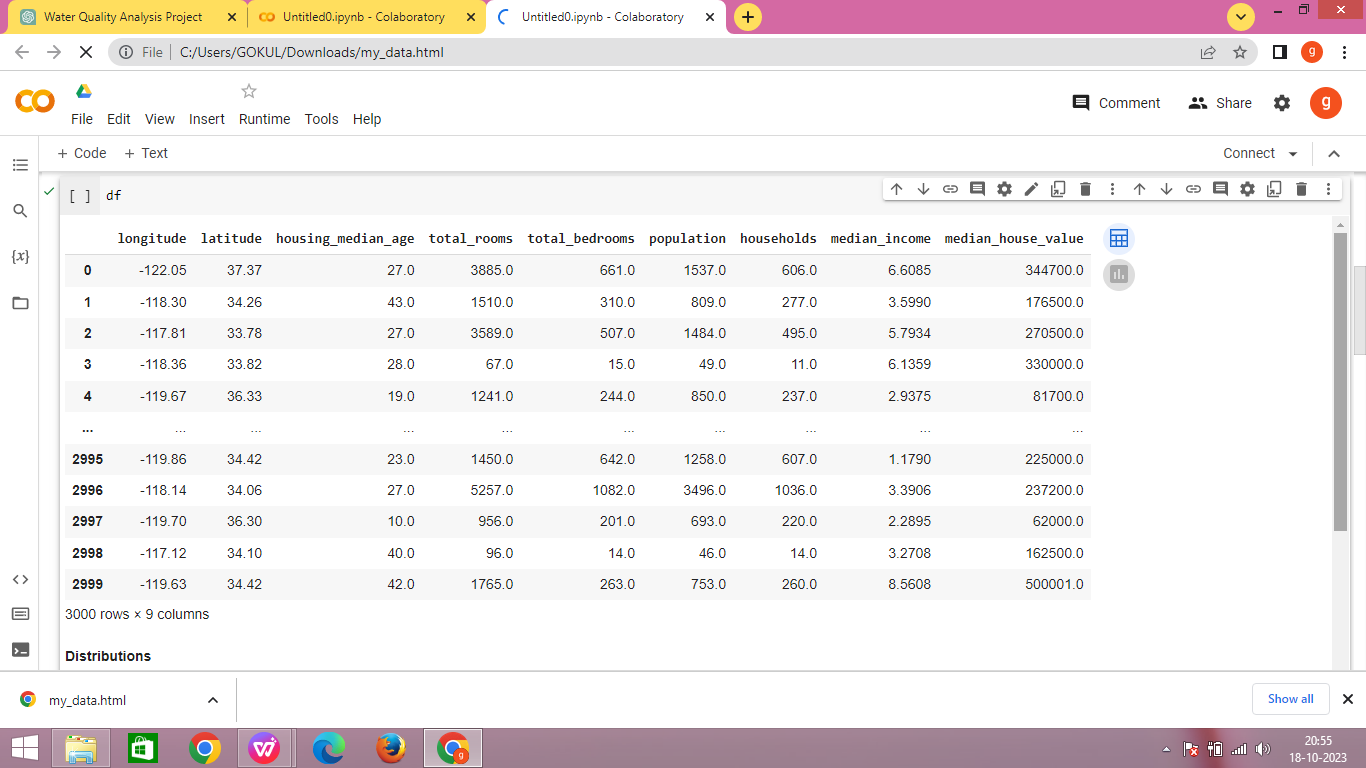
## WATER QUALITY ANALYSIS

**INTRODUCTION:**

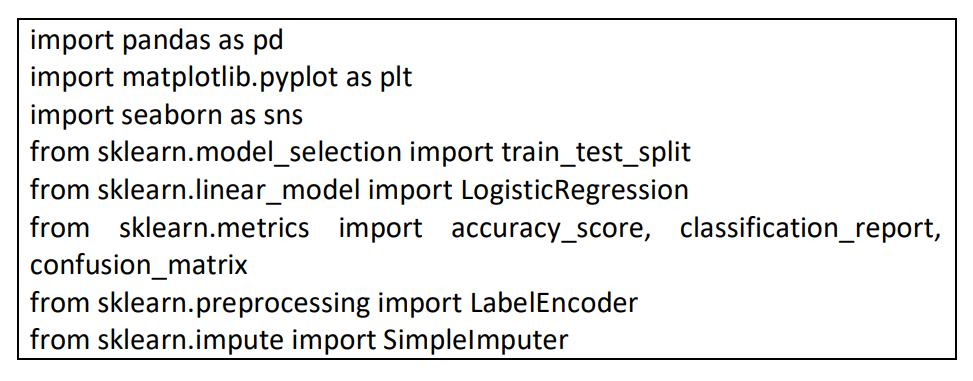
Water is essential for life, and its quality is a critical concern for both environmental sustainability and public health. The "Water Quality Analysis Project" aims to investigate and assess the condition of local water sources. Through this project, we seek to understand the factors influencing water quality, detect potential contaminants, and provide insights that can guide informed decisions for maintaining and improving the quality of our water resources. In this brief report, we will outline the objectives, methodologies, and significance of this vital initiative.

