**Tittle : Water Quality Analysis**

# Introduction:

* Transportation efficiency is a critical factor in urban planning and sustainability. This document initiates the process of analyzing public transportation efficiency using IBM Cognos for visualization. Beginning with an exploration of the concept of transportation efficiency, we aim to collect, process, and clean relevant data to facilitate in-depth analysis. This analysis will provide valuable insights for improving public transportation system.

# *Analysis Objectives:*

* The primary objectives of this project are to assess and improve public transportation efficiency. This involves evaluating factors such as ridership trends, route optimization, on-time performance, and environmental impact. We seek to leverage IBM Cognos for data visualization to gain actionable insights, enhance decision-making for transportation authorities, and contribute to more sustainable and effective urban mobility systems.
* At present we tried visualisations that show how NumberOfBoardings is distributed across routes, stops and a week.

# *Data Cleaning and Preprocessing:*

In [1]:

import numpy as np import pandas as pd

import osfor dirname, \_, filenames **in** os.walk('/kaggle/input'):

for filename **in** filenames:

print(os.path.join(dirname, filename))

/kaggle/input/unisys/Water Quality Analysis.doc

* **Ph** is the general of common of the ph rate
* **Hardness** is the amount is dissolved
* **Solids** is the object
* **Conductivity** is the electricity conducting

# **Step-1:** Load the data set from the above link

[**https://www.kaggle.com/datasets/adityakadiwal/water-potability**](https://www.kaggle.com/datasets/adityakadiwal/water-potability)

# import pandas as pd = pd.read\_csv('/kaggle/input/unisys/survey.CSV', low\_memory=False)data.shapedata.head(10)



*# Step 2: Drop duplicates and Check data types of columns*data = data.drop\_duplicates()import seaborn as snsprint(data.dtypes)

Ph float

Hardness float

Solids float

Chllramines float

Sulfate float

Productivity int

*# Step 3: Check data types of columns*print("**\n**Check data types of columns")print(data.dtypes)

Ph float

Hardness float

Solids float

Chllramines float

Sulfate float

Productivity int

*# Step 4: Handle mixed data types#’Ph' column has mixed types, convert it to numeric*data['Ph'] = pd.to\_numeric(data[Ph], errors='coerce')print("Handle mixed data types")print(data.shape)

Handle mixed data types

(10857234, 6)

*# Step 5: Handle missing values# Drop rows with missing values or fill them based on your project required data*= data.dropna()print("**\n**Handle missing values")print(data.shape)

Handle missing values

(6414906, 6)

*#Step 6 : Unique values for each column in the DataFrame*print(data.nunique())

