# 

**INT217:INTRODUCTION TO DATA MANAGEMENT**

**PROJECT REPORT**

(Project Semester January-April 2025)

***(“Breathe NYC: An Interactive Dashboard for Monitoring Urban Air***

***Pollution")***

Submitted by

Anshika Dubey

Registration No: - 12316623

Programme and Section: K23GS

Course Code: INT 217

Under the Guidance of   
 Baljinder Kaur(32130)

**Discipline of CSE/IT**

**Lovely School of computer science and engineering**

**Lovely Professional University, Phagwara**

**CERTIFICATE**

This is to certify that Anshika Dubey bearing Registration no. 12316623 has completed INT-217

Project titled, (“Breathe NYC: An Interactive Dashboard for Monitoring Urban Air Pollution")

“under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

**Signature and Name of the Supervisor**

**Designation of the Supervisor**

**School of …………………………………………….**

Lovely Professional University

Phagwara, Punjab.

Date: 10-04-2025

**DECLARATION**

I Anshika Dubey student of Computer science and Engineering under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 10-04-2025 Signature

Registration No. 12316623 Name: -Anshika Dubey

**Acknowledgement**

I would like to express my sincere gratitude to Ms. Baljinder Kaur, our subject teacher for *Data Management*, for her invaluable guidance and support throughout the course of this project. Her encouragement, insightful suggestions, and detailed feedback have been instrumental in shaping this report and helping me develop a deeper understanding of data visualization and analysis.

I would also like to thank the developers and curators of open data platforms like [data.gov](https://catalog.data.gov/) for providing access to high-quality datasets that made this project possible. Lastly, I extend my appreciation to my peers and family for their motivation and constant support during the development of this dashboard.

**Table of Contents**

1. Introduction
2. Source of Dataset

3.Exploratory Data Analysis (EDA) Process

4Analysis on Dataset(for all the objectives)

4.1. Introduction  
4.2. General Description  
4.3. Specific Requirements, Functions, and Formulas  
4.4. Analysis Results  
4.5. Visualization

5.Conclusion

6.Future Scope

7.References

# Abstract

This project presents an in-depth analysis of nitrogen dioxide (NO₂) pollution levels across New York City using a comprehensive Excel-based dashboard. Air pollution poses a significant threat to public health and the environment, particularly in urban centers like NYC. Nitrogen dioxide, a harmful pollutant primarily produced from vehicular and industrial emissions, plays a crucial role in air quality deterioration. This study utilizes publicly available air quality data sourced from [data.gov](https://catalog.data.gov/dataset/air-quality/resource/f3ed1638-92da-4f88-bb6b-7d3940514574) and converts it into an interactive visualization dashboard. The dashboard enables users to observe trends, identify pollution hotspots, understand seasonal variations, and compare measurement metrics, offering valuable insights for urban planners, environmental analysts, and the general public.

Through five well-defined objectives, the project investigates temporal changes, spatial disparities, seasonal fluctuations, and metric-based analysis of NO₂ levels. This report details the methodology, data processing steps, dashboard design, visual interpretation, and outcomes of each objective. The resulting interactive Excel tool highlights the potential of low-cost data visualization platforms in supporting environmental decision-making and promoting awareness.

# Introduction

Air quality monitoring has emerged as an essential aspect of urban planning and public health. Among the various air pollutants

monitored, nitrogen dioxide (NO₂) is a significant indicator due to its adverse health impacts and its association with combustion processes. NO₂ can aggravate respiratory diseases, reduce lung function, and contribute to the formation of ground-level ozone and particulate matter.

New York City, one of the largest metropolitan areas in the world, faces air quality challenges due to its dense population, heavy traffic, and industrial activities. Monitoring and analyzing NO₂ levels can help identify patterns and hotspots that require policy attention. This project is an effort to harness the power of Excel to build a dynamic, interactive dashboard that supports the exploration of NO₂ pollution trends and distribution.

The dataset used is sourced from the official US government open data portal and contains records of pollutant levels across various community districts in NYC. Key fields include date of observation, pollutant type, measurement method (mean or max), community district (Geo Place Name), and season.

The project follows a step-by-step approach:

1. **Data Cleaning and Preprocessing**
2. **Creation of Derived Columns (e.g., Season)**
3. **Pivot Tables and Charts for Analysis**
4. **Use of Slicers for Interactivity**
5. **Dashboard Design and Layout**

Each section of the dashboard corresponds to a specific objective that addresses an analytical question related to NO₂ pollution in NYC.

# II. Source of the Dataset

Air pollution is one of the most important environmental threats to urban populations and while all people are exposed, pollutant emissions, levels of exposure, and population vulnerability vary across neighborhoods. Exposures to common air pollutants have been linked to respiratory and cardiovascular diseases, cancers, and premature deaths. These indicators provide a perspective across time and NYC geographies to better characterize air quality and health in NYC. Data can also be explored online at the Environment and Health Data Portal: <http://nyc.gov/health/environmentdata>.

# III. Data Processing

**Cleaning Steps**

1. Load and Preview Data
2. Remove Duplicate Rows
3. Handle Missing Values
4. Standardize Column Names
5. Fix Data Types
6. Format Dates

In the Context of the Dataset the cleaning that was required is as follows:

**BEFORE CLEANING THE DATASET**

1.No missing values in important columns like Data Value, Geo Place Name, and Time Period.

2. Start\_Date was in a string format converted it to datetime following the steps:

Selected the Start\_Date column.

