Problem Statement

Air pollution data given to us is a stock time series data or a flow time series data? Can air pollution data be both? If you are given two air pollution data what are the factors which allow you to differentiate them as stock time series data or flow time series data?

Answer

The air pollution data given to us in particuar is a stock time series data as it refers to measuring characteristics at a specific moment, much like a static image of the data as it was. Because we were given the density of the pollutants at some particular timestamp and not the flow of variation in density of the pollutants or some timely emission of any pollutant.

In general, air pollution data can be considered as both stock time series data and flow time series data, depending on the specific context and variables being measured.

Stock Time Series Data: Air pollution data can be treated as stock time series data when measuring the concentration or level of pollutants at a specific point in time. For example, measurements of particulate matter (PM2.5), nitrogen dioxide (NO2), or ozone (O3) levels taken at regular intervals (e.g., hourly, daily, monthly) represent stock data. Each data point represents the stock or quantity of the pollutant at that particular time.

As the data given to us is similar to above definition for stock time series data so we can say that the data given to us is stock time series data.

Flow Time Series Data: Air pollution data can also be considered as flow time series data when measuring the flow or rate of pollutants over time. This typically involves measurements of pollutant emissions or pollutant concentrations over a specific time period. For example, hourly emissions of carbon monoxide (CO) from a specific source or the daily average concentration of sulfur dioxide (SO2) in a particular region represent flow data. Each data point represents the flow or rate of the pollutant during that time interval.

In practice, air pollution data can include both stock and flow components. For example, pollutant concentration measurements at monitoring stations represent stock data, while emission inventories or modeled pollutant dispersion data represent flow data. If we are given two dataset representing air pollution then we can differentiate them as stock time series data or flow time series data depending upon the specific variables being measured and the context in which the data is analyzed.

Problem Statement

Do the explanatory analysis for the air pollution data given to you for each pollutant and explain the features in the data. Explain the reason for missing values in the data and give some methonds to find them. Also explain whether the actual blasting time follows the estimated blasting time, if not then state reasons why?

Answer

Air pollution data given to us is a stock time series data which represent the density of the pollutants in air at a particular time interval of 15 minutes

Pollutants released during open pit mining are :

Dust and Particulate Matter (PM): Open pit mining involves excavation and movement of large amounts of soil, rock, and ore. This activity generates dust and particulate matter, which can include both coarse particles (PM10) and fine particles (PM2.5). These particles can cause air pollution and have potential health impacts when inhaled.

Sulfur Dioxide (SO2): Mining operations involving sulfide ores, such as copper, lead, and zinc, can release sulfur dioxide into the air. Sulfur dioxide is produced when sulfide minerals are exposed to air and react with oxygen. It is a major contributor to acid rain and can have adverse effects on human health and the environment.

Nitrogen Oxides (NOx): Equipment and machinery used in open pit mining, such as diesel-powered vehicles and machinery, can emit nitrogen oxides. Nitrogen oxides are formed during the combustion process and contribute to air pollution. They are known to contribute to the formation of smog, respiratory issues, and environmental degradation.

Volatile Organic Compounds (VOCs): Various mining activities, including blasting, drilling, and processing of ores, can release volatile organic compounds. These compounds can include hydrocarbons and other chemicals that evaporate easily into the air. VOC emissions contribute to air pollution and can have adverse effects on human health and the environment.

Heavy Metals: Mining operations can release heavy metals such as mercury (Hg), lead (Pb), cadmium (Cd), arsenic (As), and others. These metals are naturally present in many ores and can be released into the environment through various mining processes, such as ore crushing, grinding, and ore processing. Heavy metals can persist in the environment, accumulate in the food chain, and pose significant health and ecological risks.

Now the data for benzene and ozone given to us comes uder VOCs released during open pit mining which are not particularly released but are formed by the reaction between other gases relaesed durig open pit mining

As inferred form the given data PM10 and PM2.5 gets the highest peak among all pollutants whereas ozone and benzene show some seasonality in the data.

So for the missing data we considered seasonality for benzene and ozone.

Now PM10 and PM2.5 are suspended particulate matter so one reason for them to be high is, during the time of blasting the gases released in air are more dense and heavier than normal so due to this reason the PM10 sensor might consider them as PM10 only but in reality there may be some other gases mixed with it such as NO, NO2, NOX or CO.

One of the main reason for missing values in the data maay be the failure of the sensors that take the reading other may be the sensors worked properly but the reading was not taken at that time due to some other failere. Also during the time of blasting gases are released at high temperature so due to high temperature sensors my malfunction and could not take the reading.

NOX formed is the mixture of NO and NO2 so we can find the correlation between NO and NO2 to find the missing values of NOX.