Ph float

Hardness float

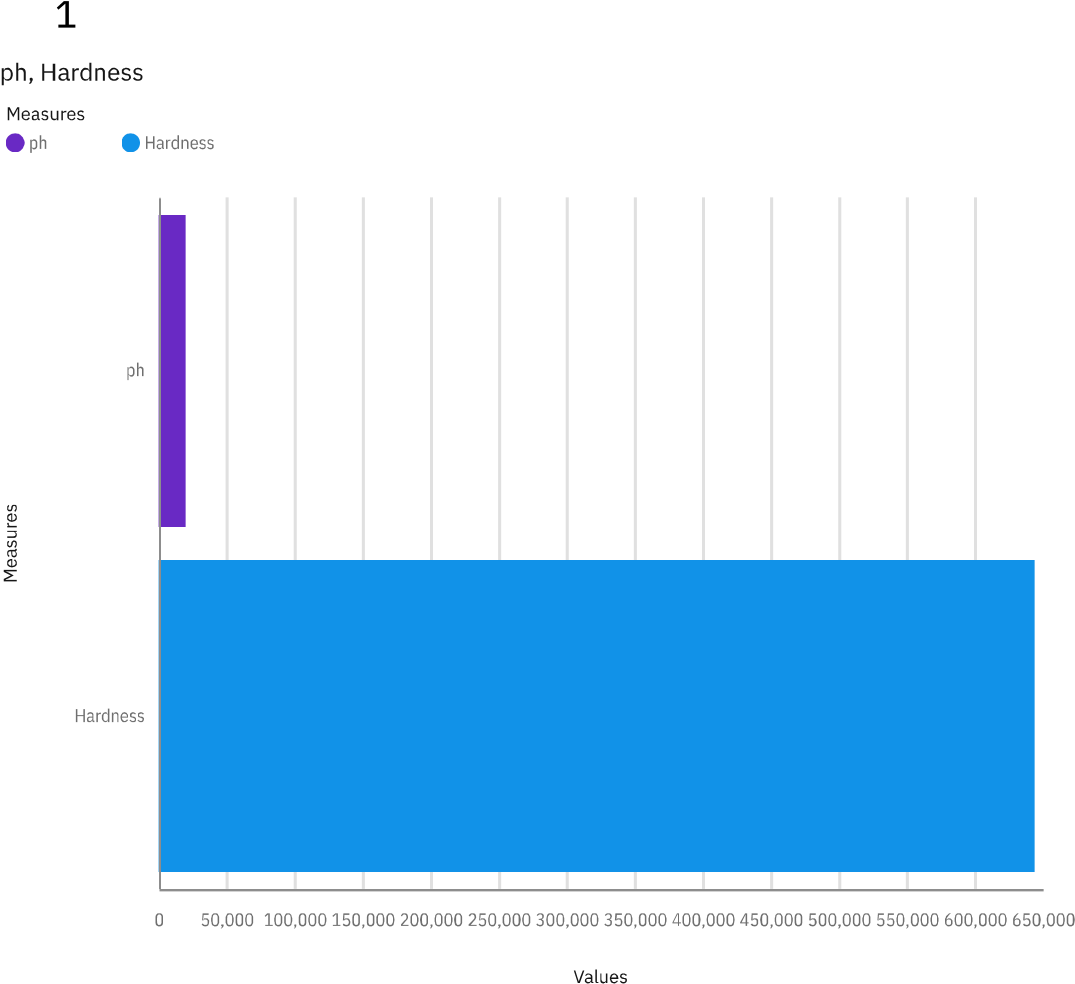
Solids float

Chllramines float

