### Project Report

### Int 217

**Project**

#### LOVELY PROFESSIONAL UNIVERSITY PHAGWARA, PUNJAB

****

**A Data-Driven Dashboard for Analyzing AIR Quality Analysis**

**SUBMITTED BY -** A Sairam Patro

**Registration Number:** 12306775

**Section** – K23GW

**Roll no**- 07

**Under the Guidance of:** Baljinder Kaur

#### DECLARATION

I, A Sairam Patro, hereby declare that the work done by me on “Excel Project” is a record of original work for the partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in Computer Science - Data Science, Lovely Professional University, Phagwara.

Signature Signature

Name: A Sairam Patro Mam Baljinder Kaur

Reg: No: 12306775 UID: 27952

# ACKNOWLEDGMENT

First and foremost, I would like to express my deepest gratitude to my college for providing me with the opportunity and resources to undertake this project.

I extend my sincere thanks to my Teacher, **Mam Baljinder Kaur,** for her invaluable guidance, constructive feedback, and constant encouragement throughout the project. Her expertise and support were instrumental in achieving the objectives of this work.

Thank you all

Table of Content

1. Introduction

2. Source of dataset

3. Dataset Preprocessing

4. Analysis on dataset (for each objective)

1. General Description
2. Specific Requirements
3. Analysis results
4. Visualization

5. Conclusion

6. Future scope

7. References

1. **Introduction** :-

In today's data-driven environment, the ability to visualize and interpret data efficiently is crucial. This project focuses on building an interactive dashboard in **Microsoft Excel** using the **AirQuality** dataset. The dataset contains key environmental indicators such as levels of pollutants, temperature, and corresponding time and date values.

The goal of this dashboard is to provide a clear, concise, and interactive overview of air quality trends and patterns. By utilizing Excel's features such as pivot tables, slicers, charts, and conditional formatting, we aim to transform raw data into a user-friendly tool for monitoring and analyzing air quality over time. This can assist environmental researchers, policy makers, and the general public in making informed decisions based on real-time or historical air quality metrics.

1. **Source of Dataset:-** <https://catalog.data.gov/dataset/air-quality>

### ****Dataset Preprocessing :-****

### Before building the dashboard, several preprocessing steps were carried out:

### Handling Missing Values NA or NAN Missing numerical values were replaced with the average value of the respective column.

### Data Cleaning Removed unnecessary columns and standardized text formatting such as location names.

### Column Selection Selected key columns relevant for analysis such as State, Location, Temperature, pH, DO, Conductivity, BOD, Nitrate, and Coliform levels.

### Data Type Fixing Ensured numeric columns were formatted correctly for accurate aggregation.

1. **Analysis on Dataset :-**

Objective 1 :- **Measure Count According to Name**

**i) General Description**

Analyze how many air quality measurements have been recorded for each **pollutant name**, helping to identify monitoring focus and data distribution across different pollutants.

**ii) Specific Requirements**

Create a **pivot table** with **Pollutant Name** (e.g., PM2.5, PM10, NO₂,O₃) in rows and **Count of Measurements** (records) in values.

**iii) Analysis Results**

Certain pollutants, such as **PM2.5 and NO2**, have significantly higher measurement counts.

**iv) Visualization**

Use a **Vertical bar chart** to show the count of measurements per pollutant name, with **slicers** for **pollutant name** filtering.

Objective 2:- **Count of Measurement Info According to Geo Place Name**

**i) General Description**

Assess how many air quality measurements have been recorded per geographic location to evaluate monitoring coverage and identify data concentration across regions.

**ii) Specific Requirements**

Generate a **pivot table** with **Geographic Place Name** in rows and **Count of Measurements** in values.

**iii) Analysis Results**

Some geographic areas show significantly higher counts, suggesting denser monitoring networks, while others have sparse data, indicating potential coverage gaps.

**iv) Visualization**

Use a **PieChart** showing measurement count by geographic area.

Objective 3:- **Count of Measurements**

**i) General Description**

Track the total number of air quality measurements recorded to understand data volume, frequency, and trends in monitoring efforts.

**ii) Specific Requirements**

Create a **pivot table** showing the **total count of measurements**.

**iii) Analysis Results**

The overall count highlights fluctuations in data collection—some periods or stations show consistent monitoring, while others have gaps or lower activity.

**iv) Visualization**

Display a **Bar** **Graph** (for total count). with **slicers** for **Measure** to explore the data dynamically.

#### Objective 4:- Count of Places Where Geo Name is Found

#### i) General Description

#### Identify the number of unique geographic locations where air quality data has been recorded.

#### ii) Specific Requirements

#### Create a pivot table or summary table with:

#### Geo Place Name in rows

#### Count of Unique Places or a distinct count of geographic names in values.

#### iii) Analysis Results

#### The dataset covers a specific number of unique locations. Some regions may have multiple monitoring sites contributing to dense data coverage, while others may have little or no presence.

#### iv) Visualization

#### Bar Chart (Optional): Display count of sites per state or zone.

#### Objective 5:- Sum of Data Values According to Geo Names

#### i) General Description

#### Calculate the total (sum) of recorded air quality data values for each geographic location to understand the overall pollutant load across regions.

#### ii) Specific Requirements

#### Create a pivot table with:

#### Geo Place Name (e.g., City, District) in rows

#### Sum of Data Values (e.g., PM2.5, PM10, NO₂, etc.) in values

#### iii) Analysis Results

#### Regions with higher cumulative pollutant values indicate areas with consistently high pollution levels or frequent monitoring. This helps identify pollution hotspots and prioritize areas for action.

#### iv) Visualization

#### Clustered Bar Chart to compare total pollutant values across places

1. **Conclusion** :-  
    The analysis highlights uneven air quality monitoring across locations, with some areas showing high pollutant levels and frequent measurements, while others lack data. Visual insights reveal pollution hotspots and data gaps, supporting informed decisions for targeted environmental action and improved air quality management.
2. **Future Scope:-**

Future work can include integrating real-time air quality data, forecasting pollution levels using machine learning, and comparing seasonal or year-over-year trends. Expanding geographic coverage and correlating health impact data will enhance insights, enabling proactive policy-making and more effective environmental interventions.

1. **References :-**

**DATASET:-** [**https://catalog.data.gov/dataset/air-quality**](https://catalog.data.gov/dataset/air-quality)

**Cleaning:-** [**https://chatgpt.com/**](https://chatgpt.com/)

**Research articles and journals related to air pollution and environmental monitoring.**

**LINKEDIN LINK:-**

**[https://www.linkedin.com/posts/sairam-patro\_dataanalytics-airquality-exceldashboard-activity-7316484892639277056-vJS\_?utm\_source=share&utm\_medium=member\_desktop&rcm=ACoAAEe-IxoB1mAiMovF4HTEYaqxS-\_Ffwi3zag](https://www.linkedin.com/posts/sairam-patro_dataanalytics-airquality-exceldashboard-activity-7316484892639277056-vJS_?utm_source=share&utm_medium=member_desktop&rcm=ACoAAEe-IxoB1mAiMovF4HTEYaqxS-_Ffwi3zag%0c)**