Visual Data Analytics
CST4060

Sir Sudhakar Camilos Fernando

Coursework 2 Report

By

Name: Muhammad Wasif Ijaz
Class: MS Data Science (Full-time 2023-2024)
Misis No.: M00975792

Department of Computer Engineering & Informatics
Block 16, Knowledge Park, Dubai

Water Quality Analysis: Findings and Anomalies

1. Trends and Anomalies in Chemical Contamination

i. General Trends

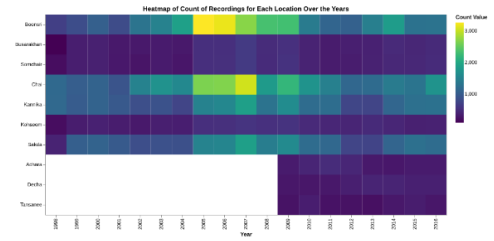
Time Period Focus (2006 - 2007): Notable peaks across various chemical measures during this period, as shown in time series graphs and heatmaps, indicate heightened chemical activity or contamination. Fluctuations in Oxygen Saturation: As per the time series data, there is variability across locations, with a significant decline at Kohsoom, suggesting deteriorating water quality.

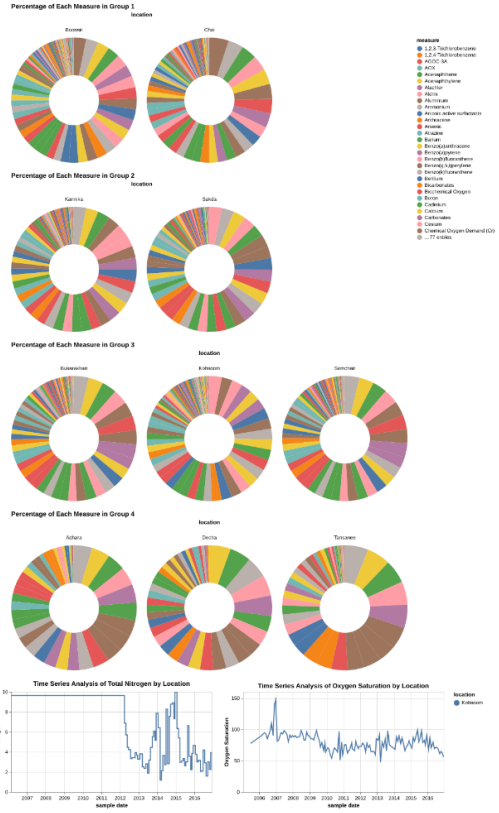
Reference Visuals:

- *Time Series Analysis of Total Nitrogen by Location*
- *Heatmap of Count of Recordings for Each Measure Over the Years*
- *Time Series Analysis of Oxygen Saturation by Location*
- *Percentage of each Measure by Group 1*
- *Percentage of each Measure by Group 2*
- *Percentage of each Measure by Group 3*
- *Percentage of each Measure by Group 4*

Dashboard 1:

1. ***Heatmaps***
    - ***Marks:*** *Colored cells.*
    - ***Channels:*** *Color intensity.*
    - ***Appropriateness:*** *Good for visualizing patterns over two dimensions (location and time).*
    - ***Effectiveness:*** *Color gradients quickly reveal trends and outliers.*
2. ***Donut Charts***
    - ***Marks:*** *Arc segments.*
    - ***Channels:*** *Angles and colors.*
    - ***Appropriateness:*** *Show proportional breakdowns within a whole.*
    - ***Effectiveness:*** *Visual comparison of parts to a whole is straightforward.*
3. ***Time Series Line Graph***
    - ***Marks:*** *Points connected by lines.*
    - ***Channels:*** *Position along axes.*
    - ***Appropriateness:*** *Ideal for showing data trends over time.*
    - ***Effectiveness:*** *Lines make it easy to follow changes and identify trends.*





