AIR QUALITY DATA ANALYSIS

INDIVIDUAL ASSIGNMENT

(300-Foot Towers Proposed to Clean Delhi's Toxic Air, 2018)
https://www.newsweek.com/futuristic-300-foot-towers-proposed-clean-toxic-air-worlds-most-polluted-city-1129169

Student Name: Abdillahi Osman Omar

Student ID: s3220699

Email: abdillahiosmanomar@student.utwente.nl

Specialization: Geoinformatics

Course Name: Acquisition & Exploration of Geo Data (2023-1B)

Case Study: Amsterdam Municipality

Month, Year: January 2024

FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION
UNIVERSITY OF TWENTE, ENSCHEDE, THE NETHERLANDS

Table of Contents

List of Figures

List of Tables

1. INTRODUCTION

One of the key issues facing environmental studies nowadays is urban air quality. Air pollution threatens people's health due to growing industry, rising energy use, and urbanization. The number of contaminants in the air can be utilized to assess the quality of the air around us. These pollutants fall into two categories: primary and secondary. In this project, the solid and liquid particles that make up particulate air pollutants like PM10 and PM2.5 are the subject of this study. These contaminants may be classified as primary or secondary.

According to the Protocol to Abate Acidification, Eutrophication and Ground-Level Ozone | UNECE, n.d.). in many places of the world, poor air quality can be fatal or seriously harmful to health, especially for young children and newborns. It can also cause respiratory ailments. In the Netherlands, it is required to comply with emissions restrictions for gases and particles that are considered air pollutants. The emissions ceiling denotes the highest number of emissions allowed in a nation. These include volatile organic compounds (VOCs), ammonia (NH_3), nitrogen oxides (NO_x), and Sulphur dioxide (SO_2) (Waterstaat, 2014)

Strict guidelines are recommended for the concentrations of particulate matter (PM) pollution in the air by the World Health Organization. It is recommended to keep PM2.5 levels at a 24-hour average of 15 μg/m³ and an annual average of 5 μg/m³. (WHO Global Air Quality Guidelines, n.d.) recommends maintaining the interim targets from 2005 and incorporating an intermediate target 4 that aligns with the guideline from the same year. Because PM2.5 has a finer particle size and has the potential to have deeper respiratory impacts, it is imperative to monitor its levels to protect public health.

The (WHO Global Air Quality Guidelines, n.d.) recommends a 24-hour average of 45 μg/m³ and an annual average limit of 15 μg/m³ for PM10. The advice is supported by studies examining the long-term mortality effects of PM10. It does, however, recognize that PM2.5 makes up a sizable amount of PM10, making the annual PM10 guideline less effective. When measurements of PM2.5 and PM10 are available, it is recommended to use PM2.5 levels due to their greater health effects. The World Health Organization recommends sticking with the interim PM10 standards from 2005 and adding an intermediate target 4 to match the 2005 air quality guideline.

Table 1: Recommended annual AQG level and interim targets for PM2.5 and PM10.

PM 2.5	**PM 10**
• Annual average: 5 μg/m³ • 24-hour average: 15 μg/m³ • At the level of the 2005 air quality guideline, the GDG (Guideline Development Group) advises keeping the interim targets from 2005 and adding an intermediate target 4.	• Annual average: 15 μg/m³ • 24-hour average: 45 μg/m³ • An assessment of research on the long-term impacts of PM10 on mortality forms the basis of the advice. But while PM2.5 makes up a large amount of PM10, the yearly PM10 AQG level of 15 μg/m³ is considered less protective than the PM2.5 AQG level.

This report will begin with a summary of the regulations governing air quality monitoring, then move on to the study's objectives along with the municipality of Amsterdam as the study region.

In addition, we will go over the specific steps of our methodology, which include data collection, data engineering, accessing several government and community sources' air quality monitoring platforms across Amsterdam, and verifying the accuracy of the data using the distribution of the sensors and responsible organization.

Using geographic technologies and appropriate statistical tools, such as R or Python, we will continue to work on data analysis for each of the three objectives. Afterward, we will discuss the analysis's findings and offer recommendations and suggestions with a summary conclusion in the last section of this report.

2. OBJECTIVES

This study aims to evaluate the Netherlands' air quality in 2022, focusing on the Amsterdam municipality. Three main objectives will be the focus of the investigation of this project:

a) What is the variation in air quality along bike paths during peak and off-peak hours in Amsterdam municipality?
b) Which neighborhoods in the city have the best and worst yearly air quality?
c) Which areas have the most and least reliable air quality data in Amsterdam?

3. STUDY AREA

The Amsterdam municipality, located in the North Holland province of the Netherlands, was chosen as the study area for the air quality analysis. The municipality has an area of approximately 220km^2, is divided into two parts with 8 districts: Centrum, West, Zuid, Oost, Noord, Nieuw-West, Zuidoost, and Westpoort (Fig. 1). Amsterdam Metropolitan Area (AMA) is the country's fastest-growing economic area, and it competes well on a global scale (*About Amsterdam Metropolitan Area*, n.d.).

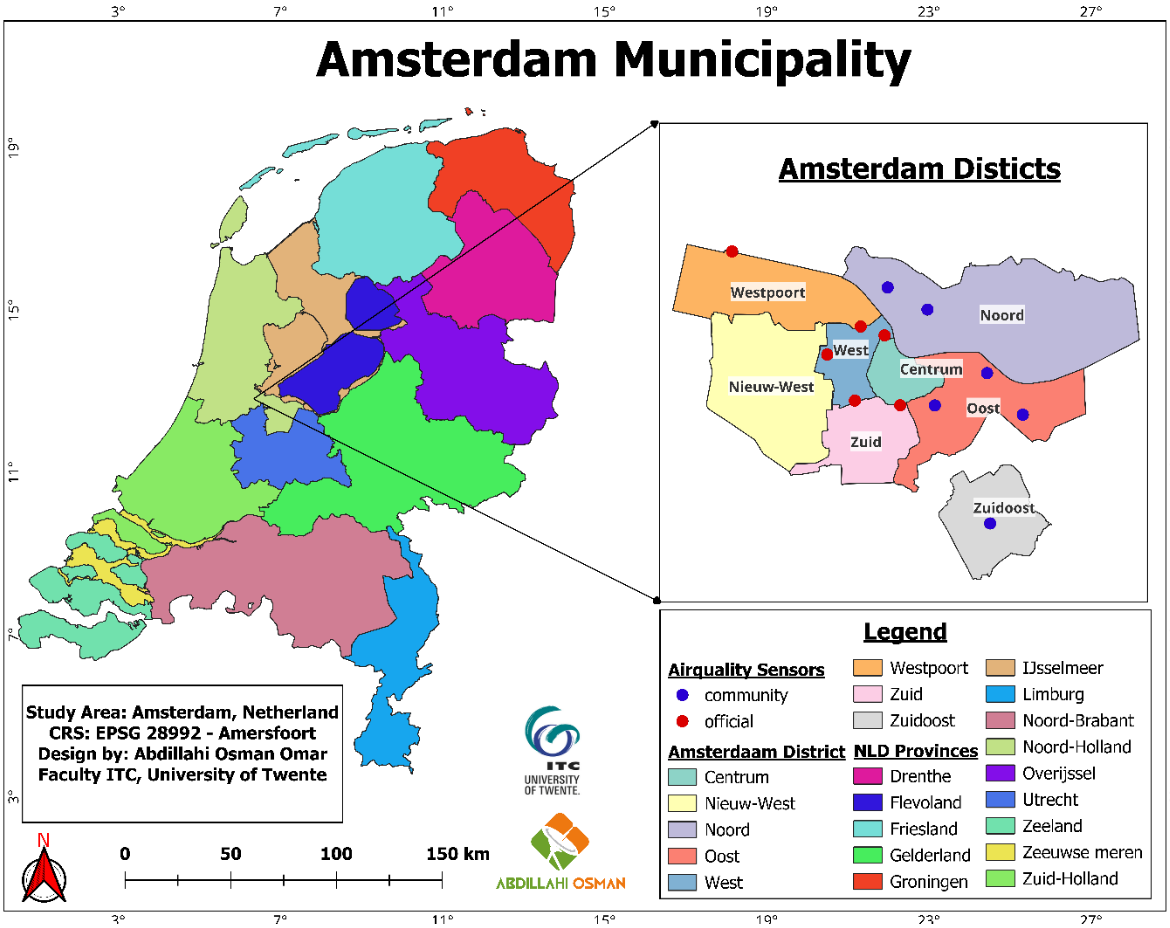


Fig.1: study area (Amsterdam Municipality) with districts and sensor distribution

Source: Author (Omar)

Table 2: Official and Community Sensor Positions

Station ID	Name	Source	Latitude	Longitude
NL49007	Amsterdam-Einsteinweg	official	52.381	4.845
NL49012	Amsterdam-Van Diemenstraat	official	52.390	4.888
NL49014	Amsterdam-Vondelpark	official	52.360	4.866
NL49016	Amsterdam-Westerpark	official	52.394	4.870
NL49017	Amsterdam-Stadhouderskade	official	52.358	4.900
NL49704	Amsterdam-Hoogtij	official	52.428	4.773
921	LTD_4969	community	52.304	4.968
2777	LTD_32517	community	52.373	4.965
2215	LTD_23231	community	52.358	4.926
720	LTD_8802	community	52.402	4.920
2768	LTD_32999	community	52.412	4.890
5964	LTD_55037	community	52.354	4.992

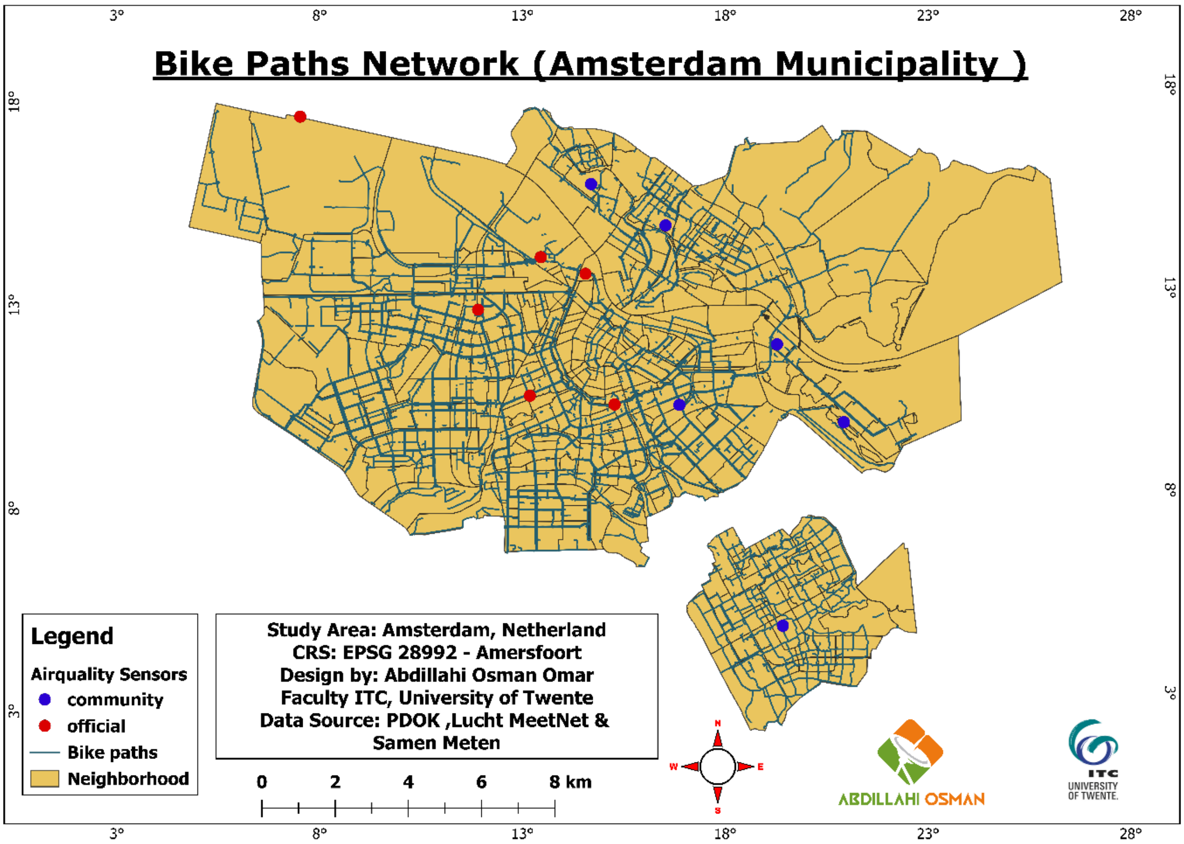


Fig.2: Bike Paths Network (Amsterdam Municipality) with neighborhoods and sensor distribution.

Source: Author (Omar)

4. METHODOLOGY

4.1 Data Collection & Management Plan

To answer the project questions, air quality datasets from both official and community sources were gathered (see Table 2); data is both spatial and temporal fit for the analysis in this study, and PM2.5 and PM10 were chosen as variables to evaluate the study's objectives. From 1 January 2022 to 31 December 2022, the two sources offered hourly air quality measurements.

Aside from air quality data, other datasets such as Amsterdam Municipality shapefile, districts, neighborhoods, and bike paths were obtained from the PDOK (Home - PDOK, n.d.) website (Figures 1 & 2).

Table 3: Metadata table source: Author, see detailed data management plan in Appendix A.

Dataset Used	**Source**	**Filename**	**Data Type/ Format**
Amsterdam Districts	PDOK	Amsterdam district	Geopackage/Shapefile.
Neighborhoods	PDOK	neighborhoods	Geopackage/Shapefile.
Bike Paths Network	PDOK	Bike paths	Geopackage/Shapefile.
Air quality Sensors	*Luchtmeetnet & Samenmeten*	Air quality Sensors	Geopackage/Shapefile.
PM2.5	*Luchtmeetnet & Samenmeten*	2022_pm25	CSV
PM10	*Luchtmeetnet & Samenmeten*	2022_pm10	CSV

This study on air quality in Amsterdam uses a large dataset collected from public and government sources. The PM2.5 and PM10 concentrations are the focus of the time- and space-dependent dataset, which offers a comprehensive analysis of variations in air quality across the municipality and its neighborhoods. A thorough assessment of atmospheric conditions in an appropriate data management plan is ensured by the inclusive method, which integrates government and community sources (see Table 3 & Appendix A).

