INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

ROBERT GORDON UNIVERSITY ABERDEEN

BSc. Artificial Intelligence & Data Science

Level 05

CM 2606 Data Engineering

COURSEWORK

Module Leader: Mr. Mohamed Ayoob

SHAMAL RATHNAYAKA

IIT ID: 20222117

RGU ID: 2309039

Table of Contents

Table of Table Figures

Table of Figures

Abstract

As a component of the Data Engineering module (CM2606), this evaluation takes the form of a strategic investigation of the fundamental Machine Learning Models and Data Engineering. Module Coordinator Mr. Mohamed Ayoob led the facilitation. This study investigates the temporal and spatial variability of formaldehyde (HCHO) concentrations in cities using data from monitoring networks in multiple cities. Monthly average HCHO concentrations were analyses along with pollutant emissions and meteorological variables to identify likely sources of HCHO variability. The results exhibit distinct seasonal patterns, with higher HCHO levels in the warmer months being attributed to higher levels of biogenic emissions and photochemical synthesis. The local emission sources, air dispersion patterns, and land use characteristics all affect the spatial variability in HCHO concentrations. The analysis highlights how important it is to take climate conditions and human activities into account when trying to comprehend the dynamics of urban air quality. The study's conclusions can help guide air quality control plans that try to lower HCHO exposure and enhance urban public health.

Acknowledgement

I would like to sincerely thank Mr. Mohamed Ayoob and Mr. Nipuna Senanayake for their crucial advice, encouragement, and assistance during the project's completion. Their perspectives and expertise have had a significant impact on how this investigation has unfolded. I am also appreciative of my classmates' and colleagues' cooperation and helpful criticism during the creation and assessment of machine learning models. I also want to express my gratitude to the people who created the scikit-learn framework for giving me the resources and tools I needed to carry out this research. In conclusion, I give my gratitude to my family and friends for their consistent support and comprehension during this effort.

Introduction

The air pollution crisis is one of the biggest issues facing the modern world. Of all the pollutants, formaldehyde (HCHO) stands out as one that should worry you the most because it is linked to several health problems, such as eye strain, lung irritation, and increased cancer risk. Comprehending both temporal and spatial trends of HCHO is essential for evaluating air quality, identifying the sources of emissions, and developing practical mitigation plans. This research uses satellite observations made possible by the European Space Agency's Sentinel-5P satellite to conduct a thorough analysis of HCHO data collected from seven cities in Sri Lanka. The dataset, which runs from January 1, 2019, to December 31, 2023, provides a wealth of information about HCHO levels over a sizable time.

Methodology

Data Acquisition

The first step in our methodology involved acquiring the HCHO dataset from the European Space Agency's Sentinel-5P satellite. The dataset spans from January 1, 2019, to December 31, 2023, and covers seven cities in Sri Lanka, providing daily measurements of tropospheric HCHO column number density. https://drive.google.com/drive/folders/1xzQ5pIEnaUN2DOyZTqYSJrxFMC8Unx73?usp=sharing

GitHub Repository

Link:- https://github.com/ShamalRthnayaka/HCHO-Prediction

Data Preprocessing

Data Collection

When examining the dataset, the seven cities are included in three CSV files. Puttalam, Kurunegala, Kandy, Monaragala, Jaffna, Colombo, and Bibile (Matara) are among them. A portion of the dataset is represented by each CSV file, which focuses on city clusters to facilitate quick analysis and interpretation. This division makes it possible to analyze HCHO levels at a finer level within the context of specific cities and to make comparison evaluations between various geographic locations.

Data Cleaning

Careful data cleansing is a necessary first step in our data analysis journey to guarantee the accuracy and consistency of our dataset. Finding and fixing any errors, inconsistencies, or missing numbers in the raw data is the task of this process. We want to improve the consistency and quality of our dataset by methodically going over each data point and using the right cleaning methods, such imputation for missing values and format standardization, to provide a solid basis for further analysis.

Outlier Handling

In this case the outlier handles separately because catches all outliers.

Kandy

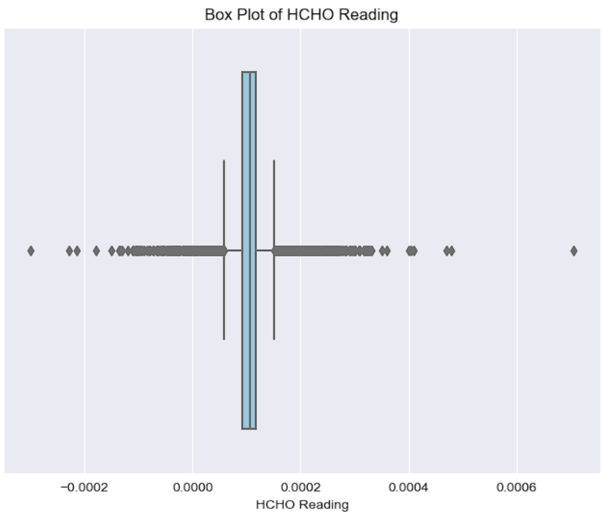


Figure 2(Kandy Outlier handling before)

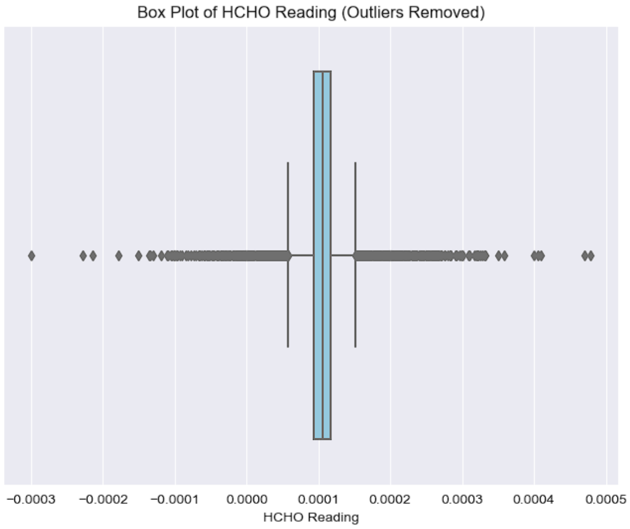


Figure 1(Kandy Outlier handling after)

The first box plot shows the spread of values and highlights any outliers in the dataset's HCHO reading distribution. It shows the quartiles (Q1 to Q3), the median as a line, and the whiskers extending out to 1.5 times the quartiles' interquartile range (IQR). Data points that are out of the range are indicated as outliers. The distribution of the filtered data is shown in the second box plot, which was created after outliers over 0.0006 HCHO measurements were eliminated. In comparison to the previous plot, it ought to show a more constrained range of values, shorter whisker lengths, and either fewer or no outliers. This indicates how outlier reduction affected the distribution of the dataset. Comparing these plots provides insights into how filtering affects the distribution and spread of HCHO readings, aiding in the assessment of data quality and potential anomalies.

Kurunegala, Monaragala, Jaffna

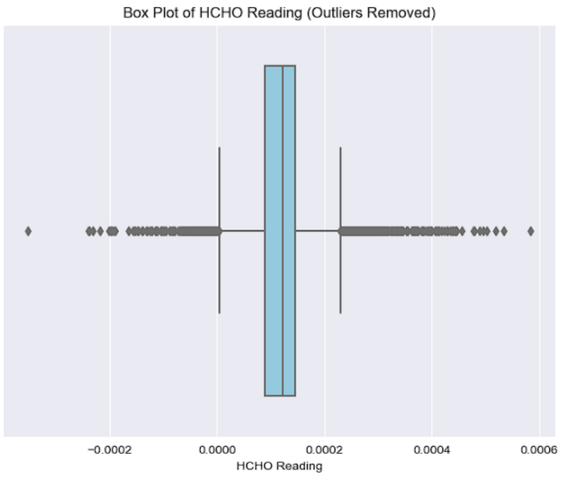


Figure 3(Kurunegala, Monaragala, Jaffna Outlier before)

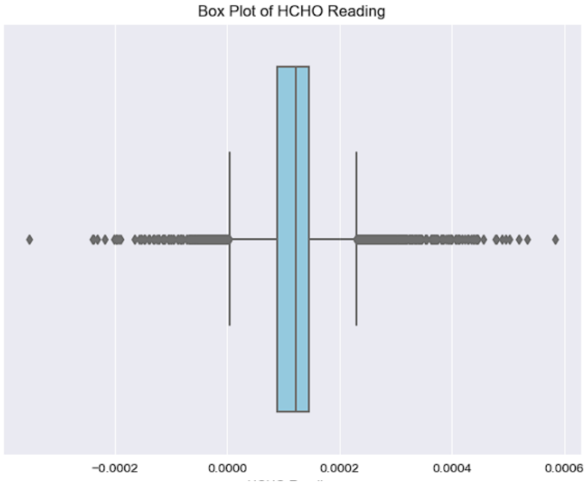


Figure 4(Kurunegala, Monaragala, Jaffna outlier handling after)

To see the distribution of formaldehyde (HCHO) readings in a dataset, the algorithm first builds a box plot. After that, it removes all HCHO values greater than 0.0006 and shows the dataset that has been filtered. The filtered data is then used to create another box plot, which illustrates how the distribution of HCHO measurements varies when outliers are eliminated. It is possible to observe how outlier removal impacts the distribution and spread of HCHO measurements in the dataset by contrasting the two box plots.

