ANALYSIS OF WATER QUALITY BASED ON SURVEY AND REGRESSION TECHNIQUE

Dr. Swapnali D. Mahadik, Assistant Professor, MCA Department, DES's NMITD, Mumbai -28.

Abstract
Mumbai, a major port city on the western coast of India, has serious problems with water quality because of growing urbanization and the release of household and commercial wastewater. By conducting a thorough survey of the city's citizens, this study seeks to assess Mumbai's water quality by concentrating on their opinions and level of knowledge about the water they use. Water makes up around 70% of the earth's surface, yet because fresh water is so scarce, water pollution is a serious global problem. Water scarcity, ecosystem degradation, and negative health impacts will result from pollution-related deterioration of the water quality from these sources. So the primary objectives of this survey-based study are to investigate people's varied viewpoints and degrees of knowledge regarding the water they usually drink as well as the regression analysis of the water's parameters.
Keywords: Water Quality, Surface water, Ground Water, Regression.

Introduction:
When dangerous materials or diseases contaminate water bodies—such as rivers, lakes, seas, and groundwater—it is referred to as water pollution. These pollutants, which may take the shape of chemicals, microbes, or other substances, can harm wildlife, the environment, and human health. Water contamination can arise from various sources such as improper disposal of waste and sewage, oil and chemical spills, air deposition, and industrial and agricultural operations. These contaminants can lead to a number of issues when they get into water bodies, including decreased oxygen levels, toxic algal blooms, and the buildup of hazardous materials in fish and other aquatic life.
In this study the survey was carried out with a representative sample of the populace in several Mumbai neighborhoods. The respondents shared information about the main sources of water they use, how they view the quality of the water, how they are aware of contamination problems, and what steps they take to ensure safe consumption. The goal of this study is to develop a more efficient technique for keeping an eye on water quality and spotting possible problems before they become serious ones.
During the testing and analysis of surface water samples in the years 2020–2021, MPCB considered several factors. Nonetheless, four factors—pH, dissolved oxygen, BOD, and FC—are taken into account while calculating WQI. This helps in conveying about water quality status effectively.
The kind of water being tested, the intended results, the kind and degree of pollution, and other factors all play a role in choosing the best testing technique. A thorough evaluation of water quality is typically obtained by combining many testing techniques.

Literature Review:
The concepts of hydrodynamic analysis, water quality analysis, environmental flow assessment, flow augmentation, and their combinations have been used in many studies carried out all around the world. The concepts of hydrodynamic analysis, water quality analysis, flow scenario creation, and environmental flow evaluation are novel to Maharashtra's network of connected rivers.

KNN Imputer and Multilayer Perceptron are the tools used by author Afaq Juna, Muhammad Umer, Saima Sadiq, et al. to implement Water Quality prediction[1].
Groundwater Quality Assessment was also conducted by Ana Elizabeth, Diego Cruz, Elena, and others utilizing K-Means Clustering, Principal Component Analysis, and Spatial Analysis [2].
The authors also investigated the real-time analysis of water quality data using the Fuzzy Comprehensive Evaluation Method (Ji-hong et al., 2020). This fuzzy logic-based approach was shown to be useful for evaluating water quality since it permits the simultaneous assessment of several parameters (Ji-hong et al., 2020; Biao et al., 2019).
Athira K R , has done a study on Water Quality Modeling and Regression Analysis of Vembanad Lake and its Primary Inflow. [3] Warsito has also used hierarchical clustering analysis to evaluate the quality of river water.[4]. Sailkrishna has suggested to evaluate the groundwater quality through drinking WQI and regression analysis.[5]
Yang Zengchuen used the shuffling frog leaping technique (SFLA) to build an artificial neural network (ANN) that would improve the performance of a combined water quality assessment model. While the SFLA is an evolutionary method that may be used for optimization, the ANN is a type of machine learning approach that can be used for pattern identification. Natural selection produced these two algorithms. It demonstrates that the authors thought this approach would be helpful in determining the water's calibre for their tests.

Research Analysis:
The study started by asking people who are in-person representatives of their respective organizations how much water they drink each day. According to the data, 44% of these heads drink more than three liters of water a day, 16% drink about three liters, 15% drink two liters, 14% drink 1.5 liters, 9% drink one liter, and 2% drink half a liter. According to WHO recommendations, the body needs to be hydrated with three to four liters of water each day; however, this also depends on the individual's weight and other factors.

