**1. The database(s) you plan to use for storage.**

I will be using Mysql, Psql or google firebase for storing my dataset. Firstly, Mysql syntax is very similar to psql so adaptability and ease of use is two main reasons behind suggesting it. Secondly google firebase is a popular db which works quite fast and efficient compared to many cloud database systems.

Finally, talking about psql there is no doubt that I am well capable of handling such a sql language very similar to human language like python. Psql is very easy to learn if someone has the interest to work with it.

**2. Where the data is loaded from.**

Raw data is taken from the below website.

https://data.world/data-society/us-air-pollution-data/workspace/project-summary?agentid=data-society&datasetid=us-air-pollution-data

and the segregated data is loaded from my local computer folder. Github is used as repository for version control.

https://github.com/sujilkumarkm/da_assignment2_semester2_2021_dkit

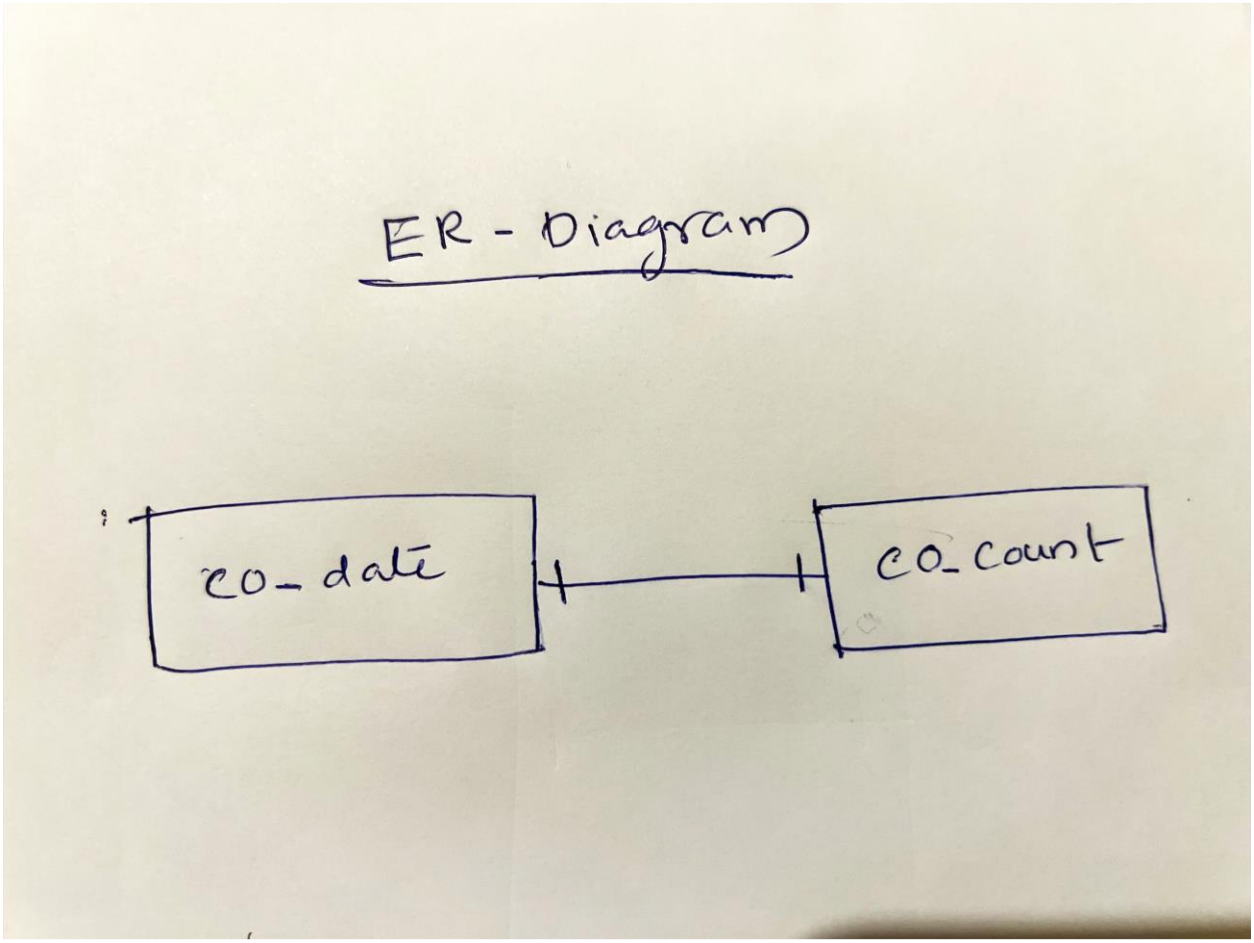
**3. Any transformations or scripts to load the data to the**

database (e.g. cleansing in pandas)

dataset has been initially cleaned using pandas library In python. For time series analysis we only needed two columns date and carbon count do rest of the columns are removed from the csv file.

Current data contained multiple carbon values on the same day so using sql I have removed repeated dates to make the analysis accurate.

**4. The database schema (e.g. E-R diagram) itself.**




From my understanding, the data we used in this project is comparatively smaller compared to previous project. The reason being that we have to work with the same time series dataset which we used for time series prediction. It only contains 2 columns. So, the schema is made according to the available dataset for one year of carbon emission in the US.

**5. Connections to/from the database for Time Series Analysis module**

So, as we know for time series analysis, we need sequential data and that is what time series is all about. In my study my project was focusing on daily carbon emissions in the US. Database contains same time series data which is taken for doing analysis in psql. Transformation techniques from time series has ben applied to this project making it a part of the time series analysis module.

I have always wondered how to make python techniques we used in time series completely in sql and this assignment gave me chance to look deeper and explore more levels of PostgreSQL.