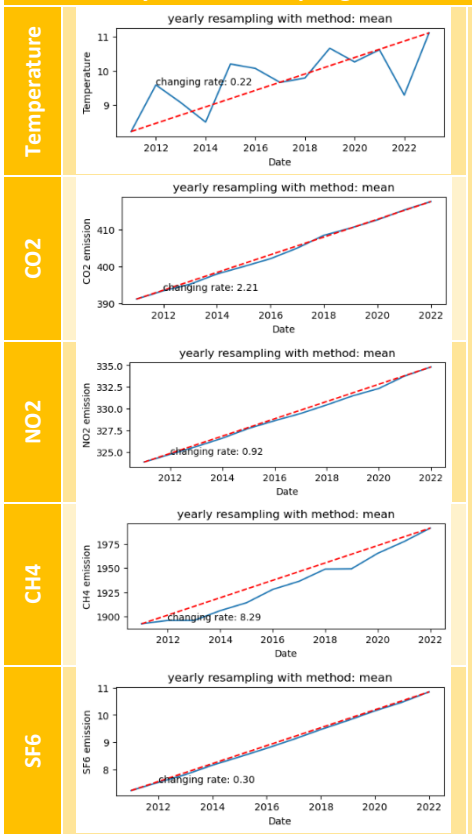


What we see,

- CH4 distribution is right skewed
- CO2 distribution is normal
- NO2 & SF6 distribution doesn't have any specific distribution
- Temperature in this area has a normal distribution between zero to 20 degrees

yearly and weekly temperature resampling graph

Yearly Mean resampling method

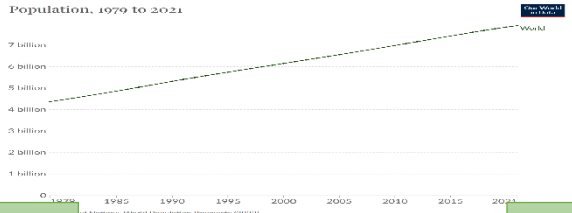


What we see,

- The temperature is increasing these years and with less gradient than the emission gases,
- The slop of releasing carbon-based gases (CH4 & CO2) is more than others and it can be directly related to the growing human population

Our growing population

The world's population is more than three times larger than it was in the mid-twentieth century. The global human population reached 8.0 billion in mid-November 2022 from an estimated 2.5 billion people in 1950, adding 1 billion people since 2010 and 2 billion since 1998. The world's population is expected to increase by nearly 2 billion persons in the next 30 years, from the current 8 billion to 9.7 billion in 2050 and could peak at nearly 10.4 billion in the mid-2080s.