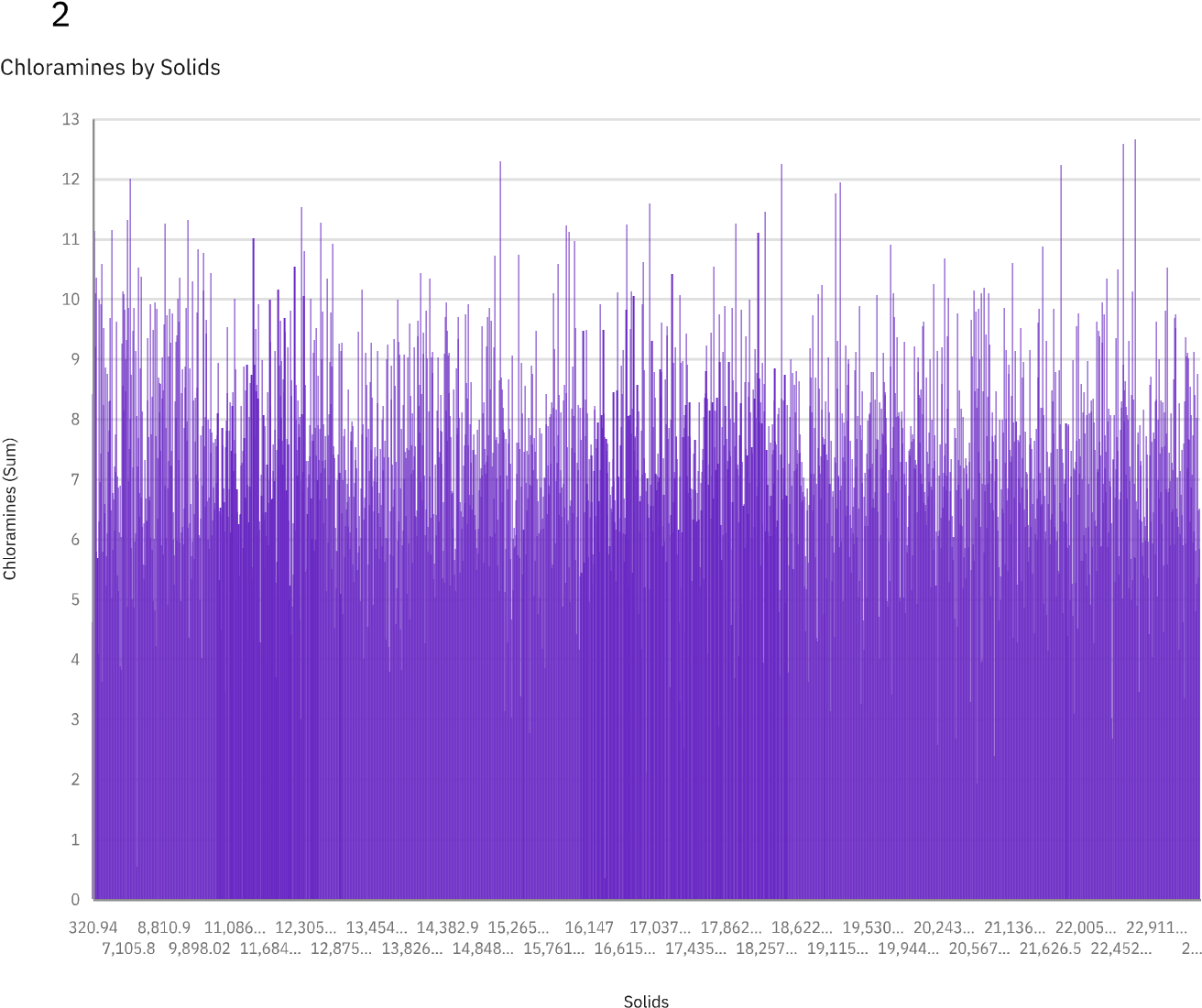
Sulfate float

Productivity int

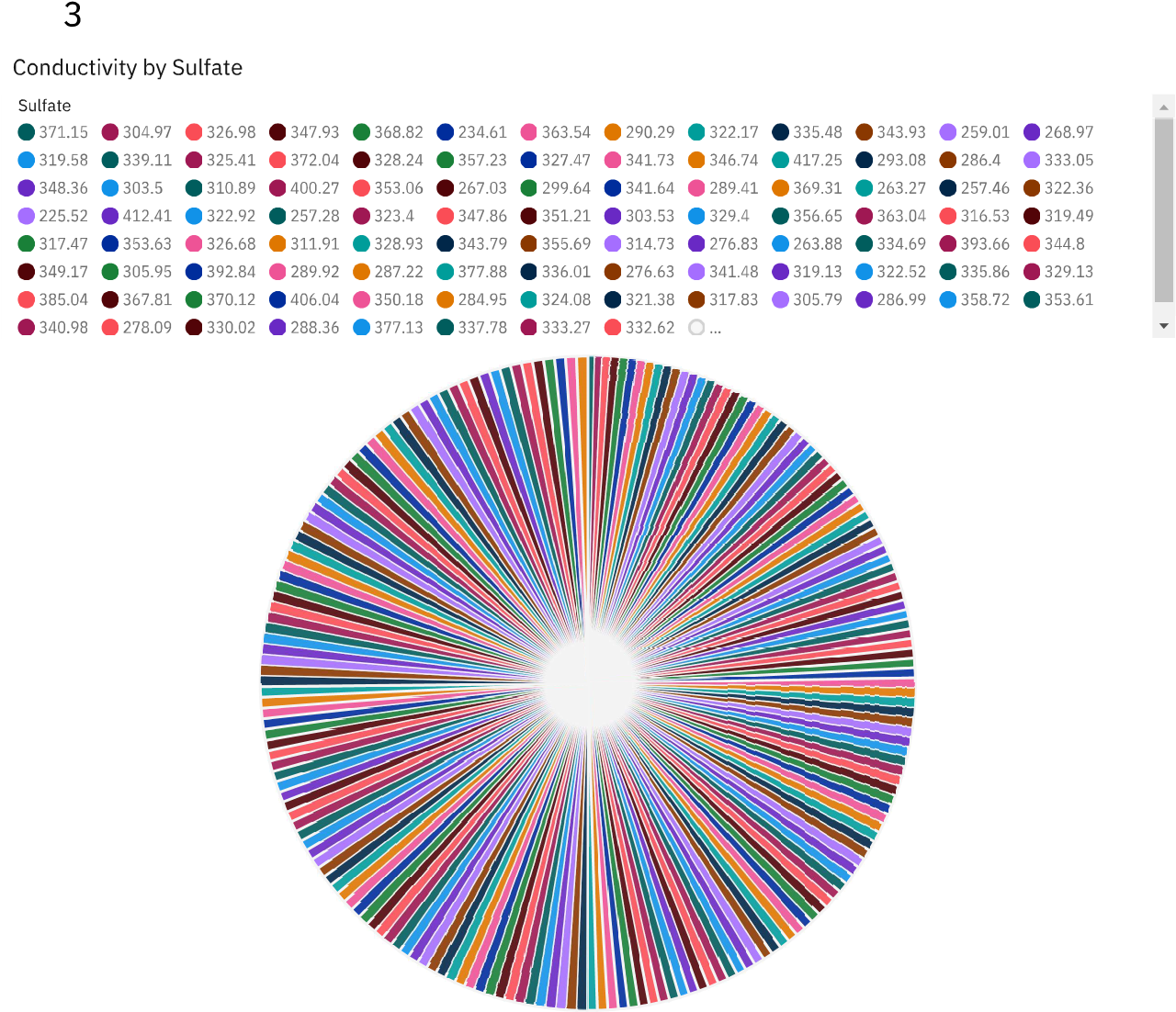
***Visualization on