**Data** tab---**Text to Columns**----**Delimited**--- **Column data format** ----**Date**----**MDY**

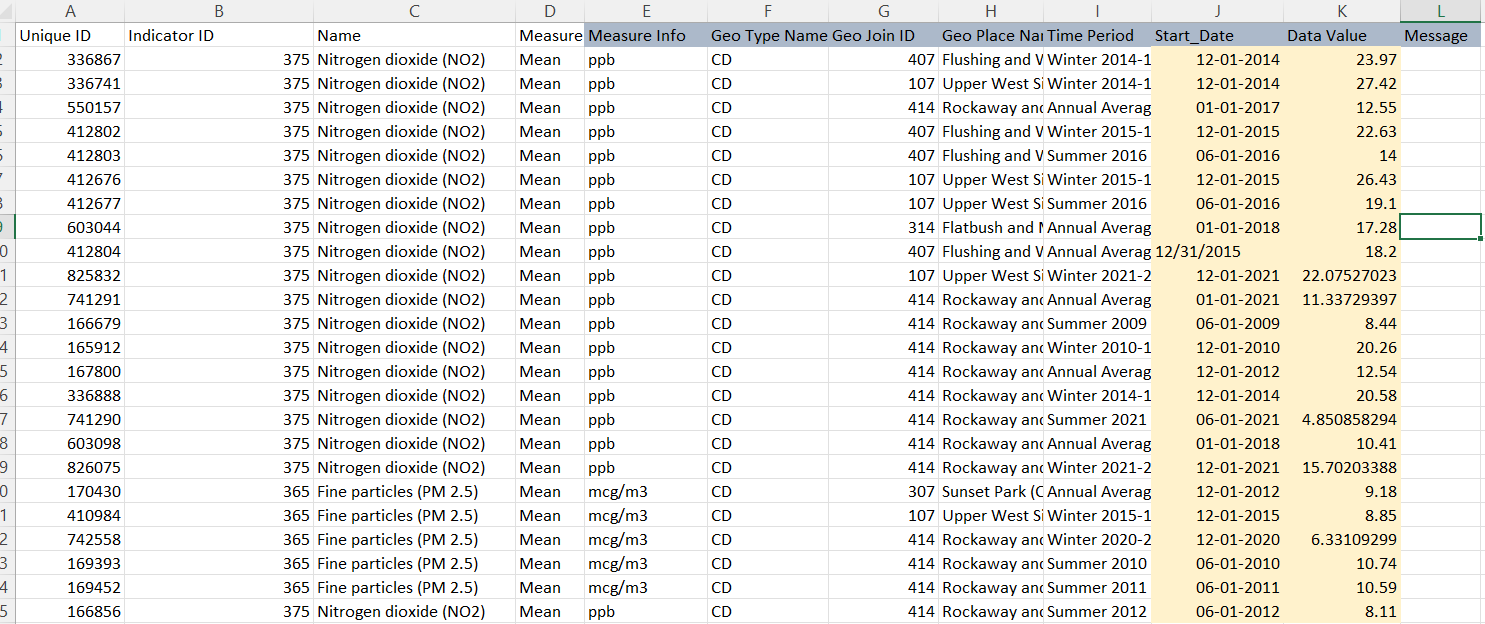
3.Rename columns: replace spaces with underscores.

4. Messgae coloumn was dropped as it was blank

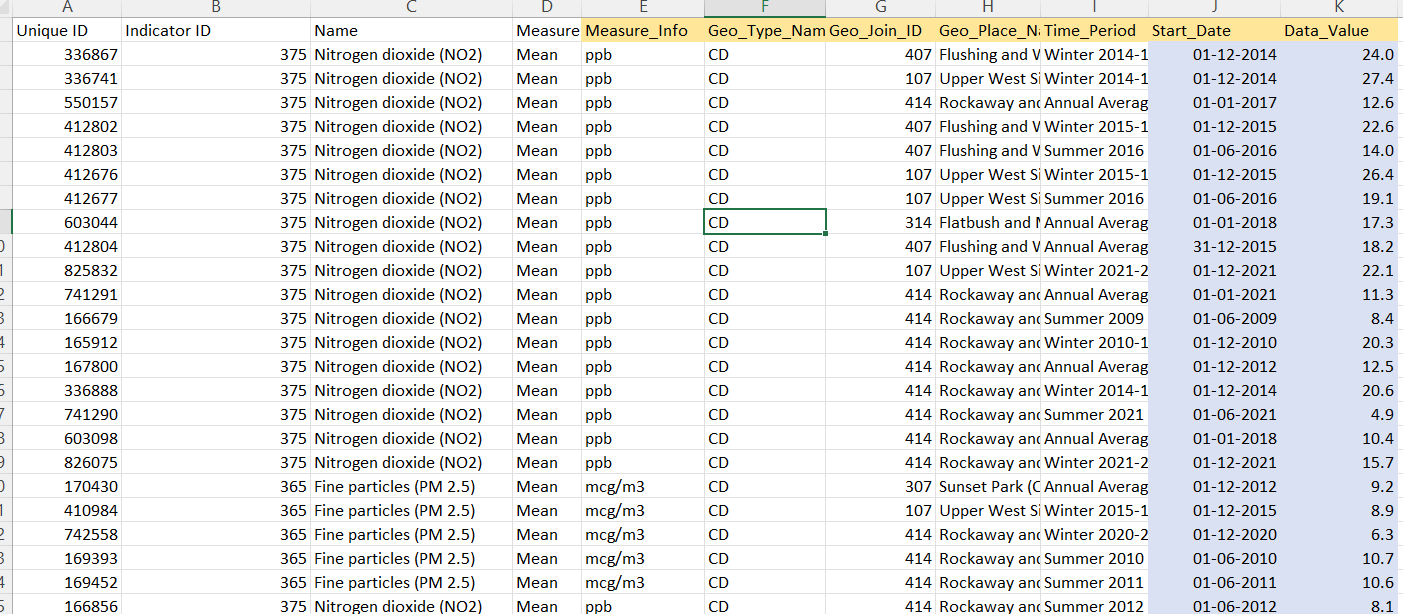
**After :**

1. No missing values: All 11 columns have complete data
2. Consistent data types
3. Converted date which was previously in string format to proper date format
4. Proper column structure

**. Analysis on Dataset**

BEFORE:

AFTER:



# (Objective 1): Analyze how air pollutant levels (e.g., NO₂) have changed over time across NYC.

## 

## General Description

Created a line chart to analyze how nitrogen dioxide (NO₂) concentration levels have changed over time. The chart plots the average NO₂ data values by year and month to identify patterns, fluctuations, or long-term trends in air quality.