OBJECTIVE 1 FLOWCHART

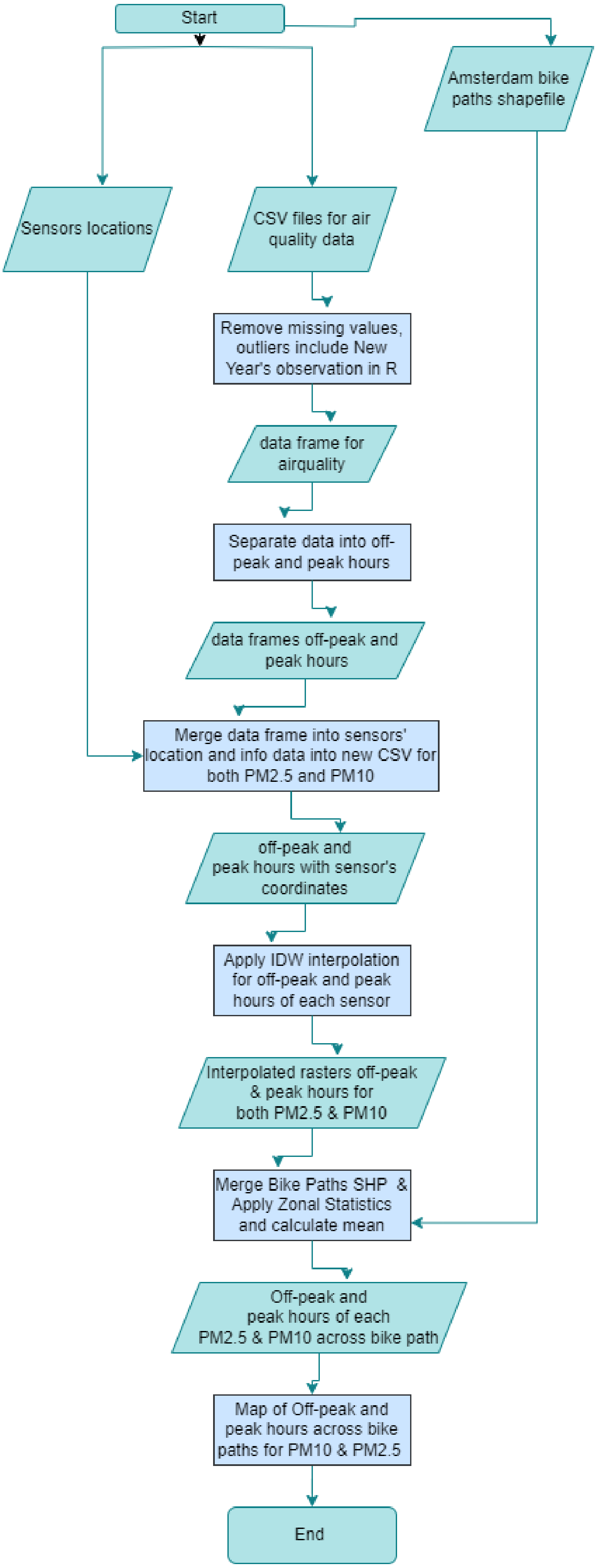
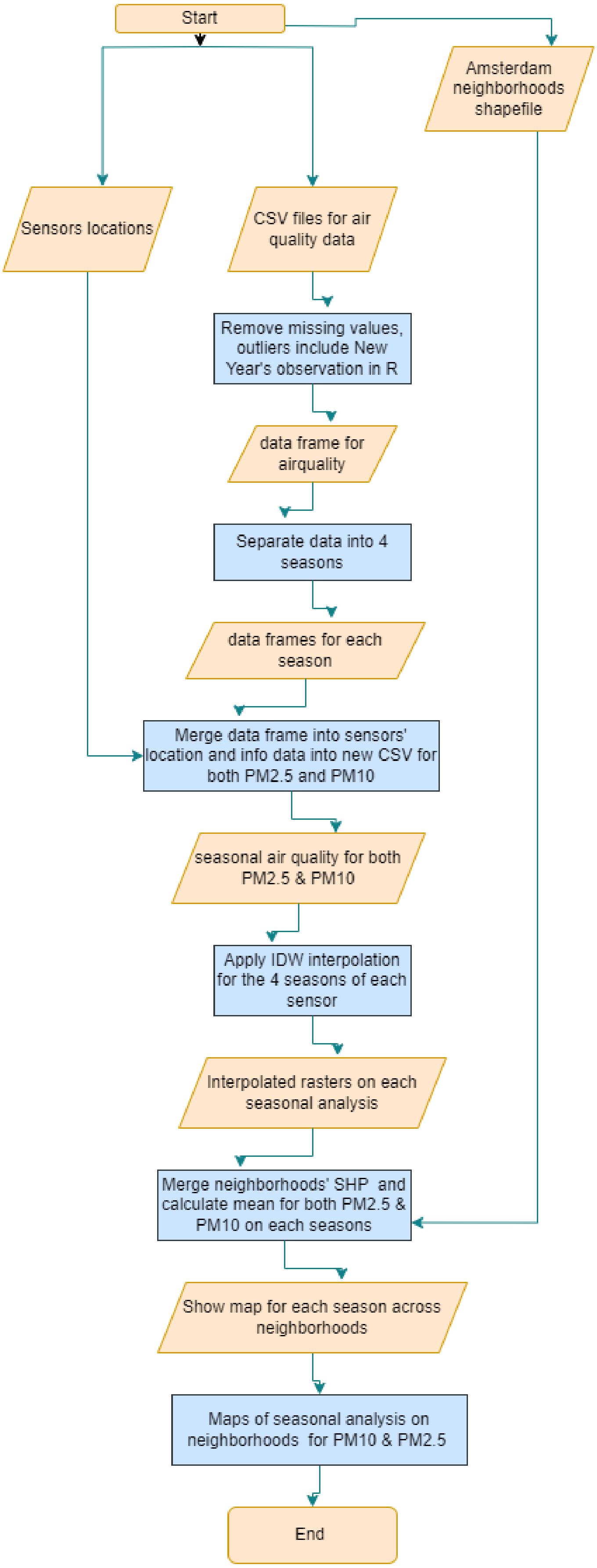


Fig.3: Flowchart for Objective 1 (Source Author)



OBJECTIVE 2 FLOWCHART

Fig.4: Flowchart for Objective 2 (Source Author)

OBJECTIVE 3 FLOWCHART

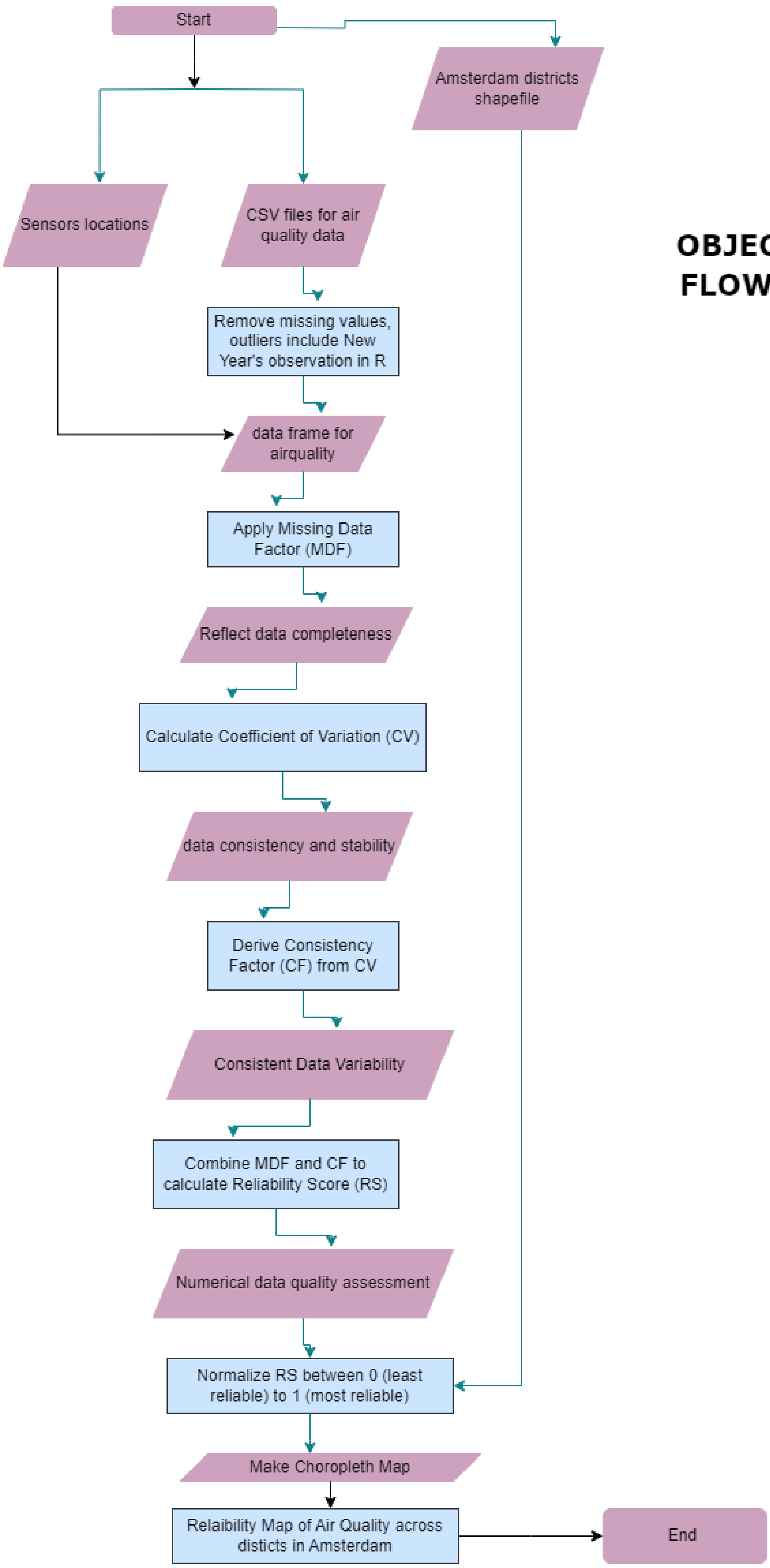


Fig.5: Flowchart for Objective 3 (Source Author)

4.2 Data Infrastructure

4.2.1 Data Scraping

The Dutch air quality monitoring network (Luchtmeetnet) portal keeps an annual record of all variable measurements including PM10 and PM2.5 (as displayed in Figure 6). I was capable of freely discovering the data in the format of a CSV file as the main emphasis was on the year 2022.

The Measuring Together program (Samen Meten) of the National Institute for Public Health and the Environment (RIVM) website offers a simple-to-use graphical user interface that enables readers to choose the time and air quality variables while downloading the data as a CSV file for each station and region (as displayed in Figure 7 and Figure 8), but sensor locations and metadata has been scraped from the portal. This website supplemented the government's sensors, allowing our research to collect data on the Amsterdam municipality area as most of the government sensors cover the center and west districts, while citizen sensors cover the rest.

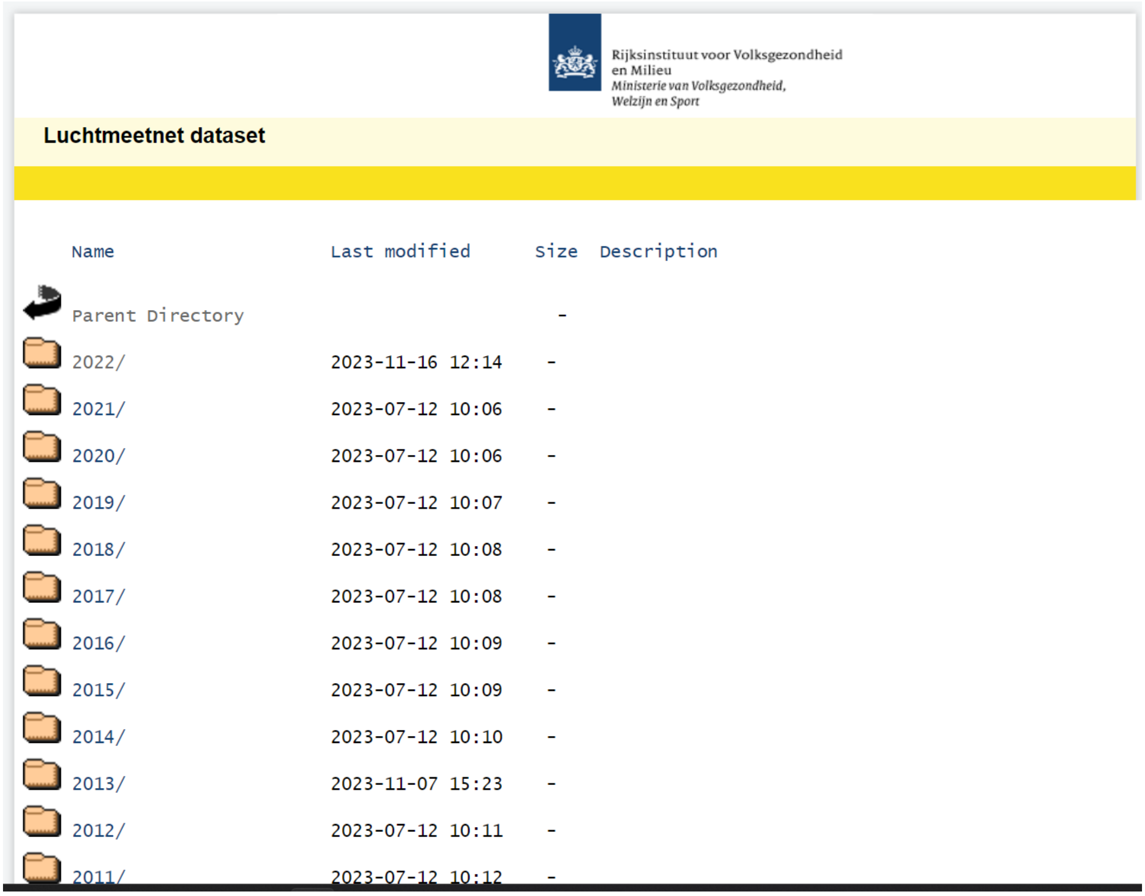


Fig.6: The Dutch air quality monitoring network (Luchtmeetnet) archive

Source: https://www.luchtmeetnet.nl/

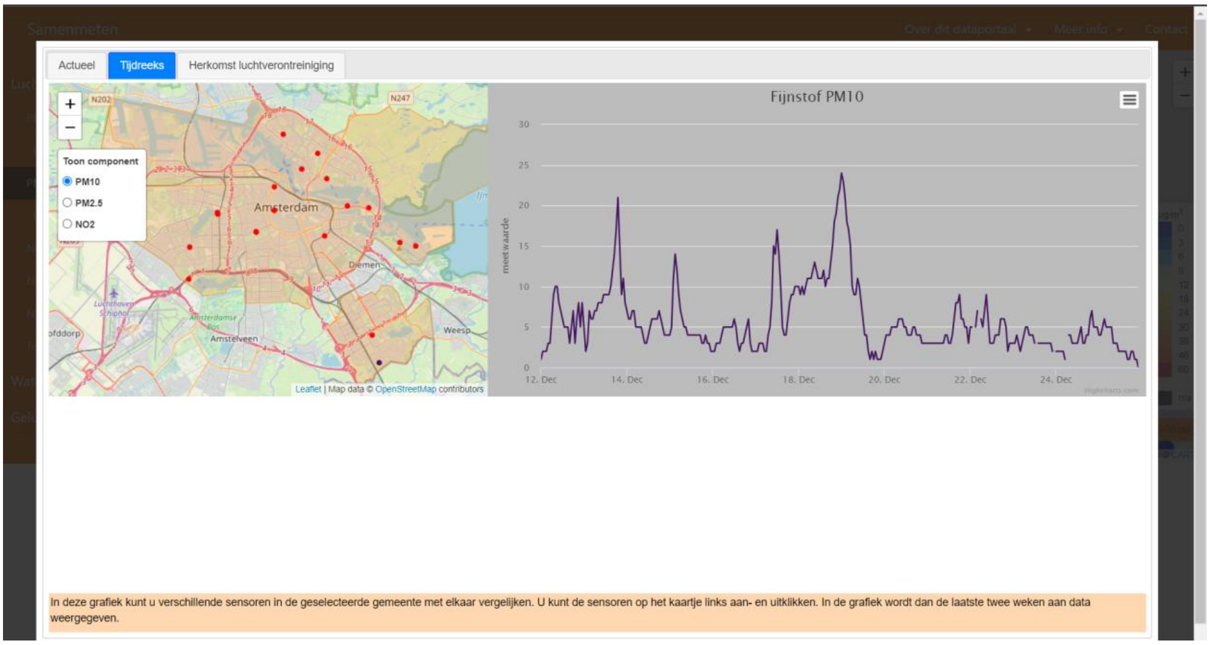


Fig.7: The Measuring Together program (Samen Meten) of the National Institute for Public Health and the Environment (RIVM).

