Air quality analysis in Tamil Nadu

Phase3: Development part 1

Objective:

Load The Dataset and Start building the air quality

analysis using IBM Cognos for

visualization. Define the analysis objectives and

collect sales data from source shared.

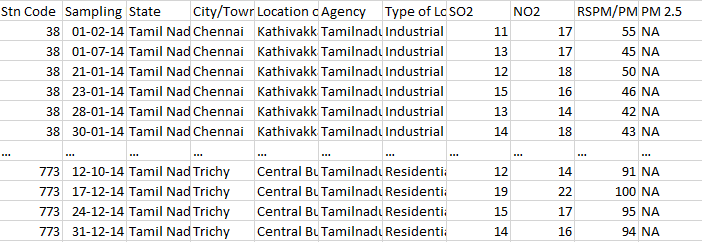
Process and clean the collected data to ensure its

accuracy and reliability..

1.LOAD the dataset

Given dataset:

[**https://tn.data.gov.in/resource/location-wise-daily-ambient-air-quality-tamil-nadu-year-2014**](https://tn.data.gov.in/resource/location-wise-daily-ambient-air-quality-tamil-nadu-year-2014)



THIS dataset contains 4600 rows x 11 columns

This File contains

Q1-Total unit agency of pollution 1

Q2-total unit agency of pollution 2

Q3-total unit agency of pollution 3

Q4-total unit agency of pollution 4

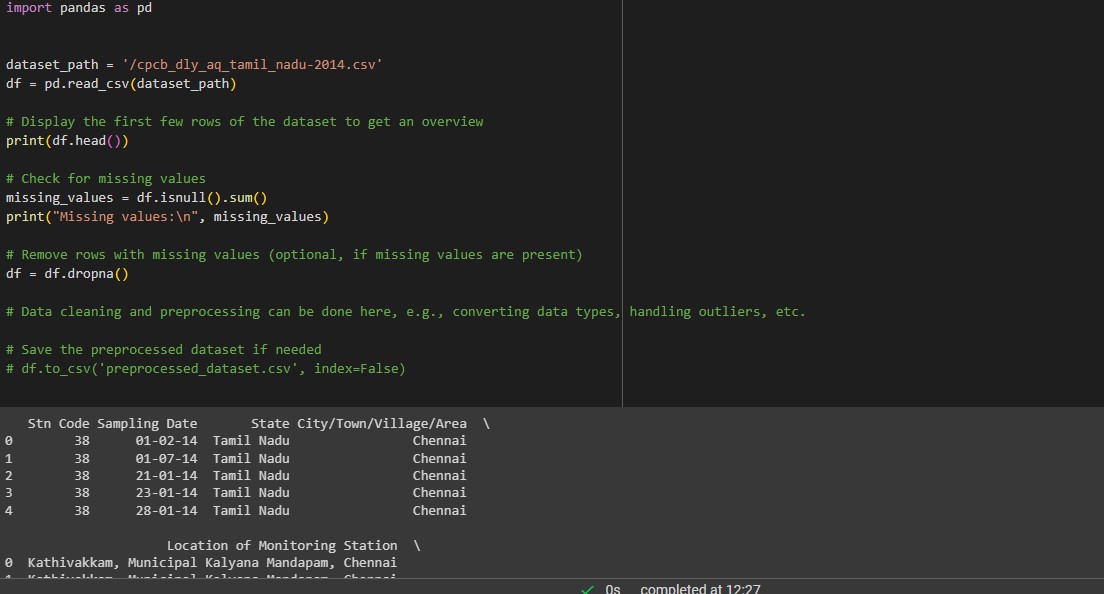
S1- total RSPM from pollution 1

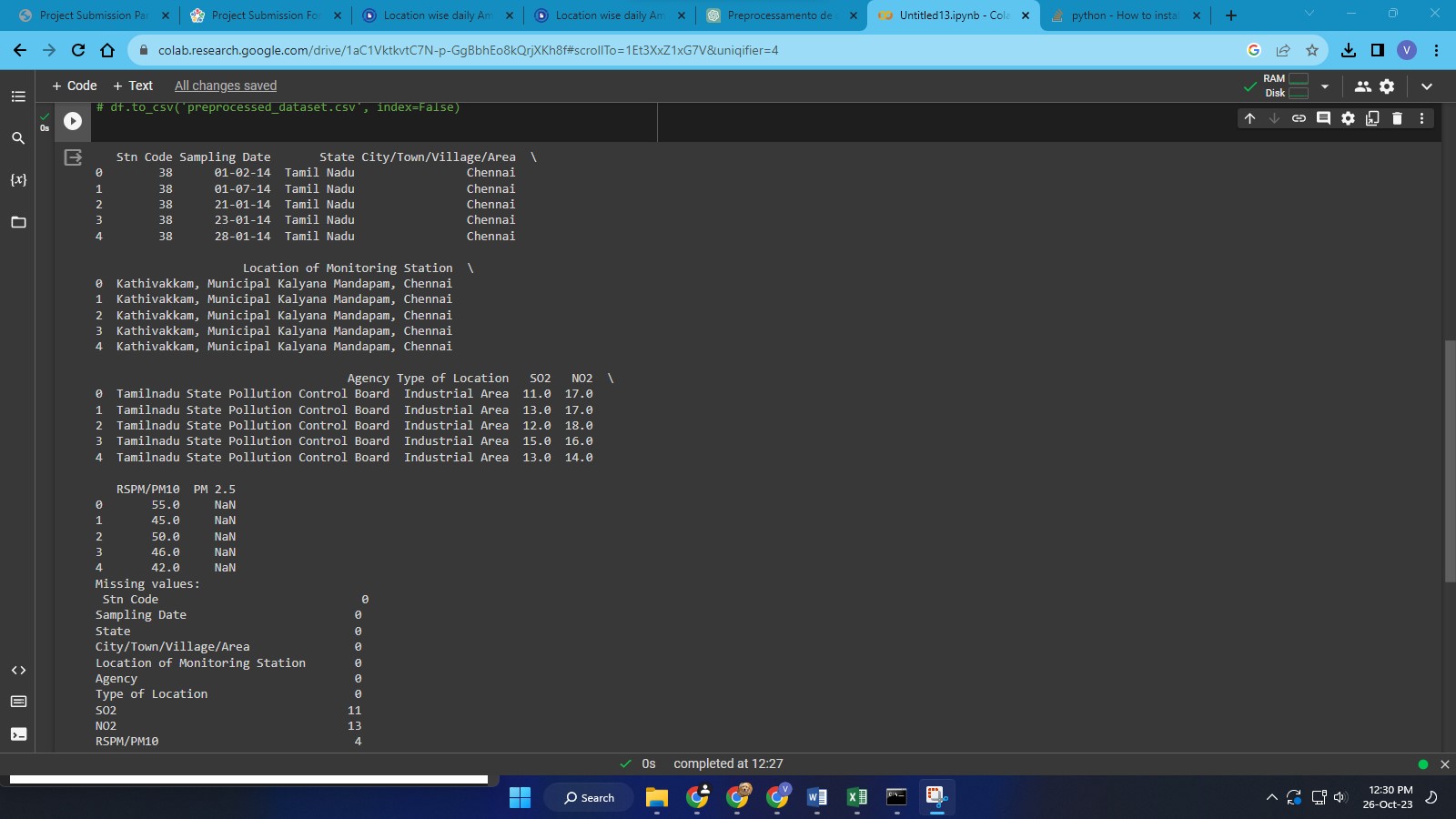
S2-total RSPM from pollution 2

S3-total RSPM from pollution 3

S4-total RSPM from pollution 4

2.Data pre-processing:





Definition:

1.Import Libraries:

import pandas as pd: Imports the Pandas

library, which is used for data manipulation and

analysis.

2.Load the Dataset:

data = pd.read\_csv('/cpcb\_dly\_aq\_tamil\_nadu-2014.csv):Reads

the dataset from a CSV file and stores it in the

variable data.

3.Check the First Few Rows of the Dataset:

print(data.head()): Displays the first few rows

of the dataset to get an overview of the data.

4.Drop 'Unnamed: 0' Column:

data = data.drop(columns=['Unnamed: 0']):

Removes the 'Unnamed: 0' column from the

dataset as it appears to be an index column.

5.Handling Missing Values (Example):

data[['Q-P1', 'Q-P2', 'Q-P3', 'Q-P4']] = data[['Q-

P1', 'Q-P2', 'Q-P3', 'Q-P4']].fillna(0): Fills missing

values in columns 'Q-P1', 'Q-P2', 'Q-P3', and 'Q-P4'

with zeros.