Colombo, Nuwara Eliya, Matara

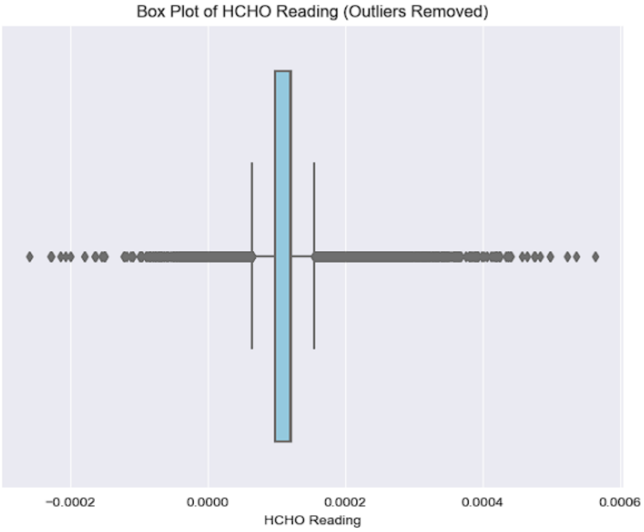


Figure 6(Colombo, Matara, Nuwara Eliya outlier before)

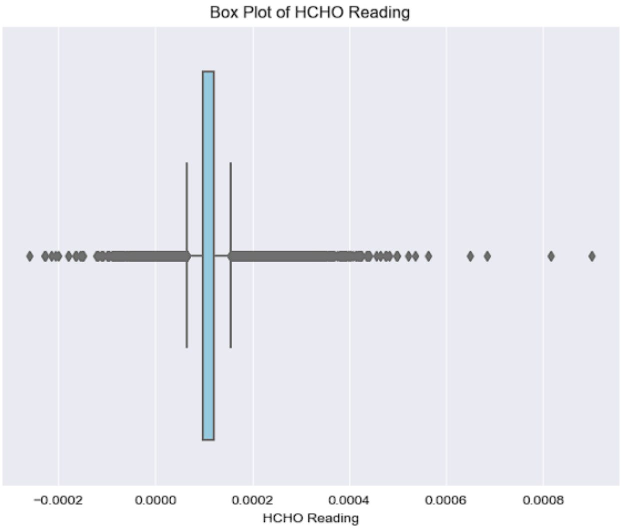


Figure 5(Colombo, Matara, Nuwara Eliya outlier after)

Two box plots are produced by this code to show formaldehyde (HCHO) readings in a dataset. Any outliers in the original distribution of HCHO measurements are depicted in the first plot. After that, the code publishes the final dataset after removing HCHO measurements greater than 0.0006. Using the filtered data, it then generates a second box plot to show how the distribution of HCHO readings varies when outliers are eliminated. Understanding the effect of outlier removal on the distribution and spread of HCHO measurements in the dataset is made easier by comparing the two box plots.

Data Merging

Data merging is required to compile the information from all the sources into a single, cohesive dataset because each of the three CSV files in our dataset contains data from a different city. To create a single dataset with HCHO readings from all seven cities, the individual CSV files can be joined together using a common identifier, such as the date or the city name. This unified dataset allows for in-depth research and cross-regional comparisons, which contributes to a more thorough knowledge of the prevalence and trends of HCHO in Sri Lanka.

Descriptive Statistics and Visualization

Monthly Trend

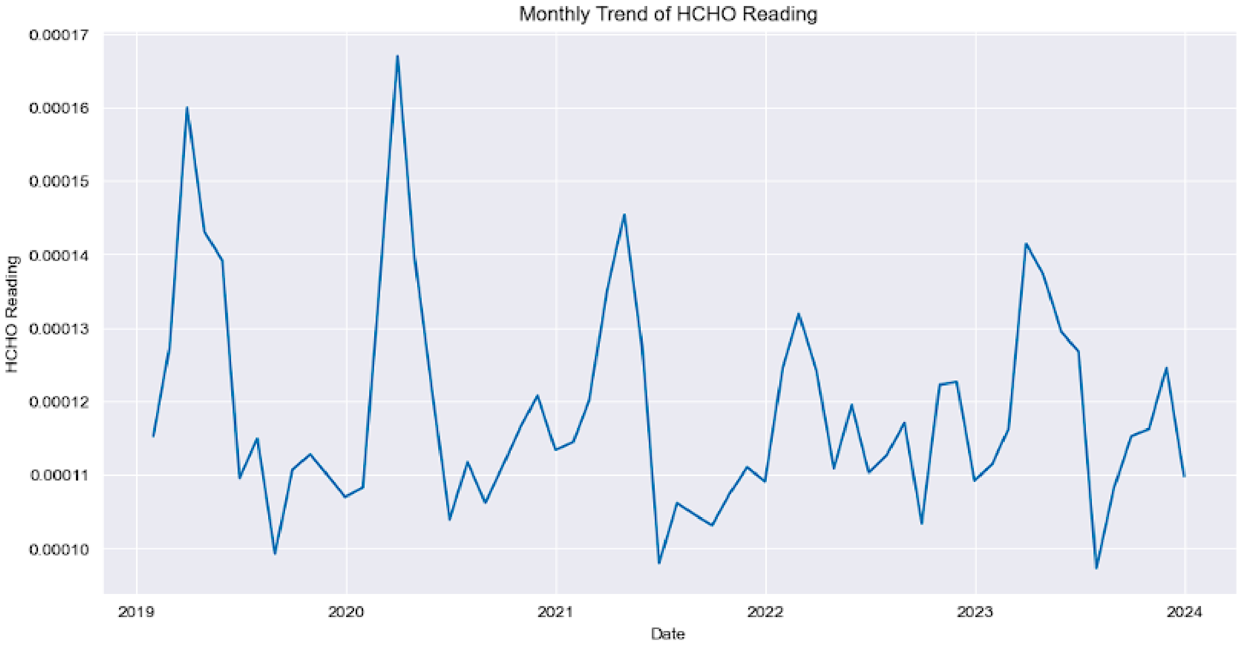


Figure 7(Monthly Trend)

The formaldehyde (HCHO) monthly trend is shown on the graph over a period. Every month, the data is combined and the average HCHO reading for that month is displayed at each location, which signifies the end of a month. The average HCHO value is shown on the y-axis, while the date is indicated on the x-axis. Finding patterns, trends, or seasonal variations in HCHO levels throughout the course of the observed time is made simpler when the data is plotted as a line plot. Accurately tracking monthly variations in HCHO values is made easier with the help of grid lines.

Seasonal Component

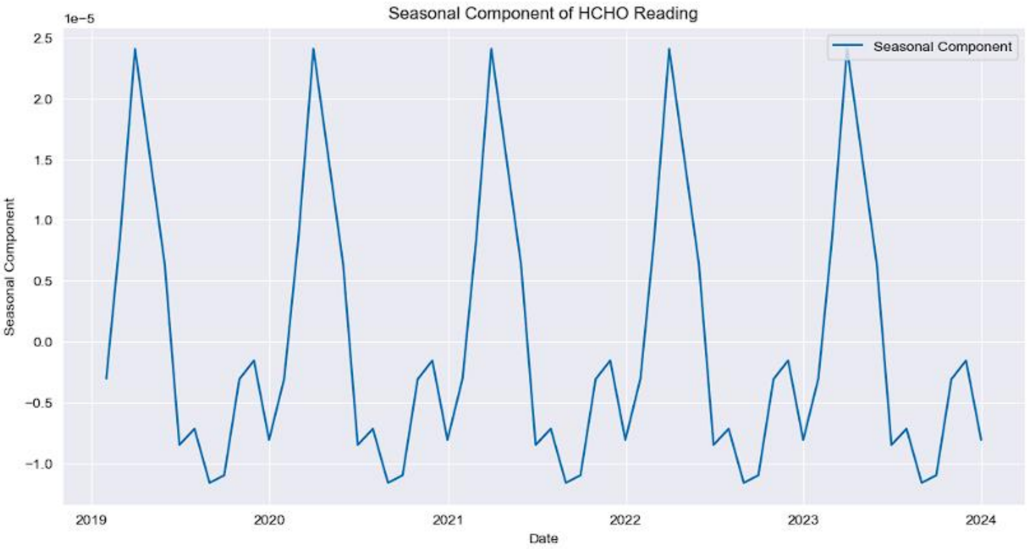


Figure 8(Seasonal Component)

Using seasonal decomposition, this graph shows the seasonal component of formaldehyde (HCHO) values across time. To identify the seasonal fluctuation inherent in HCHO levels, the graph breaks down the data into its component parts: trend, seasonal, and residual. The date or time is shown on the x-axis, and the seasonal component of the HCHO measurements is indicated on the y-axis. Every dot on the graph represents a seasonal variation related to periods of the year. By examining this graph, it is possible to spot seasonal variations or recurrent patterns in HCHO levels, which can provide information about external causes or other factors that affect HCHO concentrations. Grid lines make interpretations easier to understand since they make it easier to see how seasonal fluctuations in HCHO values are cyclical.('Consensus Seasonal Weather Outlook February, March and April (FMA) Seasonal Rainfall and Temperature for Sri Lanka', no date)

Monthly Across Cities

The monthly trend of formaldehyde (HCHO) readings in several cities is depicted in this graph. The ability to compare HCHO levels across different locales is made possible by the data's grouping by city. The x-axis shows the date, and the y-axis shows the average HCHO reading for each month. Each line on the plot indicates the monthly trend of HCHO readings for a particular city. It is feasible to compare the trends in HCHO levels between the cities graphically by plotting many time series on the same graph. Finding any changes or patterns in the trends between cities can give important information about possible variations in the environment, pollution levels, or other variables affecting the concentrations of HCHO in various metropolitan regions.

