

DATA SCIENCE MINI PROJECT

CGWB - CHANGES IN DEPTH TO WATER LEVEL

PRESENTED BY :
GUNJAL CHAUHAN -
22070521024
7TH A

PROBLEM STATEMENT

1

Groundwater is vital for agriculture, domestic, and industrial use in India .Rising groundwater depletion poses serious environmental and sustainability challenges.

2

Regional and seasonal fluctuations in water levels remain poorly analyzed. Outdated or fragmented data limits effective policy and planning.

3

There is a strong need for data-driven insights using authentic CGWB datasets to identify:

- Water-stress regions
- Long-term groundwater trends
- Impacts of monsoon and human activity

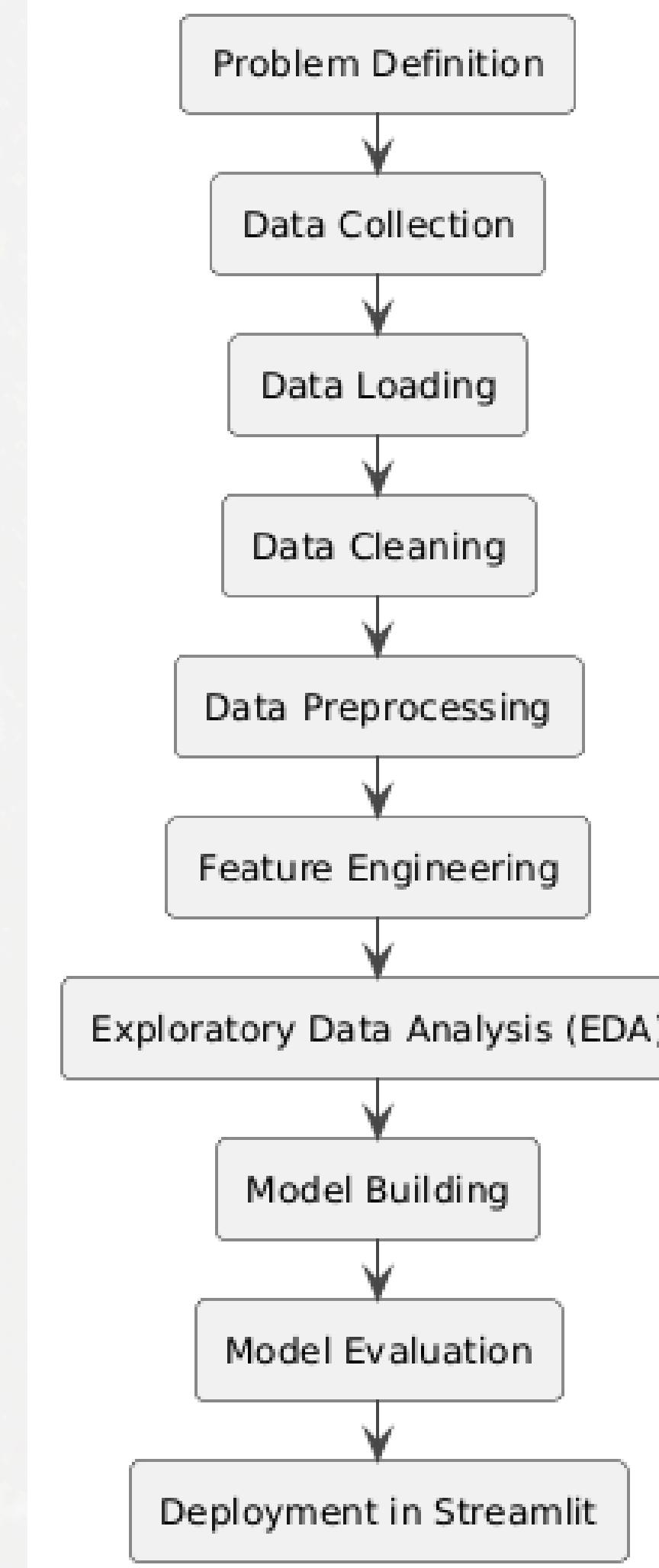
ABOUT

- This dataset provides detailed measurements of changes in groundwater levels from a network of 22,965 observation wells across India.
- It records the fluctuations in groundwater depth over time, offering crucial insights for water resource management and sustainability studies.

- Time Period Covered: 2013 – 2022
- Frequency: Quarterly (January, March/April/May, August, and November)
- Granularity: Station-level data (individual observation wells)

- Dataset Size: 550,850 rows and 14 columns
- Geographic Coverage: 32 Indian states, 565 districts, 23,000+ observation stations
- Key Variables: currentlevel (current groundwater depth in meters), level_diff (change from previous observation), along with spatial and temporal identifiers

Flowchart



Preprocessing

Loading Data: The dataset was loaded from a CSV file into a pandas DataFrame.

Missing Values: The code checked for and confirmed there are no missing values in the dataset.

Outlier Visualization: Box plots were generated to visualize outliers in the numerical columns `currentlevel` and `level_diff`.

