Water Quality Data Analysis – Summary Report

# 1. Introduction

This summary highlights key findings from the preliminary analysis of stream-based drinking water quality data. The analysis focused on data preparation, exploration, and insights into categorical associations and reporting patterns.

# 2. Methodology

Python-based processing included data cleaning, type conversions, and encoding of categorical values. Exploratory analysis used frequency distributions, cross-tabulations, and correlation metrics such as Cramér’s V and Spearman rank correlation to identify key patterns and associations.

# 3. Key Findings

- Certain zones showed consistently high levels of unprotected water sources.  
- Significant association was observed between water source type and indicators like taste, color, and odor.  
- Correlation heatmaps revealed clusters of related survey responses.  
- Basic quality indicators vary significantly across regions, suggesting a need for zone-specific interventions.

# 4. Visual Highlights

Key visual outputs supported the analysis, including bar plots of category frequencies and a correlation heatmap highlighting relationships among categorical indicators.

[Figure 1: Frequency of Water Source Types by Zone]

[Figure 2: Spearman Correlation Heatmap of Categorical Variables]

# 5. Conclusion

This preliminary analysis identified critical patterns in water quality data. It establishes a foundation for further exploration, including advanced statistical modeling and integration with geographic data to support targeted water quality interventions.