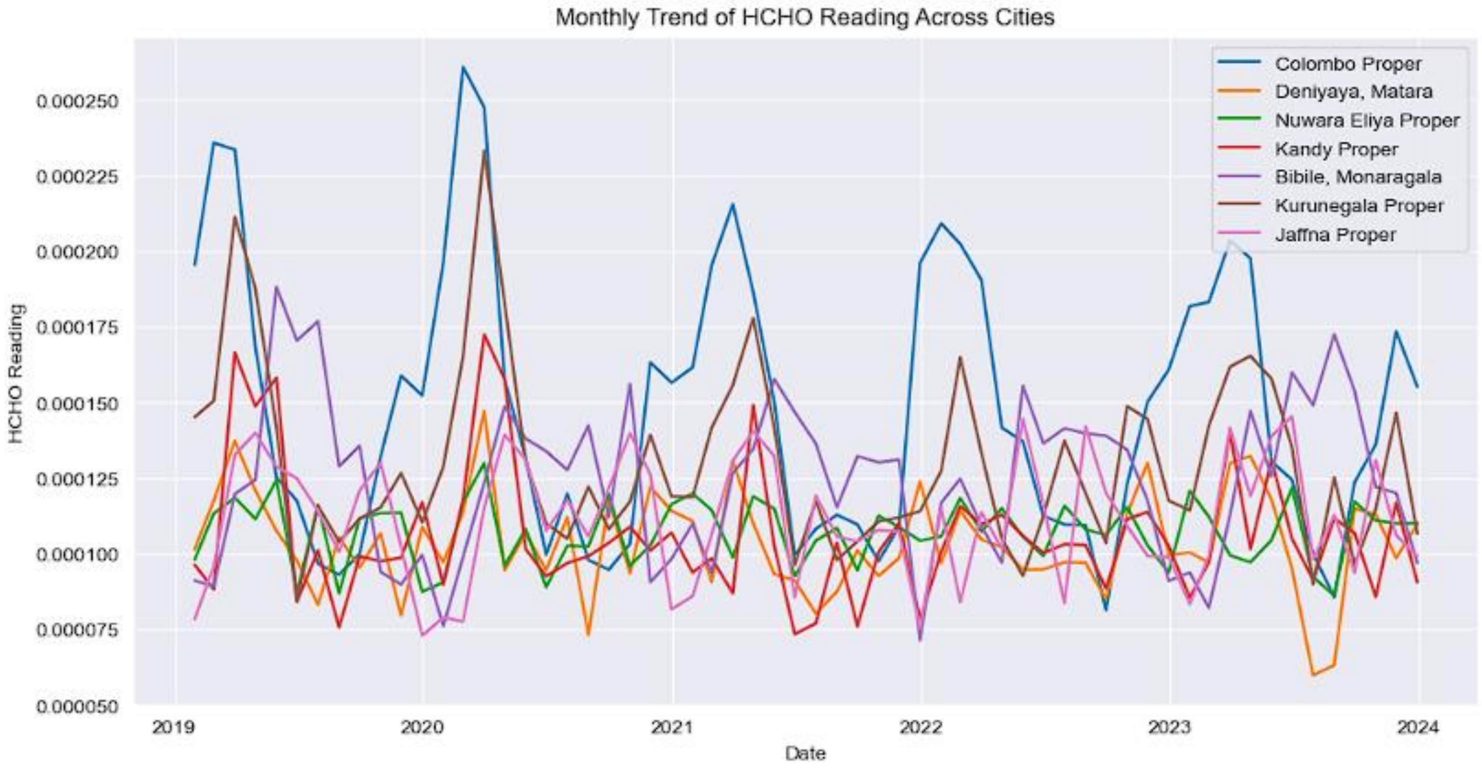


Figure 9(Monthly trends across cities)

The inclusion of a legend helps in identifying which line corresponds to each city, while grid lines aid in accurately interpreting the fluctuations in HCHO readings over time.

Impact of Covid Lockdown

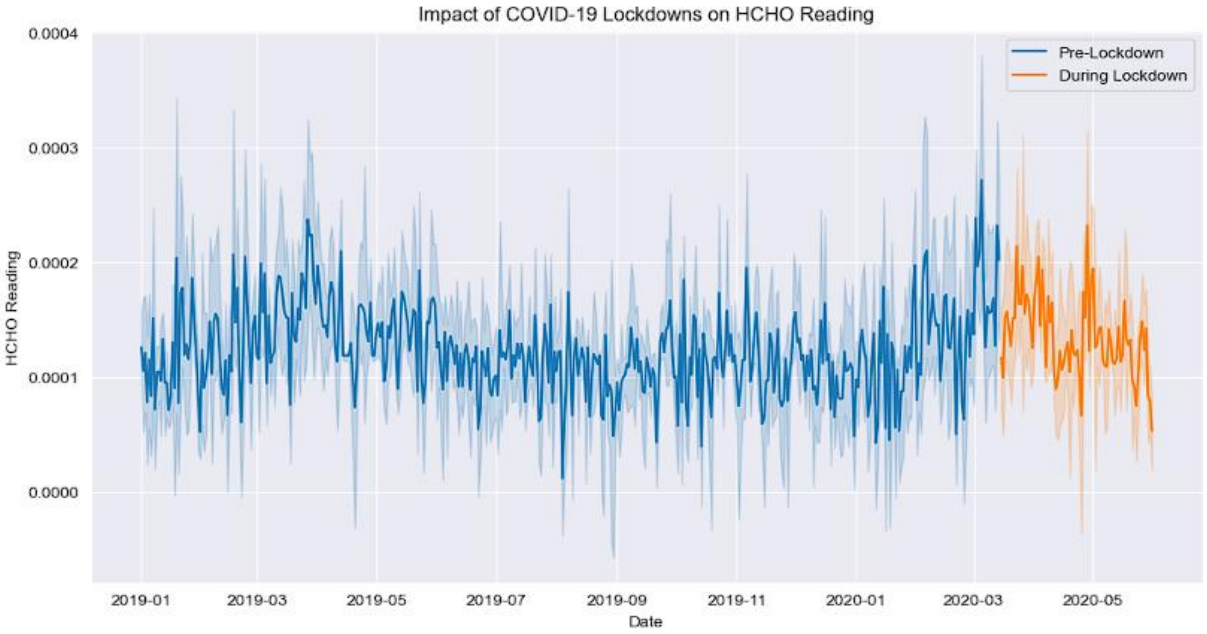
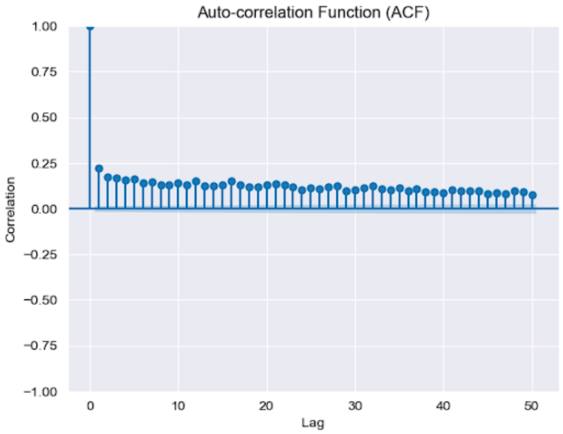
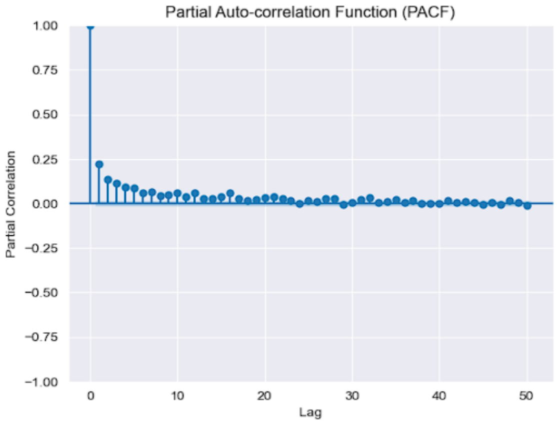


Figure 10(Impact of covid)

The effect of COVID-19 lockdowns on formaldehyde (HCHO) readings over time is shown in this graph. Two time periods are distinguished in the data: before and during the lockdowns. The date range is shown by the x-axis, which runs from prior to the start of the lockdowns until they end. The recorded HCHO values for each time are displayed on the y-axis. It allows for a direct comparison of the two times by putting HCHO values on the same graph before and during the lockdowns. This comparison makes it easier to spot any variations or patterns in HCHO levels related to the COVID-19 lockdown measures. Grid lines make it easier to understand how HCHO values change over time, and the legend makes it easier to distinguish between the times before and during the lockdown. Analyzing this graph provides insights into how human activities and environmental factors, potentially affected by lockdown measures, may have influenced formaldehyde levels in the observed areas.

Correlation function

Table 1(Correlation function)

Auto Correlation Function		Knowing how each observation in a time series connects to its previous observations is made easier with the use of the auto-correlation function (ACF) graphic. Plotting the correlation coefficients for various time lags allows one to see any patterns or dependencies in the data. Possible associations between previous and current values in the time series are shown by peaks or substantial correlations at lags. Finding temporal structures and guiding modelling choices, such picking the best forecasting methods or comprehending the underlying dynamics of the data, depend on this research.
Partial Auto Correlation Function		When determining the direct link between data in a time series, the partial auto-correlation function (PACF) plot eliminates the impact of additional observations made at intermediate lags. The PACF isolates only the direct effect of a particular lag on the current observation, in contrast to the auto-correlation function (ACF) plot, which analyses correlations incorporating both direct and indirect effects. In especially for autoregressive (AR) models, this distinction is useful for determining the exact lag at which the correlation between observations becomes significant, which helps with model selection and forecasting accuracy.

Spatio-Temporal Analysis

Correlation Matrix

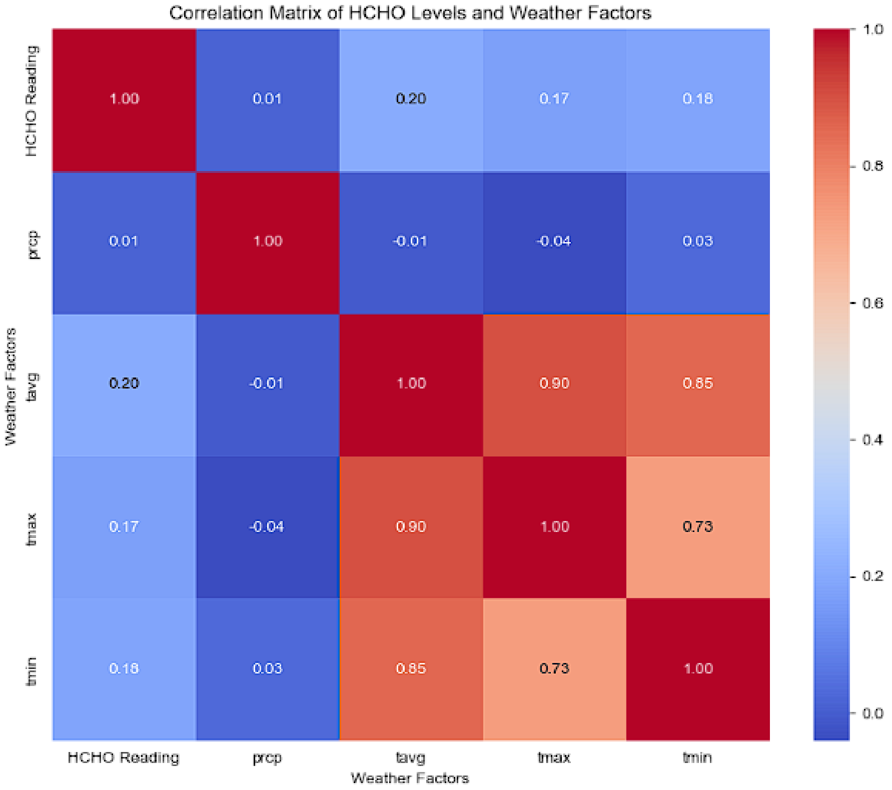


Figure 11(Correlation matrix)