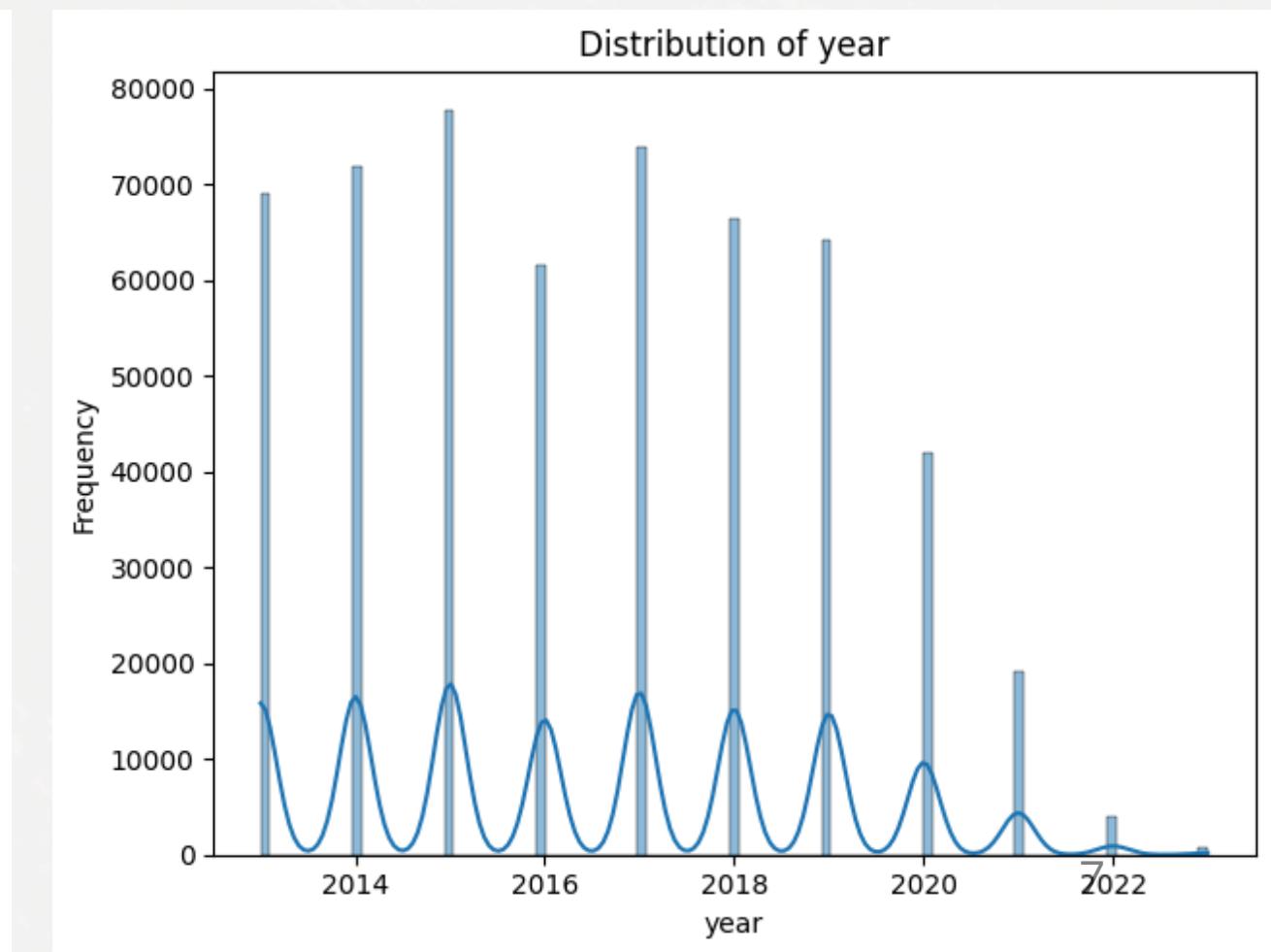
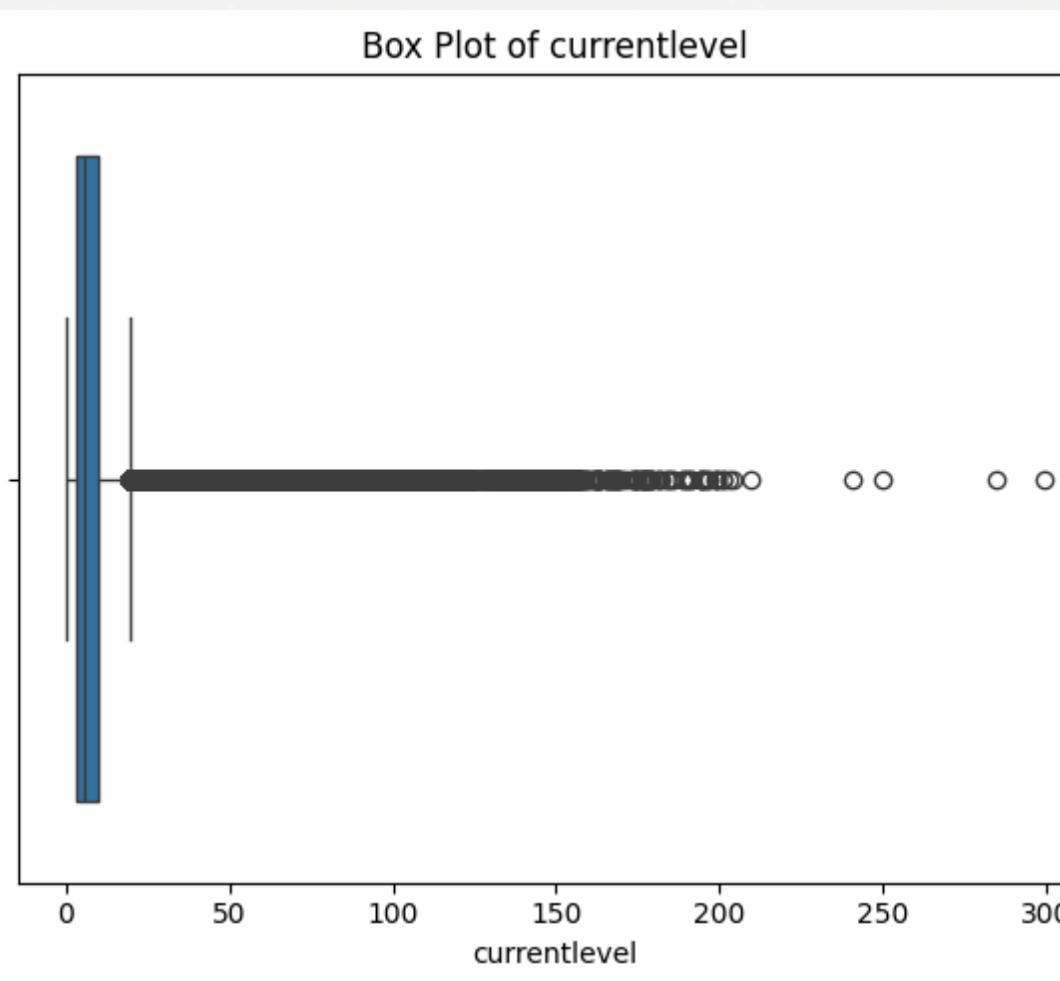