Source: https://samenmeten.rivm.nl/dataportaal/

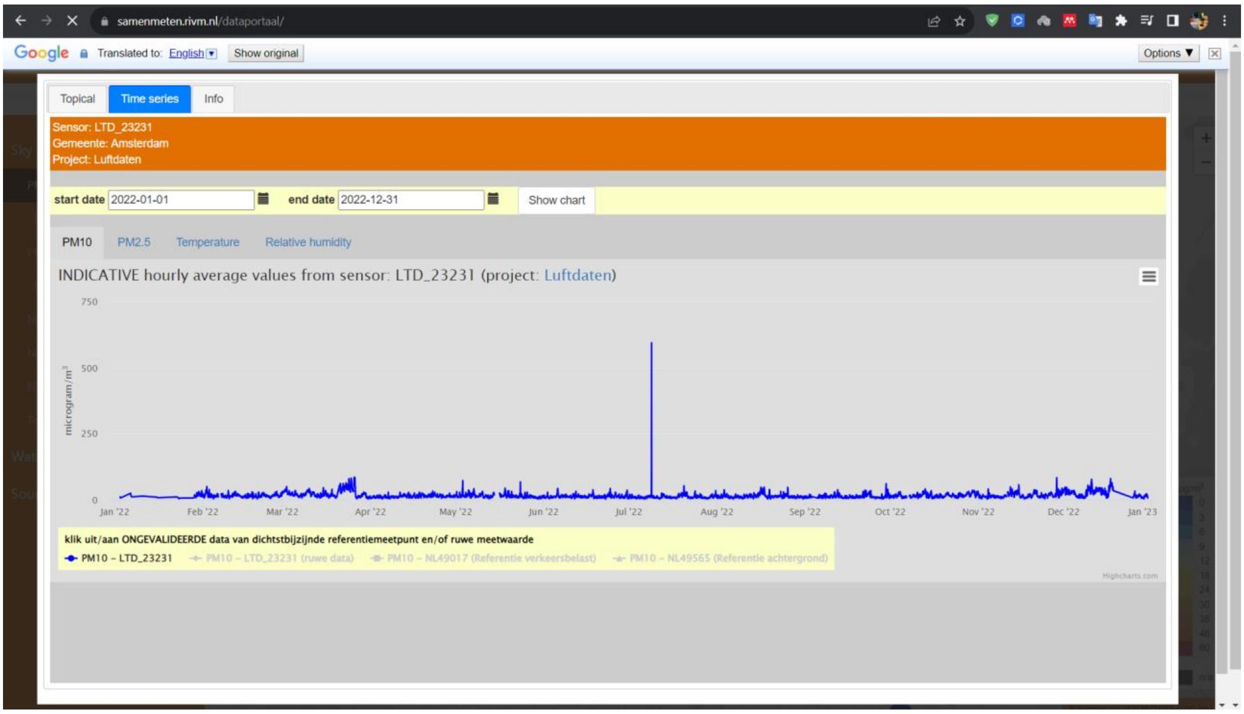


Fig.8: Sensor LTD_23231 from Samen Meten.

Source: https://samenmeten.rivm.nl/dataportaal/

The data was organized using the CSV (comma-separated values) format (Figures 9 and 10). There were two columns for PM2.5 and PM10 for each station, with the time stamp appearing in the first column. The hourly data for the year 2022 were represented by the rows. The coordinates were acquired from the official and community sensors application programming interfaces (APIs) at the two portals both community and official by supplying the IP addresses of the sensors.

As seen in Figure 2, a separate CSV file with the geographic coordinates (latitude and longitude) of each station was created to combine the findings of our study with the locations and visualize the results.

					NL49007	NL49012	NL49014	NL49016	NL49017	NL49704
					(52.38133:	(52.38998:	(52.35971·	(52.39397:	(52.35803!	(52.428017,4.773478)
					NL49007	NL49012	NL49014	NL49016	NL49017	NL49704
PM10	uur	µg/m³	20220101 00:00	20220101 01:00	89.6	157.2	168.2	52.8	157.6	39.3
PM10	uur	µg/m³	20220101 01:00	20220101 02:00	99.7	119.4	98.1	87.6	95.4	20
PM10	uur	µg/m³	20220101 02:00	20220101 03:00	115.5	114.4	97	114.2	103	32.5
PM10	uur	µg/m³	20220101 03:00	20220101 04:00	73.2	73.5	63.4	76.1	54.8	16.2
PM10	uur	µg/m³	20220101 04:00	20220101 05:00	37.2	49	46.8	49	44.2	14.5
PM10	uur	µg/m³	20220101 05:00	20220101 06:00	26.6	37.6	27.9	30.7	33.3	20.7
PM10	uur	µg/m³	20220101 06:00	20220101 07:00	28.4	29.9	24.6	20.2	28.3	14.1
PM10	uur	µg/m³	20220101 07:00	20220101 08:00	22.4	20.9	22.6	20.3	23.8	23.7
PM10	uur	µg/m³	20220101 08:00	20220101 09:00	19.3	25.7	17.1	19.8	17.9	20
PM10	uur	µg/m³	20220101 09:00	20220101 10:00	11.5	16.2	13.4	12.7	14.3	11.7
PM10	uur	µg/m³	20220101 10:00	20220101 11:00	13.1	19.4	14.3	15.5	17.3	12.2
PM10	uur	µg/m³	20220101 11:00	20220101 12:00	17.5	21.5	14	9.9	15.3	15.6
PM10	uur	µg/m³	20220101 12:00	20220101 13:00	9.5	15.8	13	7.8	16.7	20.1
PM10	uur	µg/m³	20220101 13:00	20220101 14:00	17.3	22.3	16.1	15.9	12.5	17
PM10	uur	µg/m³	20220101 14:00	20220101 15:00	7.1	14.7	16	14.2	11.6	8.4
PM10	uur	µg/m³	20220101 15:00	20220101 16:00	7.1	17.3	8.2	14.4	16.6	9.1
PM10	uur	µg/m³	20220101 16:00	20220101 17:00	7.2	16.4	9.2	8.7	16.6	9.1
PM10	uur	µg/m³	20220101 17:00	20220101 18:00	8.7	14.5	7.8	7.7	10.7	8.7
PM10	uur	µg/m³	20220101 18:00	20220101 19:00	6.3	16.5	7.4	7.6	11.8	8.3
PM10	uur	µg/m³	20220101 19:00	20220101 20:00	5.3	18.8	9	9.9	12.6	8.8
PM10	uur	µg/m³	20220101 20:00	20220101 21:00	8.5	15.2	9.1	11.3	11.4	7.4
PM10	uur	µg/m³	20220101 21:00	20220101 22:00	9.4	19.9	8.5	10.4	12.6	8.3
PM10	uur	µg/m³	20220101 22:00	20220101 23:00	11	10.6	8.1	8.6	17.1	10.2
PM10	uur	µg/m³	20220101 23:00	20220102 00:00	15.9	11.8	11.6	12.1	13.1	6.8
PM10	uur	µg/m³	20220102 00:00	20220102 01:00	12.6	18	13.1	14.8	24.5	19.2
PM10	uur	µg/m³	20220102 01:00	20220102 02:00	13.8	15.4	10.2	12.3	16	14.4
PM10	uur	µg/m³	20220102 02:00	20220102 03:00	13.7	15.7	11.2	14.2	14.9	14.6
PM10	uur	µg/m³	20220102 03:00	20220102 04:00	12.9	20.3	13.7	13.8	14.4	14.9
PM10	uur	µg/m³	20220102 04:00	20220102 05:00	16.2	19.1	18.2	13.5	15.8	16.7

	NL49007	NL49012	NL49014	NL49016	NL49017	NL49704
	(52.38133:	(52.38998:	(52.35971·	(52.39397:	(52.35803!	(52.428017,4.773478)
	NL49007	NL49012	NL49014	NL49016	NL49017	NL49704
PM2.5	66.8	124.5	158	40.3	116	27.8
PM2.5	83.7	96.5	89.3	70.1	75.2	14.6
PM2.5	102.3	105.9	86.7	106.4	88.2	30.3
PM2.5	58.9	62.5	50.9	63.2	48.6	12.3
PM2.5	28.5	30.6	28.9	35.3	28.6	12
PM2.5	15.8	18.9	11.1	19.3	15.4	17.3
PM2.5	20.4	18.5	19.1	12.9	23.5	7.5
PM2.5	15.3	20.9	12.8	15.8	10.2	18
PM2.5	13.7	7.4	13.4	10.6	10.5	17.3
PM2.5	8.7	6.4	11.6	8.4	12	12.7
PM2.5	7.1	7.5	9.5	21.8	11.6	12
PM2.5	8.6	11.3	11.2	11.2	12.1	13.4
PM2.5	8.8	18.5	13.7	13.2	16.1	19
PM2.5	10.7	18.2	14.7	11	21.1	16.9
PM2.5	10	14.3	12.1	6.6	12.3	9.7
PM2.5	6.7	10.4	8.2	5.7	10.8	10.2
PM2.5	9	7.6	6.6	3	11.9	11.2
PM2.5	10.2	5.1	7.4	3.5	11.4	11
PM2.5	4.6	3.4	9	5.3	10.3	10.5
PM2.5	5.9	6.2	10.6	5.6	10.6	7.4
PM2.5	7	8.1	8.9	5.3	11.1	4.9
PM2.5	4	6.5	8.3	5.7	12.2	8.8
PM2.5	6.3	7.4	7.1	7	11.3	12.4
PM2.5	8.7	6.9	5.8	8.4	10.9	16.6
PM2.5	11.2	7.9	6.4	10.1	8.9	12.6
PM2.5	8.8	11	8.3	8.9	9.6	11.9
PM2.5	4.7	8.5	8.7	8.1	11.4	11
PM2.5	3.3	6.5	7.5	8.5	9.7	8.7
PM2.5	5.1	8.3	7.5	7.4	7.9	6.7

luchtneetnet_clean_2022

Fig.9: All Luchtmeetnet sensors' data structured in CSV file.
Source: Author (Omar)

variable	DateTime	LTD_4969	LTD_32999	LTD_55037	LTD_8802	LTD_23231	LTD_32517
PM10	01/01/2022 00:00	49	35	14	16	11	15
	01/01/2022 01:00	69	79	91	63	8	68
	01/01/2022 02:00	77	90	82	76	7	57
	01/01/2022 03:00					25	
	01/01/2022 04:00	48	127	54	58	14	51
	01/01/2022 05:00	24	41	27	23	11	22
	01/01/2022 06:00	20	28	16	15	14	14
	01/01/2022 07:00	18	27	19	14	10	11
	01/01/2022 08:00	13	22	17	17	8	9
	01/01/2022 09:00	10	18	11	10	11	10
	01/01/2022 10:00	15	14	10	7	5	6
	01/01/2022 11:00	26	16	8	9	4	8
	01/01/2022 12:00	36	13	9	8	6	8
	01/01/2022 13:00	21	11	8	8	7	7
	01/01/2022 14:00	25	21	13	15	7	14
	01/01/2022 15:00	15	11	8	8	9	7
	01/01/2022 16:00	19	10	9	8	9	8
	01/01/2022 17:00	25	9	7	7	18	7
	01/01/2022 18:00	26	8	6	6	17	6
	01/01/2022 19:00	31	7	7	6	19	6
	01/01/2022 20:00	28	8	7	6		6
	01/01/2022 21:00	20	8	9	6	18	5
	01/01/2022 22:00	23	7	9	6	15	5
	01/01/2022 23:00	29	9	9	6	14	6
	02/01/2022 00:00	20	11	9	7	13	6
	02/01/2022 01:00	82	14	10	8	13	7
	02/01/2022 02:00	108	15	8	6	14	6
	02/01/2022 03:00	65	14	8	7	12	6
	02/01/2022 04:00	40	15	8	8	10	7
	02/01/2022 05:00	28	18	9	9	10	8
	02/01/2022 06:00	20	15	8	8	8	7

variable	DateTime	LTD_4969	LTD_55037	LTD_8802	LTD_23231	LTD_32517	LTD_32999
PM2.5	01/01/2022 00:00	44	6	12	4	9	19
	01/01/2022 01:00	57	62	68	3	57	52
	01/01/2022 02:00	64	47	85	3	46	58
	01/01/2022 03:00				7		
	01/01/2022 04:00	50	35	80	5	46	108
	01/01/2022 05:00	20	13	21	7	16	23
	01/01/2022 06:00	14	7	11	7	8	13
	01/01/2022 07:00	15	10	12	10	9	15
	01/01/2022 08:00	9	8	18	11	6	13
	01/01/2022 09:00	7	5	6	7	6	8
	01/01/2022 10:00	13	5	5	1	5	7
	01/01/2022 11:00	20	5	6	1	5	8
	01/01/2022 12:00	30	5	7	2	5	8
	01/01/2022 13:00	24	7	12	5	8	13
	01/01/2022 14:00	25	10	17	2	13	19
	01/01/2022 15:00	17	7	11	4	7	10
	01/01/2022 16:00	13	5	6	4	5	6
	01/01/2022 17:00	19	4	6	9	5	5
	01/01/2022 18:00	23	4	5	9	4	5
	01/01/2022 19:00	23	3	4	8	4	4
	01/01/2022 20:00	22	4	5		4	5
	01/01/2022 21:00	18	6	5	7	4	5
	01/01/2022 22:00	21	7	5	7	4	5
	01/01/2022 23:00	25	5	5	7	4	5
	02/01/2022 00:00	18	6	6	7	4	6
	02/01/2022 01:00	59	5	5	5	4	7
	02/01/2022 02:00	98	5	6	5	4	9
	02/01/2022 03:00	60	4	5	5	4	7
	02/01/2022 04:00	37	4	5	4	4	6
	02/01/2022 05:00	23	4	5	4	4	6
	02/01/2022 06:00	19	4	5	5	4	7

samenten_clean_2022

Fig.10: All Samen Meten sensors' data structured in CSV file.
Source: Author (Omar)

In this study, Python was the programming language of choice for data gathering and analysis, as it is crucial to the efficacy of coding and the clarity of visualizations. However, after careful consideration, I realized that R was the best language for handling and analyzing datasets in the areas of numerical data manipulation and statistical analysis. This is because R has a wide range of specialized packages and concise syntax, which enables us to write fewer lines of code and produce reliable results, but I prefer to do my statistical analysis using Python libraries like NumPy.

Python was chosen due to its efficiency in coding and its capacity to produce engaging visualizations. Python excels in producing publish-ready graphs and charts, which makes it simpler to represent complex data patterns in a clear and visually compelling manner. (*I have attached both Python scripts for this study in the submission zip file*).

4.2.2 Gaps and Outliers in Data

Another crucial point to note is that most of the sensors recorded the New Year's celebrations at a level higher than usual (see Figure 12). After examining the datasets and verifying the relevant data regarding air pollution during the fireworks, we decided to eliminate these measurements because they would have an adverse effect on and bias the entire dataset.

According to Greven et al. (2019), from 1993 to 2012, urban monitoring stations in the Netherlands recorded an average PM10 level of approximately 550μg/m^3 in the first hour following New Year's Day. Throughout the rest of the year, hourly PM10 concentrations did not often surpass 100μg/m^3, and in 2011, the average annual background PM10 concentration was 24μg/m^3.

After completing the data harvesting process and gaining access to the information needed to meet the project's goals, I began looking through the data and discovered that certain sensors' missing data varied in periods and hours (see Figure 11).