The association between formaldehyde (HCHO) levels and other meteorological conditions is revealed by the correlation matrix and heatmap. The intensity and direction of the association between HCHO levels and a particular meteorological component are shown by each correlation value in the matrix. According to this analysis, the minimum temperature (tmin), maximum temperature (tmax), and average temperature (tavg) all show a somewhat positive link with HCHO levels, with correlation values of approximately 0.20, 0.17, and 0.18, respectively. This implies that elevated temperatures might be associated with elevated HCHO concentrations.(Hadavand-Siri and Deutsch, 2012) Nevertheless, there is only a very weak positive association (correlation coefficient of about 0.01) between HCHO levels and precipitation (prcp), suggesting that rainfall has little effect on HCHO concentrations. By understanding these correlations, we gain valuable insights into how weather conditions may impact HCHO levels, informing further research or mitigation strategies to address potential environmental factors affecting air quality.

Spatial Distribution

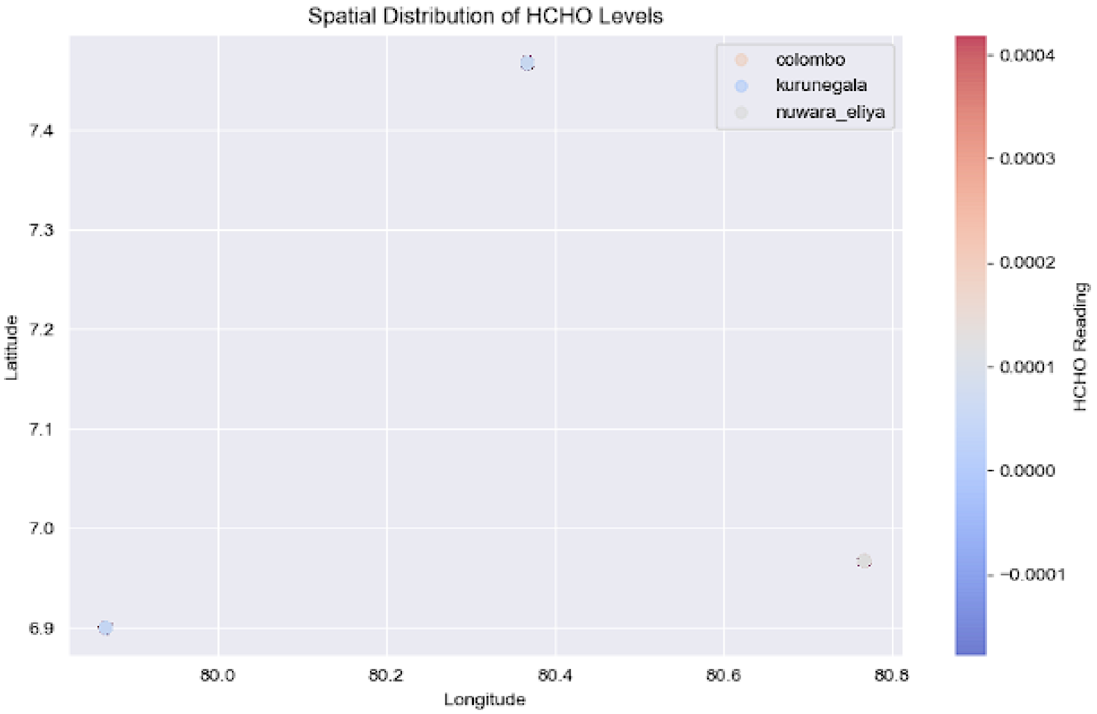


Figure 12(Spatial Distribution)

The spatial distribution of formaldehyde (HCHO) levels in three cities—Kurunegala, Nuwara Eliya, and Colombo—is represented graphically in this graph. A scatter plot is used to represent each city, with the geographical coordinates shown by the longitude and latitude axes. Each point's colour corresponds to its associated HCHO value; greater levels are shown by warmer colours, and lower concentrations are indicated by cooler colours.(Oregon State University, no date) We can see any spatial trends or differences in HCHO concentrations within and across the cities by looking at this graph. For example, regions with higher HCHO levels may be shown by areas with warmer points, and regions with lower concentrations may be indicated by areas with colder points. This visualization provides valuable insights into the geographical distribution of HCHO levels and can aid in understanding potential sources or environmental factors influencing air quality across the studied cities.

Time Series

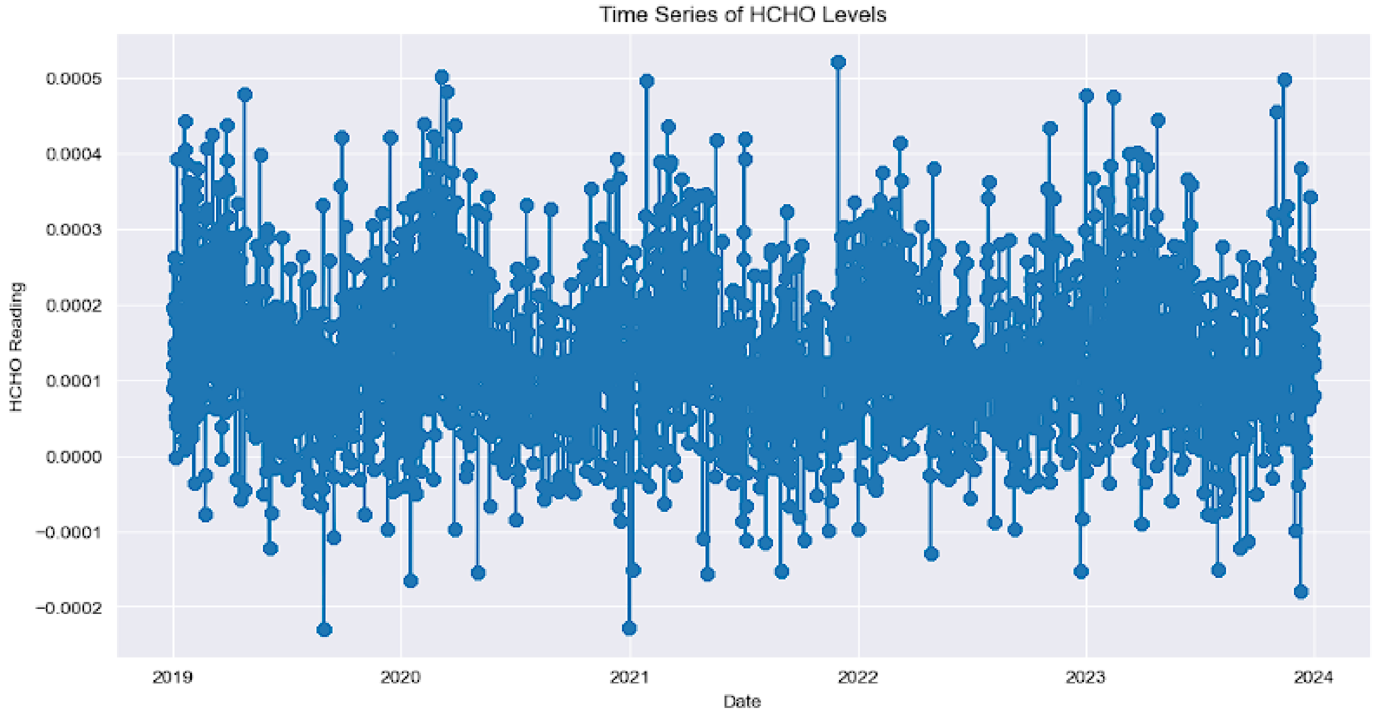


Figure 13(Time series)

The formaldehyde (HCHO) levels for the specified timeframe are shown as a time series in this graph. Dates are plotted on the x-axis, and HCHO measurements are plotted on the y-axis. Every data point in the plot represents a particular day and the associated HCHO level. This kind of visual representation of the time series data allows us to see any trends, patterns, or variations in HCHO concentrations over time. This figure offers a thorough summary of the variations in HCHO levels over the course of the recorded period, making it possible to see any trends or anomalies in the data.