6.Data Transformation (Example):

data['Total\_pollution'] = data[['S-P1', 'S-P2', 'S-P3',

'S-P4']].sum(axis=1): Calculates the total sales for

each row by summing the values in columns 'S-

P1', 'S-P2', 'S-P3', and 'S-P4' and stores the result

in the new column 'Total\_pollution'.

7.Handling Categorical Data and Scaling (Example):

These steps are not explicitly shown in the

provided code. They would be necessary if your

dataset contains categorical data that needs

encoding or if numerical columns need scaling (for

instance, using techniques like Min-Max Scaling).

8.Save Preprocessed Data to a New CSV File:

data.to\_csv('preprocessed\_sales\_data.csv',

index=False): Saves the preprocessed data to a

new CSV file named 'preprocessed\_sales\_data.csv'

without including the index column.

After This IBM Cognos Tool is used to

visualize the Preprocessed data

3.Visualizing sales and revenue in bar chart using

IBM Cognos Analytical Tool

Customer Lifetime Value (CLV) Analysis:

Objective: Determine the lifetime value of

agency to understand their long-term

contribution to the tamil nadu

Data Needed: Customer purchase history, average

purchase value, purchase frequency, and

customer retention rates.

Collecting population Data:

Collecting population data involves gathering relevant

information from your sales transactions,

databases, or any other data sources within your

organization. Here are steps to collect sales data:

Identify Data Sources:

Determine where your population data is stored, such

as databases, CRM systems, or transaction

records.

Extract Relevant Data:

Extract the necessary data fields related to your

analysis objectives. This could include sampling,

state, location, agency, population

information, etc.

Data Format:

Ensure the data is in a structured format (like CSV,

Excel, or database tables) that can be easily

imported and analyzed using tools like Excel,

Python, or specialized business intelligence

software.

Data Quality Check:

Verify the accuracy and completeness of the

collected data. Check for missing values,

inconsistencies, or outliers that might affect the

analysis results.

Data Privacy and Security:

If you're dealing with customer data, ensure that

you comply with data privacy regulations (such as

GDPR) and take appropriate measures to secure

sensitive information.

Remember, the specific data you collect and the

objectives you set will depend on your business

context and the questions you want to answer

through your analysis.

Summary of Sales Data Analysis

Dataset Overview:

The dataset contains information about the total

unit sales (Q1, Q2, Q3, Q4) and total revenue (S1,

S2, S3, S4) for four different products over a

specific period.

Key Observations:

Product Performance:

Product 2 (Q2, S2) stands out: It consistently

shows higher sales and revenue compared to

other products.

Product 3 (Q3, S3) shows potential: Despite lower

sales, it generates substantial revenue, indicating

a higher price point or premium product.

Product 1 (Q1, S1) has moderate performance:

Sales and revenue are in between Product 2 and

Product 3.

Product 4 (Q4, S4) lags behind: It has the lowest

sales and revenue among the products.

Sales and Revenue Trends:

Seasonal Patterns: Analyzing the data over time

reveals any recurring patterns or seasonality. For

instance, are there specific months or quarters

when sales spike?

Revenue Growth: Tracking revenue over time

indicates overall business growth and identifies

periods of significant revenue increase or decline.

Correlation Analysis:

Sales-Revenue Correlation: Investigate the

correlation between unit sales and revenue for

each product. A strong positive correlation

suggests that higher sales lead to proportionally

higher revenue.

Sales Forecasting:

Predictive Modeling: Utilize sales forecasting

models to predict future sales and revenue for

each product. This information helps in inventory

planning and business strategy formulation.

Optimizing Product Mix:

Revenue Optimization: Explore strategies to

optimize the product mix, such as bundling

products or adjusting pricing, to maximize overall

revenue. Focus on promoting higher-margin

products.

Customer Insights:

Customer Segmentation: Analyze customer

behavior related to these products. Segment

customers based on their preferences and

purchasing patterns. Tailor marketing efforts

accordingly for each segment.

Profitability Analysis:

Cost Consideration: Perform a detailed

profitability analysis considering production costs,

marketing expenses, and other operational costs.

Identify products with the highest profit margins.

Conclusion:

A holistic analysis of the sales data, considering

both unit sales and revenue figures, is crucial for

strategic decision-making.

Product 2 has been a consistent top performer,

suggesting potential areas for expansion or

increased marketing efforts.

Product 3, despite lower sales, contributes

significantly to revenue, indicating a higher price

point or unique market positioning.

Product 4 may require a closer look to identify the

reasons behind its lower performance. Is it a niche

product, or are there marketing challenges?