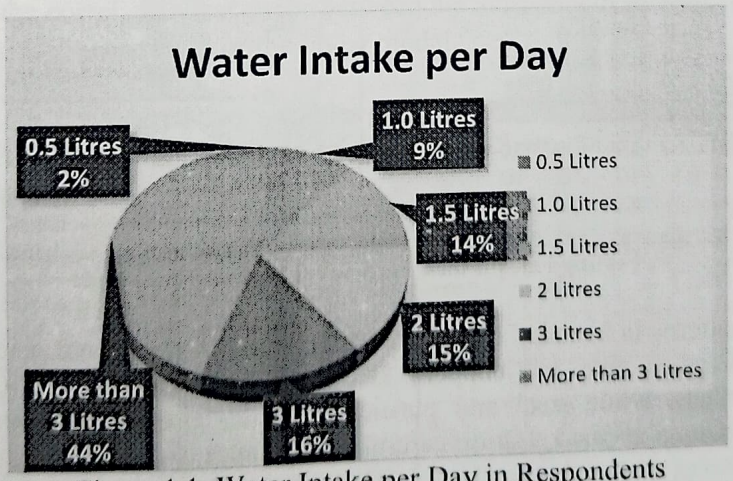

Figure 1.1. Water Intake per Day in Respondents

Each source of drinking water in Mumbai has unique benefits and limitations. While tap water is cost-effective and widely used, its quality can be inconsistent, driving many residents to use filtered or packaged water. Filtered water systems offer a balance between safety and cost but require proper maintenance. Packaged water is convenient but expensive and environmentally taxing. Outsourced water, though crucial for underserved areas, poses significant quality risks. Addressing these challenges through improved infrastructure, stringent quality controls, and

public education is essential for ensuring safe and reliable drinking water for all Mumbai residents.

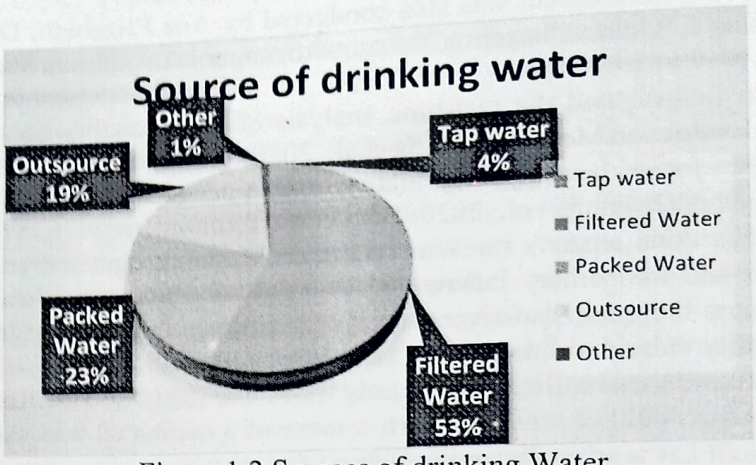

Figure 1.2 Sources of drinking Water

Another method to assess the quality of drinking water is to measure the PH level. A water's ideal pH range to drink is between 6.5 and 8.5. 83 count individuals, however, have never tested and have no idea what their PH level is. And the second commonly used term in water quality assessment is TDS level. The majority of the organizations were found to be aware of TDS and to maintain the level of TDS in accordance with the rules.

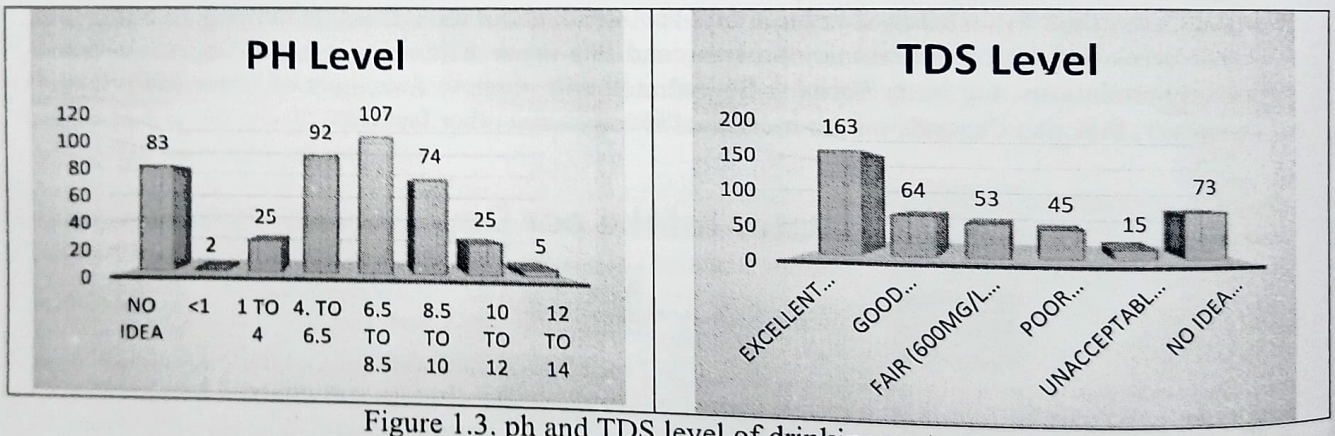

Figure 1.3. ph and TDS level of drinking water

Water storage patterns in Mumbai reflect a balance between traditional practices and modern conveniences. Earthen pots and copper vessels offer natural benefits and are rooted in traditional practices, while steel and plastic containers provide practical and economic solutions. Each storage method requires careful maintenance to ensure water safety and quality. This storage patterns obviously the pH and TDS values.