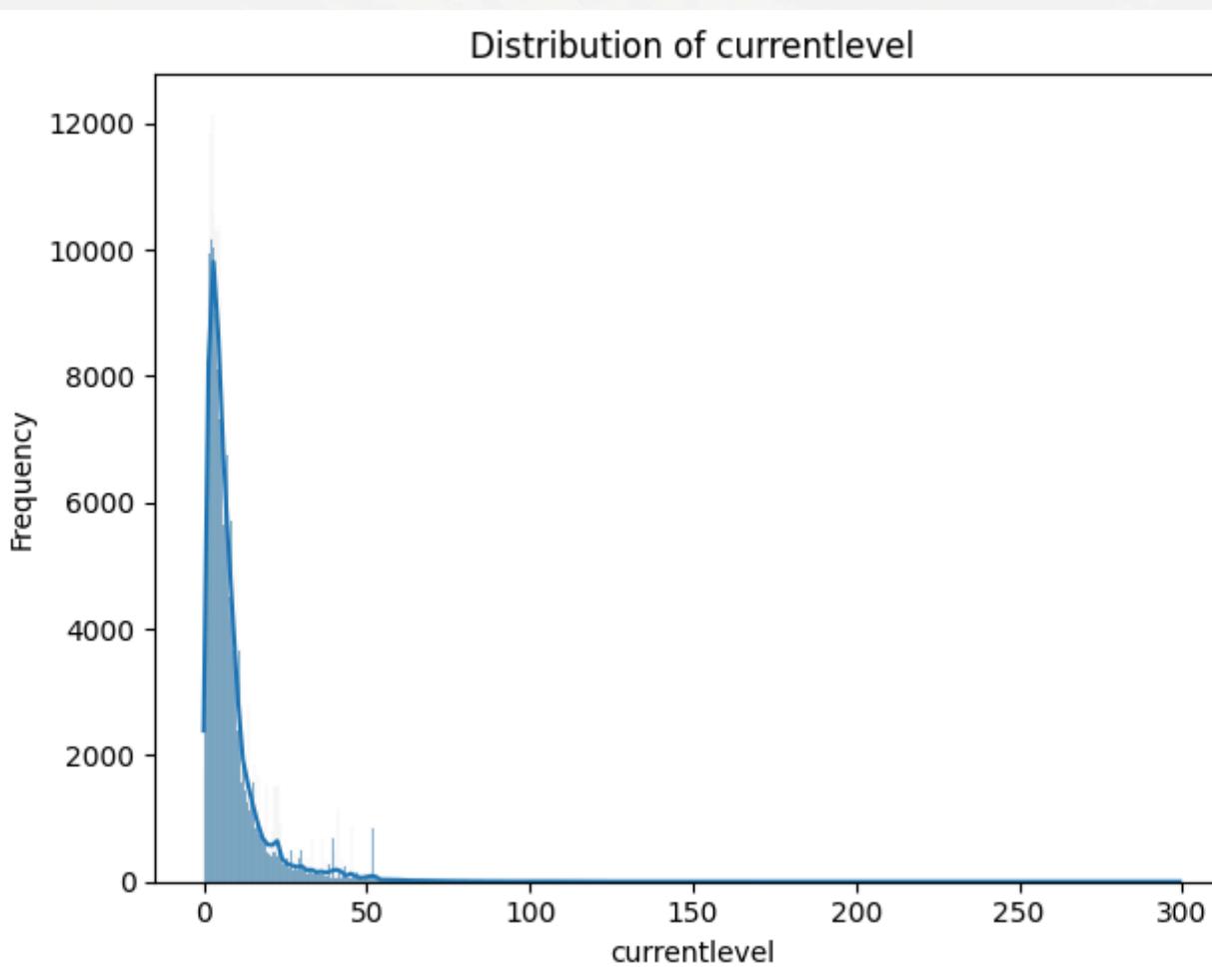
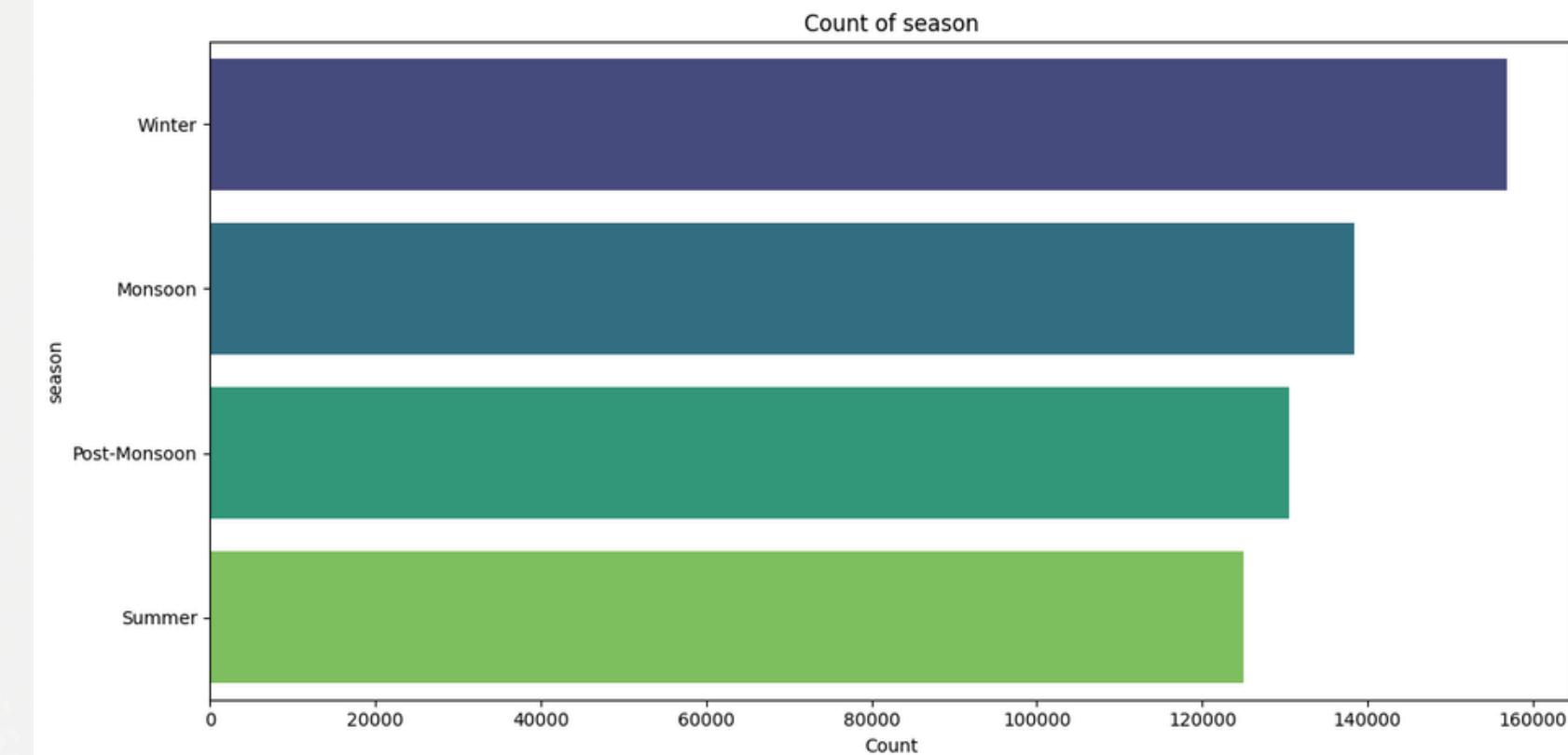
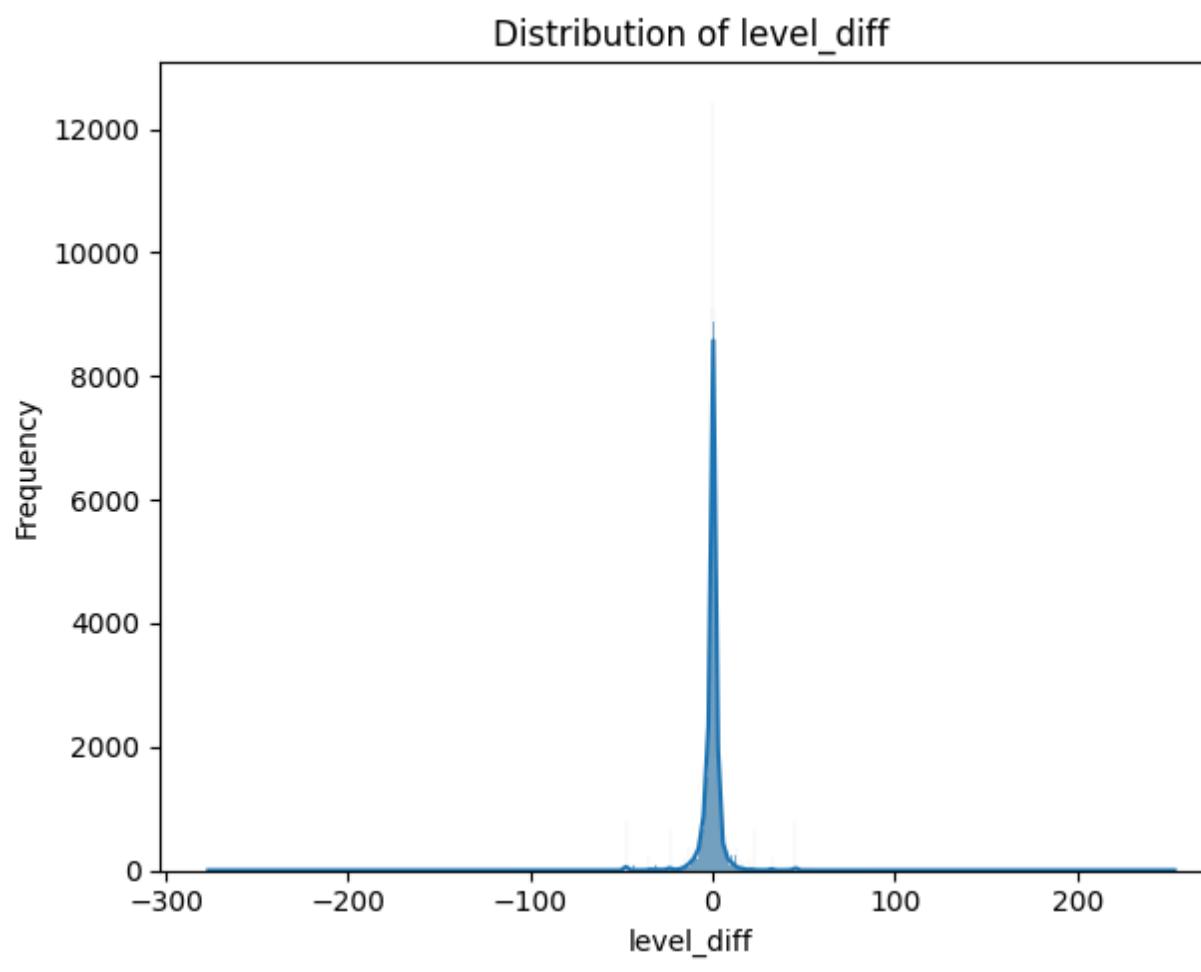
Data Type Conversion: The 'date' column was converted to datetime objects.