Interpolation is one of the best technique to find the missing data, checking cyclicity and seasonality in the data can also be helpful as done for benzene and ozone.

When we extimated the blasting time and compared it with the actual blasting time we saw major variation in them this could be due to various reasons :

The time it takes for pollutants to release after blasting in open pit mines can vary depending on several factors, including the type of blasting method used, the characteristics of the rock or ore being blasted, and the atmospheric conditions at the site. Here are some considerations:

Instantaneous Release: Some pollutants, such as dust and particulate matter (PM), are typically released instantaneously or within seconds after the blast. The explosion and subsequent fragmentation of rock or ore can generate a cloud of dust and particulates that disperses rapidly into the surrounding air.

Gaseous Emissions: Gaseous pollutants, such as sulfur dioxide (SO2), nitrogen oxides (NOx), and volatile organic compounds (VOCs), can also be released during blasting. The time it takes for these gases to be emitted and disperse depends on various factors, including the chemical composition of the rock or ore, the extent of mineral oxidation or reaction, and the ventilation conditions in the mining area.

Dispersion and Transport: Once the pollutants are released, their dispersion and transport in the air depend on atmospheric conditions, such as wind speed and direction. Pollutants can disperse and travel different distances from the source, influenced by factors like air stability, topography, and nearby structures.

Duration of Impact: The duration of pollutant impact after blasting can vary. Dust and particulate matter can settle relatively quickly, depending on particle size and environmental conditions. Gaseous emissions may persist for a longer duration, particularly if there are ongoing processes or activities at the mining site that continue to release pollutants.

Traffic : Due to traffic and movement of vehicles in the region may increase or decrease the density of the air pollutants as dust is being carried away and the transportation of coal from open pit mines in India can occur throughout the day, depending on various factors such as operational requirements, logistical considerations, and local regulations. Coal mining operations in India typically involve round-the-clock activities to meet the demand for coal. Therefore, coal transportation can take place at any time, including day and night shifts. This can also result in the wrongly estimated time for blasting

Actual Blasting Time : It may also be the case that the actual blasting time was not the preferred blasting time and the data given to us itself is faulty for some days. This could happed due to various reasons such as failure of blasting equipment like dynamite or non-avaivalibality of those equipment or delay by the workes or the whether condition which which were not suitable for blasting.

Problem Statement

Do the descriptive analysis of the data give to you find trends like seasonality and the distribution of the residue ,measures of frequency, central tendency, dispersion or variation, and position of air pollution data.

Answer

To conduct descriptive analysis for air pollution data classified as stock time series data, you can perform several techniques to gain insights into the data distribution, patterns, and trends. Here are some approaches to consider:

Data Visualization:

Plotting Time Series: Create line plots or area plots to visualize the pollution levels over time. This will help you observe the general trends, seasonality, and any irregular patterns in the data as done above. Heatmaps: Create heatmaps to display pollution levels across different time intervals or geographical locations. This can help identify any spatial or temporal patterns. We could not plot heatmaps as we were not povided any information about the particular location of blasting or mining but just the density of pollutants in the nearby region

Summary Statistics:

Measure Central Tendency: Calculate measures like mean, median, and mode to understand the typical pollution levels. Assess Dispersion: Compute measures such as standard deviation, variance, or range to evaluate the variability or spread of pollution levels. This is also done above in TSA with nan = 0 and nan = mean Identify Outliers: Detect any extreme values or outliers in the data that may require further investigation. Instead of further investigation we discarded the extreme data which we were getting due to interpolation as it was of no use to us.

Seasonal Decomposition:

Decompose the time series data into its constituent components: trend, seasonality, and residual. This can be done using techniques like seasonal decomposition of time series (e.g., STL decomposition) or moving averages. Analyzing these components will provide insights into long-term trends and seasonal variations in pollution levels. We checked for seasonality in the Interpolation and Seasonality part and estimated the missing data points for some pollutants which show seasonality. We also did the residue and density plot for various pollutants and estimated that most of the residue was white noise as it has mean 0 and constant variance. Whereas the density mostly followed normal and cauchy distribution.

Autocorrelation Analysis:

Calculate Autocorrelation: Measure autocorrelation to assess the relationship between past and current pollution levels. This can be done using autocorrelation plots (ACF) or partial autocorrelation plots (PACF). Significant autocorrelation at specific lags may indicate temporal dependencies or patterns in the data. This was also done above

Statistical Tests:

Stationarity Test: Perform tests for stationarity, such as the Augmented Dickey-Fuller (ADF) test, to determine if the pollution data exhibits a constant mean and variance over time. We conducted ADF test for all the pollutants separately and together also to test and train our ARIM model above.