31/01/2022 06:00	14		22	13	10	17
31/01/2022 07:00	14		18	12	12	18
31/01/2022 08:00	12		18	9	13	13
31/01/2022 09:00	16		17	14	12	20
31/01/2022 10:00	24		20	19	14	30
31/01/2022 11:00	30		19	25	12	36
31/01/2022 12:00	30		15	22	10	32
31/01/2022 13:00	32		12	23	12	34
31/01/2022 14:00	41		7	27	14	43
31/01/2022 15:00	34		10	24	15	30
31/01/2022 16:00	29		10	24	12	30
31/01/2022 17:00	28		8	22	10	27
31/01/2022 18:00	31		5	23	11	29
31/01/2022 19:00	33		6	27	12	26
31/01/2022 20:00	28		12	18	13	21
31/01/2022 21:00	20		12	14	13	16
31/01/2022 22:00	12		10	8	13	12
31/01/2022 23:00	17		11	12	13	15
01/02/2022 00:00	17		8	11	14	15
01/02/2022 01:00	18		5	11	17	14
01/02/2022 02:00	55		2	9	21	13
01/02/2022 03:00	35		3	9	26	7
01/02/2022 04:00	39		4	10	18	9

06/02/2022 11:00			17	11
06/02/2022 12:00			15	11
06/02/2022 13:00			14	11
06/02/2022 14:00			11	7
06/02/2022 15:00			12	5
06/02/2022 16:00			11	7
06/02/2022 17:00				7
06/02/2022 18:00			15	5
06/02/2022 19:00			14	6
06/02/2022 20:00			13	7
06/02/2022 21:00			13	7
06/02/2022 22:00			10	7
06/02/2022 23:00			10	7
07/02/2022 00:00			10	8
07/02/2022 01:00			9	7

Fig.11: Data Missing

					NL49007	NL49012	NL49014	NL49016	NL49017	NL49704			NL49007	NL49012	NL49014	NL49016	NL49017	NL49704
					(52.38133:	(52.38998:	(52.35971·	(52.39397:	(52.35803!	(52.428017,4.773478)			(52.38133:	(52.38998:	(52.35971·	(52.39397:	(52.35803!	(52.428017,4.773478)
					NL49007	NL49012	NL49014	NL49016	NL49017	NL49704			NL49007	NL49012	NL49014	NL49016	NL49017	NL49704
PM10	uur	µg/m³	20220101 00:00	20220101 01:00	89.6	157.2	168.2	52.8	157.6	39.3		PM2.5	66.8	124.5	158	40.3	116	27.8
PM10	uur	µg/m³	20220101 01:00	20220101 02:00	99.7	119.4	98.1	87.6	95.4	20		PM2.5	83.7	96.5	89.3	70.1	75.2	14.6
PM10	uur	µg/m³	20220101 02:00	20220101 03:00	115.5	114.4	97	114.2	103	32.5		PM2.5	102.3	105.9	86.7	106.4	88.2	30.3
PM10	uur	µg/m³	20220101 03:00	20220101 04:00	73.2	73.5	63.4	76.1	54.8	16.2		PM2.5	58.9	62.5	50.9	63.2	48.6	12.3
PM10	uur	µg/m³	20220101 04:00	20220101 05:00	37.2	49	46.8	49	44.2	14.5		PM2.5	28.5	30.6	28.9	35.3	28.6	12
PM10	uur	µg/m³	20220101 05:00	20220101 06:00	26.6	37.6	27.9	30.7	33.3	20.7		PM2.5	15.8	18.9	11.1	19.3	15.4	17.3
PM10	uur	µg/m³	20220101 06:00	20220101 07:00	28.4	29.9	24.6	20.2	28.3	14.1		PM2.5	20.4	18.5	19.1	12.9	23.5	7.5
PM10	uur	µg/m³	20220101 07:00	20220101 08:00	22.4	20.9	22.6	20.3	23.8	23.7		PM2.5	15.3	20.9	12.8	15.8	10.2	18
PM10	uur	µg/m³	20220101 08:00	20220101 09:00	19.3	25.7	17.1	19.8	17.9	20		PM2.5	13.7	7.4	13.4	10.6	10.5	17.3
PM10	uur	µg/m³	20220101 09:00	20220101 10:00	11.5	16.2	13.4	12.7	14.3	11.7		PM2.5	8.7	6.4	11.6	8.4	12	12.7
PM10	uur	µg/m³	20220101 10:00	20220101 11:00	13.1	19.4	14.3	15.5	17.3	12.2		PM2.5	7.1	7.5	9.5	21.8	11.6	12
PM10	uur	µg/m³	20220101 11:00	20220101 12:00	17.5	21.5	14	9.9	15.3	15.6		PM2.5	8.6	11.3	11.2	11.2	12.1	13.4
PM10	uur	µg/m³	20220101 12:00	20220101 13:00	9.5	15.8	13	7.8	16.7	20.1		PM2.5	8.8	18.5	13.7	13.2	16.1	19
PM10	uur	µg/m³	20220101 13:00	20220101 14:00	17.3	22.3	16.1	15.9	12.5	17		PM2.5	10.7	18.2	14.7	11	21.1	16.9
PM10	uur	µg/m³	20220101 14:00	20220101 15:00	7.1	14.7	16	14.2	11.6	8.4		PM2.5	10	14.3	12.1	6.6	12.3	9.7
PM10	uur	µg/m³	20220101 15:00	20220101 16:00	7.1	17.3	8.2	14.4	16.6	9.1		PM2.5	6.7	10.4	8.2	5.7	10.8	10.2
PM10	uur	µg/m³	20220101 16:00	20220101 17:00	7.2	16.4	9.2	8.7	16.6	9.1		PM2.5	9	7.6	6.6	3	11.9	11.2
PM10	uur	µg/m³	20220101 17:00	20220101 18:00	8.7	14.5	7.8	7.7	10.7	8.7		PM2.5	10.2	5.1	7.4	3.5	11.4	11
PM10	uur	µg/m³	20220101 18:00	20220101 19:00	6.3	16.5	7.4	7.6	11.8	8.3		PM2.5	4.6	3.4	9	5.3	10.3	10.5
PM10	uur	µg/m³	20220101 19:00	20220101 20:00	5.3	18.8	9	9.9	12.6	8.8		PM2.5	5.9	6.2	10.6	5.6	10.6	7.4
PM10	uur	µg/m³	20220101 20:00	20220101 21:00	8.5	15.2	9.1	11.3	11.4	7.4		PM2.5	7	8.1	8.9	5.3	11.1	4.9
PM10	uur	µg/m³	20220101 21:00	20220101 22:00	9.4	19.9	8.5	10.4	12.6	8.3		PM2.5	4	6.5	8.3	5.7	12.2	8.8
PM10	uur	µg/m³	20220101 22:00	20220101 23:00	11	10.6	8.1	8.6	17.1	10.2		PM2.5	6.3	7.4	7.1	7	11.3	12.4
PM10	uur	µg/m³	20220101 23:00	20220102 00:00	15.9	11.8	11.6	12.1	13.1	6.8		PM2.5	8.7	6.9	5.8	8.4	10.9	16.6
PM10	uur	µg/m³	20220102 00:00	20220102 01:00	12.6	18	13.1	14.8	24.5	19.2		PM2.5	11.2	7.9	6.4	10.1	8.9	12.6
PM10	uur	µg/m³	20220102 01:00	20220102 02:00	13.8	15.4	10.2	12.3	16	14.4		PM2.5	8.8	11	8.3	8.9	9.6	11.9
PM10	uur	µg/m³	20220102 02:00	20220102 03:00	13.7	15.7	11.2	14.2	14.9	14.6		PM2.5	4.7	8.5	8.7	8.1	11.4	11
PM10	uur	µg/m³	20220102 03:00	20220102 04:00	12.9	20.3	13.7	13.8	14.4	14.9		PM2.5	3.3	6.5	7.5	8.5	9.7	8.7
PM10	uur	µg/m³	20220102 04:00	20220102 05:00	16.2	19.1	18.2	13.5	15.8	16.7		PM2.5	5.1	8.3	7.5	7.4	7.9	6.7

Fig.12: New Year's 2022 Values

Asthma and other respiratory conditions have been linked to brief exposure to extremely high ambient particulate matter (PM) during fireworks displays. Furthermore, exposure to PM2.5, both short- and long-term, has been linked to higher mortality rates from lung cancer and cardiovascular disease (Laden et al. 2006; Hamad et al. 2016).

4.2.3 Data Mastery for Problem Solving

The goals of the study are well-supported and valid by the carefully gathered and arranged processed PM10 and PM2.5 air quality data collected from public and government sensors, sensor locations, Amsterdam bike paths network, municipality districts, and neighborhoods. While setting up a monitoring plan for the analysis is a first step, the real job will be to get the word out to decision-makers and the public about pollution values in our project.

Despite the lack of official sensors in the Nieuw-West district (see Figure 13), our equitable arrangement of sensors may guarantee coverage of the area. To fill in spatial data gaps when readings from community sensors were unavailable, we intended to cross-reference measurements from municipalities located nearby but the complete year dataset was not available in nearby areas, hence the possible approach was to integrate with the nearest official sensors in other districts.

A straightforward approach was used to fill up the temporal data gaps by merging information from government and community sensors. Using hourly values, we employed interpolation techniques to close gaps in the datasets and correct temporal discrepancies. As a result, our methodology ensures the consistency and correctness of our dataset, giving the public and decision-makers an informed and lucid perspective on the condition of the air in the examined area.

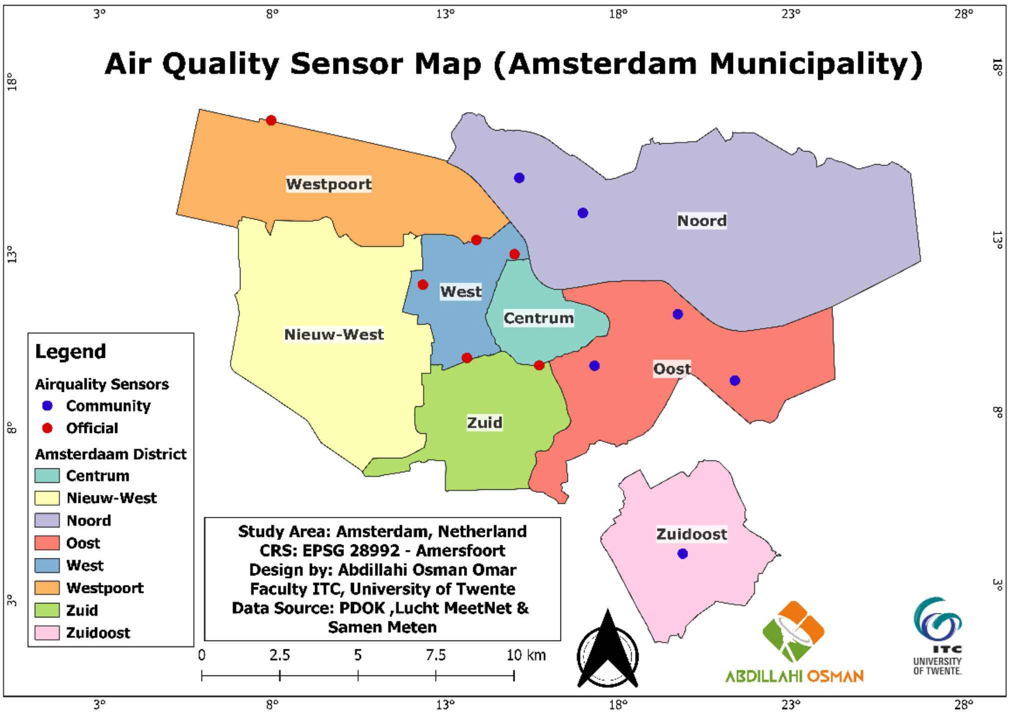


Fig.13: Air Quality Selected Network in Amsterdam Districts
Source: Author (Omar)

4.3 Data Quality

4.3.1 Quality of Community Data

The required degree of data quality determines the potential uses for sensor data. One important consideration when assessing a sensor's performance is data quality. It examines utilizing verified reference data to assess the performance of a single data source. The data quality can be evaluated using temporal and thematic data accuracy and sensor measurement accuracy.

The official sensors provided the air quality data were measured using a highly calibrated apparatus. However, some data is absent. This is a reliable and helpful source of information. However, data collected by community sensors needs to be verified and cross-referenced with appropriately calibrated reference sensors.

4.3.2 Limitations

Accuracy of measurement and position: The problem with positional accuracy is the difficulty in determining the exact location of the sensors. While sensor data provides the concentration of air pollutants at a particular spot, the crowdsource location gives a nearby position within a few hundred meters. The air quality reading displayed the wrong location, and several of the values were negative, high, and inappropriate.

Temporal accuracy: There were problems with the community data's temporal information and the first problem was determining the sensors' consistency over the time under consideration. The municipality does, however, have community sensors available. Only a few of them were used for this assignment. This is because some sensors experienced periodic problems, a few had data gaps that lasted for several months, and a few still had missing data in the daily record.

4.3.3 Usefulness

Community-provided air quality data is useful since it can quickly reveal specific, localized features and address current issues. Additionally, there are limitations, such as potential biases, unequal spatial distribution, and reliability problems. The quality of population-generated air quality data may increase due to initiatives to encourage community involvement and educate the public on data collection standards.

We evaluated the data quality by examining the types of sensors (some provide more accurate measurements than others), their location, and the actor (the community or the government) involved in handling the data once the data had been arranged for analysis and processing. To check the air quality of the dataset, we have collected an equal number of sensors for both the official and community, hence we selected NL_49007 from the official and LTD_4969 from the community and examined the distribution trend of the data from the two sources across every station using statistical tools in Python.

Based on their characteristics, we investigated the quality of the sensors. The LTD_4969 community sensor is composed of the low-cost NOVA SDS011 PM sensor. According to Božilov et al.'s 2022 experimental research, which used the NOVA SDS011 sensor in outdoor air in Bor, Serbia for two years, the results showed an RMSE >80% and that the PM10 readings overestimated the values obtained by the reference instrument (GRIMM EDM180).

Furthermore, a dashboard that tracks each sensor's plausibility score—that is, the chance that the sensor will measure something accurately—is being developed by the Samen Meten National Institute for Public Health and the Environment (RIVM) in the Netherlands (Plausibility of the Particulate Matter Measurement | Measuring Air Quality Together, n.d.). After that, we can investigate the sensor in greater detail to identify the problem's location and, if necessary, take specific action down the road.

4.3.4 Means of Improvement

Even though it could be challenging, some things can be done to improve the sensor's quality. According to Measuring Air Quality Together (Plausibility of the Particulate Matter Measurement | n.d.), it is crucial to ensure that the sensor is suspended correctly and has unrestricted flow. It is advised to relocate the sensor away from areas that are prone to standing water, pools, and emission sources like radiators or air conditioning units. Routine sensor cleaning is also necessary to ensure accurate data. In summary, although it varies on the organizational structure and preferences of the relevant organizations, the authorities must regulate the overall quality of data in both community and official settings through good administration and policy.

Apart from the above-mentioned limitations of the harvested data of community sensors, one point to mention is that all scraped data have uniform data formats. Data from all community sensors were provided as comma-delimited files (CSVs). Because we exported the structured data from community sensors in ArcGIS Pro and QGIS for additional analysis, they can now be used on any platform alongside official data without needing to be converted into a different format as long have coordinates columns.

4.4 Air Quality Distribution - Descriptive Statistics

PM 2.5 and PM10 variables will be analyzed to determine how they are distributed. We employ visual tools and descriptive statistics to look for features like skewness, symmetry, and outlier presence. By comparing these factors, we can develop an understanding of the patterns and variations found in the air quality data.

We used merged CSV files for both datasets derived from a variety of sensors and each dataset includes measurements of PM2.5 and PM10 and the measurement time. The NL49007 and LTD_4969 sensors will be our main emphasis. For both, we have combined CSV files.

Table 4: Comparative Data on PM2.5 and PM10 Concentrations for LTD_4969 and NL49007 Sensors

Variable	Sensor	Count	Mean	Standard Deviation	Minimum	25%	50%	75%	Maximum
PM2.5	NL49007	3299	9.2	8.1	-4.1	4.1	7.1	11.35	102.3
PM10	NL49007	3224	18.4	10.8	-1.2	11.7	16.4	22.6	275.6
PM2.5	LTD_4969	3275	11.8	10.2	1	5	8	16	117
PM10	LTD_4969	3275	14.1	9.9	1	7	12	18	122

All measurements are in µg/m³unit except the count.