Feature Engineering: New features ('year', 'month', 'day', and 'season') were extracted from the 'date' column.

Categorical Target Variable Creation: The continuous 'level_diff' variable was converted into a categorical variable('level_diff_category') with 'Low', 'Medium', and 'High' categories based on quantile thresholds.

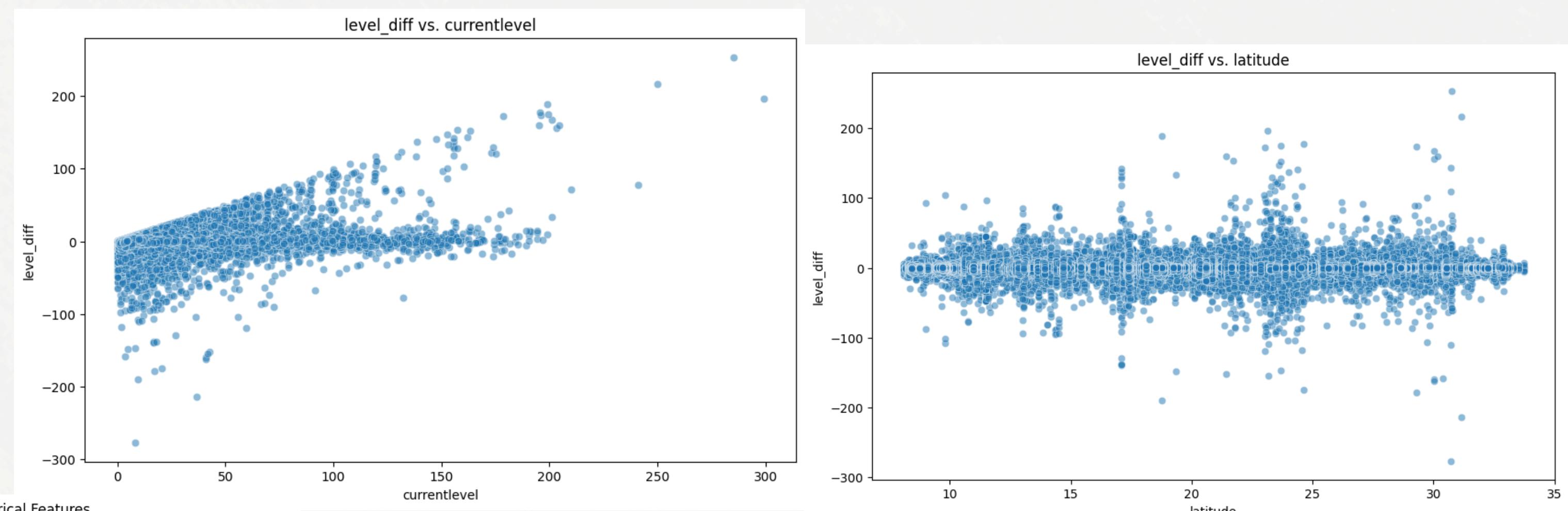
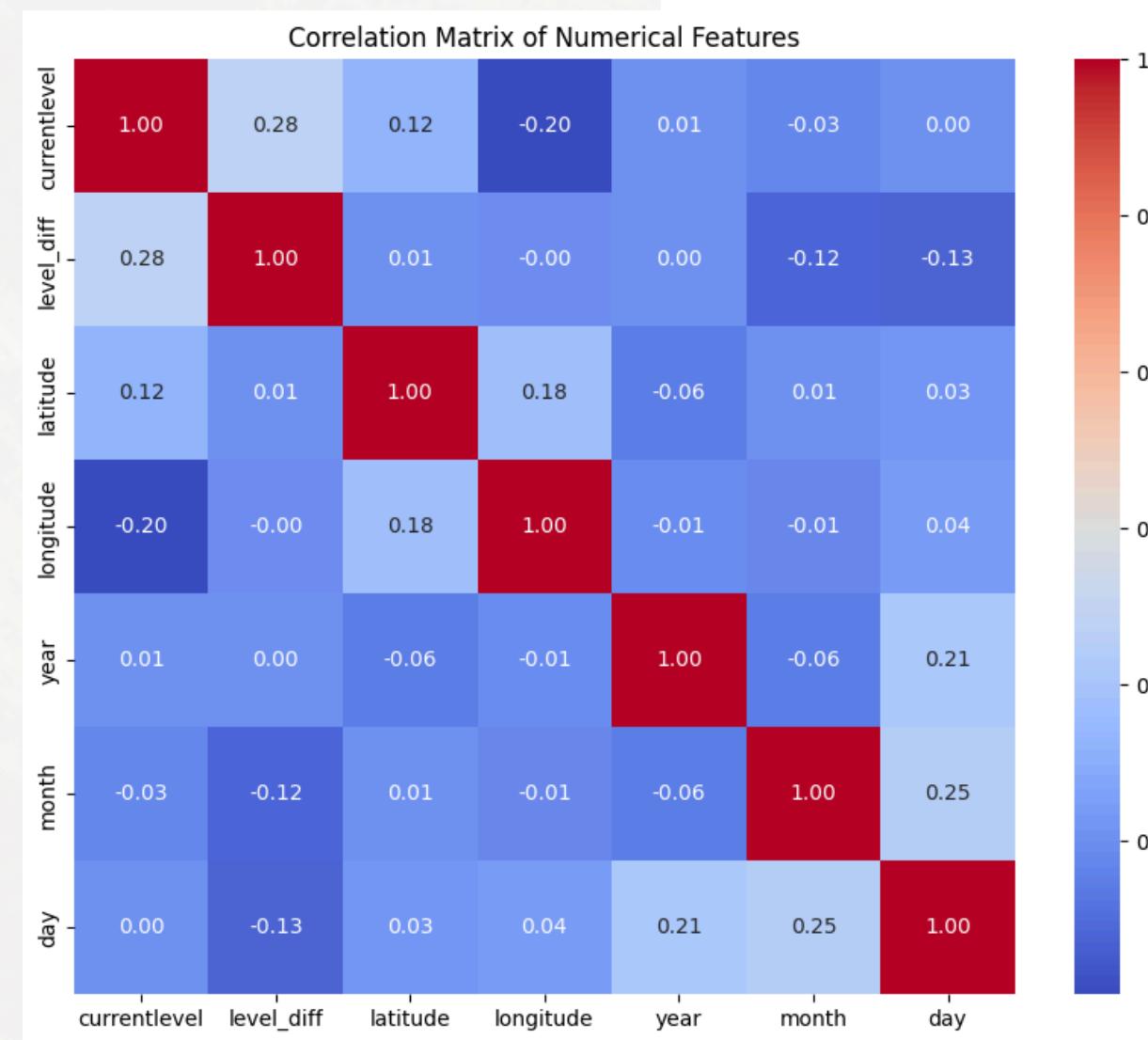
EDA (1/3)