ii. Trends: Changes Over Time and Location

Chemical Specific Trends

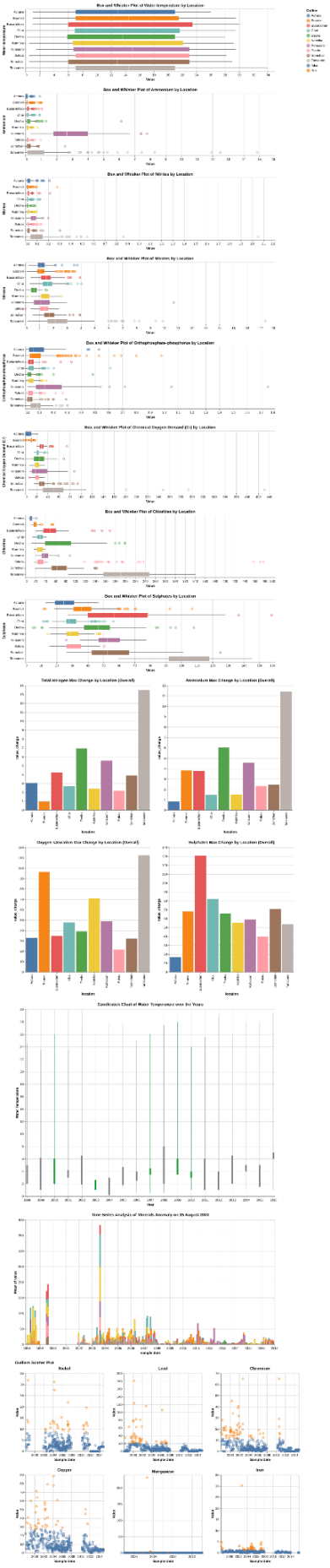
- **Ammonium:** Kohsoom shows a constant uptrend in Ammonium levels, peaking in 1998. The Ammonium concentration in Tansanee has been increasing, with a notable peak in November 2014.
- **Chlorodinine:** Specific abnormalities in Chlorodinine levels noted in Kohsoom.
- **Magnesium:** A consistent uptrend in Magnesium levels across all locations, suggesting a widespread increase.
- **Total Nitrogen:** Tansanee shows high spikes in recent years, particularly in the PCT5 bands.
- **Water Temperature:** Exhibits seasonal patterns, as reflected in the scatter plots and histograms.
- **Orthophosphorus-Phosphorus and Total Phosphorus:** Kohsoom has higher baselines and peaks, particularly in December 2013.
- **Chemical Oxygen Demand (COD), Chlorides, and Sulphates:** Elevated baselines in Tansanee and higher average sulphates in Busarakhan.

Location Specific Trends

- **Chai & Boonsri (Group 1):** These locations recorded high counts of data entries for measures like Ammonium, Nitrites, Nitrates, and Orthophosphates-phosphorus, indicating prevalent chemical presence or contamination concerns.
- **Kannika & Sakda (Group 2), Kohsoom, Somchair & Busarakhan (Group 3):** These groups show variations in dominant water quality measures, suggesting different contamination profiles or natural chemical distributions.

Reference Visuals:

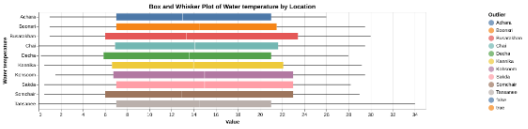
- *Box and Whisker Plot of Ammonium by Location*
- *Box and Whisker Plot of Chlorodinine by Location*
- *Box and Whisker Plot of Magnesium by Location*
- *Box and Whisker Plot of Total Nitrogen by Location*
- *Box and Whisker Plot of Chemical Oxygen Demand (Cr) by Location*
- *Box and Whisker Plot of Orthophosphate-phosphorus by Location*
- *Outliers Scatter Plot*
- *Time Series Analysis of Minerals Anomaly on 15 August 2003*
- *Sulphates Max Change by Location*
- *Total Nitrogen Max Change by Location*
- *Ammonium Change by Location*
- *Oxygen Saturation Max Change by Location*
- *Candlestick Chart of Water Temperature over the Years*

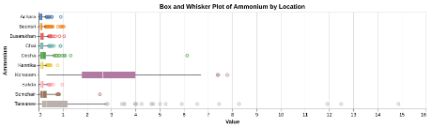


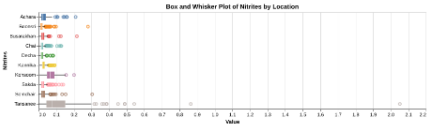
Dashboard 2:

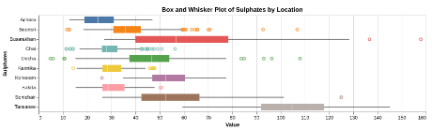
1. ***Bar Charts***
    - ***Marks:*** *Bars.*
    - ***Channels:*** *Length and Color.*
    - ***Appropriateness:*** *Compare quantities across categories.*
    - ***Effectiveness:*** *Easy to interpret differences in lengths.*