In our next analysis, we want to acquire a better understanding of PM10 and PM2.5 patterns that the sensors LTD_4969 and NL49007 captured. For every one of the four datasets, box plots, histograms, and linear models will be created. With the aid of these graphic depictions, we will be able to analyze the distributions of PM10 and PM2.5 readings, determine whether the data is skewed, and spot any abnormalities. Box plots show variation in data and outlier existence, while histograms give a clear picture of data occurrence within metric ranges.

Based on the patterns in Figure 14, PM10 Measurements with the NL49007 Sensor. With most of the values clustered at the lower end and a tail extending towards higher values, the histogram displays a right-skewed distribution. Some outliers on the higher side are visible in the box plot, indicating the possibility of sporadic elevated PM10 readings.

Like PM10, the distribution of PM2.5 measurements is right skewed, with a peak at lower values and a tail towards higher values. The box plot displays outliers, which have intermittently high PM2.5 concentrations.

We can recognize properties like skewness, modality, and outliers by looking at the histograms' shapes and the presence of the Kernel Density Estimate (KDE) curve shows us some of the characteristics in the dataset, but for the relationship between the two variable check detailed regression result in Appendix D and scatter plot in Appendix E.

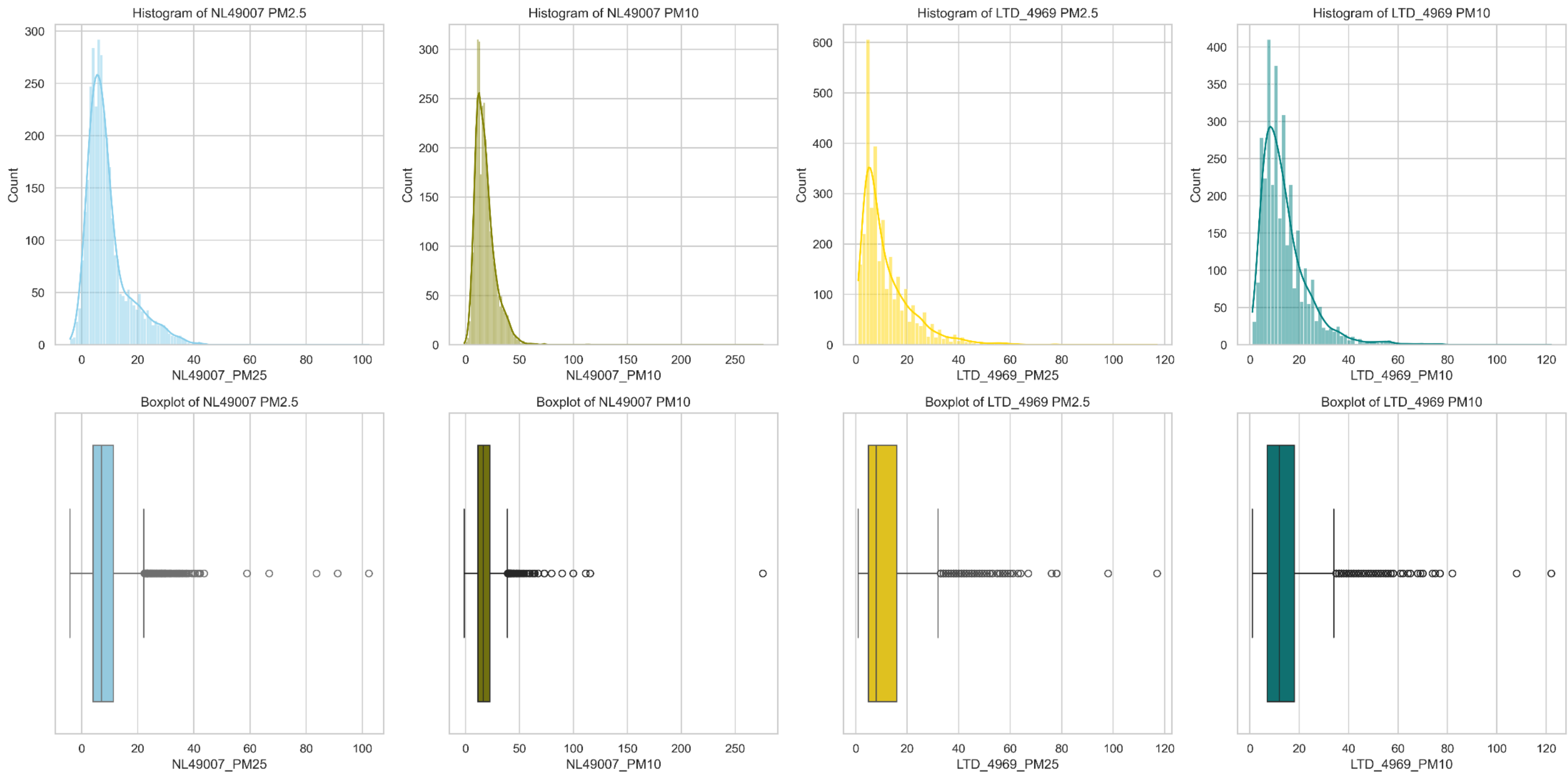


Fig.14: PM2.5 and PM10 Levels from NL49007 and LTD_4969 – Histogram and Box Plot

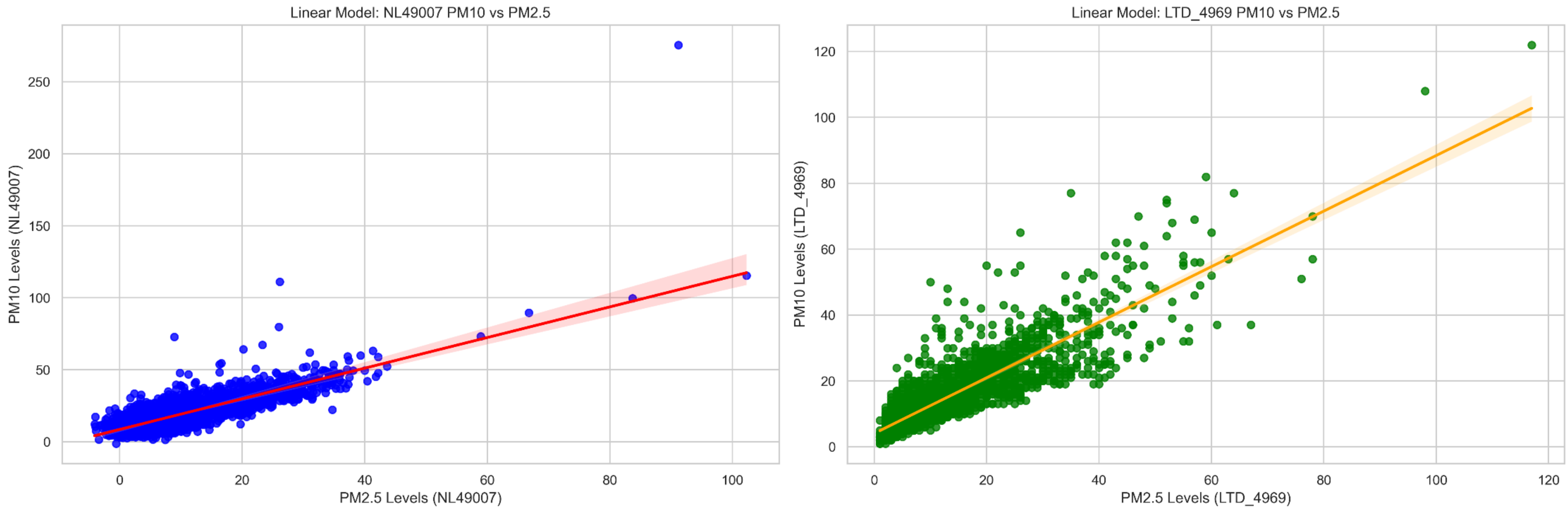


Fig.15: PM2.5 and PM10 Correlation at NL49007 and LTD_4969

A few presumptions are necessary for a linear regression model to be valid: the relationship must be linear, the errors must be independent, the errors must have equal variances, and the error terms must be normal. Valid results can be obtained by linear regression if these presumptions are met. Although real-world data frequently deviates from these assumptions to some extent, the validity may be compromised by the high R-squared values, which indicate a strong linear relationship, and in my dataset for the two variables the R-squared is 0.635 for NL49007_PM10 and 0.749 for LTD_4969_PM10 *(see Appendix D for regression results, Appendix E for detailed Scatter Plot and regression_model_summary.txt in the submission zipped file).*

The information distribution for the PM10 observations from the LTD_4969 sensor shows a right-skewed trend; however, it is not as noticeable as it is for the NL49007. surprisingly, the box plot shows outliers, indicating occasional reading spikes that are consistent with NL49007.

The histogram likewise shows a right-skewed distribution for PM2.5 readings from the same sensor, which is consistent with the PM10 data. The box plot also shows several outliers on the higher end, which indicate times when PM2.5 levels were greater. These display the features of the sensor's data, emphasizing the skewed distributions and irregular high readings for both PM10 and PM2.5 observations.

Consequently, for both PM10 and PM2.5 readings, sensors display right-skewed distributions, meaning that most readings are low and there are few high-value readings. Outliers in the data from both sensors point to sporadic times when particulate matter concentrations were noticeably higher. For every sensor, the distribution patterns are the same for both kinds of particulate matter, although the range of values and the degree of skewness vary.

Basic descriptive statistics offer a trustworthy summary of the data. If the data is accurately entered and devoid of systemic flaws, they are genuine this can be validated to draw a histogram. A precise depiction of the data point distribution is provided by histograms. For them to be valid, there must be enough data to create a clear image of the distribution as we did in Figure 14.

Boxplots offer a reliable and instructive overview of the variability and distribution of the data. However, if the data is not symmetrically distributed or has many outliers, they may be deceptive.

We quantified the correlation between PM2.5 and PM10 levels using linear regression models. The relationship's strength and direction are indicated by the slope of the regression line (see Figure 15), and the percentage of the dependent variable's variance that can be accounted for by the independent variable is indicated by the R-squared value.

The correlation between PM2.5 and PM10 levels at NL49007 (blue) and LTD_4969 (green) is displayed in the scatter figure in Appendix E. Comparing how these pollutants correspond in various locations is helpful. The scatter plot's association between PM2.5 and PM10 levels points to a favorable correlation. Higher fine particulate matter (PM2.5) is consistently correlated with higher coarse particulate matter (PM10), as seen by the tendency for PM10 levels to rise as PM2.5 levels do.

5. DATA ANALYSIS & RESULT

5.1 Examining Differences in Air Quality Between Off-Peak and Peak Hours Along Amsterdam's Bike Paths

How off-peak and peak hours are classified might change depending on the situation, but for the objective of this research, peak hours relate to more activity or higher demand for transport (Bikes), and off-peak hours are usually thought of as periods of lower demand or less activity. Off-peak and peak hours are frequently linked to the times of day when there is either a lower or higher demand for public transport services in the context of transport in Amsterdam.

Our goal is to investigate how the air quality varies along Amsterdam's bike lanes during peak and off-peak hours. There are two datasets loaded (see peak_offpeak_script.py) *; one with sensor locations and the other with air quality data. The air quality dataset's 'Time' column is transformed into Date Time format. There are designated peak hours, which are 7-9 AM and 4-6 PM. After writing a script, we separated the data on air quality into two sets: one for peak and one for off-peak hours. The average PM2.5 and PM10 readings are computed for both peak and off-peak hours, and the results **(see Table 5) are** arranged into distinct data frames. For uniformity, the station IDs are formatted to match the sensor locations' data frame.

Finally, each sensor's location was added as latitude and longitude coordinates to the summary file, and a CSV file was created from the summary files so that QGIS could read it later. Lastly, we have mapped 4 different maps for each variable in peak hours and off-peak hours.

Table 5: Peak and Off-Peak Hour Air Quality Data for Amsterdam Bike Paths Source: Author

station id	source	latitude	longitude	peak_pm25	offpeak_pm25	peak_pm10	offpeak_pm10
NL49007	official	52.381	4.845	9.5	9.1	19.6	18.1
NL49012	official	52.39	4.888	10.1	10.5	18.9	18.4
NL49014	official	52.36	4.866	8.9	9.4	16.5	16.1
NL49016	official	52.394	4.87	8.0	8.0	16.1	15.6
NL49017	official	52.358	4.9	10.0	10.1	18.2	17.5
NL49704	official	52.428	4.773	9.8	9.8	16.0	15.9
LTD_4969	community	52.304	4.968	10.2	12.4	12.6	14.6
LTD_32517	community	52.373	4.965	7.6	7.7	16.2	16.5
LTD_23231	community	52.358	4.926	7.0	7.1	14.5	14.2
LTD_8802	community	52.402	4.92	7.7	8.0	9.3	9.2
LTD_32999	community	52.412	4.89	5.8	6.0	15.0	15.3
LTD_55037	community	52.354	4.992	7.7	7.8	9.9	9.8

See script peak_offpeak_script.py (in the submission zipped file)