IBM Cognos:***



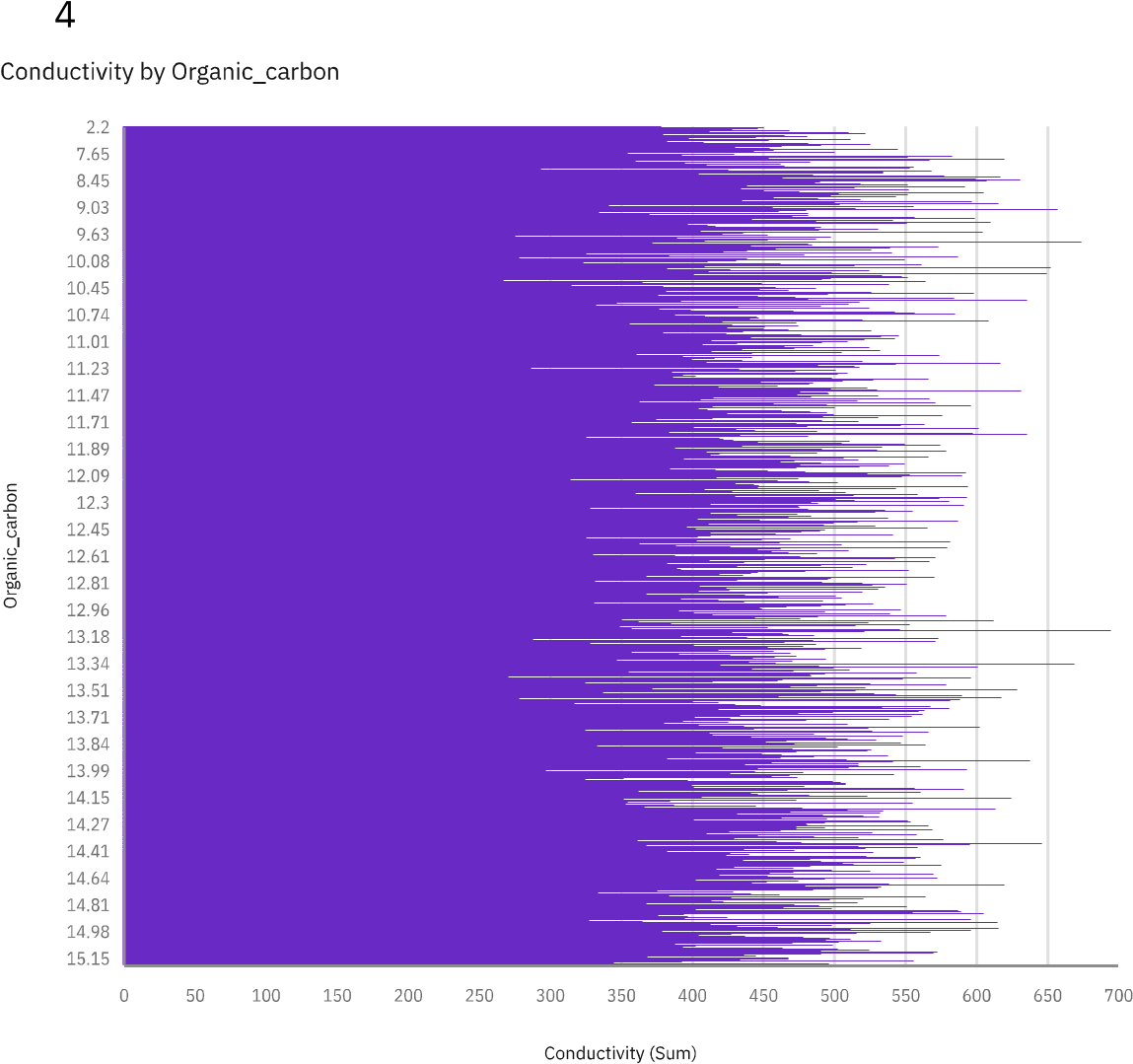
**Ph and Hardness is how the chemicals used in water**