**ii. Specific Requirements:**

* Filter: Selected only rows where Pollutant = NO₂ (Name column)
* Pivot Table Configuration:
  + Values: Average of Data\_Value
  + Rows: Start\_Date broken down into Year(Row) and Month(Column)
* Chart Type: Line chart

Connected to pivot table

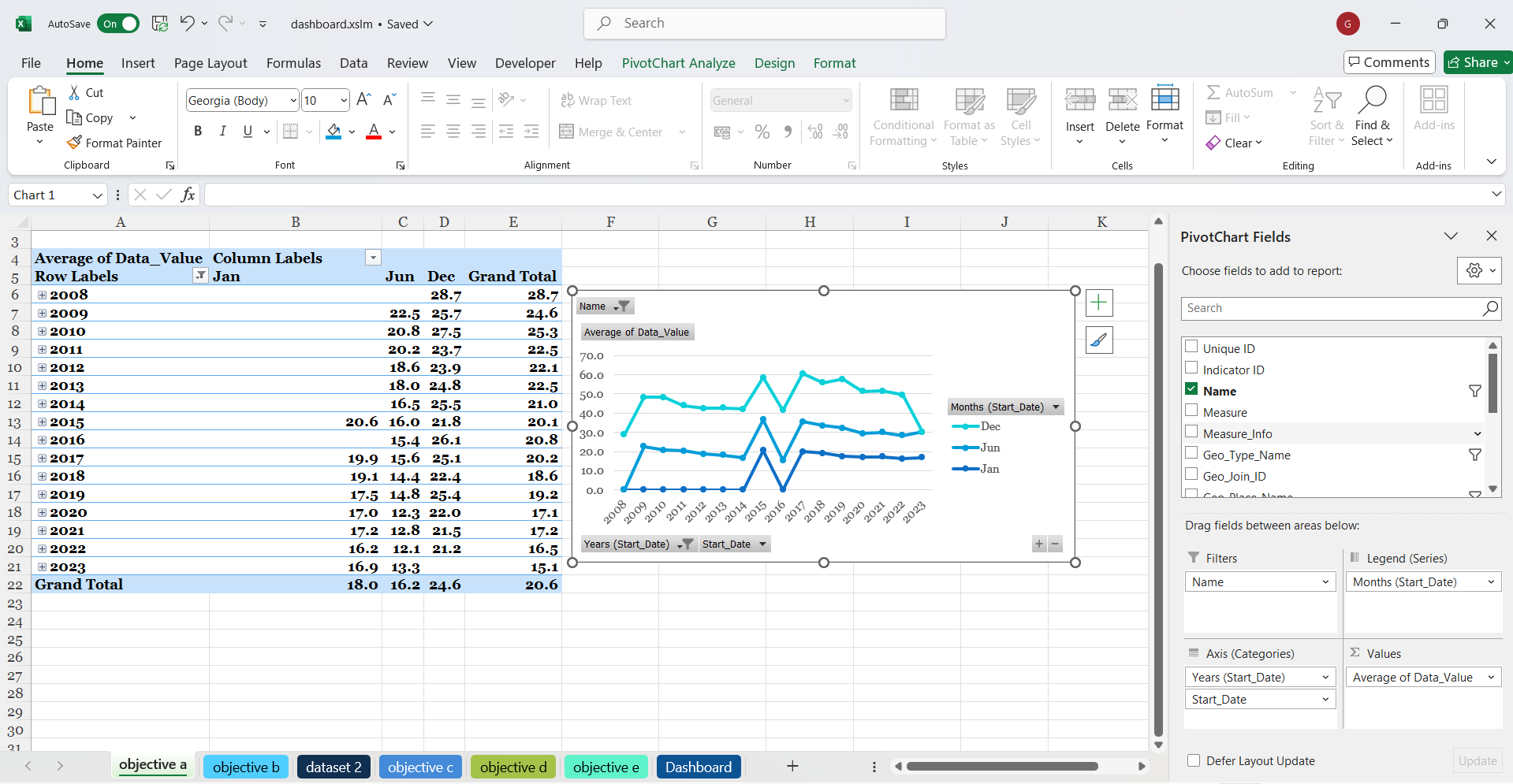
 Each line represents a **month’s average NO₂** level across different years

* Slicers Used:
  + Geo\_Type\_Name
  + Month

Analysis Results

* **Clear Seasonal Trends:** NO₂ levels tend to peak during colder months (especially January and December), likely due to heating and increased vehicular emissions**.**
* ** Year-over-Year Trends:** A gradual decline in average NO₂ levels over the years indicates improved air quality, potentially due to policy measures or decreased traffic emissions.
* Some fluctuations still occur, especially during winter.
* ** Summer Dip**: Across most years, the lowest NO₂ concentrations were recorded in June through August, reflecting seasonal dispersion due to higher temperatures and lower fuel combustion.

## v. Visualization

****

# (Objective 2): Compare Air Quality Levels Spatially Across Different Community Districts

**i. General Description:**

To evaluate the spatial distribution of air quality by comparing NO₂ levels across various community districts. The goal is to identify geographical disparities and potential pollution hotspots that may warrant targeted environmental interventions or public health initiatives.