**GIVEN DATA:**

DATA SOURCE : https://www.kaggle.com/datasets/adityakadiwal/water-potability

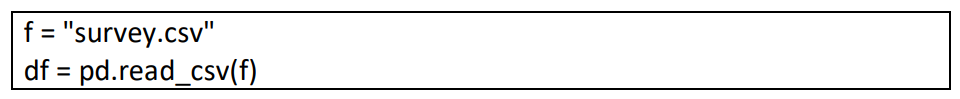


STEP 1 : IMPORT LIBRARIES



In this step, the necessary libraries are imported to work with data, visualize it, and build a machine learning model. These libraries include pandas for data manipulation, matplotlib and seaborn for data visualization, and scikit-learn for machine learning.

STEP 2 : LOAD DATA



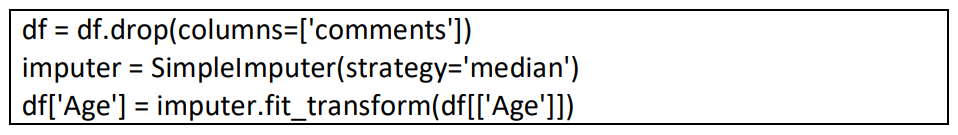
This step loads a dataset from a CSV file named "survey.csv" into a pandas DataFrame called 'df'.

STEP 3 : DISPLAY BASIC INFORMATION ABOUT THE DATASET



This step provides an overview of the dataset's structure, including the number of rows, columns, data types, and information about missing values.

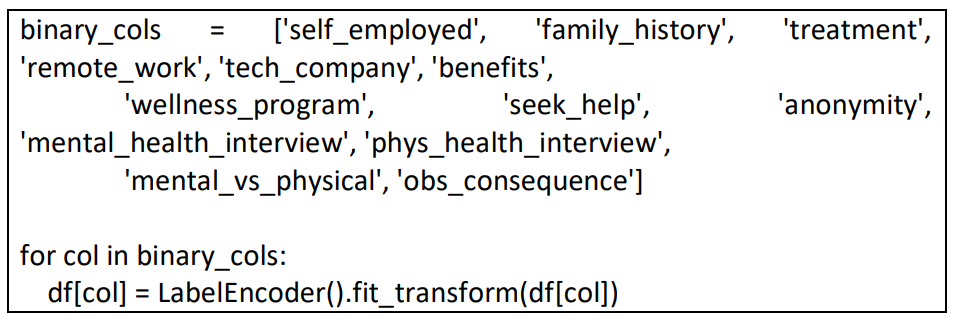
STEP 4 : HANDLE MISSING VALUES



This step first drops the 'comments' column as it's considered not useful for analysis. Then, it handles missing values in the 'Age' column by filling them with the median value of the 'Age' column.

STEP 5 : ENCODE BINARY COLUMNS

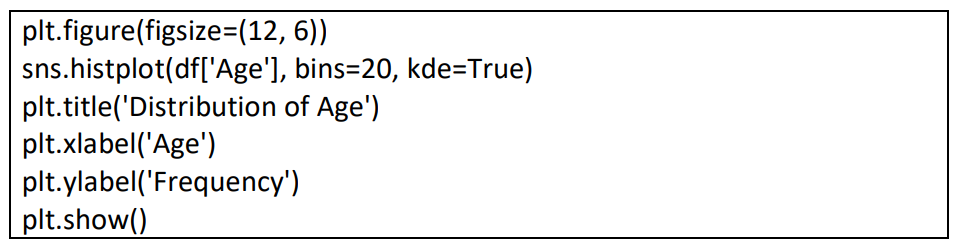
This step encodes binary columns with '0' and '1' values using LabelEncoder. These columns typically contain 'Yes' or 'No' responses, and they are transformed into numerical values.



STEP 6 : DATA VISUALIZATION

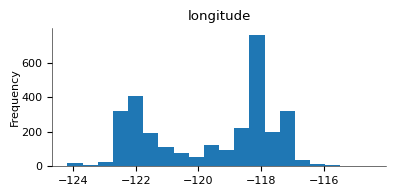
This section includes various data visualization steps using matplotlib and seaborn for understanding the dataset.