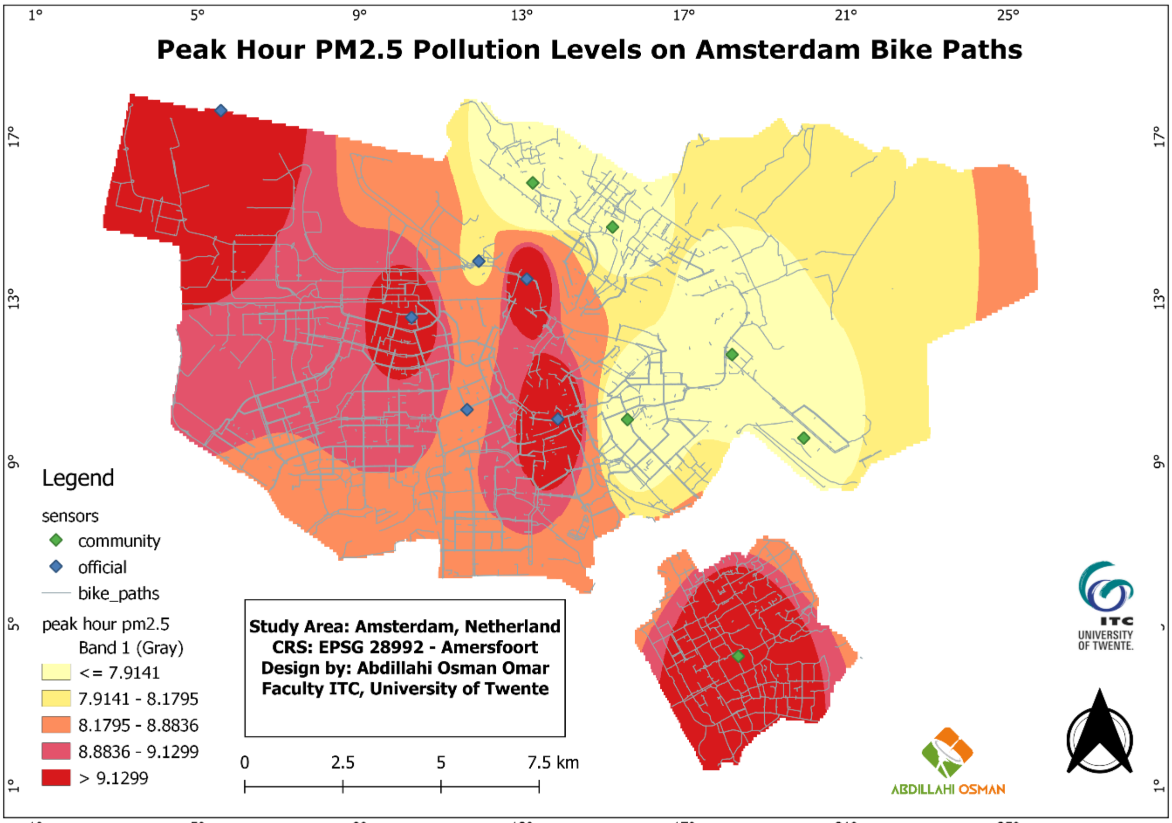


Fig. 16: Peak Hour PM2.5 Pollution Levels on Amsterdam Bike Paths

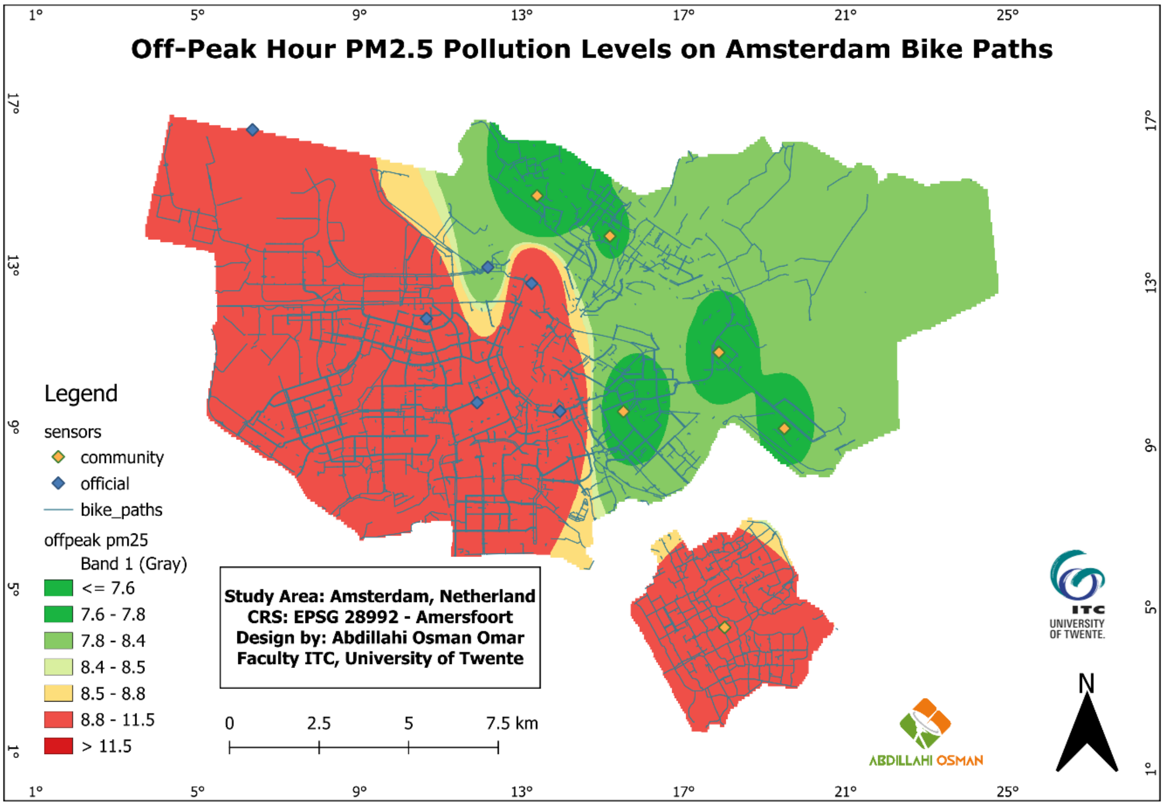


Fig. 17: Off-peak Hour PM2.5 Pollution Levels on Amsterdam Bike Paths

The outcome demonstrates that PM2.5 is consistently measured during peak and off-peak hours on the bike routes close to the city center. Figures 16 and 17 show that bike paths in the west and south districts have greater concentrations of PM2.5 than the other districts, while bike paths in the north district had lower concentrations. A key point worth mentioning is that, following the guidelines set forth by the World Health Organization, measurements of PM2.5 from both official and community sensors were not within a reasonable range (see Table 1), which represents the annual average: 5 µg/m³.

While there is evidence linking prolonged exposure to PM10 to respiratory mortality, the effects of long-term exposure to PM10 are less certain (Inhalable Particulate Matter and Health (PM2.5 and PM10) | California Air Resources Board, n.d.). Prolonged exposure to PM2.5 has been linked to decreased lung function growth in children and early death, especially in those with chronic heart or lung disorders.

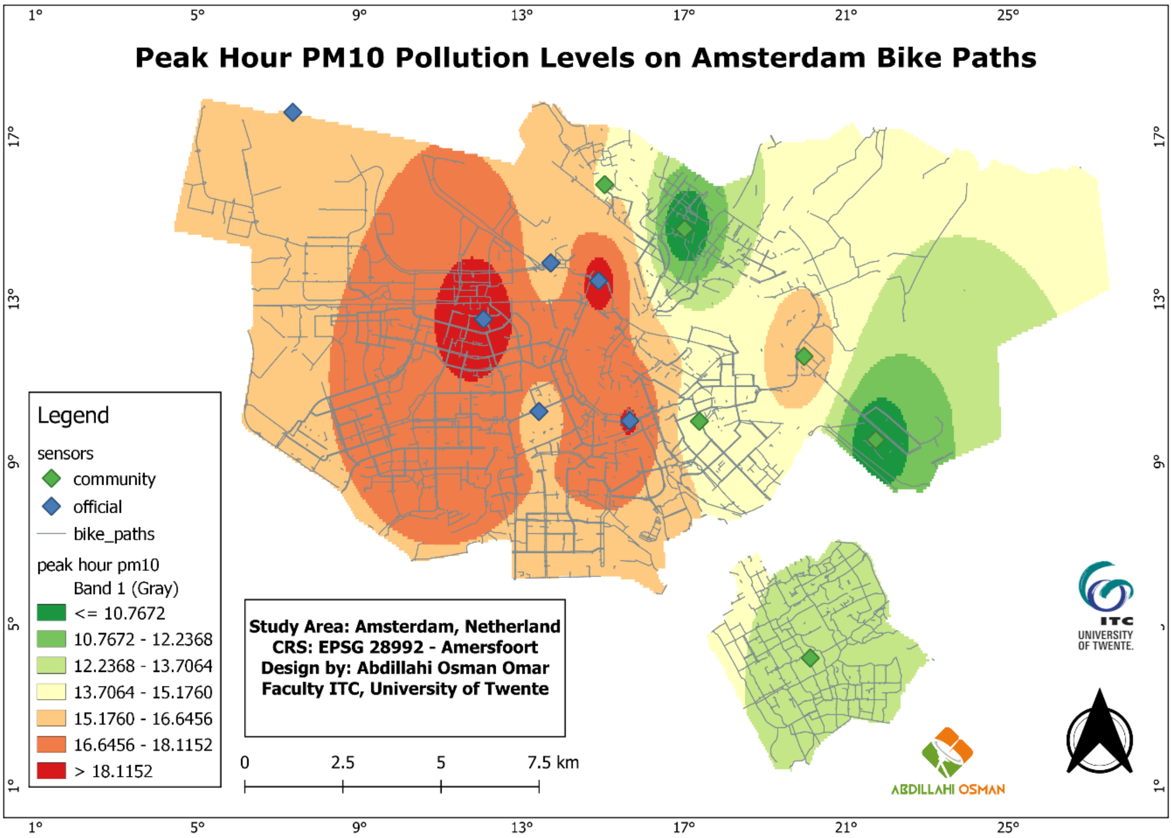


Fig. 18: Peak Hour PM10 Pollution Levels on Amsterdam Bike Paths

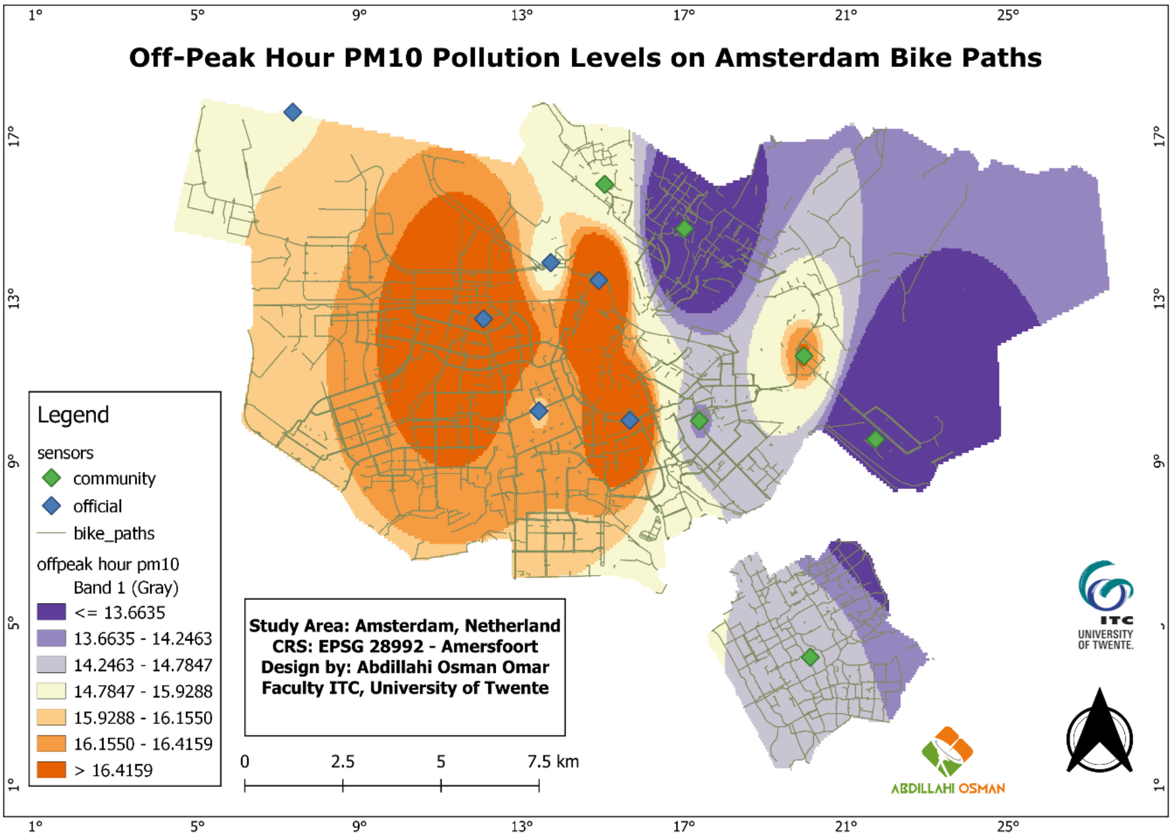


Fig. 19: Off-peak Hour PM10 Pollution Levels on Amsterdam Bike Paths

This emphasizes how critical it is to address the poor air quality in the impacted areas, put policies in place to lower PM2.5 pollution, and educate the public about the possible health hazards of prolonged exposure to this kind of pollution across bike paths which is common transportation in the Netherlands. It might also call for more research into the precise origins of PM2.5 in the south and west districts and the causes of the north district's lower levels.

However, the results of Figures 18 and 19 demonstrate that, in 2022, the bike paths in Amsterdam's Centrum, West, and Zuid districts have the highest PM10 readings of any district. This is because of the industrial activity that occurs in these areas, particularly in the West, and tourist activities in the city.

The WHO guidelines state that the annual average of PM10 is 15 µg/m³. It is evident that the air quality on bike paths in the city center was higher for both PM10 and PM2.5 than it was in the other districts; Oast and Nord showed lower readings because they are primarily made up of water and have fewer sensors installed.

5.2 Evaluating Annual Variations in Air Quality in Amsterdam's Neighborhoods

The purpose of this study was to use PM2.5 and PM10 to compare the variations in air quality for the 2022 four seasons across Amsterdam’s neighborhoods. Winter (January–March), Spring (April–June), Summer (July–September), and Autumn (October–December) were the four seasons of the year.

We have created a script that analyses the variations in Amsterdam, Netherlands' air quality throughout four seasons. The `get_season` function is defined first*, classifying each month into one of the four seasons (Spring, Summer, Autumn, or Winter). The 'Time' column is then converted to DateTime format, the code loads the air quality data from a CSV file, and it is assigned as the index for time-based analysis. Depending on which month each timestamp falls within, a new 'Season' column is added to the DataFrame. The script determines the seasonal means of the mean air quality (PM2.5 and PM10) for each sensor individually for each season and stores them in new DataFrames.

It merges the seasonal means data frames on the 'Season' column and renames columns to reflect the component and sensor in the column name. The resultant data is saved to a new CSV file called "final_seasonal_means.csv" see tables 5 and 6 which is merged manually with sensor location CSV and exported in QGIS to generate maps for analysis. This investigation contributes to our understanding of the seasonal variations in Amsterdam neighborhood's air quality as well as the differences between sensor locations.

See script fourSeasonsAnalysis.py (in the submission zipped file)

Table 6: Variations of Air Quality (PM2.5) Across Amsterdam Neighborhoods, Source (Author)

station id	source	latitude	longitude	Winter_PM2.5	Spring_PM2.5	Summer_PM2.5	Autumn_PM2.5
NL49007	official	52.381	4.845	12.9	8.2	5.2	10.8
NL49012	official	52.39	4.888	13.8	9.4	6.2	12.0
NL49014	official	52.36	4.866	12.0	8.3	6.6	10.0
NL49016	official	52.394	4.87	11.9	7.3	4.1	8.7
NL49017	official	52.358	4.9	12.8	9.1	7.7	10.8
NL49704	official	52.428	4.773	13.6	8.6	6.7	10.8
LTD_4969	community	52.304	4.968	13.1	10.2	10.6	13.4
LTD_32517	community	52.373	4.965	8.2	6.1	4.6	4.8
LTD_23231	community	52.358	4.926	8.8	8.0	6.8	8.0
LTD_8802	community	52.402	4.92	9.4	8.2	4.1	6.7
LTD_32999	community	52.412	4.89	7.5	7.2	7.5	8.9
LTD_55037	community	52.354	4.992	8.2	8.2	8.1	6.3

Table 7: Variations of Air Quality (PM10) Across Amsterdam Neighborhoods, Source (Author)

station id	source	latitude	longitude	Winter_PM10	Spring_PM10	Summer_PM10	Autumn_PM10
NL49007	official	52.381	4.845	22.3	16.5	15.0	20.1
NL49012	official	52.39	4.888	22.4	17.5	15.3	18.7
NL49014	official	52.36	4.866	19.2	14.4	13.7	17.3
NL49016	official	52.394	4.87	20.4	14.4	11.9	16.2
NL49017	official	52.358	4.9	21.2	16.7	14.9	17.9
NL49704	official	52.428	4.773	20.5	14.4	13.2	15.6
LTD_4969	community	52.304	4.968	15.2	12.5	13.2	15.3
LTD_32517	community	52.373	4.965	12.8	9.9	8.4	8.1
LTD_23231	community	52.358	4.926	16.3	15.4	13.6	15.5
LTD_8802	community	52.402	4.92	11.7	10.1	6.8	8.3
LTD_32999	community	52.412	4.89	16.3	15.5	16.0	17.7
LTD_55037	community	52.354	4.992	19.4	13.4	11.8	12.3