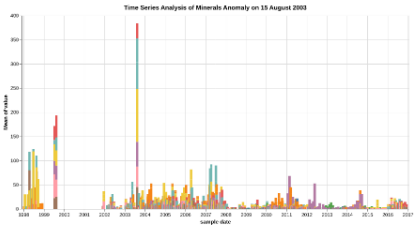
2. ***Box and Whisker Plots:***
    - ***Marks:*** *Boxes with lines (whiskers).*
    - ***Channels:*** *Position on the scale.*
    - ***Appropriateness:*** *Summarize distribution of data.*
    - ***Effectiveness:*** *Show median, quartiles, and outliers concisely.*
3. ***Histogram:***
    - ***Marks:*** *Bars.*
    - ***Channels:*** *Height of bars and Colors.*
    - ***Appropriateness:*** *Display frequency distribution.*
    - ***Effectiveness:*** *Easy visualization of data spread and skewness.*
4. ***Candlestick Chart:***
    - ***Marks:*** *Rectangles with lines.*
    - ***Channels:*** *Position and color.*
    - ***Appropriateness:*** *Show high, low, opening, and closing values over time.*
    - ***Effectiveness:*** *Compact representation of a range of data points.*
5. ***Outliers Scatter Plot:***
    - ***Marks:*** *Points.*
    - ***Channels:*** *Position on two axes.*
    - ***Appropriateness:*** *Identify outliers and patterns.*
    - ***Effectiveness:*** *Spot clusters and deviations easily.*

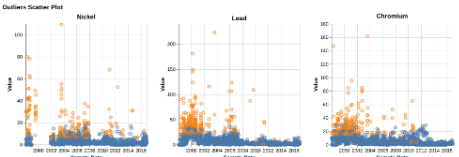


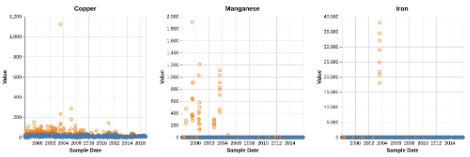












iii. Anomalies: Sudden Change Over Time or One Site Significantly Different from Others

- **Extreme Values in Mineral Concentrations:** The data points provided show significant spikes in mineral levels on specific dates, which are highly anomalous:
    - **Nickel:** Exceptionally high levels recorded in Boonsri, Kannika, Chai, and Somchair on different dates.
    - **15 August 2003 Effect:** Spikes in Chromium, Manganese, Copper, and Iron across multiple locations, suggesting possible calibration errors or extraordinary events.
- **Lead Contamination:** A steep uptrend in lead levels observed in Somchair.
- **Hot Effluents in Tansanee:** An unnatural uptrend in water temperature, particularly in August 2016, hinting at possible industrial influences.
- **Total Hardness Data Gaps:** The timeline visualization from 2005 to 2017 shows red bars indicating data entries and grey areas denoting periods without data, which may be attributed to sensor errors or malfunctions, deliberate data tampering or omission (possibly to hide environmental disturbances or comply with regulations), or intentional data alteration. The nature and regularity of these gaps suggest various possibilities, ranging from routine maintenance to tampering or technical issues, necessitating further investigation for conclusive understanding.

iv. Key Findings besides Trends and Anomalies

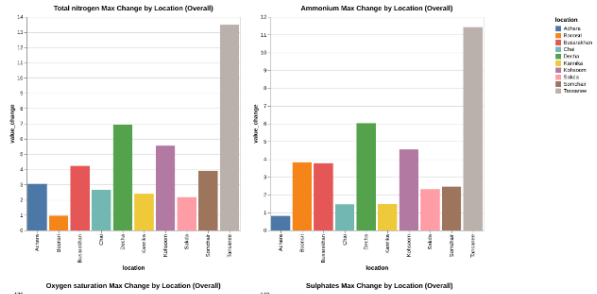
- **Oxygen Saturation Decline in Kohsoom and Tansanee:** The time series data for Kohsoom and Tansanee shows a notable declining trend in oxygen saturation, pointing to a potential degradation in water quality over time due to chemical contamination.
- **Seasonal Temperature Patterns:** Indicative of natural environmental cycles, as well as potential disturbances.
- **Chemical Spikes and Uptrends:** Sudden spikes in various chemicals on specific dates, like August 2003, point to either major contamination events or significant data recording errors.