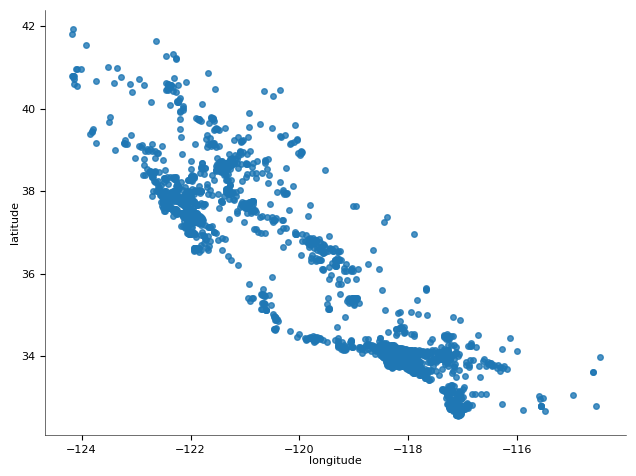
Distribution of Age

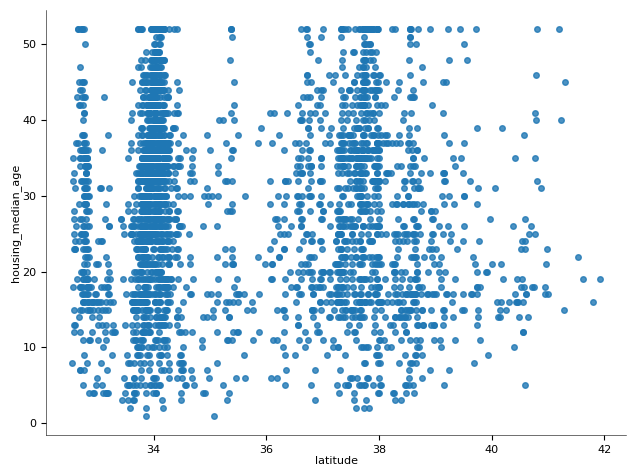


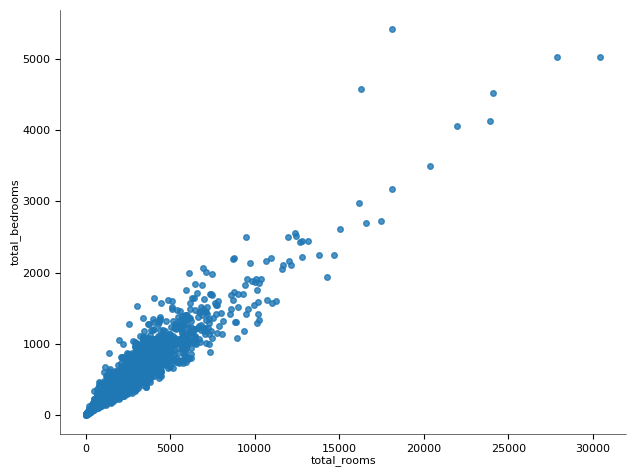
This step creates a histogram and kernel density plot to visualize the distribution of ages in the dataset.

**Visualization:**

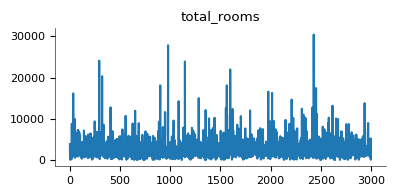
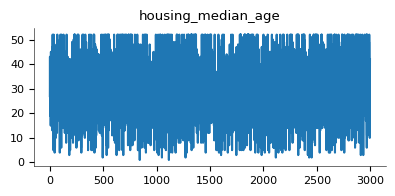
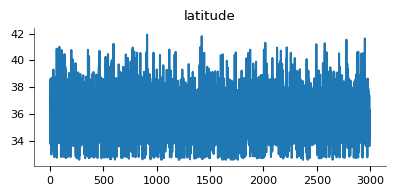
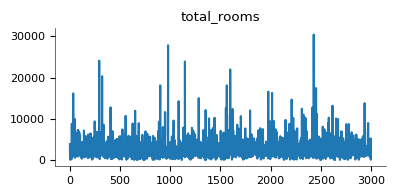








**Visualize prameter:**

\

**Conclusion:**

In conclusion, IBM Cognos is a powerful tool that revolutionizes water quality analysis through its advanced analytics capabilities. By integrating data from various sources, providing real-time monitoring, and offering a range of analytical techniques, it empowers organizations to make informed decisions for the conservation and management of water resources. The platform's user-friendly interface and comprehensive reporting features ensure that stakeholders at all levels can access and interpret water quality data effectively.