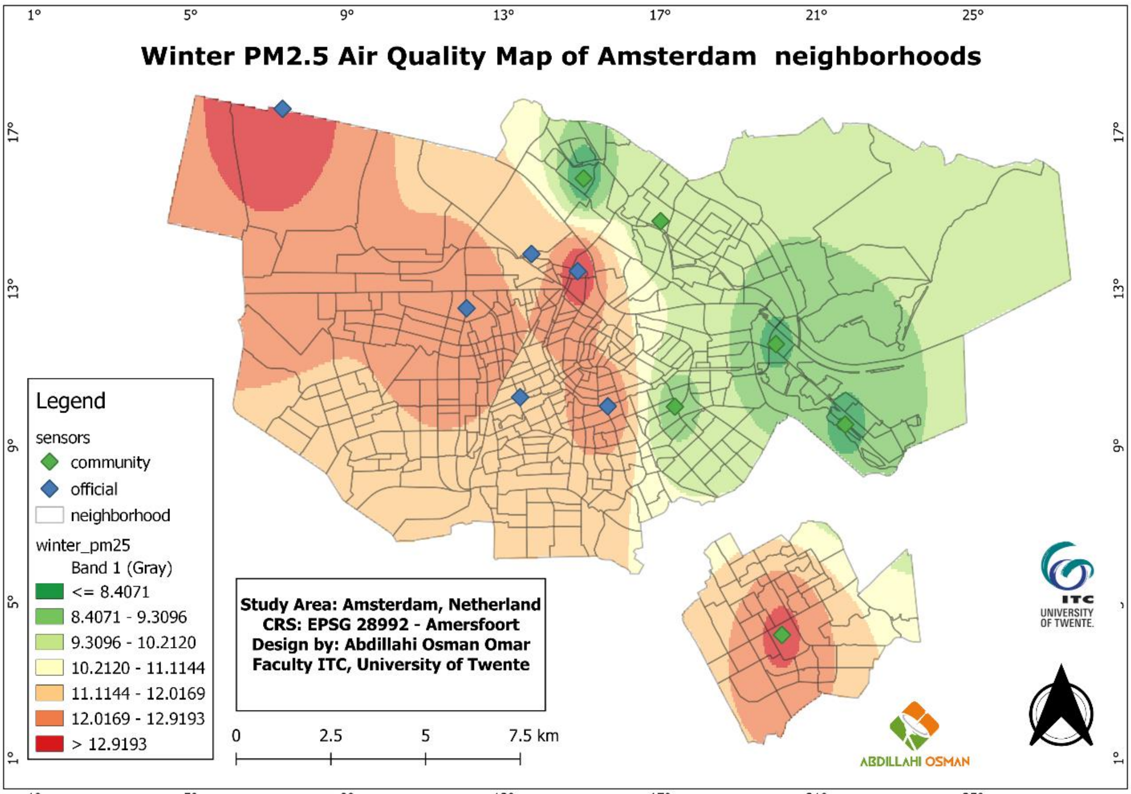


Fig. 20: Winter PM2.5 air quality levels across various neighborhoods in Amsterdam, Source (Author)

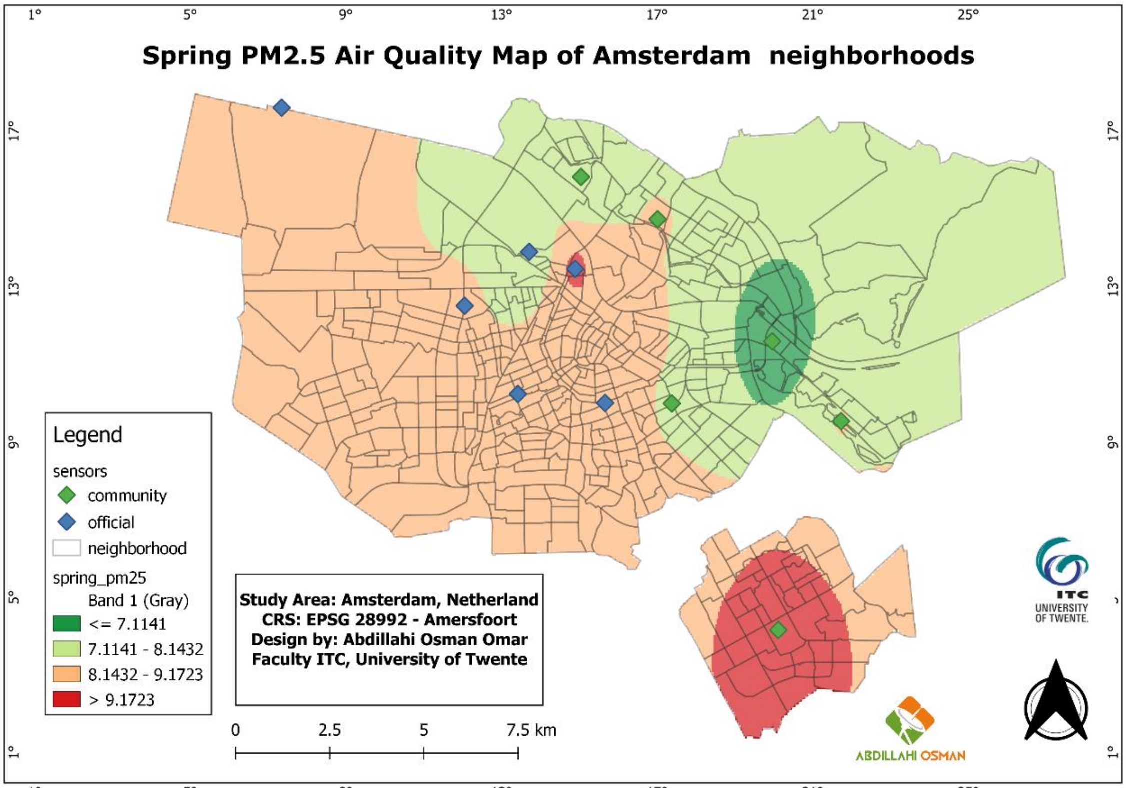


Fig. 21: Spring PM2.5 air quality levels across various neighborhoods in Amsterdam, Source (Author)

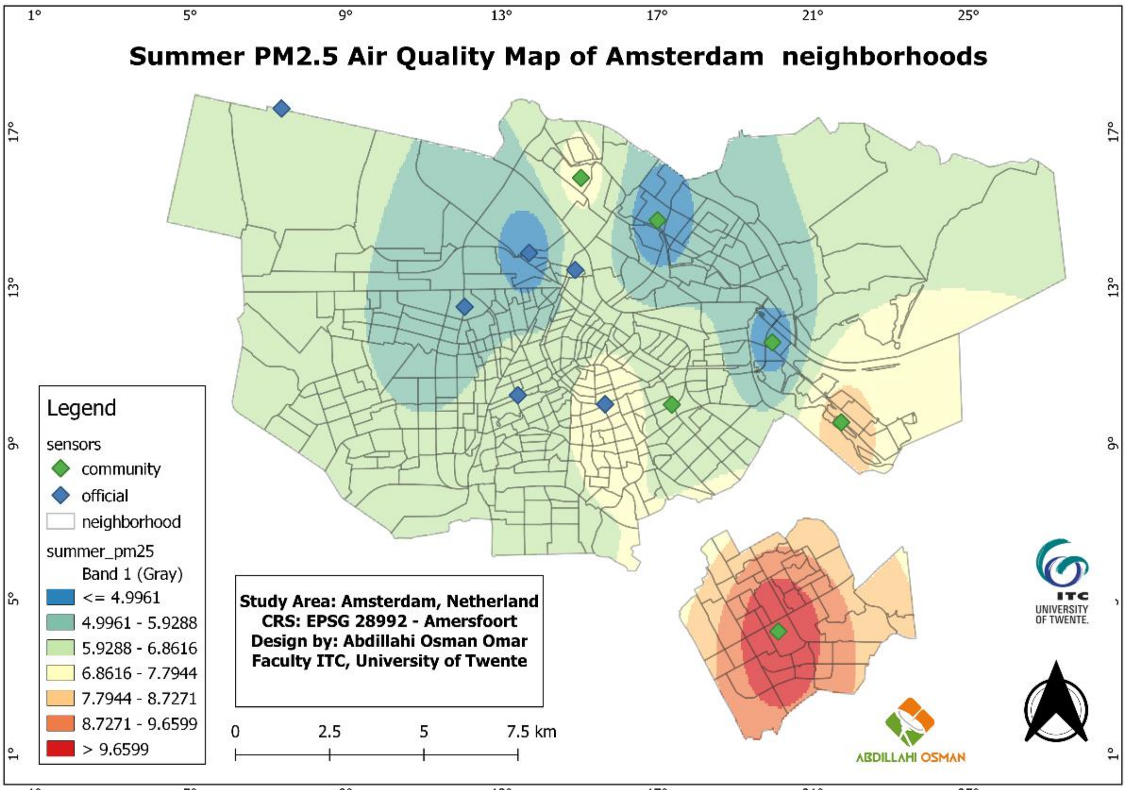


Fig. 22: Summer PM2.5 air quality levels across various neighborhoods in Amsterdam, Source (Author)

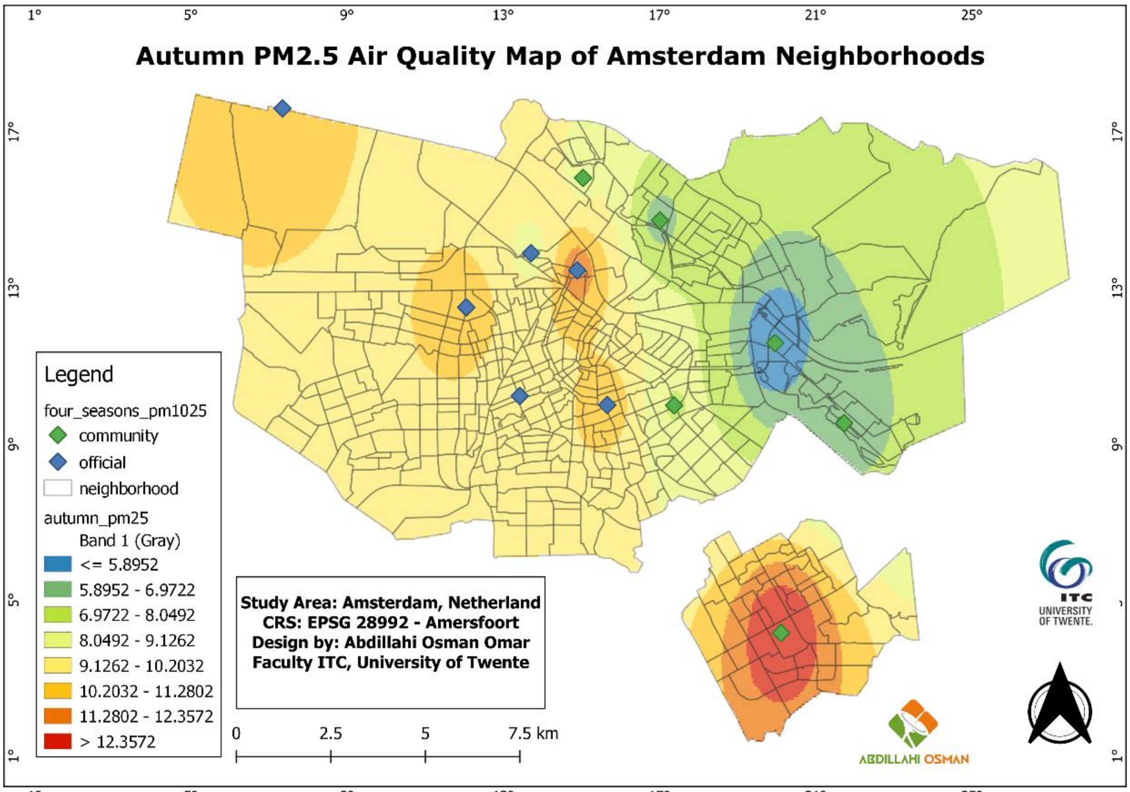


Fig. 23: Autumn PM2.5 air quality levels across various neighborhoods in Amsterdam, Source (Author)

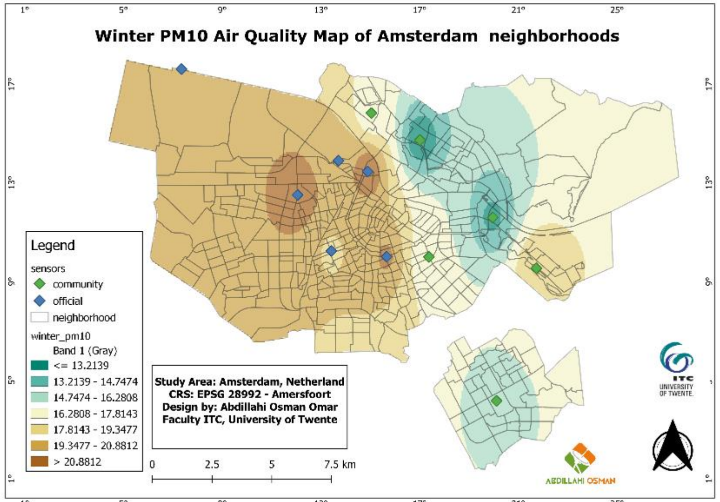


Fig. 24: Winter PM10 air quality levels

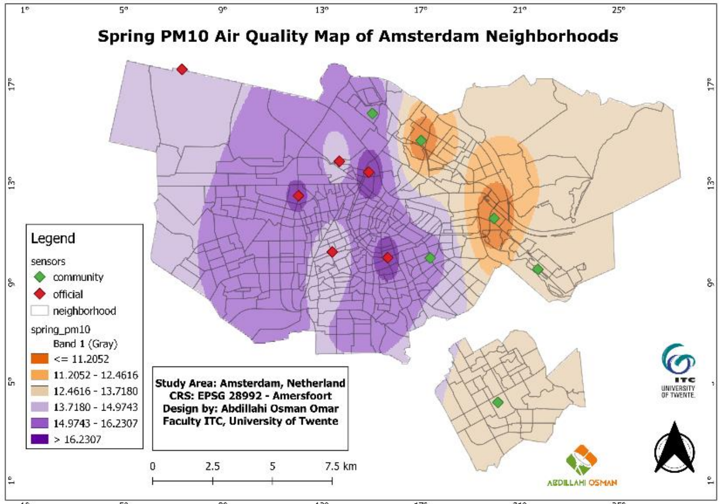


Fig. 25: Spring PM10 air quality levels

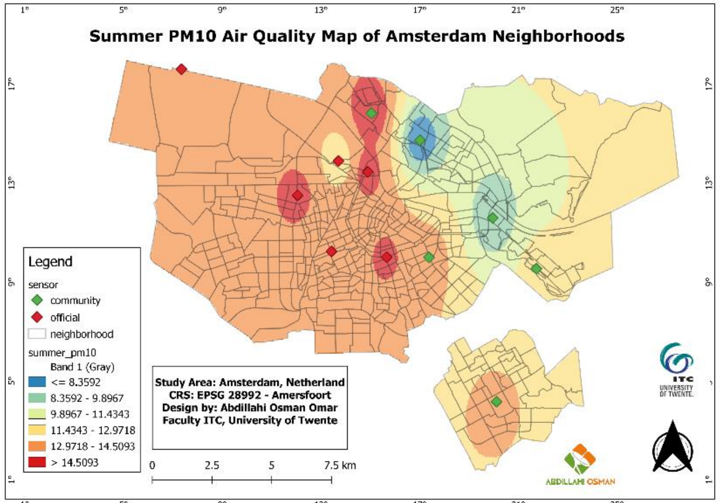


Fig. 26: Summer PM10 air quality levels

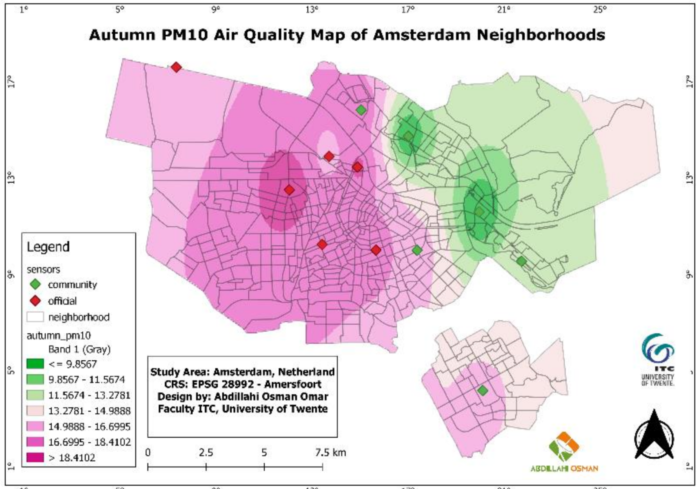


Fig. 27: Autumn PM10 air quality levels

To accomplish this goal, missing data from the official and community sensors across the neighborhoods of the municipality in 2022 was interpolated using seasonal analysis (see Tables 6 and 7) using various sensors around the municipality. The result was:

- **Winter (January - March) Air Quality Amsterdam**

While wintertime PM10 air quality levels across the city were higher than WHO (World Health Organization) guidelines, with PM10 measurements in Centrum neighborhoods reaching 22.4 μg/m³, which is higher than the annual average of PM10, which is 15 μg/m³, wintertime PM2.5 air quality levels across various neighborhoods were acceptable, with high levels of PM2.5 experienced by neighborhoods in the West and Centrum districts (see Figures 20 & 24). In Winter, neighborhoods across Nord and Oast had the best air quality in 2022.