Machine Learning Modelling

Model Selection

ARIMA (Autoregressive Integrated Moving Average)

- A popular statistical model for time series forecasting is called ARIMA. Moving average (MA), integrated (I), and autoregressive (AR) are its three constituent parts. The link between one observation and multiple lag observations is captured by the autoregressive component, which illustrates how previous values affect the current value. In order to make the time series stationery, which is required for ARIMA models, the integrated component uses differencing. The dependence between an observation and a residual error from a moving average model applied to lag observations is finally modelled by the moving average component. ARIMA can be used to make time series data stable through differencing, making it appropriate for data with non-seasonal patterns or trends. Short-term forecasting of stock prices, economic indicators, and other non-seasonal time series data is frequently done using it.(Chang *et al.*, 2012)

SARIMA (Seasonal Autoregressive Integrated Moving Average)

- SARIMA is an addition to the ARIMA model that is intended to manage time series data that exhibits variations or seasonal trends. To reflect the seasonal patterns in the data, SARIMA incorporates seasonal moving average (SMA) and seasonal autoregressive (SAR) terms in addition to the ARIMA components. While the seasonal moving average component models the dependency between an observation and the residual error from a moving average model applied to seasonally lagged observations, the seasonal autoregressive component captures the relationship between an observation and its seasonally lagged observations. When a forecasting model needs to take into consideration the seasonal component, like in the case of quarterly sales, monthly weather data, or annual production levels, SARIMA is a good fit. It is particularly useful when there are clear seasonal patterns in the data that cannot be adequately captured by standard ARIMA models.(Chang *et al.*, 2012)

Model Development

Colombo

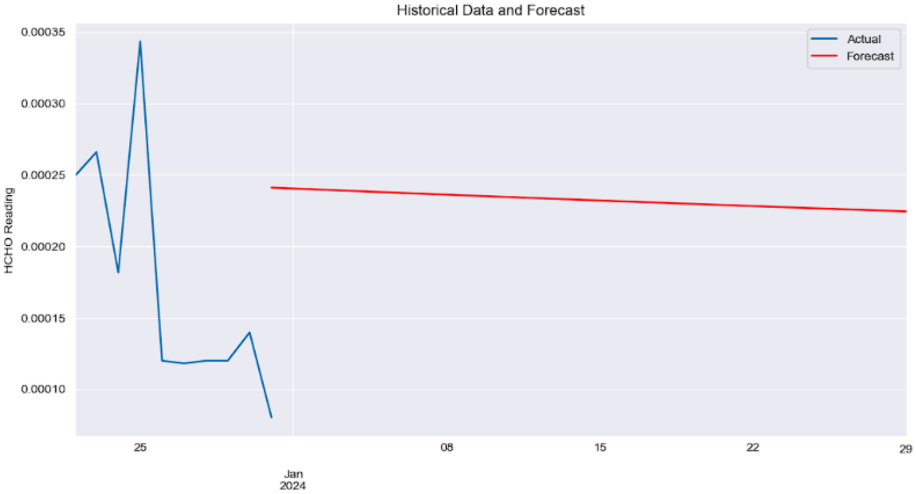


Figure 14(ARIMA Colombo)

The forecast covers a 30-day window into the future. We may evaluate how well the ARIMA model predicts HCHO levels over time by contrasting the real and predicted values. The ARIMA model is successfully capturing the underlying patterns and dynamics in the data if the predicted values closely resemble the observed trend. On the other hand, differences between the predicted and actual values can indicate regions for model improvement or locations where outside influences might be affecting HCHO levels in a different way than expected. All things considered, this plot offers insightful information on the precision and dependability of the ARIMA model's predictions for HCHO measurements in Colombo.

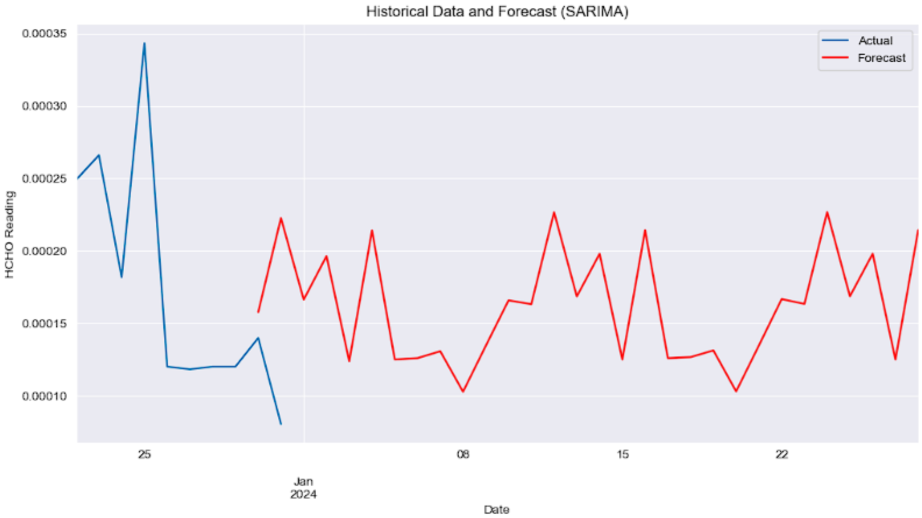


Figure 15(SARIMA Colombo)

The formaldehyde (HCHO) measurements for the city of Colombo throughout history are displayed in this plot along with the predicted values produced by a SARIMA (Seasonal Autoregressive Integrated Moving Average) model. The red line shows the expected HCHO values for the next 30 days, while the blue line shows the actual HCHO readings over time. We may assess how well the SARIMA model predicts HCHO concentrations by contrasting the predicted and real values. The SARIMA model effectively captures the underlying seasonal patterns and dynamics in the data if the predicted values closely match the observed trend. Conversely, deviations between the actual and forecasted values may signal areas where the model could be improved or where external factors may be influencing HCHO levels differently than anticipated. Overall, this plot provides valuable insights into the SARIMA model's forecasting performance for HCHO levels in Colombo.

Matara

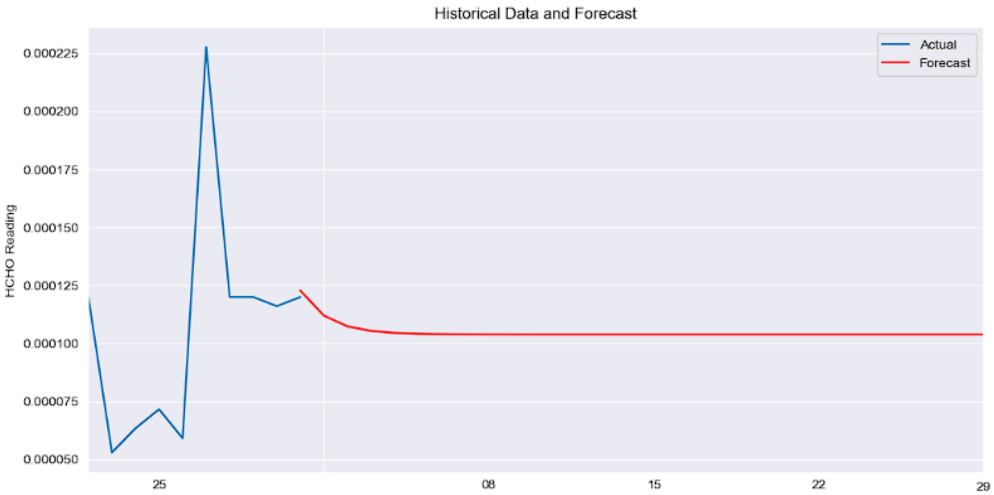


Figure 16(ARIMA Matara)

This graph shows the actual and the forecasting values of Matara using the ARIMA model, there is a simple slope to decreasing the HCHO gas level in 30 days.

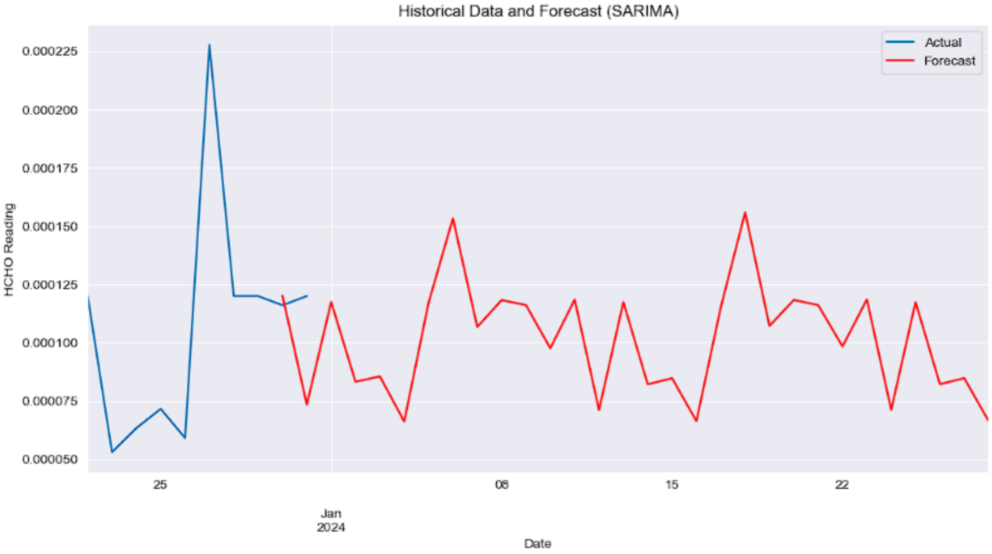


Figure 17(SARIMA Matara)

This graph shows the actual and the forecasting values using SARIMA model, there is a simple slope for decreasing the HCHO level according to the pattern.

Nuwara Eliya

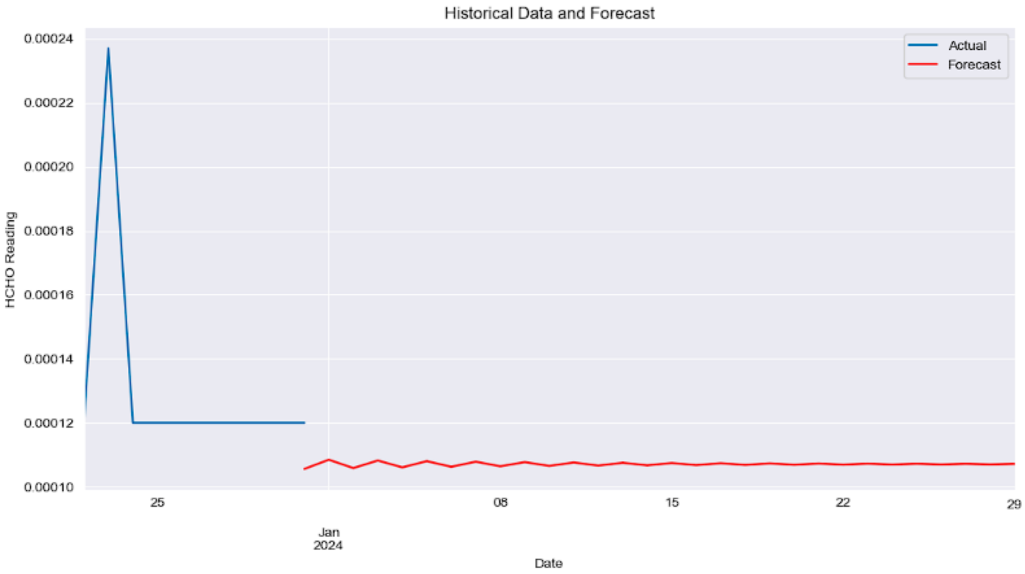


Figure 18(ARIMA Nuwara Eliya)

According to this graph, it says that the HCHO level will be lower than the 2023. HCHO level simply changed but there is no huge difference.