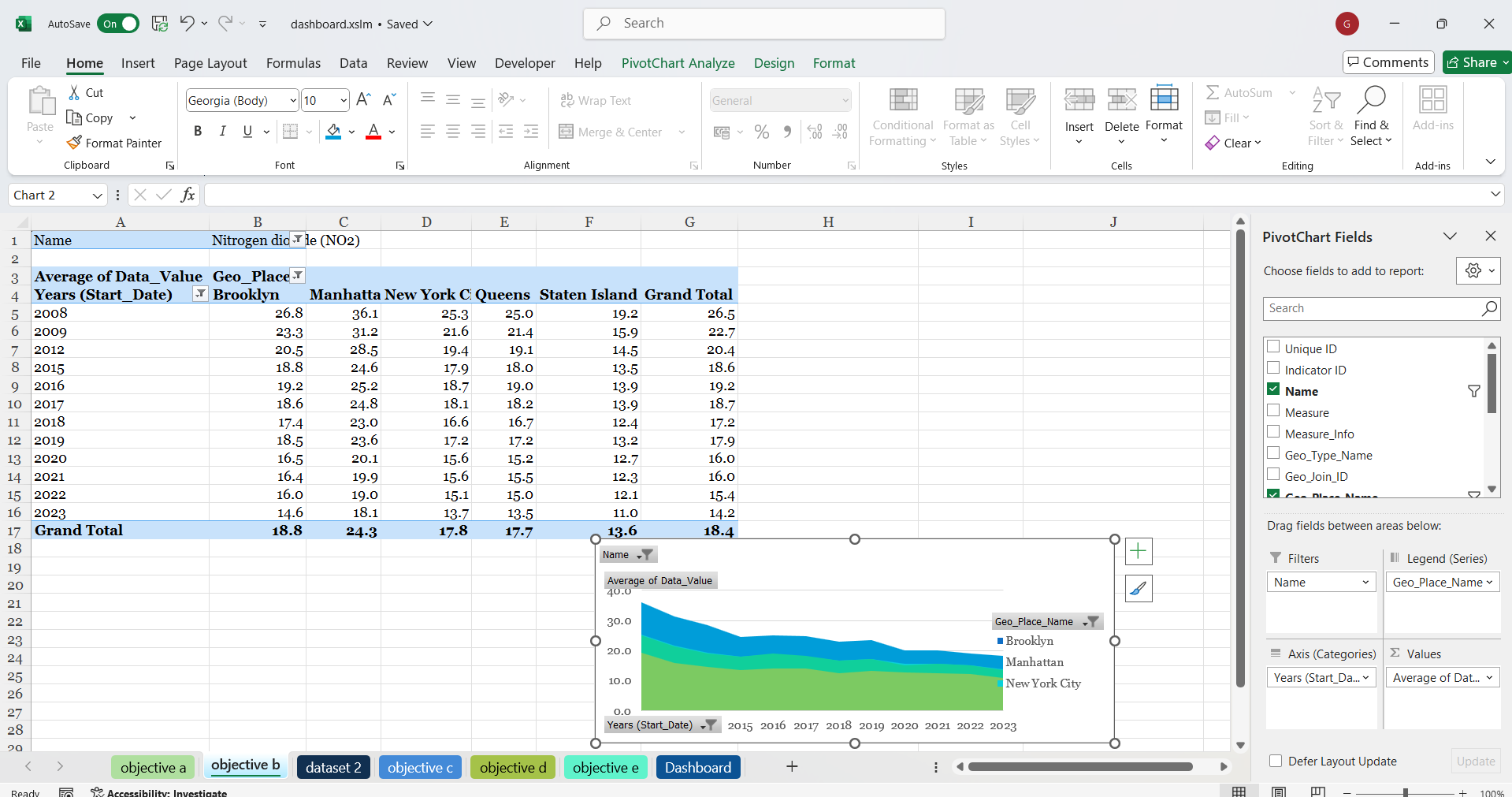
**ii.Specific Requirements:**

* **Pollutant Filter**: Name=NO₂
* **Geo Place Name Filter**: Selected only boroughs/cities within NYC (e.g., Manhattan, Queens, Staten Island, Brooklyn, etc.)
* **Pivot Table Configuration**:
  + **Columns**: Geo\_Place\_Name
  + **Rows**: Year
  + **Values**: Average of Data\_Value
* **Chart Type**: Area map
* **Slicers Used**:
  + Year
  + Season(objective 3)

**iii. Analysis Results:**

* Districts like **Manhattan** and **Bronx** showed consistently **higher average NO₂ levels**, indicating prolonged exposure to air pollution.
* Less densely populated areas like **Staten Island** had **lower average values**, implying better air quality.
* This analysis helps identify **high-risk regions** for targeted air quality interventions.

## 

****

# Objective 3: Distribution & KDE

## i. Introduction

To assess the influence of seasonal changes on ambient NO₂ concentrations by analyzing and comparing pollutant levels recorded during different periods (e.g., winter, summer, annual averages). This objective seeks to determine whether specific seasons consistently exhibit higher or lower pollution levels.

## General Description

A stacked column chart was created to show how NO₂ pollution levels vary by season and year. This helps to understand seasonal trends in air quality across NYC.

**ii. Specific Requirements:**

* Rows: Years (to observe annual trend)
* Columns: Season (Winter, Spring, Summer, Fall)
* Formula used to get the Season

