INT 375

PROJECT REPORT

(Project Semester January-April 2025)

Predicting Water Quality Analyse In India (2012-2024)

Submitted by

Allada Pavan Venkata Satya Chowdary

Registration No: 12314335

B.Tech CSE - K23GD

Course Code: INT 375

Under the Guidance of

Ms. Baljindaer kaur (UID: 28968)

DISCIPLINE of CSE/IT

School of Computer Science and Engineering

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that **Pavan** bearing Registration no**. 12314335** has completed **INT 375** project titled, "Predicting Water Quality Analyse In India (2012-2024)" under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Signature and Name of the Supervisor

Ms. Baljindaer kaur

School of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab.

Date: 11-04-25

DECLARATION

I **pavan**, student of **B.Tech CSE - K23GD**under CSE Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 11-04-25 Signature

Registration No. **12314335**

Name-**pavan**

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my faculty guide **Ms. Baljindaer kaur.**for her valuable guidance, support, and encouragement throughout this project. I am thankful to **Lovely Professional University** for providing the necessary resources and environment for conducting this project successfully. I also extend my appreciation to all those who directly or indirectly contributed to the completion of this report.

TABLE OF CONTENTS

1. Introduction

1. Introduction ▪ Water Contamination is a day by day increasing problem due to industrialization and releasing of water into river streams. Industries releasing large amounts of wastewater which are containing pollutants such as heavy metals, chemicals and other waste. Rainfall also plays important role influencing water quality by diluting contaminants. Tracking and predicting levels of water contaminants can inform early warning systems, policy making and efforts to reduce environmental damages. ▪ Water pollution has become one of the important environmental and public health issues as the levels of industrial discharge and unstable rainfall patterns have increased due to earth heating, contaminating water bodies across the world. ▪ Public health and safety – Contaminated water causes serious health problems, diseases such as cholera and dysentery. By analyzing which functions become active a predictive model can enable highers to prevent from infections. ▪ Causal Factors - Environmental Protection, Discharge of excess industrial waste causes damage to aquatic ecosystems. Environmental conservation efforts can be aided by predicting contamination levels. ▪ Industrial Regulation insights could help government agencies and environmental bodies monitor and regulate industries which helps them in following water discharge. Data Driven Decision Making - Advanced analytics utilize AI to offer businesses accurate, real time insights and discover better water management methods.

2. Source of Dataset

The dataset used for this analysis was provided through the INT 375 coursework and contains detailed transaction-level sales data. This includes information such as Staes, Ph valus, Villages, Blocks, Checals, TDS, Watrer Qality Index, and other attributes useful for analytical purposes. The data is assumed to be part of a retail or wholesale business management system.

3. EDA Process

Exploratory Data Analysis (EDA) is the initial step in data analysis that focuses on summarizing the main characteristics of a dataset. The following EDA techniques were used:

- Handling missing or inconsistent data entries
- Cleaning column names and formatting
- information about data set
- Water Quality Index And TDS, Ph values
- Visualizing key performance indicators using Python
- Identifying trends, outliers, and relationships in the dataset

4. Analysis on Dataset

1. Problem Statement

- Predicting Water contamination level by taking the data from discharge of waste water from industries into the rivers and other water bodies, considering the rainfall.
- Challenges and significance

- Data Availability and quality - Real life water quality datasets are limited or inconsistent, thereby requiring high data preprocessing.
- Feature Selection and Engineering - Relevant input parameters such as rainfall, discharge, PH, COD, etc that extremely effect contamination of water.
- Variability with Environmental conditions - Contamination levels become unpredictable due to seasonal variations or industrial activity and geographical changes.
- Accuracy and Generalization of the model -Ensuring the model generalized all locations and time period.
- Treatment of missing Data - The data most frequently presents gaps, missing values, and errors in the measurement.