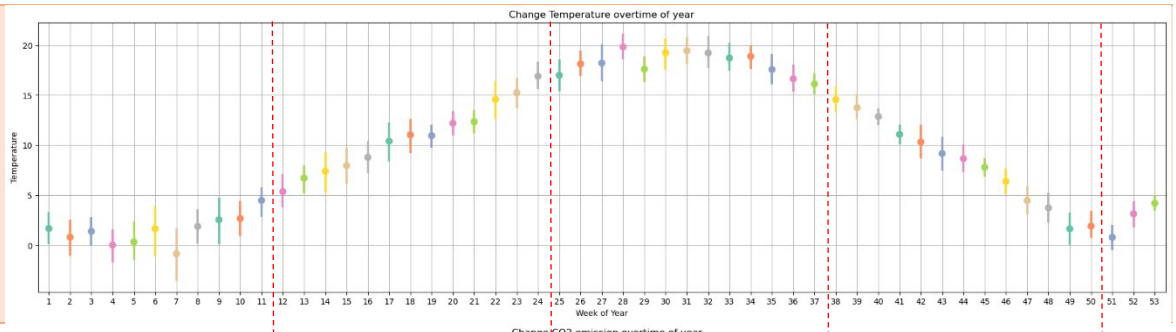


graph data source

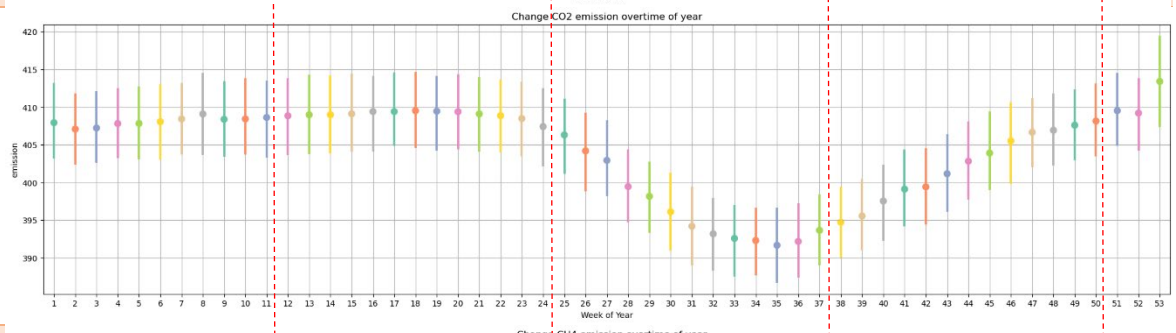
Information Box

Weekly Mean resampling method

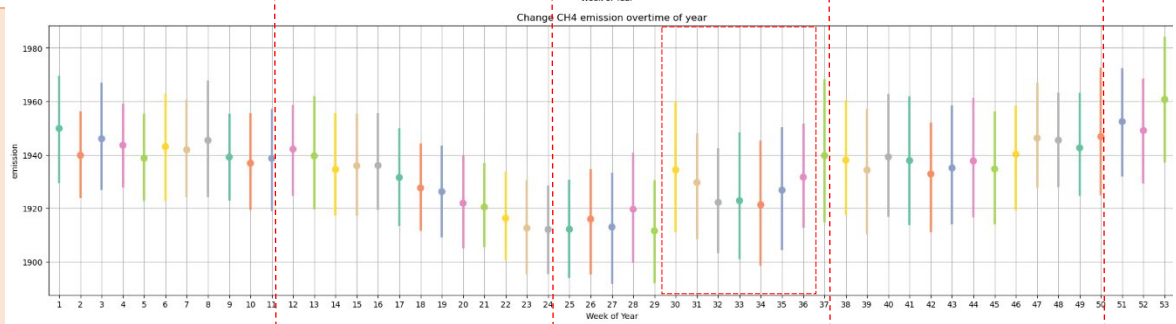
Temperature



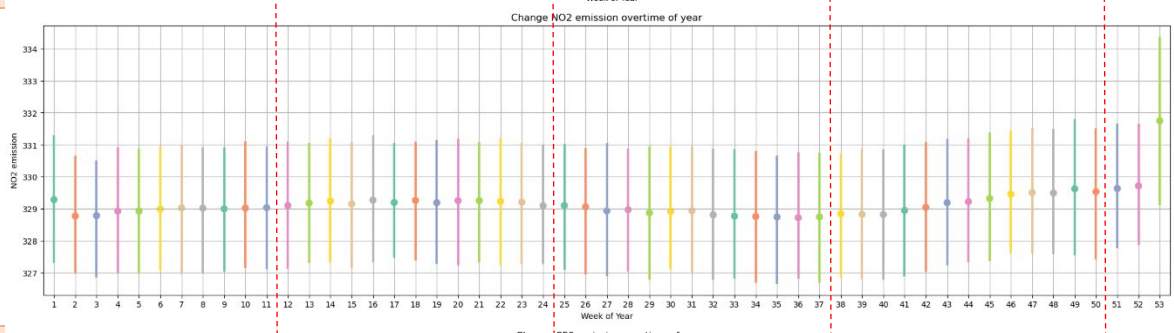
CO2



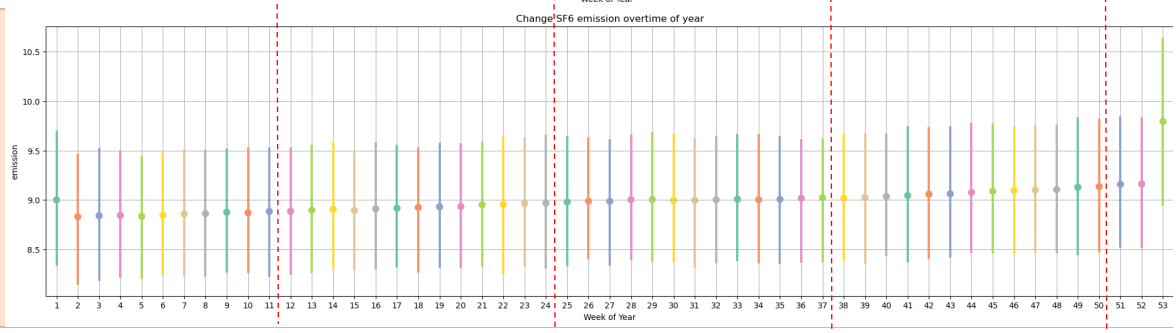
CH4



NO2



SF6



In the above graphs, seasons are separated with lines and according to them we can understand,

- emission of carbon-based gases is back to human behavior
    - emission of CO2 increases in autumn and decreases in summer
    - emission of CH4 fluctuates but related to the high season for tourists