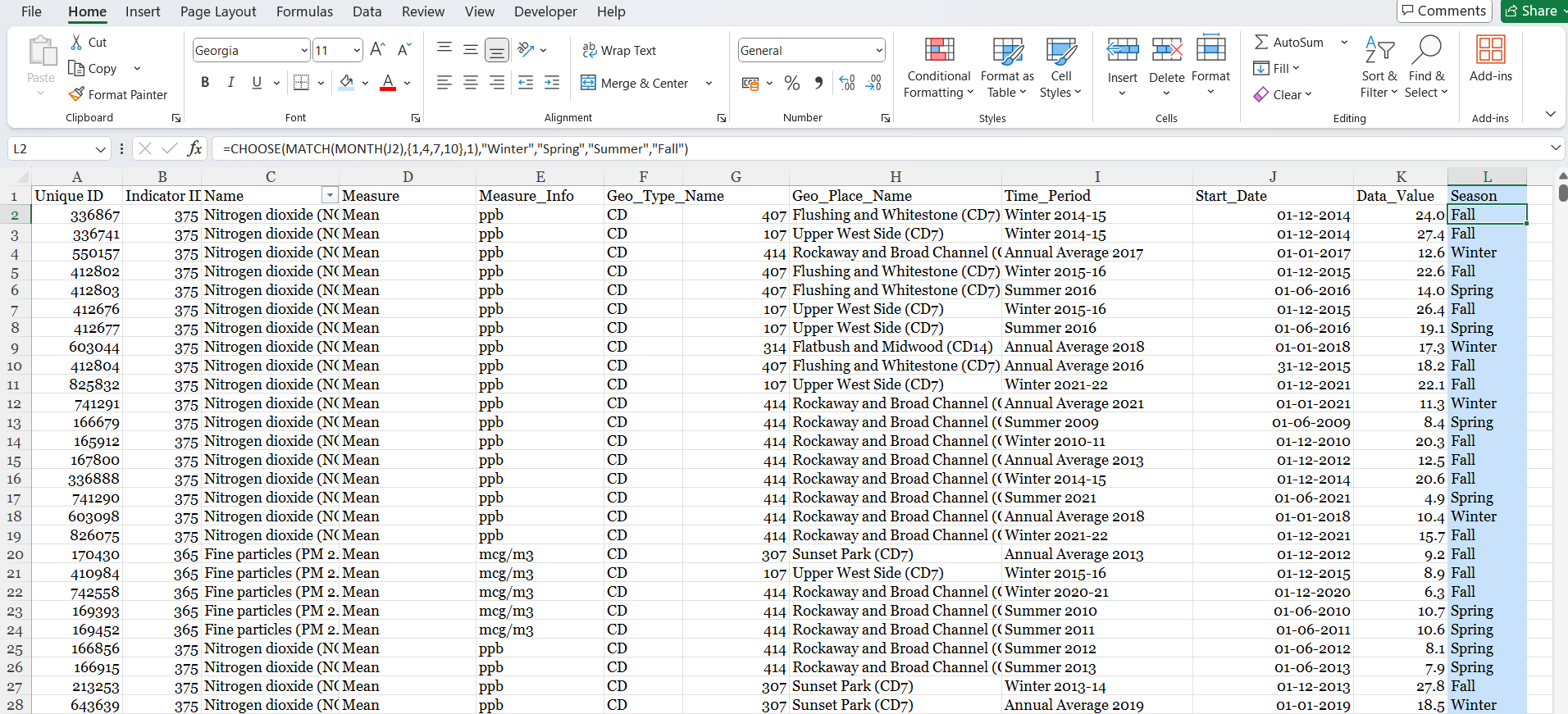
=CHOOSE(MATCH(MONTH(J2), {1,4,7,10}, 1), "Winter", "Spring", "Summer", "Fall")

* Extracts the month number from the date in cell J2.
* **MATCH(MONTH(J2), {1,4,7,10}, 1)**

This finds the **position** where the month fits within the array {1, 4, 7, 10} using **approximate match** (since the third argument is 1).

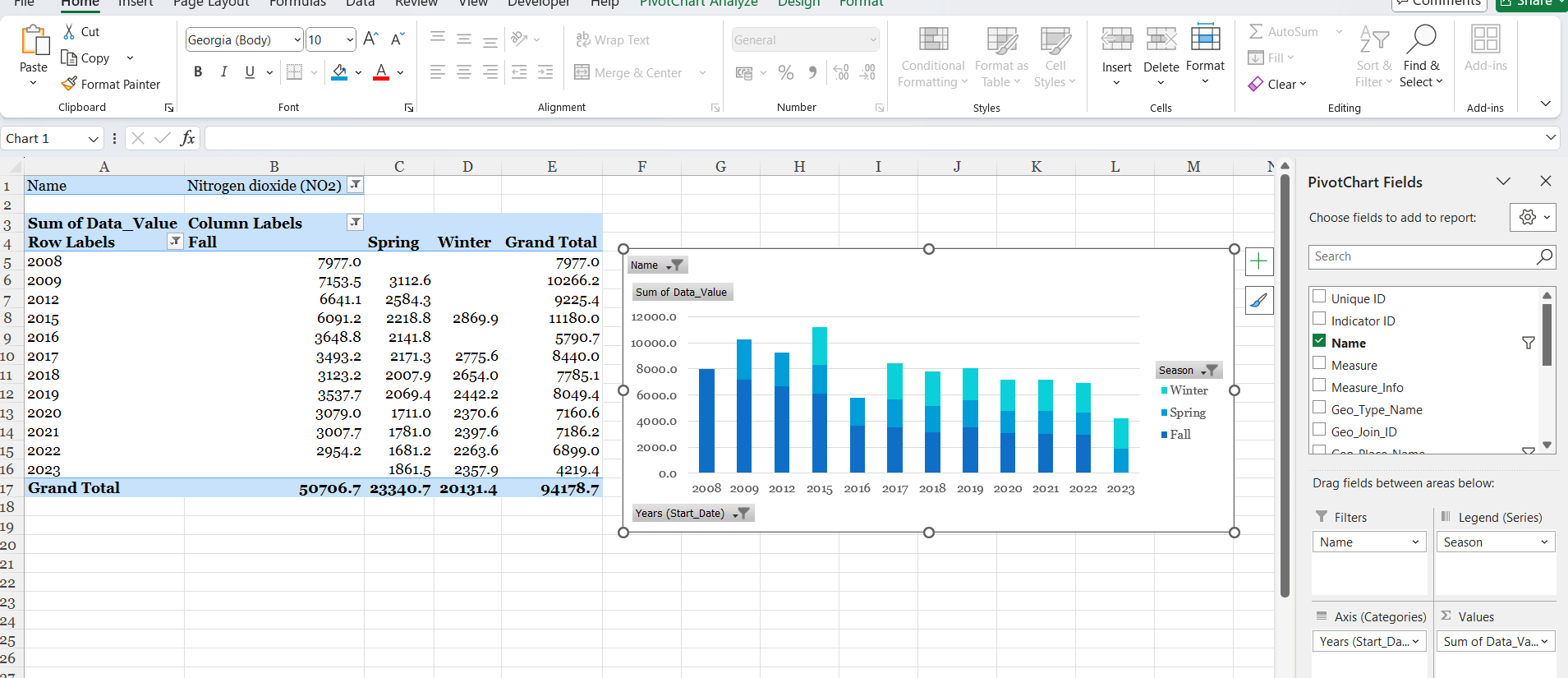
| **Value Range** | **Match Returns** | **Season** |
| --- | --- | --- |
| * Jan (1) – Mar (3) | 1 | Winter |
| * Apr (4) – Jun (6) | 2 | Spring |
| * Jul (7) – Sep (9) | 3 | Summer |
| Oct (10)– Dec (12) | 4 | Fall |