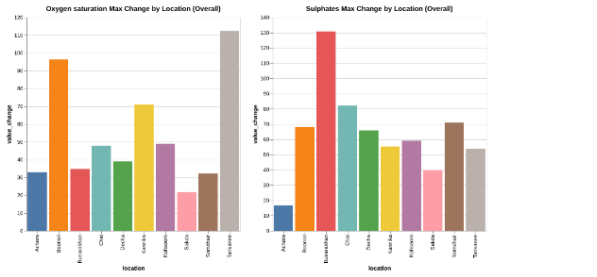
Reference Visuals:

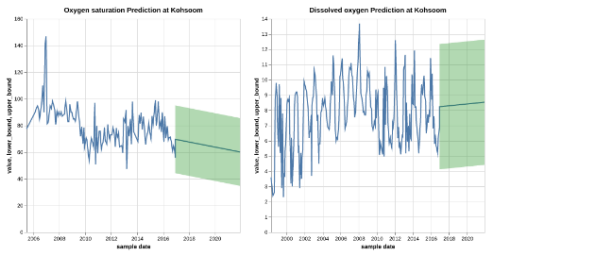
- *Oxygen Saturation Max Change by Location*
- *Oxygen saturation Prediction at Kohsoom*
- *Dissolved oxygen Prediction at Kohsoom*
- *Oxygen saturation Prediction at Tansanee*
- *Dissolved oxygen Prediction at Tansanee*
- *Correlation between Dissolved Oxygen and Water Temperature*
- *Histogram of Annual Average Water Temperature in Kohsoom*
- *Histogram of Annual Average Standardized Dissolved Oxygen in Kohsoom*
- *Series Analysis of Environmental Contamination Factor*

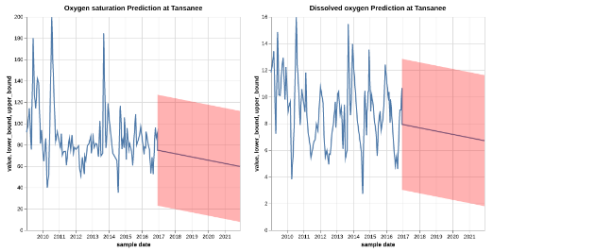
Dashboard 3:

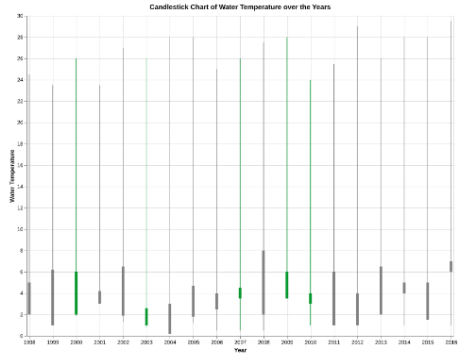
1. ***Bar Chart (Gantt-like Chart)***
    - ***Marks:*** *Horizontal bars.*
    - ***Channels:*** *Length and position along axis, color for categories..*
    - ***Appropriateness:*** *Show activity or occurrences over time by location.*
    - ***Effectiveness:*** *Spot clusters, outliers, and trends in data.*











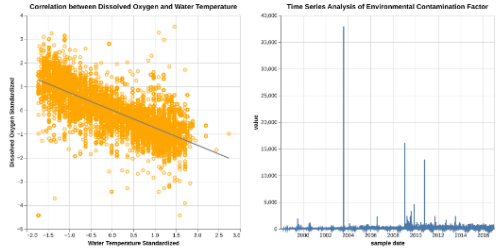
2. Data Quality and Uncertainty Issues

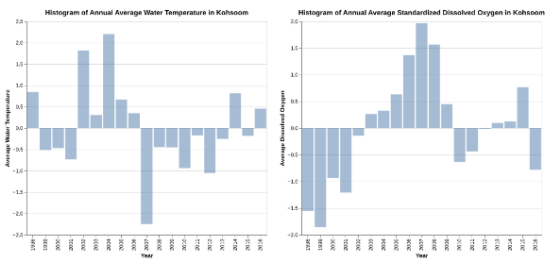
i. Missing Data

The absence of complete years of data for several locations revealed by heatmap and timeline of entries indicates gaps in monitoring, affecting the reliability of long-term trend analysis.

