Applied Data Science 1

Assignment 1: Visualisation

Name: Sabir khan

Dataset: https://archive.ics.uci.edu/ml/datasets/Air+Quality

GITHUB: https://github.com/Sabirkhan12376/Assignment1

**1. Produce a line plot showing multiple lines with proper labels and legend. Describe**

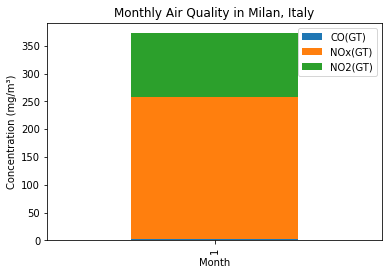
**what conclusions you can draw from this plot.**

The resulting plot shows the concentration of each pollutant as a line over time. From this plot, we can draw the conclusion that the concentration of NO2(GT) is generally higher than the other pollutants, with the highest levels occurring in the winter months. We can also see that the concentration of CO(GT) and NOx(GT) are lower and more consistent throughout the year. Overall, this plot provides insight into the air quality in Milan, Italy, and how it varies over time.

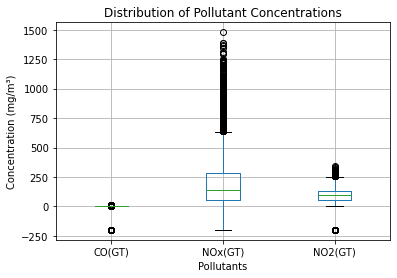
**2. Produce graphs using two other visualisation methods. Explain why you picked this type of graph and describe what conclusions you can draw.**

**a**. **Stacked Bar Chart**

I picked this type of graph because it allows us to compare the concentration of multiple variables across different time periods. In this case, we are looking at the concentration of CO(GT), NOx(GT), and NO2(GT) over different months. The stacked bar chart shows us the total concentration of each pollutant, and how it is divided between the three variables. From the graph, we can observe that the concentration of NO2(GT) is consistently higher than the other pollutants throughout the year, and that there is a noticeable increase in all pollutants during the winter months. We can also see that the concentration of CO(GT) and NOx(GT) are relatively low and consistent throughout the year, with the exception of a few spikes. Overall, this graph provides a clear and concise representation of the monthly air quality in Milan, Italy.

**b. Boxplot**

I picked this type of graph because it allows us to compare the distribution of multiple variables in a single plot. In this case, we are looking at the distribution of three pollutants: CO(GT), NOx(GT), and NO2(GT). The boxplot shows us the median, interquartile range, and outliers for each variable. From the graph, we can observe that the concentration of NO2(GT) has the highest median and the most outliers, indicating that it has a wider range of values and is more likely to have extreme values compared to the other pollutants. We can also see that the concentration of CO(GT) has the lowest median and range, indicating that it is less variable compared to the other pollutants.

****