5. Specific Requirements, functions and formulas

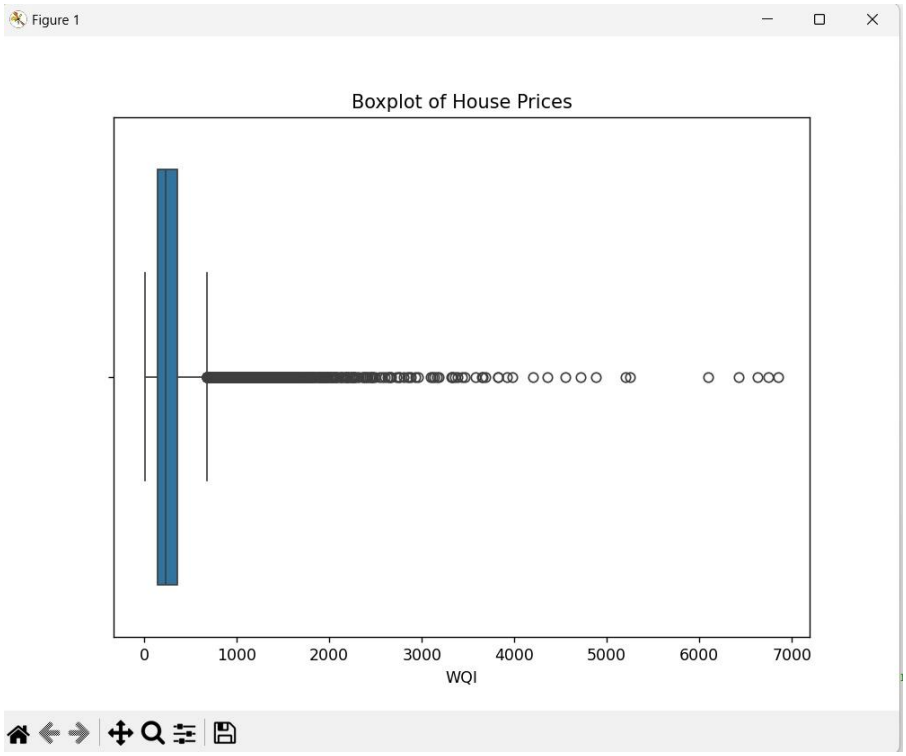
Synthetic Data – used to fill gaps and support the robustness of the model.

- WQI (Water Quality Classifivcation) [Target Variable]
- pH
- EC
- Mg, Na, K
- TDS
- Water quality classification