Univariate Analysis

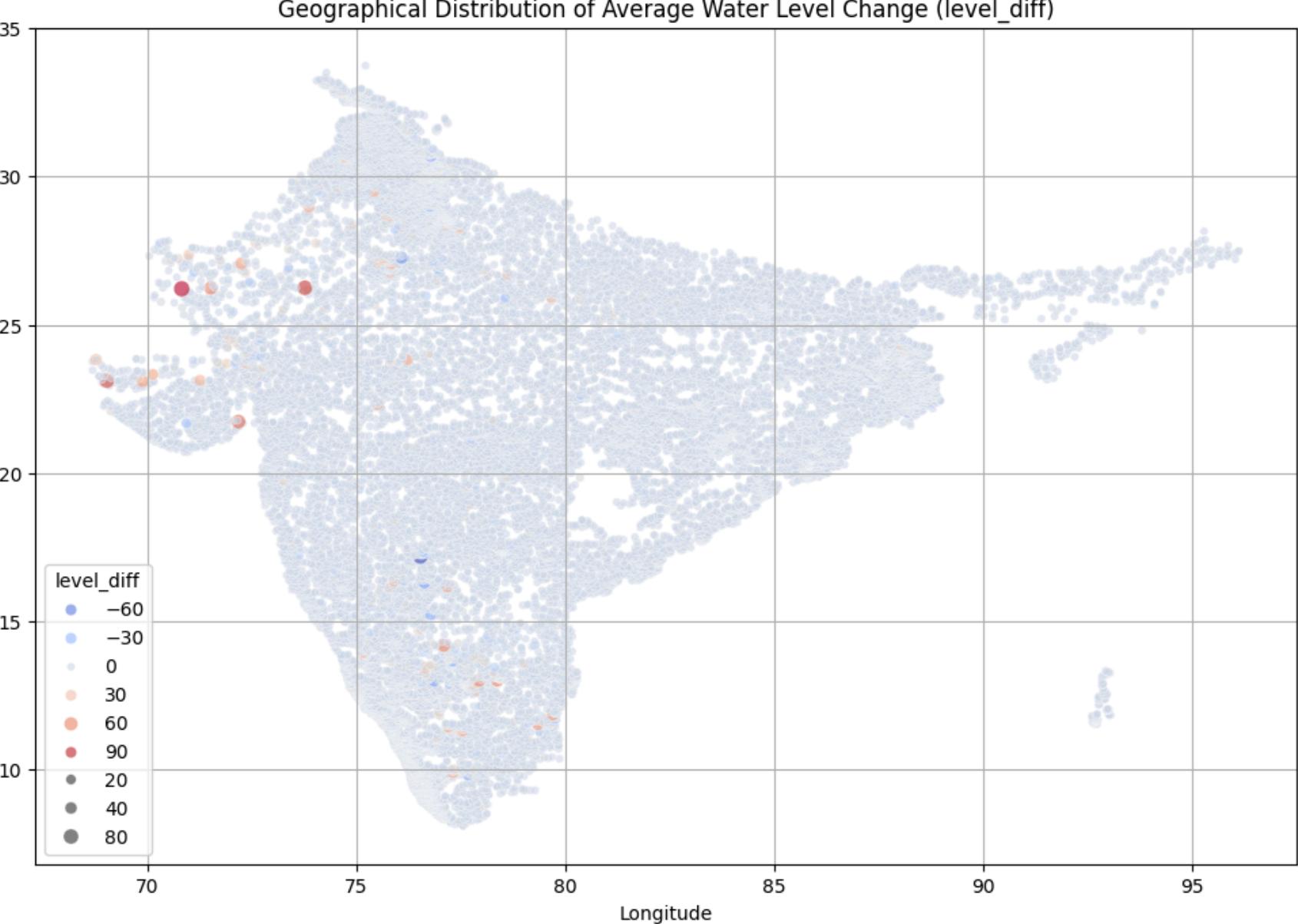
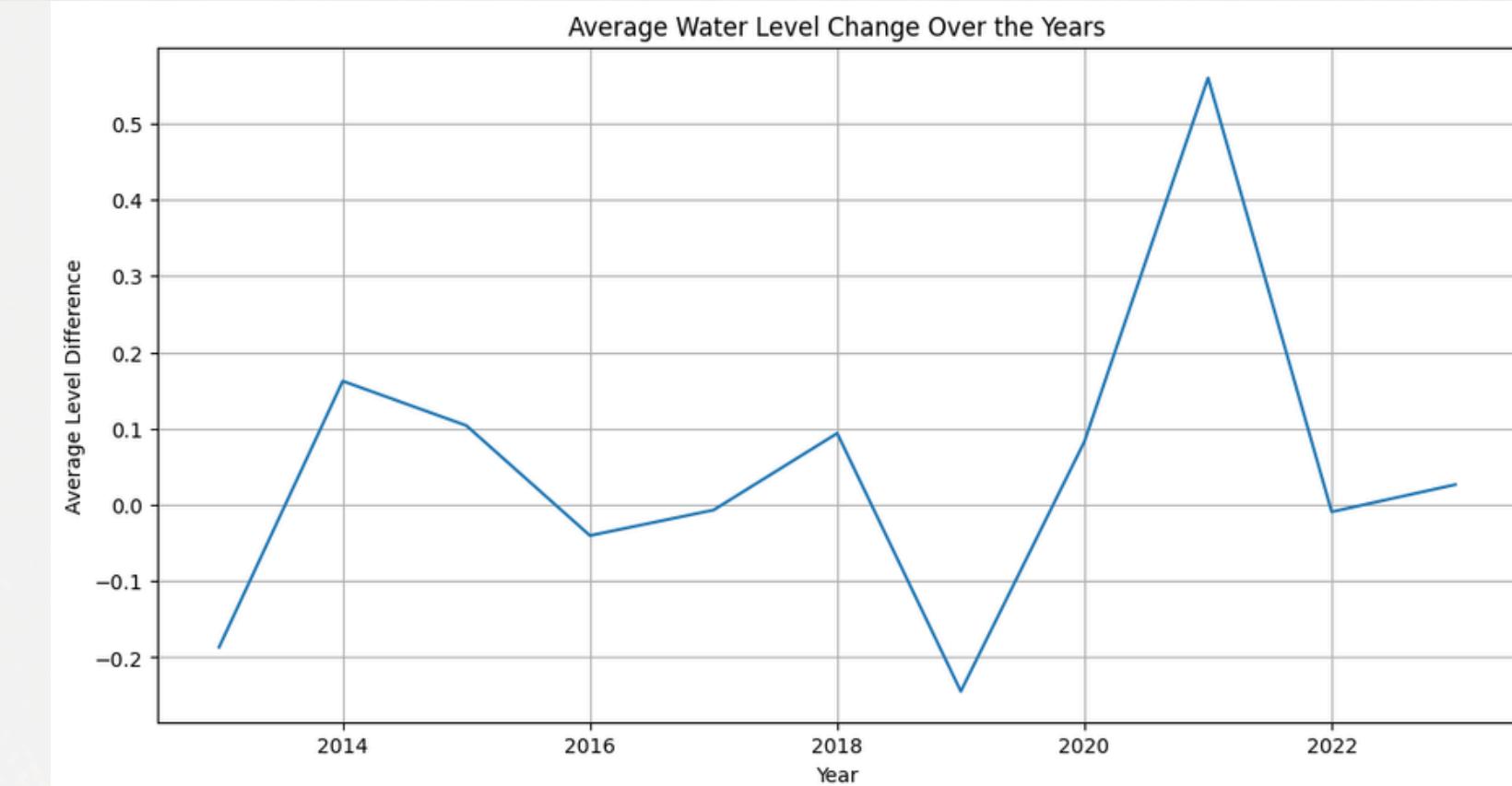
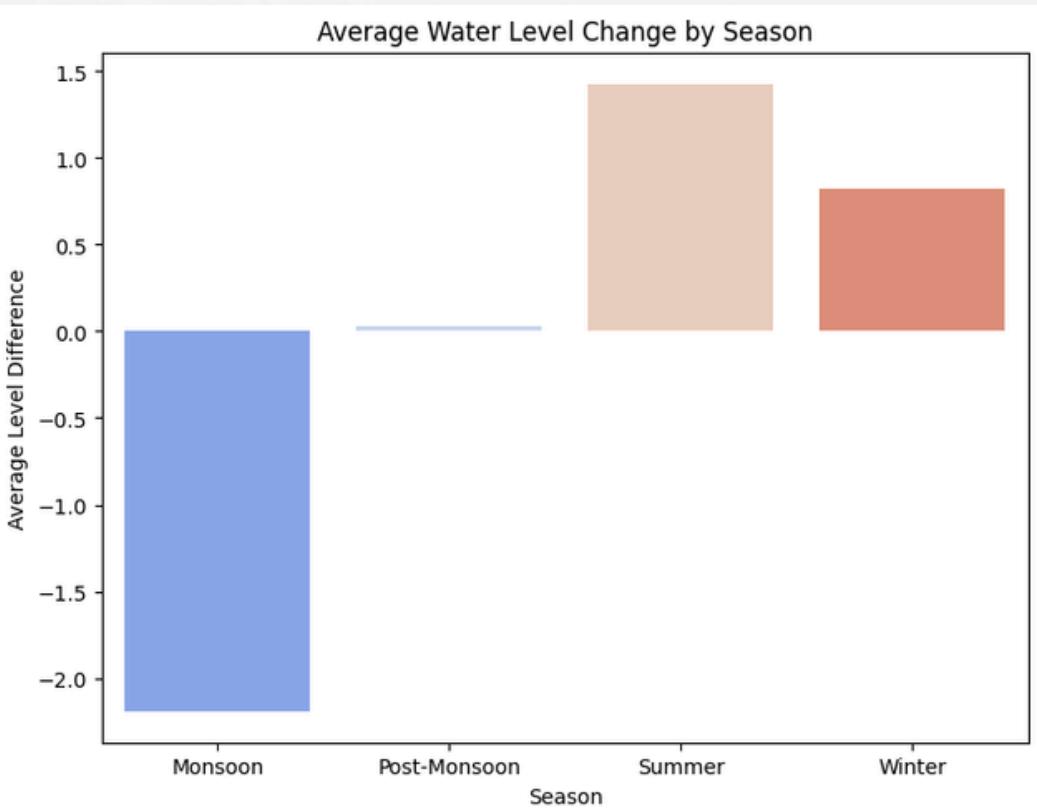