* 01-12-2014-----Fall as the date falls in 4th category (Oct-Dec)
* Values: Sum of Data\_Value (total NO₂ levels per season per year)
* Filter Applied: Pollutant = NO₂

****

**iii. Analysis Results:**

* Winter and Fall consistently recorded higher NO₂ levels, likely due to increased heating usage and poor dispersion.
* Spring and Summer generally showed lower pollution, possibly due to better air circulation and reduced emissions.
* Seasonal differences highlight the need for season-specific policy actions, especially in colder months.

** OBEJECTIVE-3**

# (Objective 4): Identify high-risk locations with consistently elevated NO₂ levels.

**i. Introduction**

To identify community districts that consistently report elevated levels of nitrogen dioxide over multiple time periods. This helped in prioritizing regions for air quality improvement programs or further investigation.

## *General Description*

You created a pivot table to identify geographic locations (districts or places) that have the highest

average concentration of Nitrogen Dioxide (NO₂) over the entire time period in the dataset  
  
Specific Requirements

T

* **Dataset Filtering:**
  + Selected only rows where the **Pollutant = NO₂**.
* **Pivot Table:**
  + **Rows:** Geo\_Place\_Name (Top 7)
  + **Values:** Average of Data\_Value to represent the typical NO₂ levels.
* **Visualization:**
  + Created a **Pie Chart** to showcase the top 7 locations with the highest NO₂ levels.
  + **Slicer :**Measure slicer to allow comparison of different metrics (

**iii. Analysis Result**

The analysis identified the top 7 community districts with the highest average NO₂ levels, signifying potential high-risk zones for long-term exposure to air pollution. This chart visually communicates which neighborhoods may require environmental monitoring, health assessments, or policy interventions.

* Using the slicer, stakeholders can switch to Max values to detect areas with dangerous spikes in pollution.
* This analysis supports targeted decision-making for improving public health and environmental quality.

# 

# Box Plot (Objective 5): Comparative Analysis of Measurement Metrics

## Introduction To examine how different measurement types (e.g., mean concentration vs. maximum concentration, if available) affect the interpretation of air quality data. This objective may help refine the criteria used for environmental assessment and reporting.

**General Description**

This objective compares how different **measurement metrics**—specifically **average** and **maximum values**—for nitrogen dioxide (NO₂) vary across different **locations** and **time periods**.

**ii. Specific Requirements**

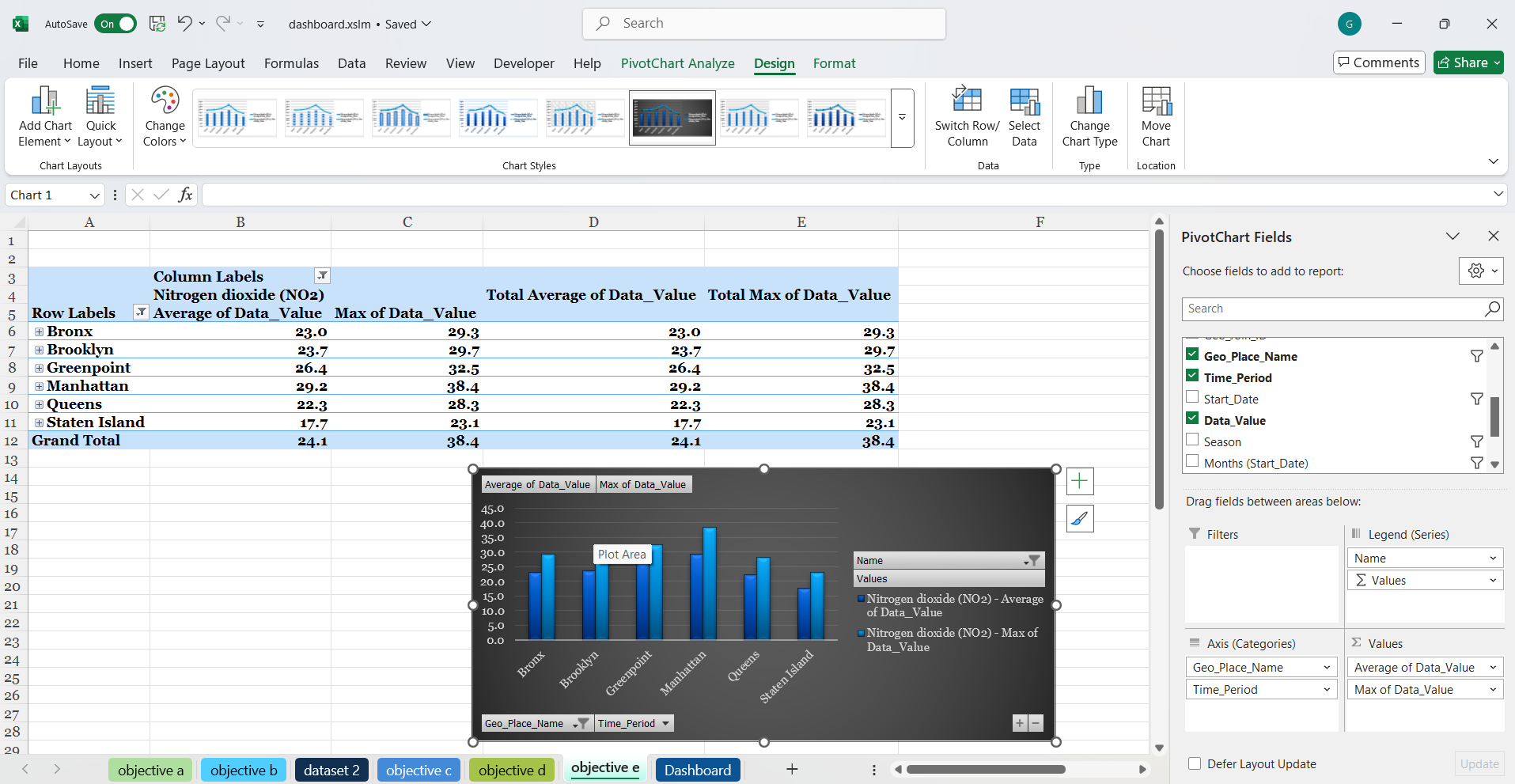
* **Filtered for Pollutant:**
  + Only **NO₂** included in the analysis.
* **Pivot Table Setup:**
  + **Rows:** Time Period (Season or Quarter) and Geo Place Name
  + **Columns:** Pollutant (though only NO₂ is selected, which is fine for this focus)
  + **Values:**
    - **Average of Data\_Value**: Shows the overall NO₂ concentration levels.
    - **Max of Data\_Value**: Indicates the highest recorded NO₂ levels in that timeframe.
* **Chart Type:**
  + A **Combo Chart** — likely a combination of  **clustered column + line** to contrast Avg and Max values visually.