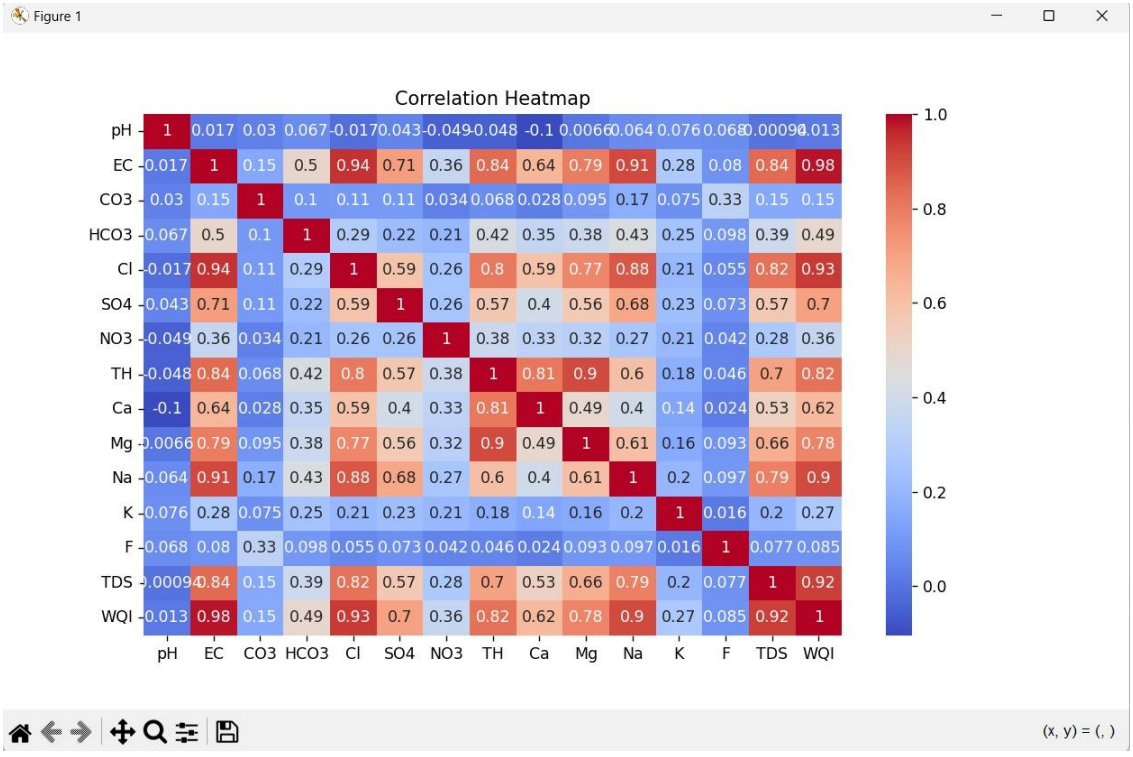
WATER QUALITY DATA OF DRAINS, STPs & WTPs MONITORED UNDER NWMP- 2023																					
STN Code	Monitoring Location	Type Water Body	State Name	Temperature (°C)		Dissolved Oxygen (mg/L)		pH		Conductivity (µmho/cm)		BOD (mg/L)		NitrateN (mg/L)		Fecal Coliform (MPN/100ml)		Total Coliform (MPN/100ml)		Fecal Streptococci (MPN/100ml)	
				Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
4481	HATHLI NALLAH LIFT WATER SUPPLY SCHEME FOR MUNICIPAL COUNCIL HAMIRPUR	DRAIN	HIMACHAL PRADESH	19	28	7.6	8.5	6.6	7.9	278	680	1.0	2.3	0.3	8.5	2	6	21	47	2	2
5253	HURLA NALLA D/S HURLA BRIDGE, VPO HURLA, TEHSIL BHUNTER, DIS. KULLU	DRAIN	HIMACHAL PRADESH	5	18	7.3	8.3	7.1	8.5	60	155	1.0	1.0	0.3	0.7	31	94	140	540	2	2
4426	RAW WATER FOR LWSS OF JAGROTI NALLAH	DRAIN	HIMACHAL PRADESH	8	14	7.6	9.3	7.6	8.3	92	156	1.0	1.0	0.3	1.0	2	130	2	1600	2	50
4454	JARANGLA NALLAH U/S WATER SUPPLY SCHEME	DRAIN	HIMACHAL PRADESH	8	15	8.8	9.4	7.5	8.3	110	485	1.0	1.0	0.4	5.3	2	20	23	3500	2	2
4466	JARANGLA NALLAH D/S WATER SUPPLY SCHEME	DRAIN	HIMACHAL PRADESH	8	15	8.8	9.4	7.5	8.4	112	524	1.0	1.0	0.4	4.2	2	20	33	3500	2	2
5257	KALAM NALLA D/S OF MSW PROCESSING SITE AT CHOWRI, DIST. CHAMBA	DRAIN	HIMACHAL PRADESH	10	16	8.5	9.1	7.0	8.1	125	282	1.0	1.0	0.5	2.0	2	22	47	1600	2	2
4477	KHALADA NALLAH NEAR WATER SUPPLY SCHEME FOR KULLU TOWN	DRAIN	HIMACHAL PRADESH	2	18	7.8	8.8	7.2	8.0	67	230	1.0	1.0	0.3	2.9	32	110	140	920	2	2
4455	KLUIN NALLAH U/S WATER SUPPLY SCHEME	DRAIN	HIMACHAL PRADESH	8	13	9.0	9.5	7.0	8.2	77	496	1.0	1.0	0.3	1.7	2	40	94	3500	2	2
4482	LIFT WATER SUPPLY SCHEME (PANJARAR) AND BHOTA ROPARI & UJHAN	DRAIN	HIMACHAL PRADESH	19	28	7.5	8.5	6.5	8.1	312	689	1.0	2.1	0.3	1.6	2	4	17	70	2	2
4036	KUNNI PUL, VILL SAMAHU, PO KUNIHAR, TEHSIL ARKI	DRAIN	HIMACHAL PRADESH	10	15	7.4	7.8	7.9	8.4	231	3640	1.0	1.8	1.2	5.3	2	12	21	350	2	2

- Data Preprocessing Techniques (Cleaning, Normalization, Feature Engineering)

1. Handling Missing Values: Use mean or median techniques to fill in pH, BOD, COD, TDS, and other similar numerical features. The categorical features include state names and river names(optional).
2. Outlier Removal: For removal of outliers Z-score filtering can be used.
3. Standardizing Features: Standardization (Z-score) can be used for Linear Regression