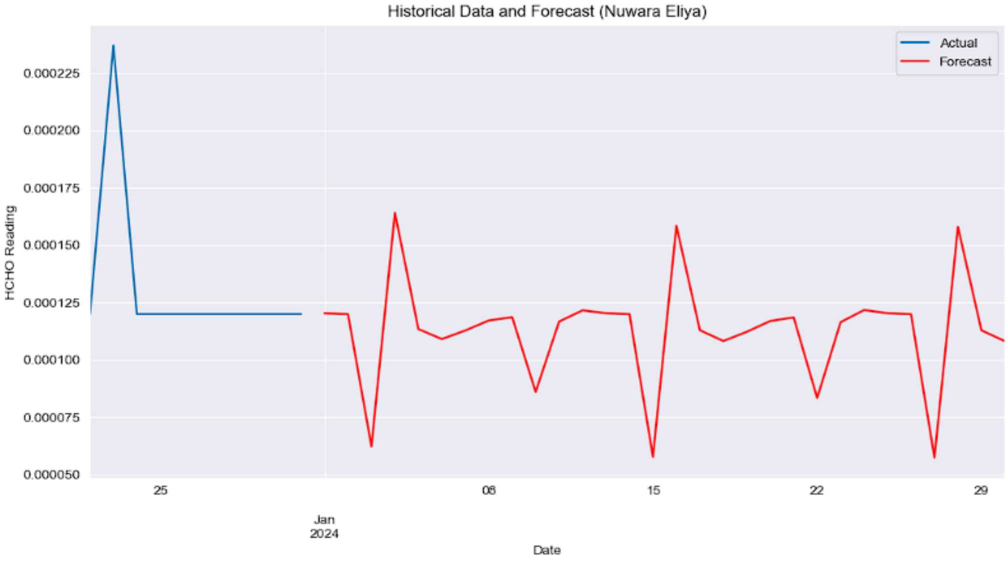


Figure 19(SARIMA Nuwara Eliya)

Above graph shows the actual and the forecasting values of the Nuwara Eliya there identified a simple decreasing pattern.

Kandy

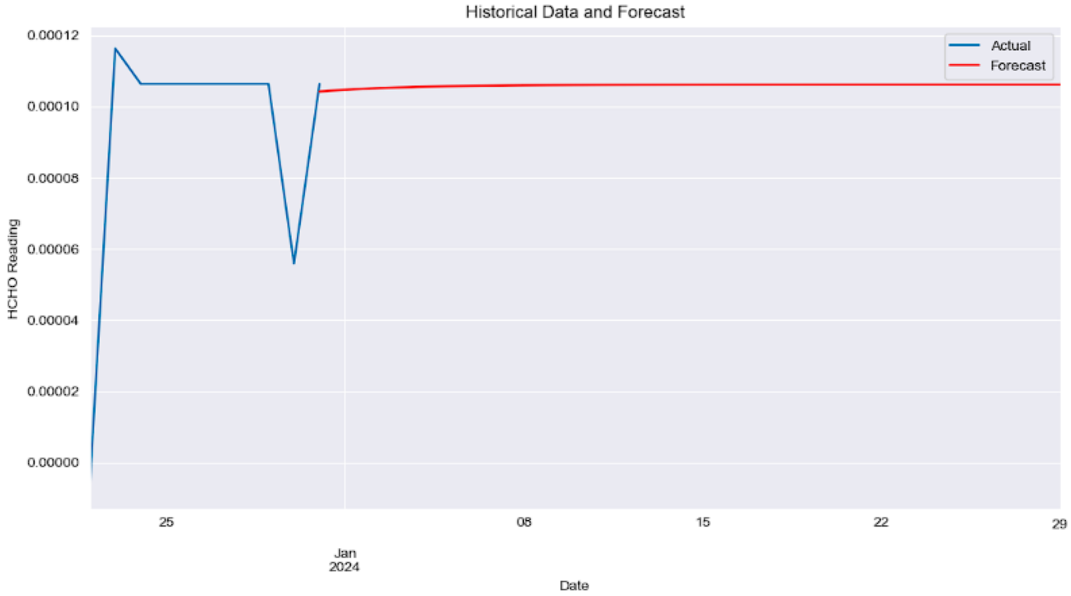


Figure 20(ARIMA Kandy)

This graph shows the Kandy actual and the forecasting HCHO values according to the ARIMA model. It says as a simple increment in the HCHO level than the 2023.

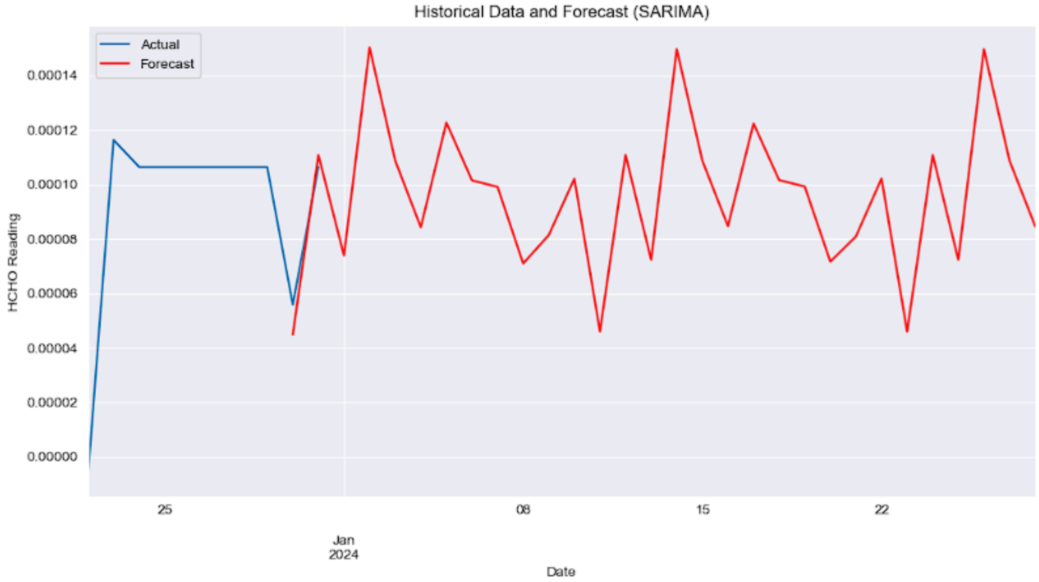


Figure 21(SARIMA Kandy)

According to the SARIMA model there is also a simple increment of the HCHO level with a pattern.

Monaragala

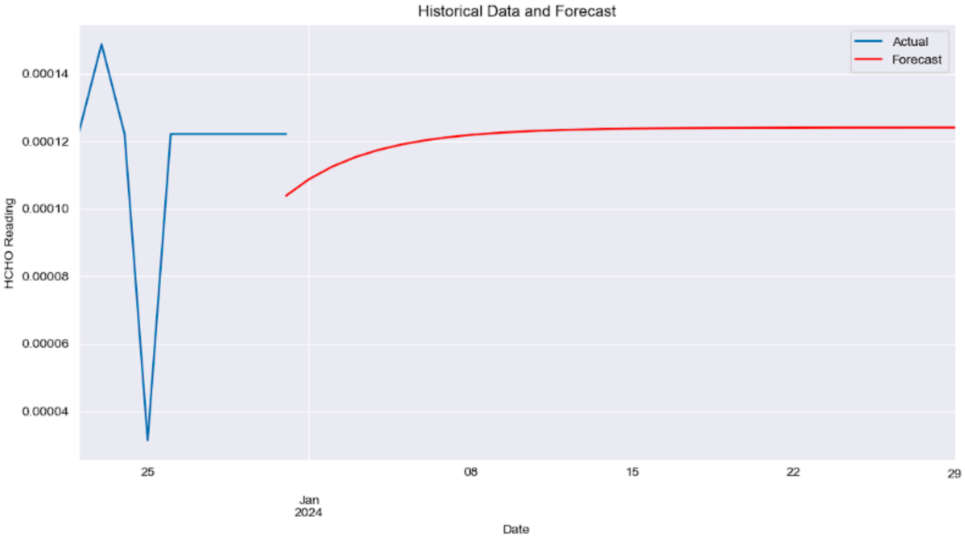


Figure 22(ARIMA Monaragala)

According to this graph HCHO level decreased at the start of the 2024 but it will be increased within month.

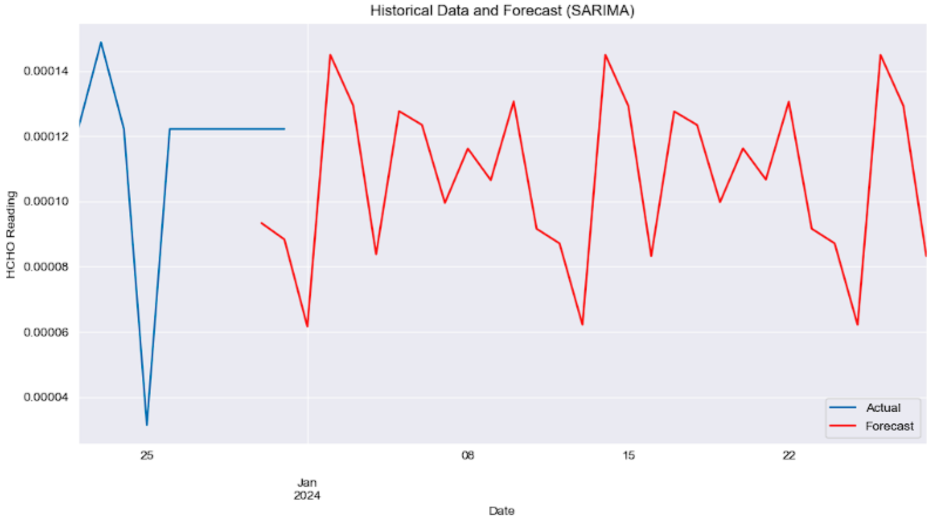


Figure 23(SARIMA Monaragala)

According to the SARIMA model this graph shows the huge low HCHO level at the start of the year but it suddenly increased.