EDA (2/3)

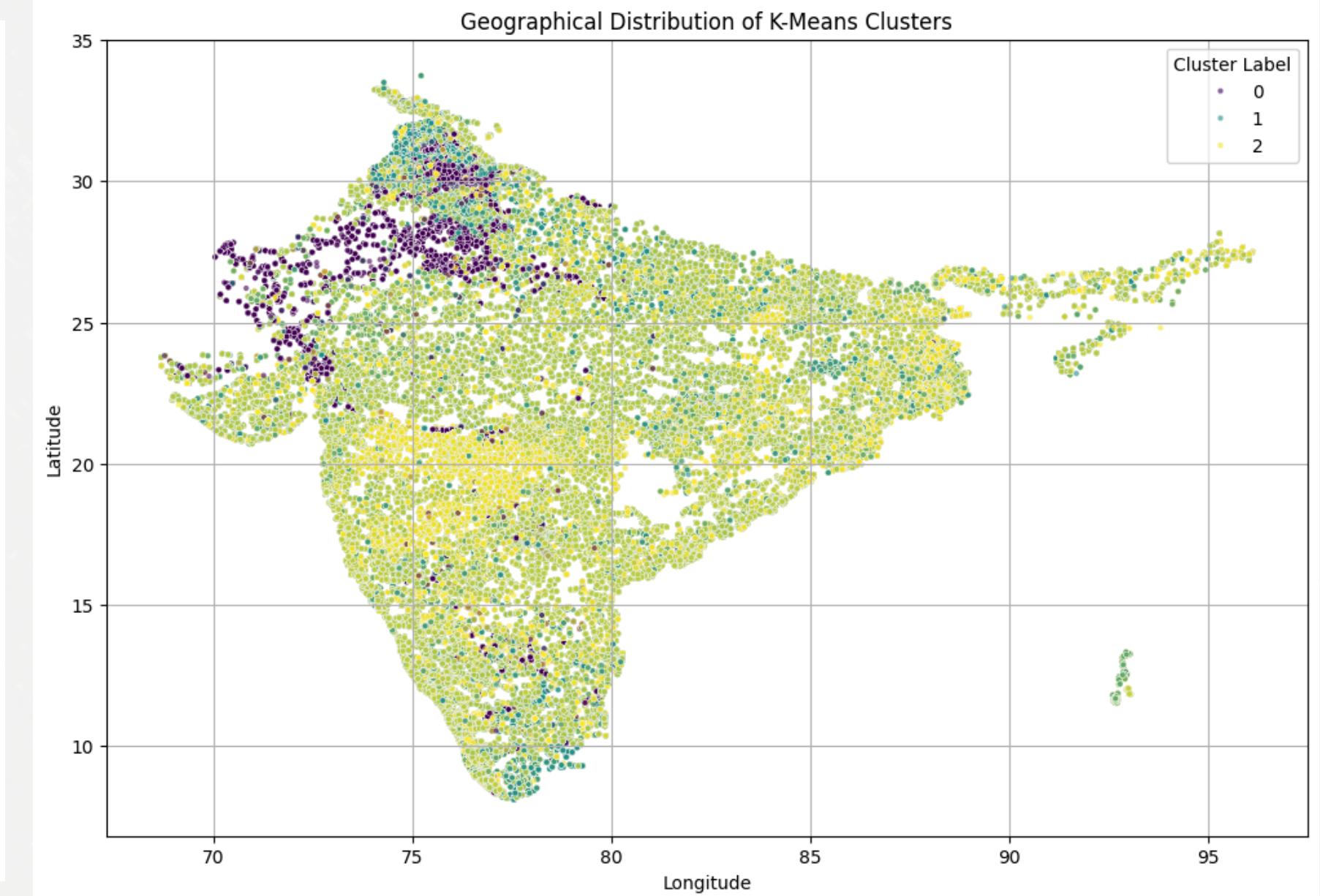
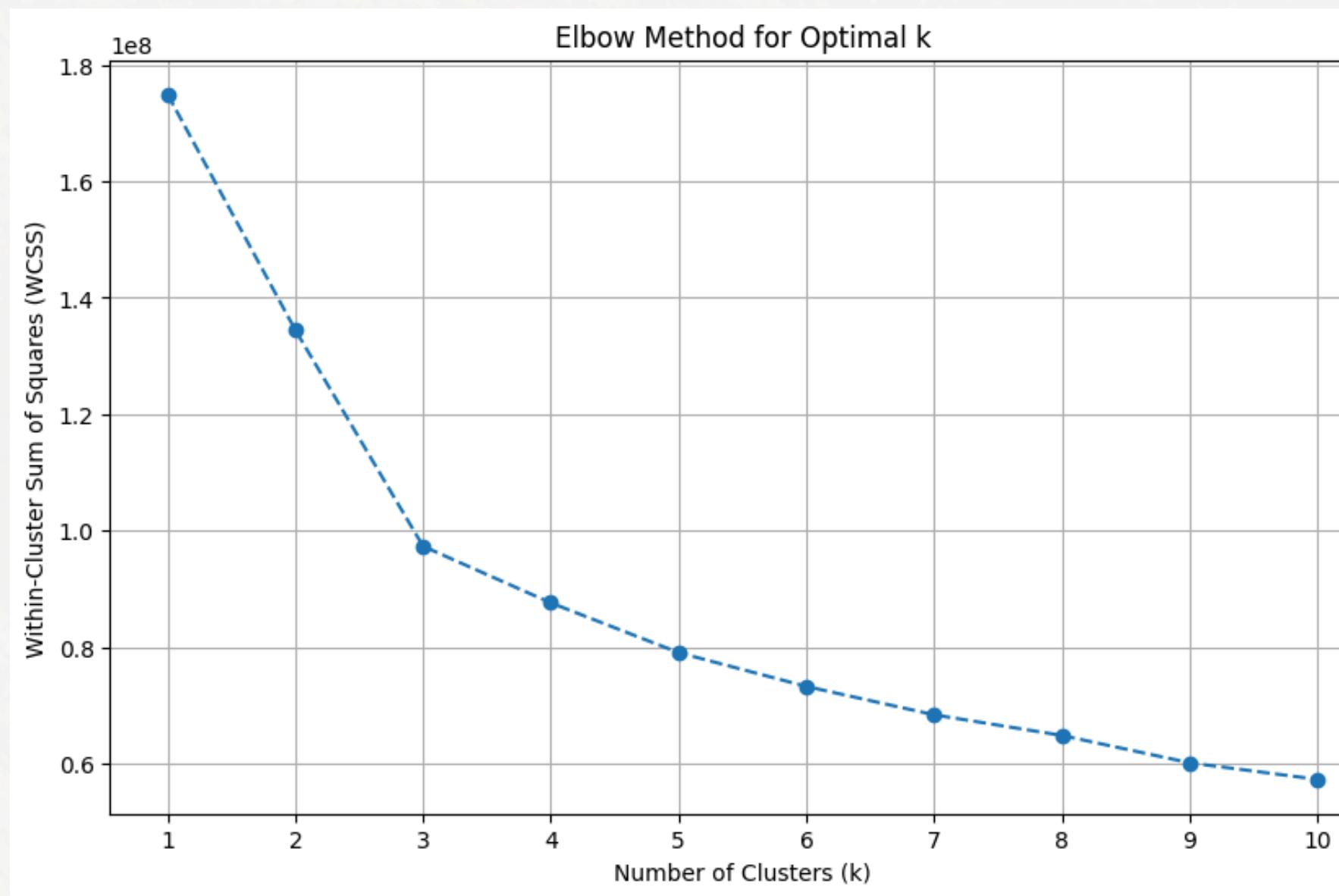
Bivariate Analysis



EDA(3/3)



Model(1/2)



Model (2/2)

- K-Means Clustering → Identified 3 distinct hydrological clusters (Elbow Method, k = 3)
- Cluster Labeling & Visualization → Added cluster insights to dataset
- Target Creation → `level_diff_category` (Low / Medium / High) for classification
- Feature Selection & Data Split (80/20)
- Classification Models Evaluated:
 1. Logistic Regression
 2. Decision Tree
 3. Random Forest
 4. SVM
 5. K-Nearest Neighbors (KNN)
- Regression Models Evaluated:
 1. Decision Tree Regressor
 2. SVR Regressor
 3. Random Forest Regressor

Executive Summary: Water Level Analysis

Navigation

- Go to
- Executive Summary
- Advanced Analytics
- Geospatial Insights
- Advanced ML Models
- Time Series Analysis

Total Observations

550,850

Date Range

2013 - 2023

Stations Monitored

23,078

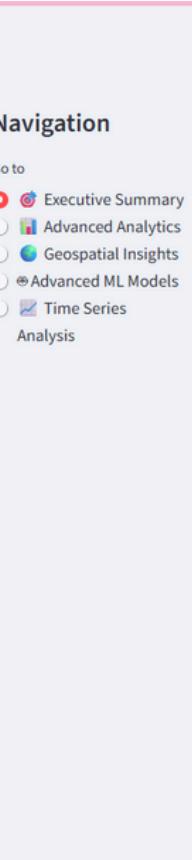
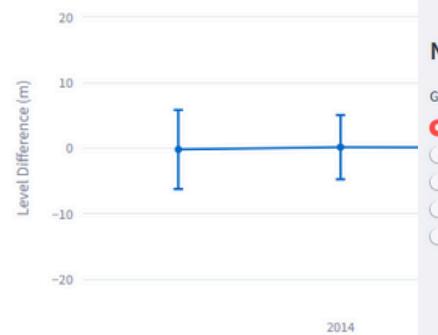
Key Insights

