Evelyn Reyes Dr. Wissam Kontar

Cive 202 – Civil Engineering Analysis II

Spring 2025

Scope of Work

Team Member: Evelyn Reyes

Date submitted for review: January 31st, 2025

Submitted to: Civil and Environmental Engineering Department Faculty

Project Goals:

The objective of this project is to conduct an in-depth statistical examination of publicly accessible drinking water data sourced from Nebraska, in response to a request for proposal (RFP) issued by our esteemed client, Zenith Data Analysis. This undertaking is designed not only to uncover insights hidden within the data but also to address three key analyses requested explicitly by the client, each aimed at providing a clearer understanding of the state's water quality landscape.

1. A detailed overview of the number of water systems in Nebraska experiencing contaminant violations, categorized by type of system and type of contaminant.

2. A statistical examination of various water system types, including essential statistical parameters such as the mean and median.

3. A targeted analysis of maximum contaminant level violations in active systems, organized by population category and source water type. The client requested a summary table for this portion of the analysis.

Project Task:

To effectively achieve the objectives established for this project, I have outlined a series of essential tasks that our proficient engineering team must execute. These tasks are critical to the successful completion of the project and will ensure that we meet our goals efficiently.

To accomplish the above project goals, I have distinguished the following tasks to be completed:

Task 1: Conduct a comprehensive assessment of the number of violations encountered by each category of the water system.

- Raw data from the Nebraska Safe Drinking Water Information System (SDWIS) is extracted.
- The data is processed in Excel to remove unnecessary fields, reformat dates, and rename columns for compatibility with Python analysis.
- Deliverables include both the original raw data and the cleaned dataset.

Task 2: Compile a summary of the types of contaminants in drinking water that most frequently coincide with violations.

- Data is categorized based on system type and contaminant type. Statistical measures (mean, median, min, max) are computed.
- Contaminant analysis by population size is conducted.
- Deliverables include the Python script and an annotated code document.

Task 3: Finding Compilation

- Analysis results are compiled into a formal report.
- Tabular and visual representations of the findings are included for clarity.
- Develop a Jupyter Notebook python script

Deliverable:

Deliverable	Files Included	Due to Client By:
Raw Data File	Nebraska_Violations_Raw.xlsx	1-31-2025
Organized Data File	Nebraska_Violations_Clean.csv	1-31-2025
Python Code	Nebraska_Violations_Analysis.py	2-13-2025
Annotated Code Document	ACD_Nebraska_Violations_Analysis.docx	2-13-2025
Final Memo	Nebraska Water Systems Evaluation.docx	2-13-2025

Finding and Analysis:

Question 1:

Table 1: Statistical summary of violations in Nebraska

Pws_type	Min	Max	Mean	Median
Community watersystem	1	5348	103.42	10
Non-Transient non-community system	1	315	29.04	1
Transient non-community system	5	2059	359.44	29

The analysis of SDWIS data reveals significant variation across different types of public water systems. Community systems show a wide distribution of violations, with a maximum of 5,348 violations. Non-transient non-community systems have lower average violations, while transient systems display high average violations.

To analyze this data, I started by reading the data from a .csv and assigned it to a Clean Data File called 'violation_report_cleaned.csv'. Next, I grouped the information by two key factors: the type of water system 'pws_type' and the specific contaminant causing violations 'contaminant_name'. This grouping allowed us to generate a summary table detailing the number of violations associated with each combination of water system type and contaminant. We then further analyzed this data by aggregating the summary statistics for each water system type for the mean, median, min, and max.

Question 2:

Table 2: Evaluation of Contaminants by PWS_Type and System_size

pws_type	contaminant_name	system_size	quantity
Community water system	1,1,1-Trichloroethane	Small	8
Community water system	1,1,1-Trichloroethane	Very Small	2
Community water system	1,1,2-Trichloroethane	Small	2

Community water system	1,1,2-Trichloroethane	Very Small	2
Community water system	1,1-Dichloroethylene	Small	7
Community water system	1,1-Dichloroethylene	Very Small	2
Community water system	1,2,4-Trichlorobenzene	Small	8
Community water system	1,2,4-Trichlorobenzene	Very Small	2
Community water system	1,2-DIBROMO-3-CHLOROPROPANE	Very Small	2
Community water system	Xylenes	Small	8
NTNCWS	Arsenic	Medium, Small, Very Small	47
NTNCWS	Coliform	Medium, Small, Very Small	254
NTNCWS	Lead and Copper	Small, Very Small	54
NTNCWS	Nitrate	Medium, Small, Very Small	145
NTNCWS	Revised Total Coliform Rule (RTCR)	Small, Very Small	117
TNCWS	Coliform (TCR)	Medium, Small, Very Small	199
TNCWS	Lead and Copper Rule	Small, Very Small	30
TNCWS	Nitrate-Nitrite	Medium, Very Small	508
TNCWS	Public Notice	Small, Very Small	213
TNCWS	Revised Total Coliform Rule (RTCR)	Medium, Small, Very Small	376

Using data extracted from Nebraska_Violation.csv, I was able to compile a list of the most common drinking water contaminants present in each water system. I further analyzed these results by categorizing the systems based on population size and the frequency of reported violations for the same contaminants. Following the classifications outlined by the USEPA, we defined populations according to the parameters detailed in Table 2.1. These findings can help identify patterns in the contaminants that frequently occur in water systems of similar size and type. Groundwater systems exhibit the highest number of MCL violations. Arsenic and nitrate are among the most common contaminants. Small community systems, particularly those relying on groundwater, are more prone to violations.

To arrive at these conclusions, I began by gathering data on the populations served from the main dataset and converting it into a Clean Data File labeled ‘population.’ Next, we established parameters for ‘population’ using an “Else-If” function to assign the appropriate population category to each water system. We then filtered the results to focus solely on active systems. Finally, we summarized the outcomes by grouping them according to water system type, contaminant name, system size, and the frequency of coinciding conditions.

Table 2.1: USEPA Population Classification

Size Classification	Population
Very Small	500 or Less
Small	501 – 3,300
Medium	3,301 – 10,000
Large	10,001 – 100,000

Question 3:

Table 3: Violation Count, Water Source and Count

Pws_type	Primary source	Contaminant	Violation Type	MCL Violation Count
CWS	Ground Water	EOX	MCL – Average	425
CWS	Ground Water purchased	EOX	MCL – Single Sample	252
CWS	Ground Water under influence of surface water	EOX	MCL – Average	109
CWS	Surface Water	SF3	MCL – Average	1
CWS	Surface Water purchased	SFJ	MCL - Average	11

The findings provide valuable insights into Nebraska's water quality, indicating a need for improved monitoring and mitigation measures for groundwater-supplied systems. The contaminant EOX is present across all population categories, with notable concentrations in very small systems. Additionally, SIF is the most frequently observed contaminant overall. Groundwater sources, especially in small systems, exhibit the greatest MCL violation counts, emphasizing the need for targeted monitoring and preventive measures.

To analyze the data specifically for small water systems with MCL violations, we filtered the dataset to include only active systems categorized under "Maximum Contaminant Level Violation, Average" or "Maximum Contaminant Level Violation, Single Sample."

We then isolated data pertaining to small systems by selecting entries where the 'system_size' column was designated as "Small." By grouping this filtered data according to primary water source, violation type, and system size, we created a summary table that outlined the number of violations for each combination. Finally, we refined the summary to focus exclusively on small systems and exported the results to a .csv file.