Kurunegala

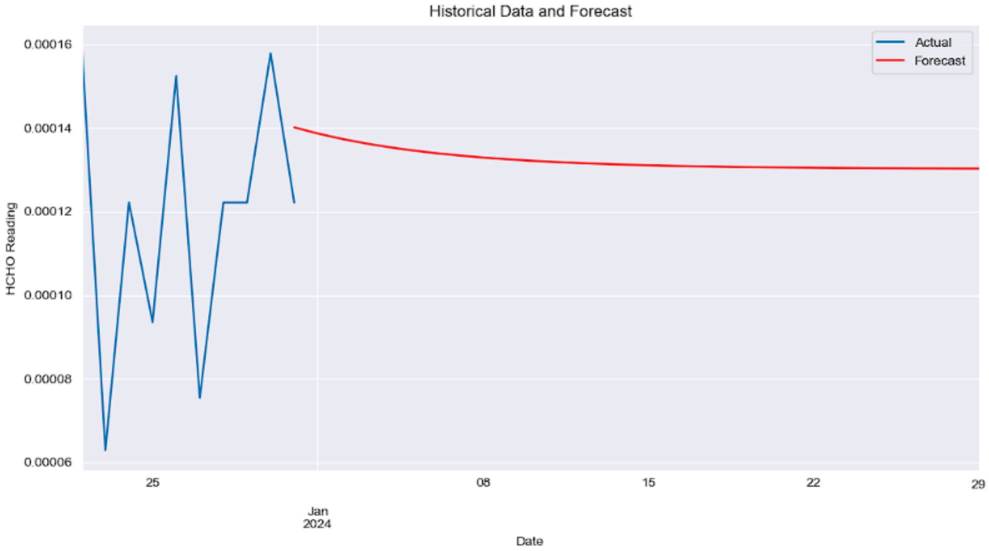


Figure 24(ARIMA Kurunegala)

According to the ARIMA model this graph shows the slope for decreasing the HCHO level in Kurunegala. Before 2024 there is a high level of HCHO gas but at the start of 2024 it slowly decreased.

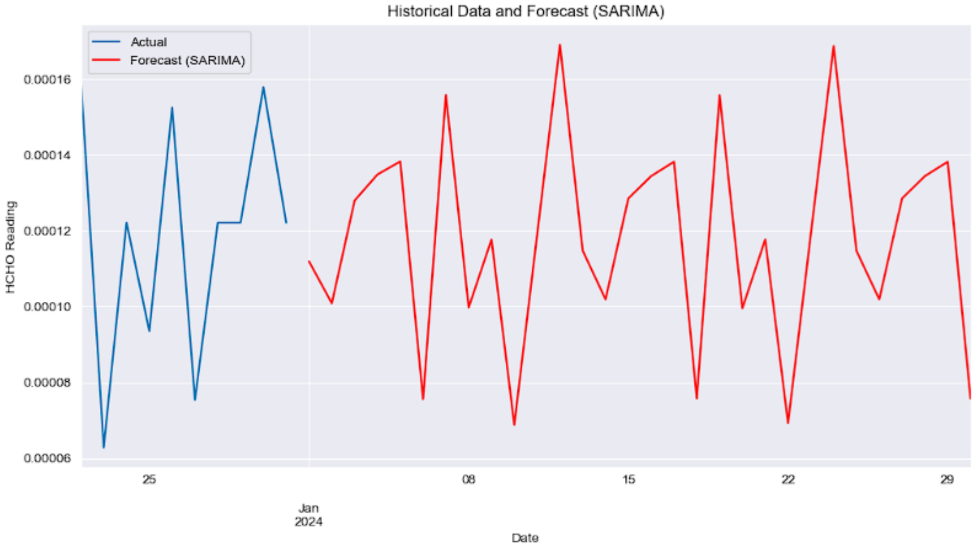


Figure 25(SARIMA Kurunegala)

According to the SARIMA model there is a simple slope for decreasing the HCHO level with a nice pattern.

Jaffna

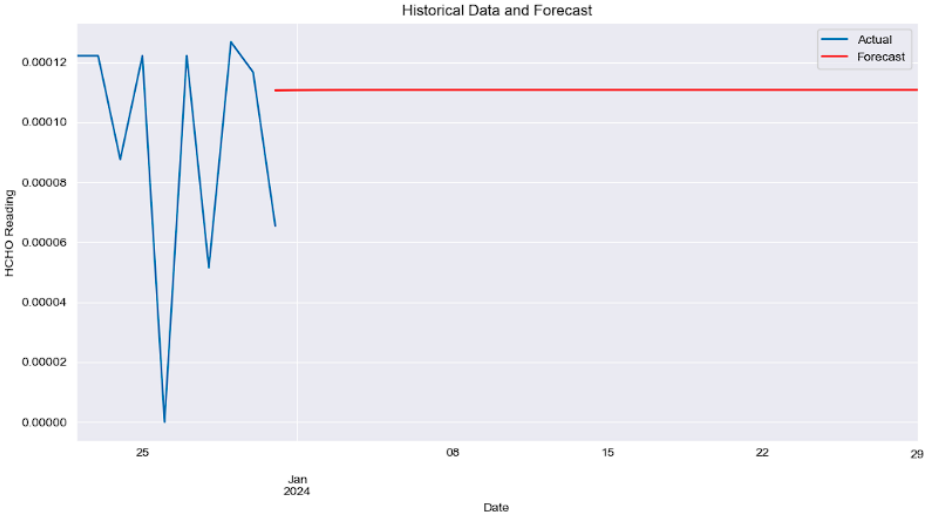


Figure 26(ARIMA Jaffna)

In Jaffna there is a low amount of HCHO gas and there is no change in the level of this gas. It will appear as a straight line.

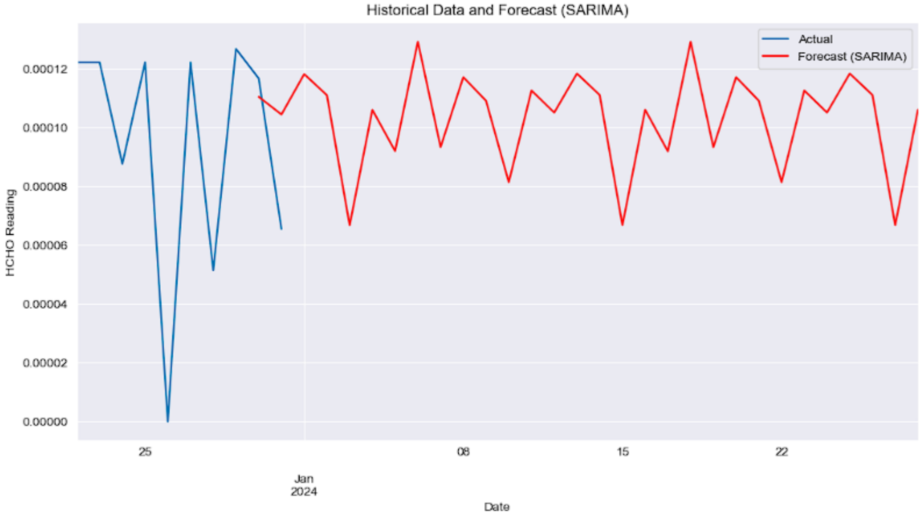


Figure 27(SARIMA Jaffna)

According to the SARIMA model there is a simple change in HCHO level when starting the 2024, it decreases slowly within a pattern.

Model Evaluation

Mean Squared Error

Mean Squared Error (MSE) is a key metric used to assess how well predictive models, including those for time series forecasting like ARIMA or SARIMA, perform in predicting future values. It calculates the average squared difference between predicted and observed values, emphasizing larger errors while providing a straightforward measure of model accuracy. Models with lower MSE values are considered better at prediction. MSE's simplicity, interpretability, and suitability for optimization algorithms make it widely adopted for comparing models and evaluating their effectiveness in various fields.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Root Mean Squared Error

Root Mean Squared Error (RMSE) is a useful metric because it provides an easily interpretable measure of the average magnitude of prediction errors. By taking the square root of the MSE, RMSE is in the same units as the target variable, making it directly comparable to the scale of the original data. This characteristic allows for straightforward interpretation, as smaller RMSE values indicate better model performance in predicting the target variable. Thus, RMSE is commonly used as a reliable measure of prediction accuracy, particularly in applications where understanding the magnitude of errors in the context of the original data is important, such as time series forecasting, regression analysis, and machine learning.

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{(\hat{y}_i - y_i)^2}{n}}$$

Table 2(Model Evaluation)

Location	Arima	Sarima
Colombo	MSE: 7.311566772027325e-09 RMSE: 8.550770007448057e-05	MSE: 8.022796349950196e-09 RMSE: 8.957006391618907e-05
Kandy	MSE: 3.920302859365808e-09 RMSE: 6.261232194517153e-05	MSE: 5.748964731697071e-09 RMSE: 7.582192777618538e-05
Kurunegala	MSE: 2.7944680055747465e-09 RMSE: 5.2862727942991614e-05	MSE: 3.907171960588082e-09 RMSE: 6.250737524955021e-05
Jaffna	MSE: 1.942016405841617e-09	MSE: 1.9898231279076447e-09
Monaragala	MSE: 5.637025791954689e-09 RMSE: 7.508012914183546e-05	MSE: 6.271229177849209e-09 RMSE: 7.919109279362932e-05
Matara	MSE: 1.1209550376568795e-08 RMSE: 0.00010587516411590018	MSE: 8.630595997210198e-09 RMSE: 9.290100105601768e-05
Nuwara Eliya	MSE: 9.233239015576148e-09 RMSE: 9.608974459106522e-05	MSE: 7.180037848133901e-09 RMSE: 8.47351039896329e-05