People are using various method to improve the water quality using the mentioned methods such as Water Neutralization, Filtration of Water, Chlorination.

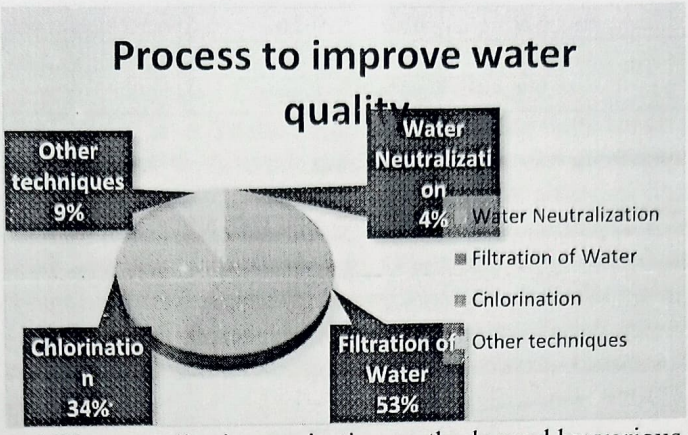

Figure 1.4 Water quality improvisation methods used by various people

Methodology and Implementation:

Regression is a machine learning method used to understand how independent variables (features) relate to a dependent variable (outcome) and to predict continuous outcomes. Linear regression aims to find a linear equation that best fits the data points, describing the relationship between the dependent variable (Y) and independent variables (X). The ordinary least squares (OLS) approach is the most common method for determining this best-fitting line. OLS minimizes the sum of the squared differences between the observed values of Y and the values estimated by the linear regression equation, effectively optimizing the line parameters for the closest fit to the data

This aids with linear regression.

- To forecast the variable that is dependant (Y)
- Compute each independent variable's (X) impact on the dependent variable (Y).
- Determine the correlation that exists between the independent and dependent variables.
- Check the significance level of the linear model.

Here in these study considered two variables pH and TDS to show the regression between these variables. Total twenty-seven readings of pH and TDS are considered which was collected from different areas.

Table 1.1 Water Parametric values of ph and TDS

Sample	pH	TDS	Sample	pH	TDS
1	7	450	15	7.3	458
2	7.4	520	16	6.8	54
3	7.9	95	17	8	62
4	6.5	754	18	7.5	45
5	6.2	523	19	7.5	41
6	7.8	55	20	7.5	352
7	7.1	850	21	7.3	158
8	7	41	22	7.1	42
9	8	55	23	6.9	42
10	7.1	59	24	6.8	56
11	6.8	45	25	7.2	82