**iii. Analysis Result**

**The** chart highlights the difference between **typical NO₂ exposure levels** (average) and **peak pollution events** (maximum) across neighborhoods and time periods.

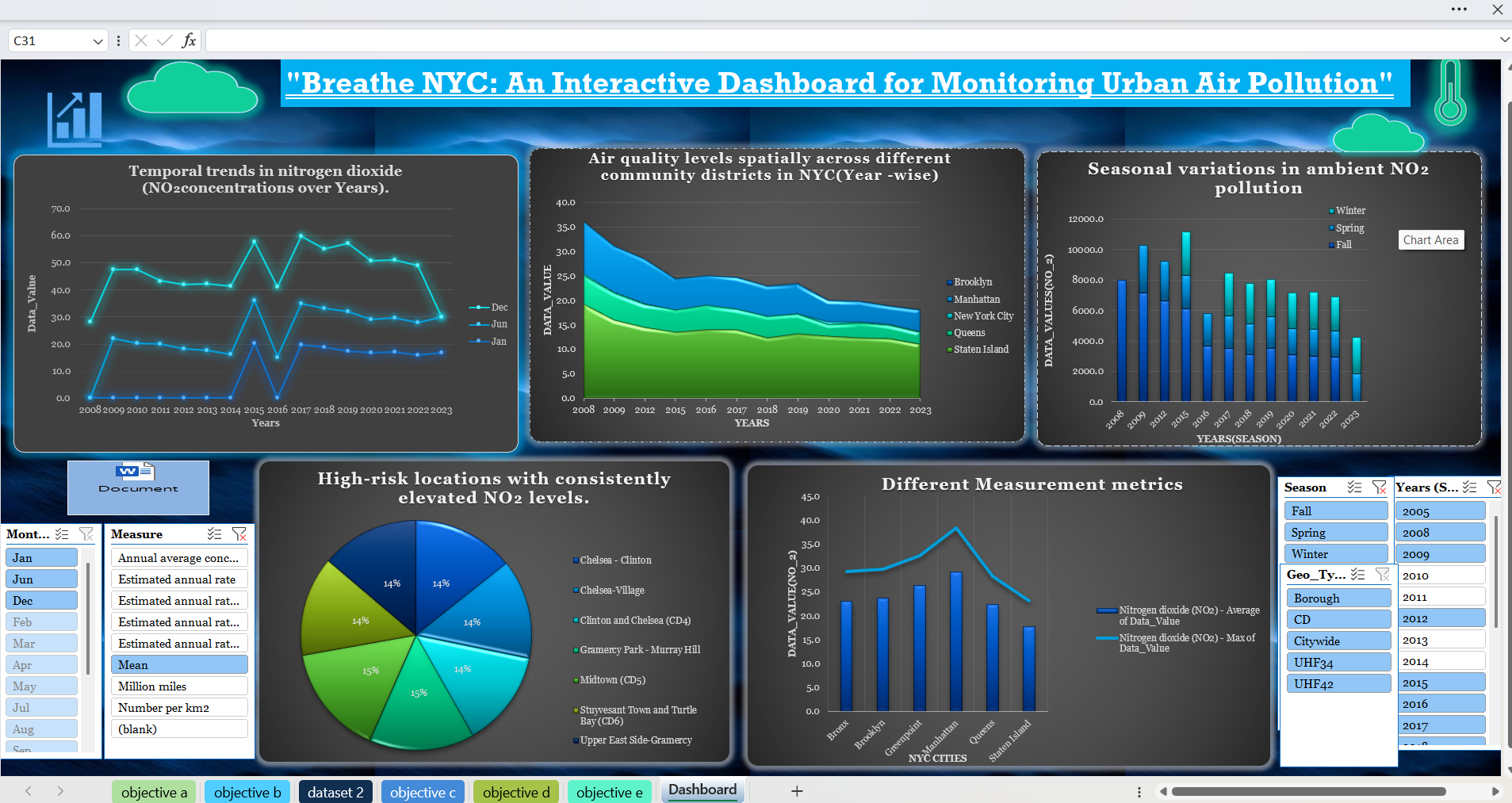
** Disparities in Peak vs. Average: Some areas might have a moderate average but very high spikes (e.g., near highways or industrial zones).**