Reference Visuals:

- *Heatmap of Count of Recordings for Each Location Over the Years*





ii. Change in Collection Frequency

Variations in collection frequency, as evidenced by the patchy distribution in the heatmap, can impact the consistency and reliability of

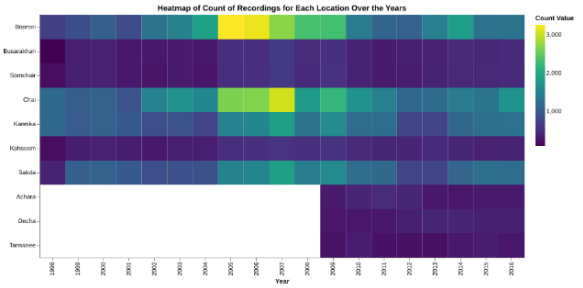
trend analysis. Total Hardness entries trend graph is evidence for inconsistency in data collection.

Reference Visuals:

- *Time Series Analysis of Minerals Anomaly on 15 August 2003*

Dashboard 4:

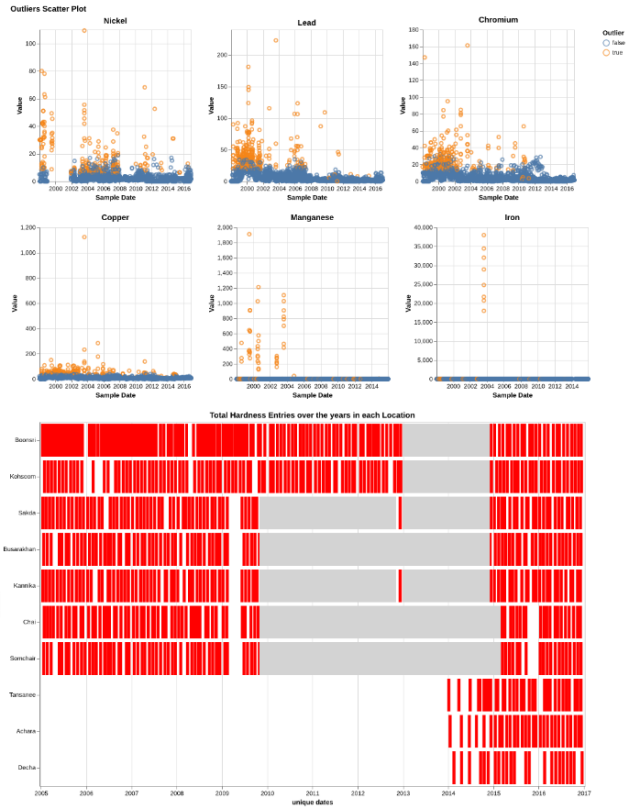
1. ***Line Graphs with Confidence Intervals***
    - ***Marks:*** *Points connected by lines; shaded areas for confidence intervals.*
    - ***Channels:*** *Position along the x-axis (time) and y-axis (measured variable); color to distinguish between different data sets.*
    - ***Purpose:*** *Predict oxygen saturation and dissolved oxygen with uncertainty estimation.*
    - ***Appropriateness:*** *Good for visualizing trends over time with an indication of variability.*
2. ***Scatter Plot with Regression Line***
    - ***Marks:*** *Points for data; line for regression.*
    - ***Channels:*** *Position along both x-axis and y-axis; color for regression line.*
    - ***Purpose:*** *Illustrate the correlation between two variables (dissolved oxygen and water temperature).*
    - ***Appropriateness:*** *Ideal for showing the relationship and trend between two continuous variables.*



iii. Unrealistic Values

Extremely high or unusual chemical concentrations on specific dates, such as the spikes observed on 15 August 2003, indicate potential contamination events or data recording errors.

- **Nickel:** Values like 109.32 in Kannika and 55.5 in Chai are unusually high and could indicate contamination incidents or data errors.
- **Lead:** Extremely high concentrations, such as 222.94 in Kohsoom, are alarming and could point to severe contamination events.
- **Chromium:** Peaks like 161.27 in Kohsoom suggest significant contamination or measurement inaccuracies.
- **Copper:** The value of 1123.13 in Kohsoom is exceptionally high, possibly indicating a major contamination event.
- **Manganese:** Extremely high levels, particularly the 1910 in Chai, raise questions about data accuracy or severe contamination.



- **Iron:** The extraordinarily high values across various locations, especially 34413.96 in Kannika, are indicative of either severe contamination events or data recording errors.