. Visualization of correlation of heap map



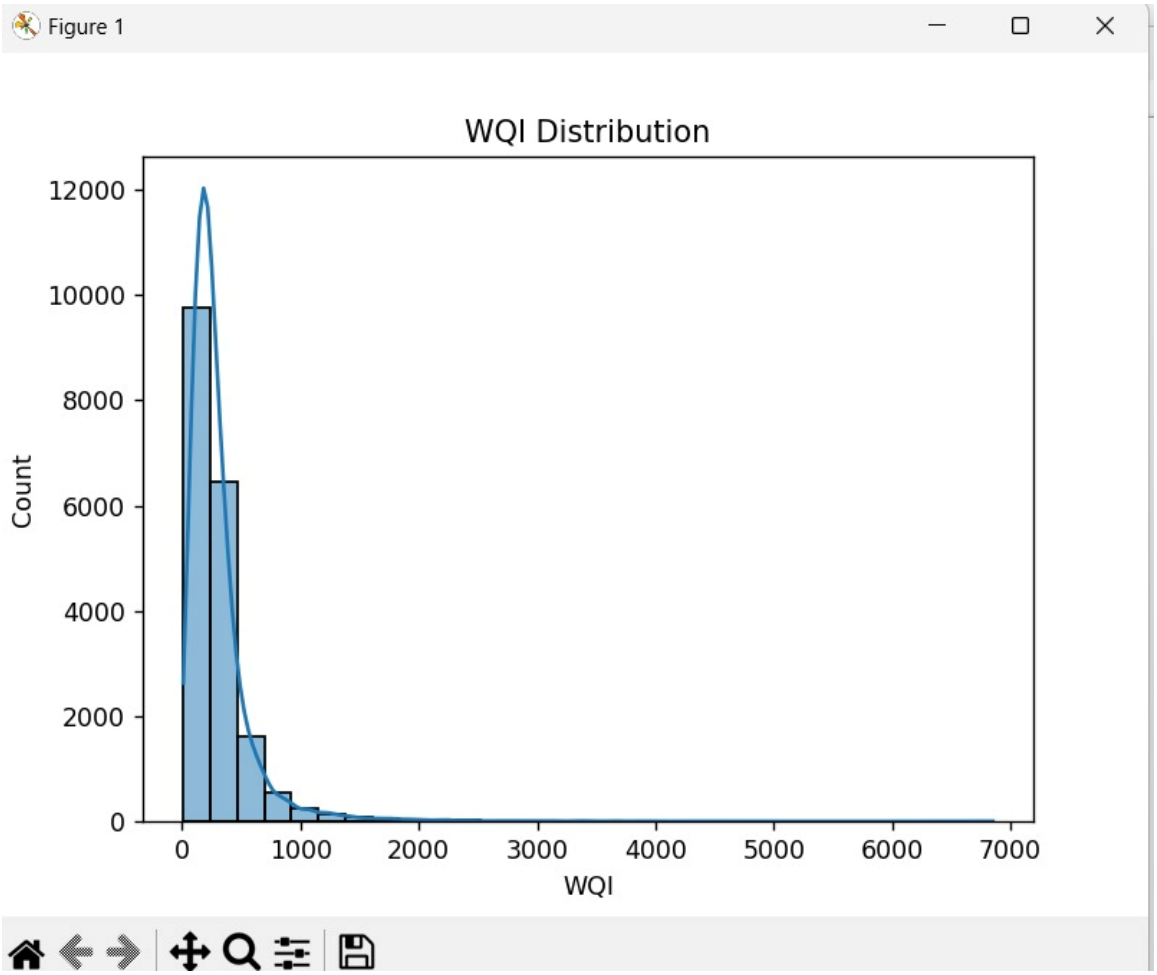
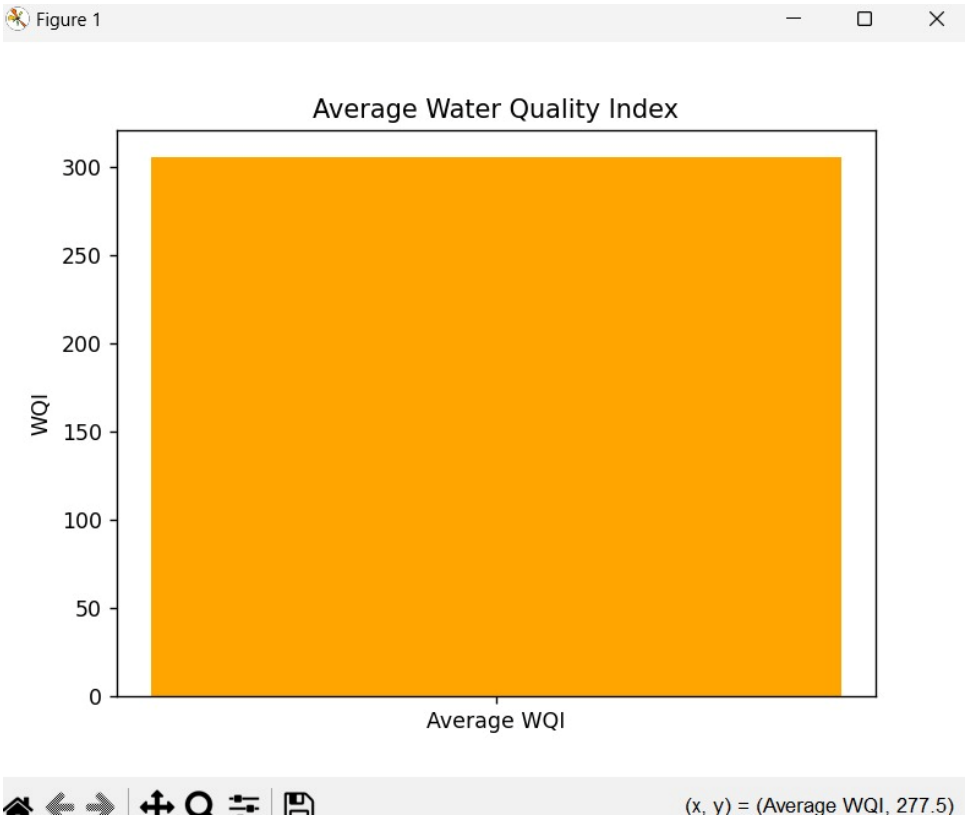
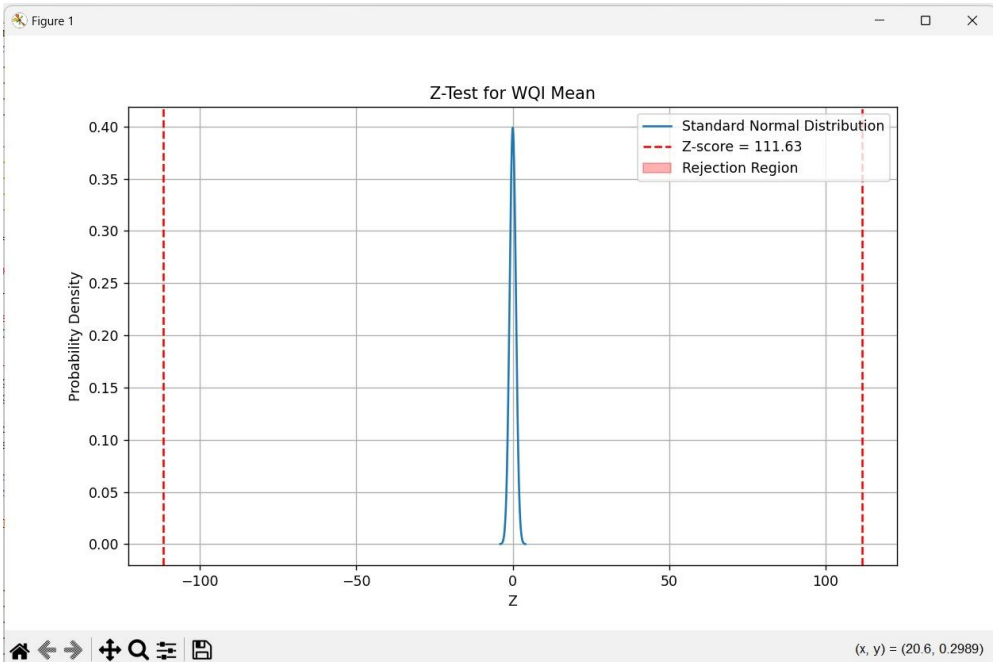


Figure 1
WQI Distribution
12000
10000
8000
6000
4000
2000
0
Count
0
1000
2000
3000
4000
5000
6000
7000
WQI



Visualization : Z testing of WQI mean



7. Conclusion

Water contamination because of industrial discharge and rain water mixing is a critical environmental issue that affects public health destroying the ecosystem balance. My project is the developing of a prediction model for water pollution levels using different kinds of regression, very important parameters including pH, COD, BOD, TDS, oxygen demand, state and river name.

8. Future Scope

Future work can expand to include predictive modeling using machine learning to forecast future Quality of Water.

9. References

1. Central Pollution Control Board (CPCB)
2. National Water Monitoring Programme (NWMP) Data
3. Open Government Data (OGD) Platform India
4. National Water Informatics Centre (NWIC)
5. Water Resources Information System (India-WRIS)
6. Water Quality Prediction Using Machine Learning Models Based on Grid Search Optimization Method, Talaat, F. M., & Zahraa, T. (2023).