Power BI Visualization

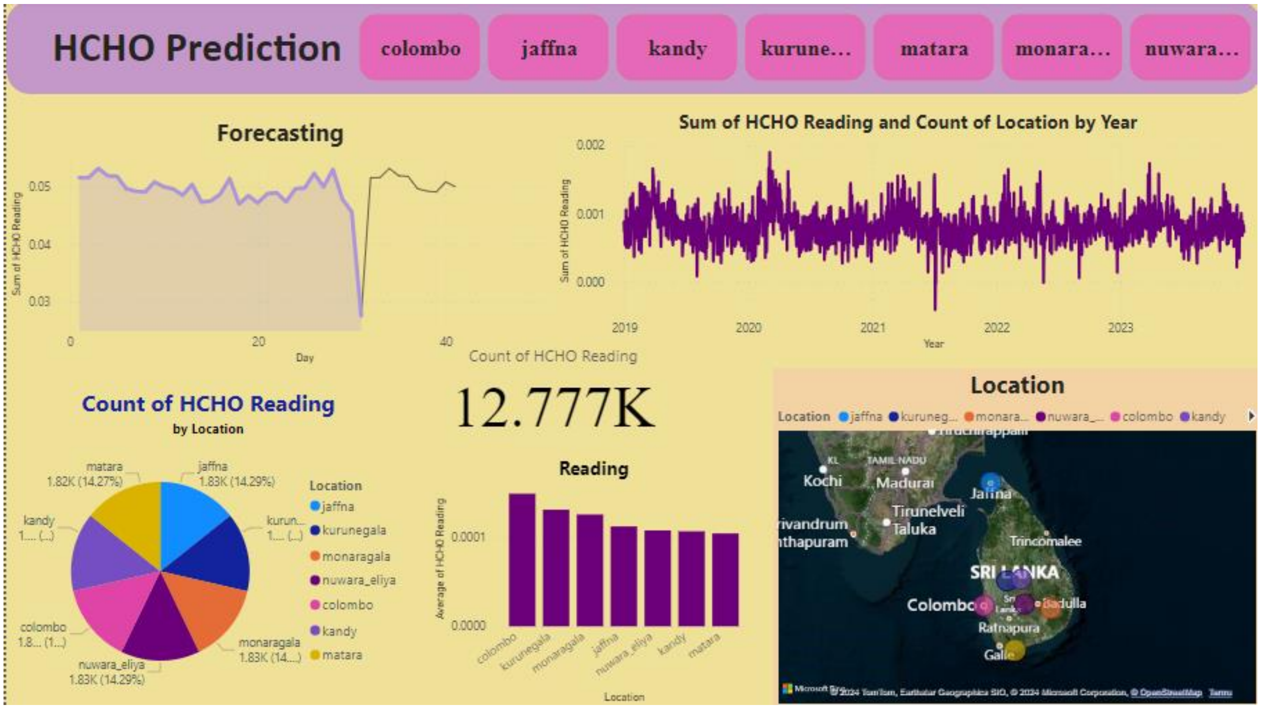


Figure 28(PowerBI)

Conclusion

To sum up, the study offers insightful information on the spatiotemporal dynamics of formaldehyde levels and how they relate to external influences. The results highlight the necessity of thorough air quality management and monitoring programs to handle the intricate interactions between man-made and natural factors that affect air pollution. The observed variances also emphasize the significance of focused policies and localized initiatives to lessen the harmful effects of air pollution on the environment and public health.

Recommendations

Based on the findings, several recommendations can be proposed to improve air quality monitoring and management efforts. These include:

1. Enhancing the spatial coverage of monitoring networks to capture localized variations in air pollution levels.
2. Implementing targeted interventions to reduce emissions from specific sources, such as vehicular traffic, industrial activities, and biomass burning.
3. Incorporating advanced modeling techniques to better understand the underlying drivers of air pollution and predict future trends.
4. Promoting public awareness and education campaigns to encourage sustainable behaviors and reduce individual contributions to air pollution.
5. Collaborating with relevant stakeholders, including government agencies, industry partners, and community organizations, to develop and implement effective air quality policies and initiatives.

Limitations

1. **Data Limitations**

   The analysis relied on available data from existing monitoring networks, which may have limitations in terms of spatial coverage, temporal resolution, and measurement accuracy. Variability in data quality and consistency across different monitoring stations could introduce uncertainties in the results.

2. **Sampling Bias**

   The representativeness of the monitoring stations and sampling protocols could introduce bias, particularly if certain locations or periods are over- or under-represented in the dataset. Biases related to site selection, instrument calibration, and data processing methods may affect the reliability of the findings.

3. **Spatial Heterogeneity**

   While the analysis considered multiple locations, it may not capture the full extent of spatial heterogeneity in air quality within each city or region. Localized sources of pollution, such as industrial facilities or traffic congestion hotspots, could influence HCHO levels in specific areas not adequately represented by the monitoring network.

4. **Temporal Variability**

   The analysis focused on monthly and seasonal trends but may not capture shorter-term variations or episodic events that could impact HCHO levels. Factors such as meteorological conditions, wildfires, or industrial accidents may lead to sudden spikes or fluctuations in pollution levels, which may not be fully captured by the available data.

5. **Interpretation Challenges**

   While efforts were made to identify potential drivers of HCHO variability, causality cannot be definitively established based on observational data alone. The analysis may highlight correlations between environmental factors and air quality but cannot conclusively determine causal relationships without additional evidence from controlled experiments or modeling studies.

6. **Extraneous Factors**

Other factors not considered in the analysis, such as socio-economic variables, population density, and land use patterns, could confound the observed relationships between environmental factors and air quality. Accounting for these extraneous factors would require more comprehensive data integration and statistical modeling approaches.

7. Generalizability

The findings may have limited generalizability beyond the study area or timeframe analyzed. Differences in geography, climate, and socio-economic conditions across regions could lead to distinct air quality dynamics not captured by the current analysis.

Improvements

1. **Expand geographic scope:** Include different areas with varying environmental conditions to get a broader understanding.
2. **Incorporate longitudinal data:** Use information collected over time to study how air quality and health change over the long run.
3. **Refine methodologies:** Improve the ways we measure air quality and health effects to make our findings more precise.
4. **Utilize advanced modeling:** Apply sophisticated techniques to better predict air quality and its impact on health.
5. **Integrate data from multiple sources:** Combine information from various places like satellites and sensors to get a more complete picture.
6. **Explore interdisciplinary approaches:** Look at how different fields of study can help us understand how air pollution affects health.
7. **Foster collaboration:** Work together with different groups, like scientists, policymakers, and communities, to tackle air quality issues.

Future Enhancement

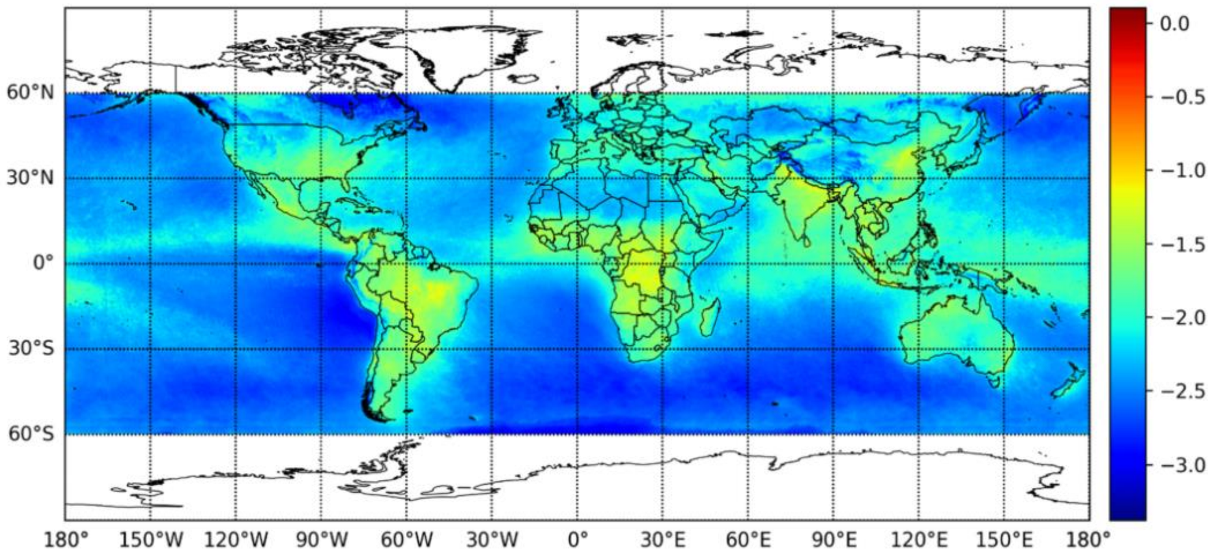
1. **Advanced sensor technology:** Develop more sensitive and affordable sensors to accurately measure air pollutants in real-time.
2. **Integration of AI and machine learning:** Utilize these technologies to improve predictive models for air quality and health outcomes.
3. **Enhanced data sharing platforms:** Create centralized platforms where researchers can access and share air quality and health data easily.
4. **Community-based monitoring programs:** Engage local communities in monitoring air quality and health outcomes to gather more comprehensive data.
5. **Targeted intervention strategies:** Develop interventions tailored to specific populations and geographical regions based on detailed air quality and health data.
6. **Policy implementation and enforcement:** Strengthen regulations and enforcement mechanisms to reduce air pollution and protect public health effectively.
7. **Investment in green technology:** Promote the development and adoption of cleaner energy sources and transportation systems to mitigate air pollution.

Similar Studies

Global Surface HCHO Distribution Derived from Satellite Observations with Neural Networks Technique

Link - https://www.mdpi.com/2072-4292/13/20/4055

This project investigates the distribution of the HCHO gas using Neural network with satellites observation.



References

Chang, X. *et al.* (2012) 'Seasonal autoregressive integrated moving average model for precipitation time series', *Journal of Mathematics and Statistics*, 8(4), pp. 500–505. Available at: https://doi.org/10.3844/jmssp.2012.500.505.

'Consensus Seasonal Weather Outlook February, March and April (FMA) Seasonal Rainfall and Temperature for Sri Lanka' (no date), pp. 1–18.

Hadavand-Siri, M. and Deutsch, C. V (2012) 'Some Thoughts on Understanding Correlation Matrices', *Centre for Computational Geostatistics Report*, 14(4), p. 408.

Oregon State University (no date) 'Chapter 13 Geovisualization Spatial Data Analysis', pp. 179–183.