12	6.9	452	**26**	7.3	74
13	6.9	48	**27**	7.1	54
14	7.3	652			

Line Fit Plot

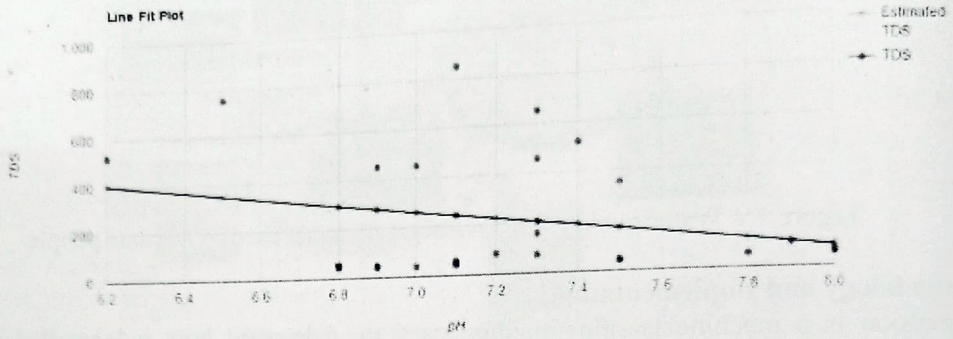

Figure 1.5: Line Fit Plot

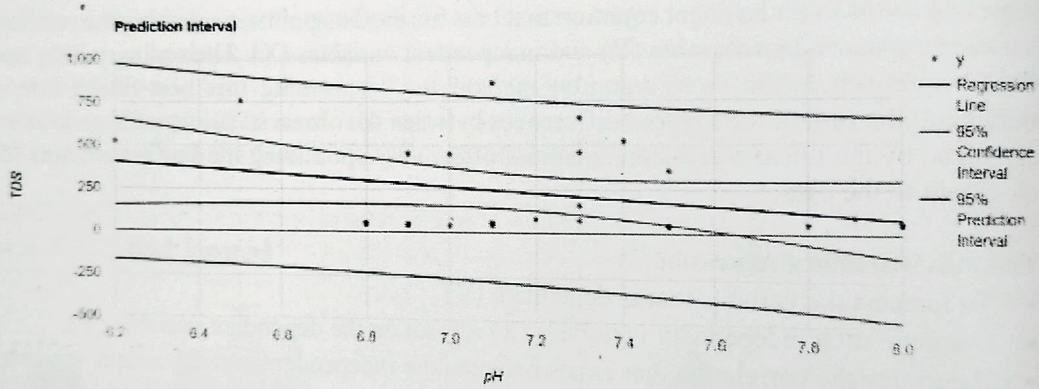

Figure 1.6: Prediction Interval

Table 1.2 Statistical Details of Regression

Source	Degree of Freedom	Sum of Square	Mean square	F Statistic (df1,df2)	P-value
Regression	1	155023.547	155023.55	2.5158 (1,25)	0.1253
Residual	25	1540526.75	61621.07		
Total	26	1695550.3	65213.473		

This represents the statistical details of a regression analysis, summarizing the sources of variability, degrees of freedom (df), sums of squares, mean squares, F-statistic, and P-value.The analysis indicates that the model explains a portion of the variability in the data. The F-statistic (2.5158) and P-value (0.1253) suggest that the regression model does not provide a statistically significant fit at the typical significance level (e.g., 0.05), implying the independent variable may not significantly predict the dependent variable in this context. So main aim is to identify the most crucial factors that have a major impact on the water quality in a variety of locations by using a variety of methodologies to analyze data on water quality.

Conclusion:

The regression analysis, implies that, under the given conditions, the independent variable does not significantly predict the outcome. Consequently, the current model may not be effective for making reliable predictions or understanding the relationship between the variables in this context. Further investigation with different models or additional data may be necessary to identify more meaningful relationships. Also survey show that although municipal supply is still the most common source of water, many homeowners are concerned about its quality and have reported problems with things like bad taste, color, and odor. The degree of awareness regarding water contamination varies greatly among various socioeconomic classes. In order to assure the safety of the water, many responders take preventative steps like boiling and filtering, which reflects a general lack of confidence in the untreated supply.

References:

1. Afaq Juna , Muhammad Umer , Saima Sadiq,et al , "Water Quality Prediction Using KNN Imputer and Multilayer Perceptron" Water 2022, 14, 2592. https://doi.org/10.3390/w14172592
2. Ana Elizabeth,Diego Cruz,Elena,et al,"Groundwater Quality Assessment: An Improved Approach to K-Means Clustering, Principal Component Analysis and Spatial Analysis: A Case Study , Water 2018, 10(4), 437; https://doi.org/10.3390/w10040437
3. Athira K R , Varsha Murali , Aswin C R , Devaprabha B, Aneesh P C ,"Water Quality Modeling and Regression Analysis of Vembanad Lake and its Primary Inflow" Volume 10, Issue 06 (June 2021) DOI : 10.17577/IJERTV10IS060185
4. B Warsito et al,"Evaluation of river water quality by using hierarchical clustering analysis",2 021IOP Conf. Ser.: Earth Environ. Sci. 896 012072
5. K.saikrishna, D.Purushotham, V.Sunitha," Data for the evaluation of groundwater quality using water quality index and regression analysis in parts of Nalgonda district, Telangana, Southern India", Elsevier, Volume 32, October 2020, 106235, https://doi.org/10.1016/j.dib.2020.106235
6. Venkaraman,Nikitha,Rishika" Water quality index estimation using linear regression model"2022, AIP Conference Proceedings 2418, 040033 (2022); https://doi.org/10.1063/5.0081794
7. Nadikatla SK, Mushini VS, Mudumba PSMK (2020) Water quality index method in assessing groundwater quality of Palakonda man- dal in Srikakulam district, Andhra Pradesh, India. Appl Water Sci 10(1):30
8. IS 10500:2012 Drinking Water Specification
9. Guidelines for Water Quality Monitoring by MPCB
10. Water Quality of Lakes, Ponds, Tanks, and Wetlands 2021 (State-wise)(CPCB)

Website:

1. https://www.who.int/teams/environment-climate-change-and-health/water-sanitation-and-health/water-safety-and-quality/drinking-water-quality-guidelines
2. https://cpcb.nic in/nwmp-data/
3. https://mpcb.gov.in/water-quality
4. https://www.usgs.gov/mission-areas/water-resources