13976 stations showing critical decline

13454 stations showing significant recovery

Trend Analysis

Annual Water Level Trends with Uncertainty



STREAMLITE

Executive Summary: Water Level Analysis

Deploy ⋮

Total Observations

550,850

Date Range

2013 - 2023

Stations Monitored

23,078

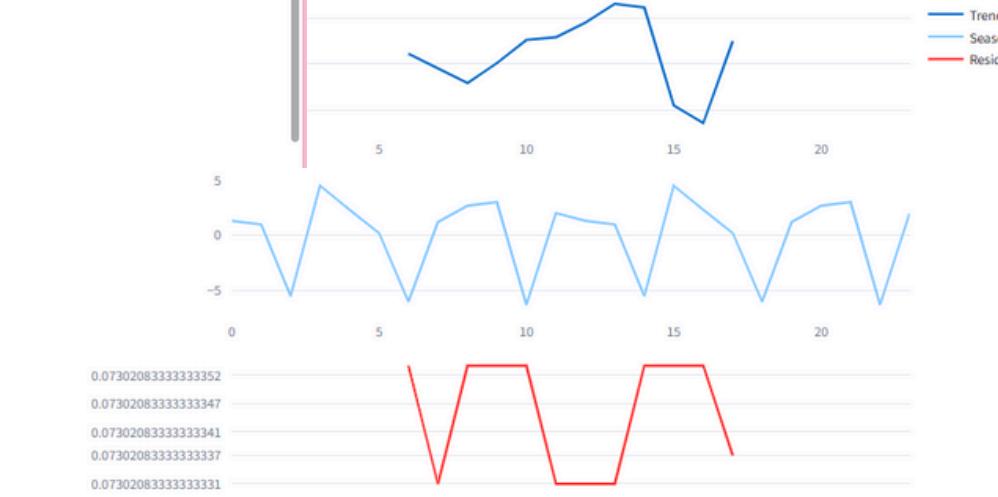
Key Insights

13976 stations showing critical decline

13454 stations showing significant recovery

Trend Analysis

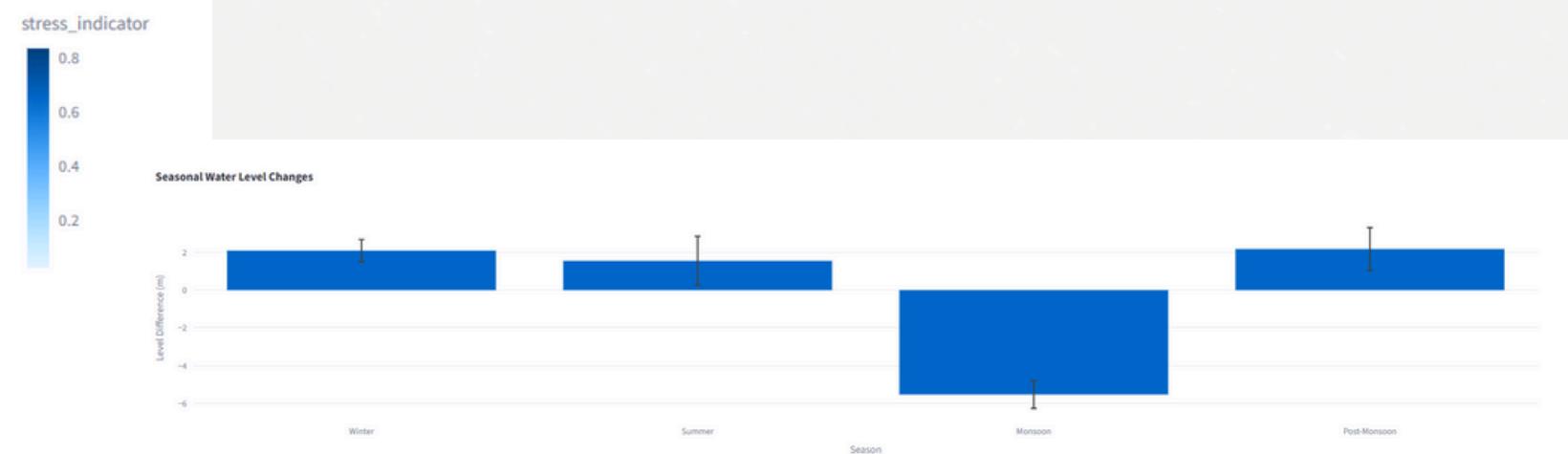
Annual Water Level Trends with Uncertainty



Pattern Analysis

Pattern Analysis

Water Stress Patterns by Region and Season

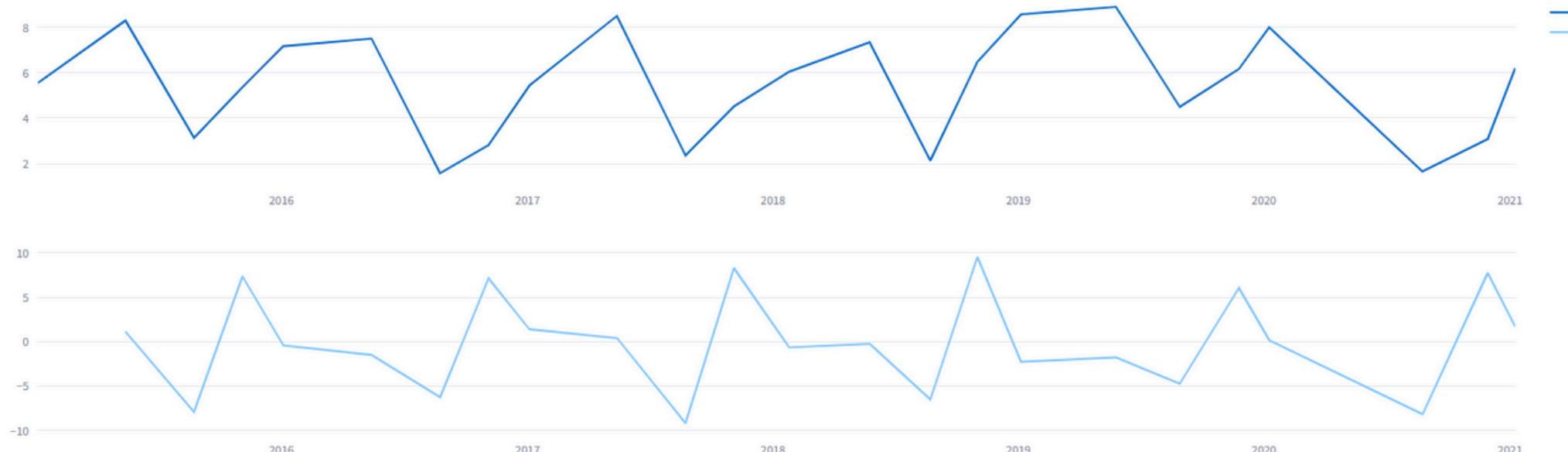


Time Series Analysis Dashboard

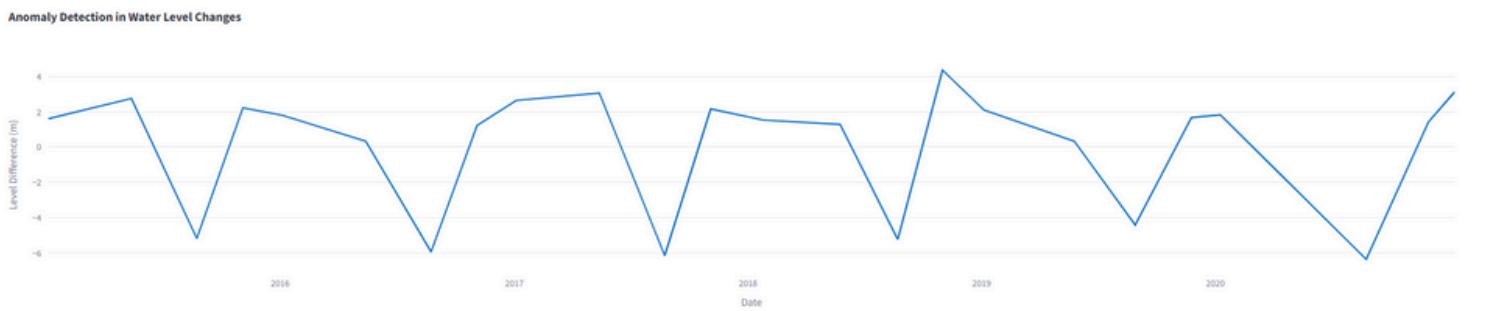
Temporal Pattern Analysis

Select Station

Time Series Analysis for Nagpur



Anomaly Detection



THANK YOU