** Consistency of Pollution: If max and avg values are close, pollution levels are more consistent.**

****

VISUALIZATION

**DASHBOARD**



ELMENTS IN DASHBOARD

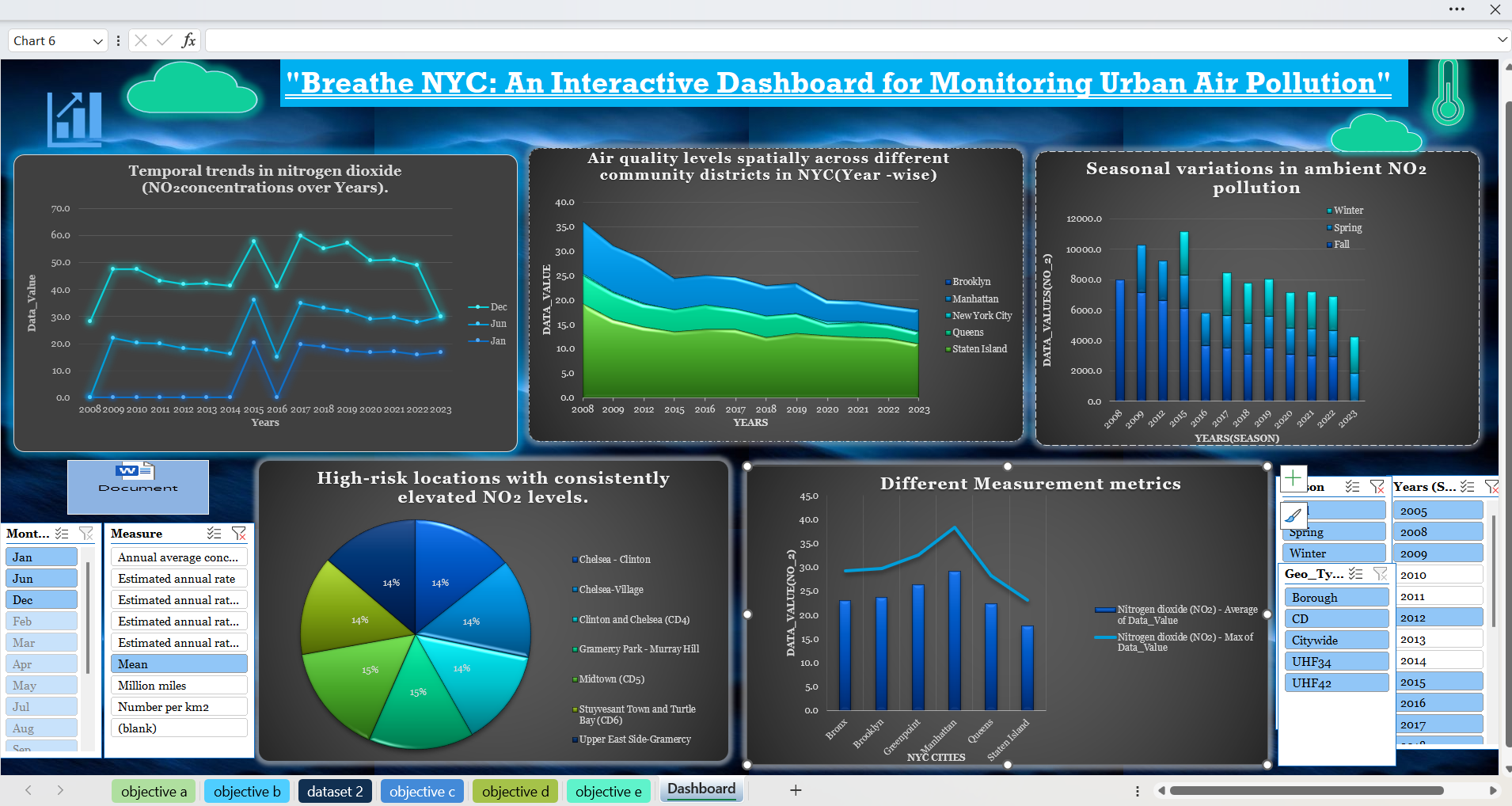
LINE CHART

LINE CHART

BAR CHART

AREA CHART

AREA CHART BAR CHART



PIE CHART

COMBO CHART

OBJECT

OBJECT PIE CHART COMBO CHART

SLICER

SLICER

# CONCLUSION

The Excel dashboard successfully showcases the capability of Microsoft Excel as a powerful data visualization and analysis tool. Through this project, multiple dimensions of NO₂ pollution in NYC have been explored:

* Temporal trends indicate fluctuating NO₂ levels over the years, often peaking during winter months due to atmospheric conditions and increased heating emissions.
* Spatial analysis reveals that certain boroughs like Manhattan and Queens have experienced consistently higher NO₂ levels, likely due to traffic density and infrastructure.
* Seasonal variations clearly show that winter months contribute to higher NO₂ concentrations, whereas summer sees relatively lower values.
* High-risk locations were identified using average NO₂ levels, enabling targeted intervention and monitoring.
* Measurement comparison using mean vs. max values provides deeper understanding of pollution intensity and exposure extremes.

These analyses provide actionable insights for policymakers, environmental agencies, and the community. The dashboard design emphasizes accessibility, interactivity, and clarity, allowing users to dynamically explore and interpret the data.

**References**

1. U.S. Environmental Protection Agency (EPA). Air Quality Criteria for Nitrogen Dioxide.
2. NYC Department of Environmental Protection. www.nyc.gov/dep
3. World Health Organization (WHO) Air Quality Guidelines. [www.who.int](https://www.who.int)
4. Data Source: [Air Quality Dataset - Data.gov](https://catalog.data.gov/dataset/air-quality/resource/f3ed1638-92da-4f88-bb6b-7d3940514574)
5. Microsoft Excel Documentation. [support.microsoft.com](https://support.microsoft.com)