Reference Visuals:

- *Total Hardness Entries over the years in each Location*

Dashboard 5:

1. ***Bar Chart-Like Visualization***
    - ***Marks:*** *Vertical lines/bars.*
    - ***Channels:*** *Position along the x-axis (year) to represent time; saturation of color or absence/presence of a mark to indicate the occurrence of data.*
    - ***Purpose:*** *Display presence of total hardness entries over years in different locations.*
    - ***Appropriateness:*** *Useful for showing activity or data entry over time and identifying gaps in data collection.*

3. Proposed Changes to Sampling Strategy

- **Regular Calibration and Maintenance:** To ensure accurate and reliable data from the monitoring equipment.
- **Standardized Data Collection Protocols:** Implementing standardized protocols across all locations for consistency.
- **Increased Frequency and Diversity of Sampling:** Focus especially on areas like Kohsoom and Tansanee, where significant trends and anomalies have been observed.
- **Long-Term Data Consistency:** Ensure a consistent set of chemicals is monitored over time for reliable trend analysis.
- **Investigate and Address Data Gaps:** Fill in and prevent future data gaps through thorough investigation.
- **Enhanced Data Review and Validation:** To identify and correct unrealistic values and anomalies in the dataset.
- **Environmental Impact Assessment:** Conduct assessments in areas with significant contamination indicators.
- **Community and Stakeholder Engagement:** Engage local communities and stakeholders for a comprehensive approach to water quality monitoring.

4. Conclusion

This analysis, backed by references to various graphs such as time series graphs, candlestick, heatmaps, scatter plots, and histograms, highlights the complexities of water quality monitoring across different locations. The trends and anomalies identified, along with the data quality issues, underscore the need for enhanced monitoring strategies and methodologies. The proposed changes to the sampling strategy are aimed at improving data reliability and facilitating better environmental management and policy decisions. Understanding these environmental challenges is vital for effective monitoring, remediation efforts, and ensuring the health of aquatic ecosystems and surrounding communities.

Story Telling

My Visual Data Analytics Coursework-2 involved investigating a pressing concern: the decline of the Rose-Crested Blue Pipit in the area surrounding the operations of Kasios Office Furniture. This decline raised questions about the possible impact of the company's environmental practices. My objective was to delve into water quality data to uncover any links between these industrial activities and the habitat changes affecting the bird's population. Here is how I approached and interpreted the situation:

My Analysis Journey: Saving the Rose-Crested Blue Pipit

Understanding the Problem

I noticed a worrying trend in the region. The once-thriving population of the Rose-Crested Blue Pipit was dwindling, coinciding with the activities of Kasios Office Furniture. My task was to find out if there was a connection between the two?

Analyzing Water Quality Data

I began by thoroughly examining the water quality data across various locations. I focused on identifying trends and anomalies in chemical concentrations over time. I carefully analyzed key parameters like Ammonium, Chlorodinine, Magnesium, and others, across different locations and over several years.

Discovering Troubling Trends

My analysis revealed several concerning patterns. For instance, a significant increase in Ammonium levels at Kohsoom, particularly noticeable in 1998, and abnormal Chlorodinine levels in the same area, indicated potential issues. A consistent rise in Magnesium levels across all locations suggested a broader environmental impact, possibly linked to industrial discharges.

Highlighting Anomalies

The data pointed to a critical date: August 15, 2003. This date marked unusual spikes in various minerals, including Chromium, Lead, Copper, and Manganese. These spikes hinted at potential extraordinary events, possibly related to industrial activities or accidents at Kasios.

The Evidence Mounts

Further analysis showed alarming gaps in total hardness data. These gaps could signify periods of intentional data omission or sensor malfunction, possibly due to industrial interference. The unusual uptrends in water temperature and declining oxygen saturation in areas like Kohsoom and Tansanee, where the Pipits were once abundant, connected the dots between the bird's disappearance and water quality deterioration.

Linking the Dots

The activities of Kasios Office Furniture were significantly impacting the water quality, leading to a disturbed habitat for the Rose-Crested Blue Pipit. The chemical contaminants and temperature variations were clear indicators of environmental stress, likely contributing to the bird's decline.

Proposing Solutions

Based on my findings, I propose a comprehensive revision of the sampling strategy. This includes regular calibration and maintenance of monitoring equipment, standardized data collection protocols, and enhanced data review and validation processes. These changes will aim at improving data accuracy, allowing for more effective monitoring and remediation efforts.

My Conclusion

The data analysis clearly indicates a strong link between the environmental practices of Kasios Office Furniture and the habitat changes affecting the Rose-Crested Blue Pipit. By addressing these water quality issues and implementing better monitoring strategies, we can work towards mitigating the impact on the Pipit's habitat and possibly reversing its population decline.