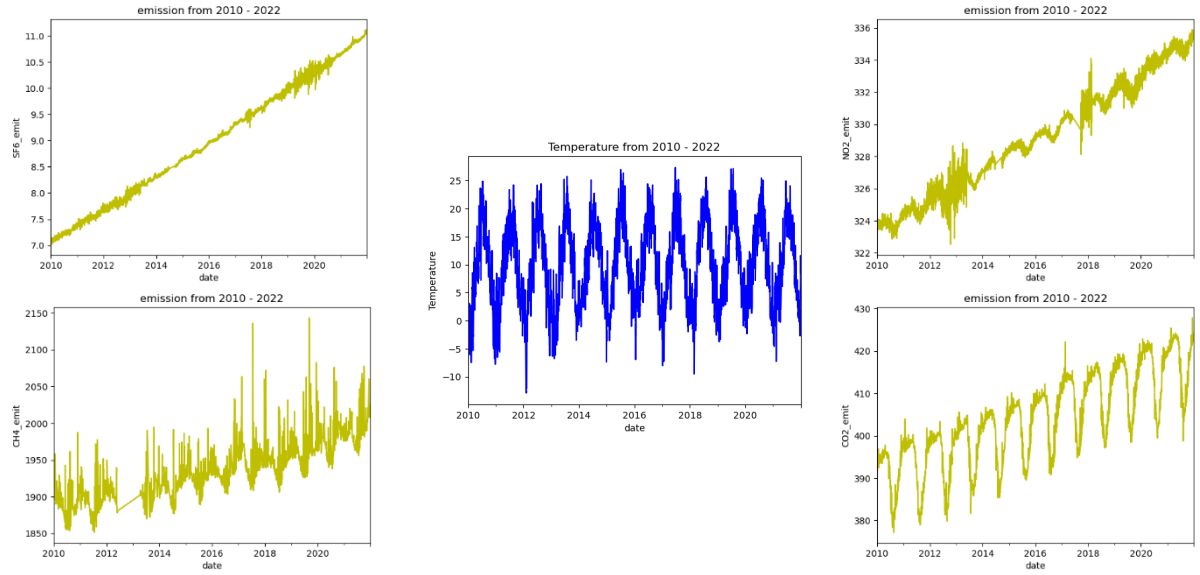
Tis the Seasons for the Tourists

High season: Ski and beach destinations aside, summer (June 15 to Sept 21 or so) is high season in most temperate zones North America, Europe, Australia, South America)—through remember: summer in the southern hemisphere means roughly November to February.

Information Box

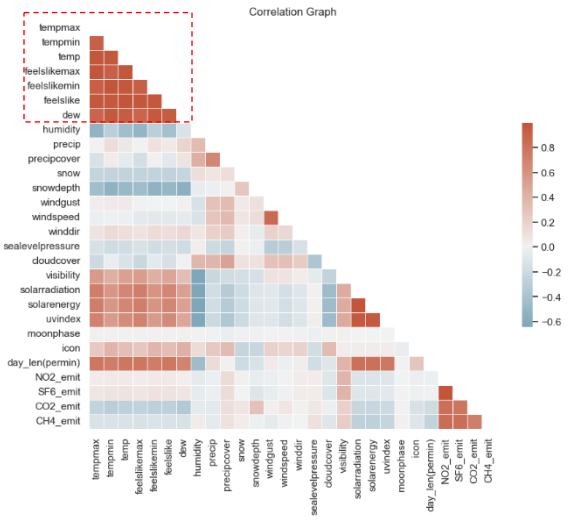
- emission of NO2 and SF6 do not have any obvious seasonality but we can see an increasing drop in autumn for NO2

Time-Series Trend graph

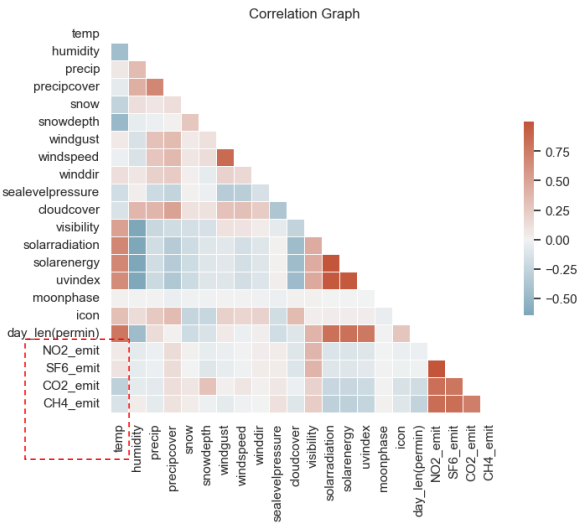


- SF6 doesn't have seasonality
- Others have positive slop and seasonality

Correlation of features graph



There is too correlation exist between 'tempmin', 'tempmax', 'feellikes', 'feellikesmin', 'feellikesmax', & 'dew' with target 'temp', and it can cause the data leak



Conclusion

I started this article with a crucial question and I followed the data shown to me. The emission of greenhouse gases has different resources and some of them are in human hands as the weekly trend shows the trend of changing temperature and carbon-based gases same as each other and based on that we can conclude that most of that is related to human activities. Here comes the next question, what can we do to decrease it? What are the main activities that cause emissions of these gases?

As I discussed in the article, the **growing population of humans** has a connection. Some governments have their birth control programs in a specific area or population. Maybe it can help, but it cannot decrease the current emissions issue. The logical way to directly target this issue is to find **the sources of carbon-based gas emissions** to control in the first place and replace them with less harmful activities for our environment, which guide us to investigate sources and find a priority list based on affection each one in emission according to valid data.

In the continuation of this article, I will work to find the models that can help to understand the possible future in the way we are now.