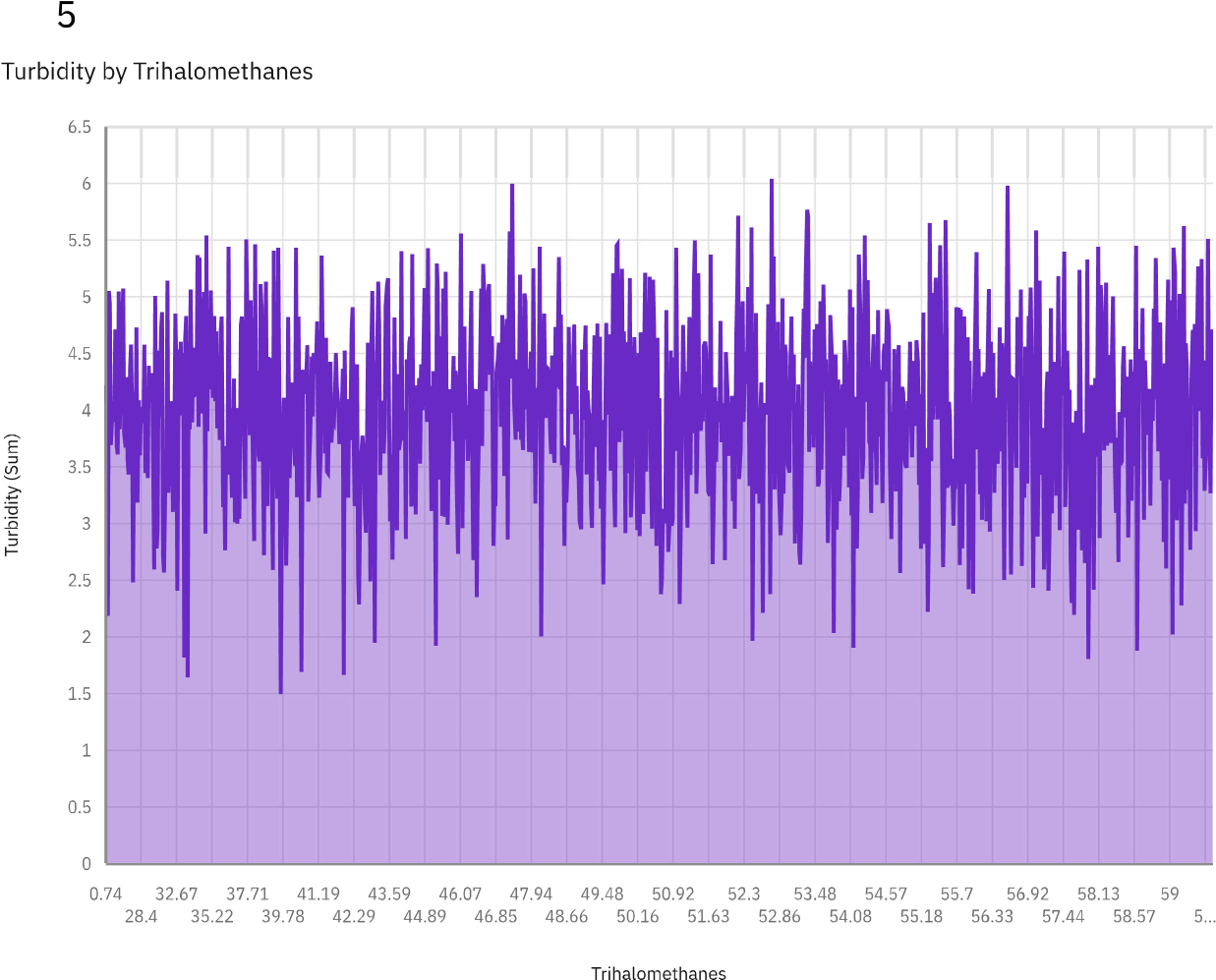
Chloramines and Solids are used in Water content



**Conductivity and sulfate in the Quality Checking**

**Conductivity is the electrical content which contains in water**

**And organic carbon also contains in water which intake to human**



**Turbidity is measure of relative clarity of a liquid.**

**Turbidity of water 0.5-1.0 NTU**

# *Conclusion:*

In this initial phase of the project, the dataset was effectively processed and cleaned to ensure its accuracy and reliability. Subsequently, compelling visualizations were generated using IBM Cognos, setting the stage for a comprehensive analysis of public transportation efficiency. These preparatory steps are essential for facilitating informed decision-making and shaping the future of urban transportation systems.