- **Spring (April - June) Air Quality Amsterdam**

In Springtime, neighborhoods in eastern neighborhoods have the worst air quality across Amsterdam (see Figures 21 & 25) to PM2.5, and PM10 concentrations in Springtime were average across the districts.

- **Summer (July - September) Air Quality Amsterdam**

Summertime brings a lot of tourists to the city, especially to the neighborhoods around the West and Centrum districts. However, according to our analysis in Figures 22 and 26, the residential areas of Zuid and Zuidoost have the worst summertime air quality, while the neighborhoods of Oast and Nord have the best summer month air quality in Amsterdam in 2022.

- **Autumn (October - December) Air Quality Amsterdam**

Communities living in the West and Nieuw-West districts have experienced the worst air quality compared to the rest of the neighborhoods in the city (see Figures 23 & 27) but one thing to note is that there were no sensors installed in Nieuw-West district from the official network unless we do interpolation from sensors across near municipalities in this districts and community sensors there was not a viable option due to lack of measurements for enough analysis to achieve objectives of the study.

5.3 Dependability Variations in Air Pollution Levels in Amsterdam's Different Districts

Data properly structured and saved is called valid data. Conversely, trustworthy data is information that can serve as a solid foundation for analysis and judgment. Although valid data is a crucial part of trustworthy data, dependability cannot be assured by validity alone (*What Is Data Reliability?* n.d.). When it comes time to use the data for analysis or practice, depending just on validity as a measure of reliability may still lead to problems since valid data may still be incomplete and uncertain.

Establishing a consistency metric that considers variables like sensor uptime, data consistency, and completeness is one way to tackle this issue. This metric can be used to identify areas of Amsterdam city with the most dependable air quality data, which indicates data that can be trusted for analysis and decision-making, as well as areas with less dependable data, which might need more work to validate the data or maintain the sensors as for decision-makers for further action.

We have established that the reliability of air pollution data is measured by considering many aspects, including the quantity of missing data points, measurement consistency, and historical sensor analysis/measurements, based on the data collected from both official and community sensors. Furthermore, in this study, we considered the reliability of air pollution data based on the missing data points from both official and community sensors.

Our methodology utilizes NumPy for numerical operations*, it identifies columns containing air quality parameters of PM10 and PM2.5 readings based on their column names and stores the results in a different data frame to evaluate the dependability of data from air quality sensors. Two key criteria were selected to assess reliability (see Table 8). Each sensor column's Missing Data Factor (MDF), which shows the percentage of non-missing data values, is computed. Next, the Coefficient of Variation (CV) is calculated to evaluate the consistency of the data; a lower CV denotes greater reliability. To make more analysis easier, the Consistency Factor (CF) is calculated from CV consistency.

See script dataReliability.py (in the submission zipped file)

Table 8: Metrics for Assessing the Reliability of Air Quality Sensor Data

Metric	**Description**	**Objective**
Missing Data Factor (MDF)	Proportion of non-missing data points in each sensor column.	Reflects the completeness of the data.
Coefficient of Variation (CV)	Measures the consistency of the data; a lower CV denotes greater reliability.	Shows the consistency and stability of the data.
Consistency Factor (CF)	Derived from CV; allows examination of infinite values.	Guarantees the measurement of data variability's robustness.
Reliability Score (RS)	Weighted MDF and CF combined metric, normalized to 0-1.	Gives a numerical assessment of the overall quality of the sensor data.

In addition, we give weights to MDF and CF to account for their relative significance in evaluating reliability. Finally, we aggregate these metrics and apply these weights to produce a single Reliability Score (RS) for every sensor, offering a numerical representation of the quality of the data.

The methodology normalizes the Reliability Score RS to a range between 0 (least reliable) to 1 (most reliable) to make it easier to understand. It presents a ranked list of the most dependable sensors by sorting the sensors according to their normalized dependability values. To send the result to QGIS for additional analysis, I combined manually RS score CSV file with the locations of sensors and exported it as a CSV file (see Table 9).

Table 9: Reliability Scores of Air Quality Sensor Stations

station id	source	latitude	longitude	RS score
NL49007	official	52.381	4.845	0.5
NL49012	official	52.39	4.888	0.7
NL49014	official	52.36	4.866	0.5
NL49016	official	52.394	4.87	0.5
NL49017	official	52.358	4.9	0.6
NL49704	official	52.428	4.773	0.3
LTD_4969	community	52.304	4.968	0.4
LTD_32517	community	52.373	4.965	0.2
LTD_23231	community	52.358	4.926	0.7
LTD_8802	community	52.402	4.92	0.3
LTD_32999	community	52.412	4.89	0.4
LTD_55037	community	52.354	4.992	0.0

The map below uses a color gradient to represent the Reliability Score (RS), which is normalized between 0 (least reliable) and 1 (most reliable). This Choropleth Map provides a clear visual depiction of the variations in data reliability throughout Amsterdam and is intended to evaluate the accuracy of air pollution statistics throughout the city based on the harvested data.

This would make it immediately clear which parts of the city have the most and least accurate air quality data. It is possible to spot trends by looking at Figure 28 over these data points on Amsterdam's neighborhoods and districts, including whether specific neighborhoods or places (such as industrial zones, urban centers, or residential areas) regularly provide trustworthy data.

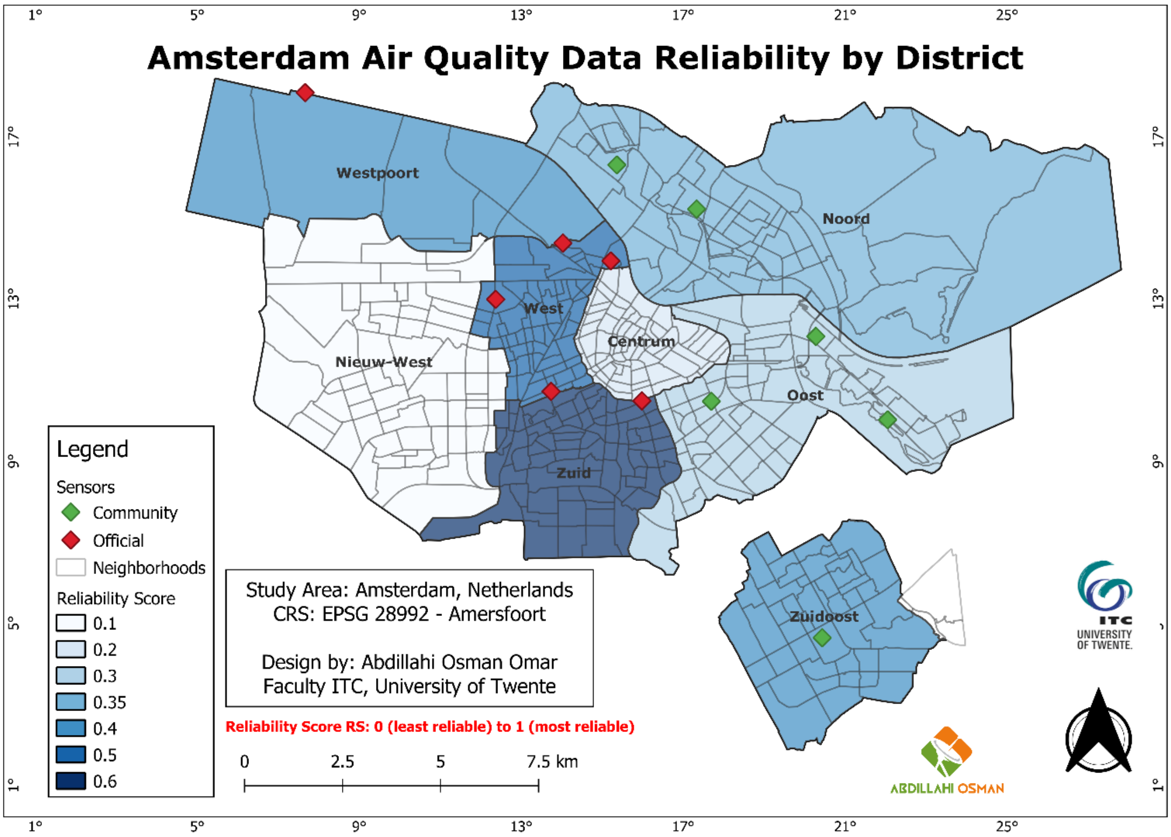


Fig.28: Amsterdam Air Quality Data Reliability by District & Neighborhood
Source: Author

We can conclude the output from this map shows us the level of reliability of air pollution data across the different districts and neighborhoods in Amsterdam. Although a lack of data from some neighborhoods and available sensors with no measurements in some areas can affect this analysis, we will need a more detailed analysis in the future for the reliability of air quality data in Amsterdam and the whole of the Netherlands.

6. DISCUSSION

6.1 Main Findings

The study reveals interesting tendencies by focusing on how the quality of the air changes along Amsterdam's bicycle paths during peak and off-peak hours. Notably, PM2.5 levels were regularly monitored in peak and off-peak times, especially near the city center on bike paths. The significance of continuously measuring air quality in high-traffic areas for urban health is shown by this measurement.

When the data is examined more closely, as shown in Figures 18 and 19, it becomes clear that in 2022, bike routes in the Centrum, West, and Zuid areas of Amsterdam had the highest PM10 readings of any district. This conclusion is noteworthy because it suggests locations where efforts to enhance air quality could have the greatest effects, particularly considering the densely populated and heavily trafficked districts in question.

For the year 2022, the air quality in the municipality's various parts remained mostly acceptable. Seasonal differences were noticeable, though, with more hot spot locations appearing in the winter and autumn. This yearly pattern points to a higher concentration of contaminants during the winter, which could be brought on by things like altered air conditions and increased emissions from heating.

Based on multiple factors, including location, the study assigned each sensor a score related to the dependability of air pollution data. It was discovered that the most trustworthy data on air pollution in Amsterdam has been identified in the Zuid and West districts. Westpoort and Nieuw-West, on the other hand, were found to have the least trustworthy statistics about air pollution. This discrepancy in data dependability emphasizes the necessity of better techniques or infrastructure to monitor.

6.2 Limitations of Study

- **Accessibility of Air Quality Data from Community Platforms:** access to air quality data from community platforms has been limited for this study, which could potentially narrow the scope of additional analysis and this limitation may result in a less complete knowledge of fluctuations in air quality.

- **The reliability of Community Sensor readings:** the validity of the study may be impacted by the lack of quality indicators for sensor readings from community platforms, which adds ambiguity to the dependability and precision of the data.

- **Social and Cultural Impacts on Data Analysis:** various target groups will understand the maps and charts in this study differently depending on their socioeconomic and cultural background which will result in differing conclusions.

- **Spatiotemporal Coverage and Data Collection Regularity:** gaps in our understanding of long-term patterns in air quality might result from irregular or intermittent data gathering, particularly when it comes to identifying periodic or quick trends though this limits the eastern district which is composed of water and much no sensors are available for data collection regularity.

- **Possibility Biassed in Sensor Location Decision:** the city's overall air quality assessment may be skewed by sensors that are concentrated in select places, such as more affluent neighborhoods or conveniently accessible sites. We have noticed that this has an immediate impact on the evaluation of air quality throughout the western districts of Amsterdam-West, Westpoort, and Nieuw-West.

6.3 Recommendations

Following the results and limitations of this study on air quality in Amsterdam, these are a few recommendations for improving future research and air quality management in Amsterdam and the whole Netherlands:

1. **Increase Collection of Data from Various Sources:** To solve the issue of data availability from community channels, initiatives should be made to integrate data from a broader range of sources, and we can collaborate with Amsterdam's local authority, environmental organizations, and civic society for deeper air quality data analysis since we need to cover wider geographic parts of the city ensuring a more detailed assessment specifically western districts of Amsterdam.

2. **Investments in modern air monitoring systems:** Investing in innovative air monitoring systems can improve the level of accuracy of data collected on air quality and involves using more sensitive and accurate sensors, along with employing current and state-of-the-art technology for real-time air quality data collecting and processing. Innovative technology could offer a more comprehensive and adaptive analysis of air quality in the Netherlands, facilitating shorter and more efficient solutions for pollution crises in the country's municipalities and future research.

3. **Public awareness and involvement in community initiatives:** Increasing public awareness of air pollution and its effects and actively interacting with local communities are vital steps in improving the quality of the air generally and participating in such projects for monitoring air quality in municipality, district, and neighborhood levels. Training efforts, sessions, and active tracking programs in which citizens collect and analyze air quality data are all initiatives, advocating communication among citizens, legislators, and academics might result in better community-informed choices and regulations in The Netherlands.

7. CONCLUSION

Using data from official Dutch sensors such as Lucht Meet Net, an assessment of the distribution of air quality in Amsterdam 2022 was effectively carried out and a definitive conclusion was reached. To complete the information lacking from government sensors, additional crowdsourcing data sources, such Samen Meten, were also employed. To examine the accuracy of the data and produce precise evaluations and findings, descriptive statistical methods were applied.

This project's purpose was to examine air quality in Holland for 2022, with an emphasis on Amsterdam city. It focused on three questions: Initially it attempted to recognize the fluctuation in air quality along bike lanes during peak and off-peak hours, which has a significant impact on a city known for its cycling lifestyle. Additionally, the study looked to determine which neighborhoods in Amsterdam had the highest and lowest levels of air quality across the year. Lastly, it investigated the credibility of air pollution data in various regions, identifying areas that offered the best and poorest reliable information.

Toward the end of this paper, the goals were met effectively. The thorough examination of pollution levels in the Netherlands, specifically in Amsterdam, provides a spotlight on the disparities in air quality across bike paths, the yearly levels of air quality across various neighborhoods of the city, and the dependability of pollution level data across areas in Amsterdam.

Lastly, by implementing real-time monitoring along well-traveled bicycle routes in the Netherlands, we can broaden the study to assess PM2.5 and PM10 or other air pollutant matter exposure for cyclists. This targeted study could improve the nation's bicycling environment by informing health and urban planning policy.

This study offers an adequate basis for everyone and government organizations. It shows the facts needed to make educated choices about establishing sustainable air quality policies. The report's findings can contribute to the formulation of environmental initiatives and approaches, leading to better air quality in The Netherlands.

8. REFERENCES

About Amsterdam Metropolitan Area. (n.d.). metropoolregioamsterdam. Retrieved 26 January 2024, from https://www.metropoolregioamsterdam.nl/about-mra/

Božilov, A., Tasić, V., Živković, N., Lazović, I., Blagojević, M., Mišić, N., & Topalović, D. (2022). Performance assessment of NOVA SDS011 low-cost PM sensor in various microenvironments. *Environmental Monitoring and Assessment*, *194*(9), 595. https://doi.org/10.1007/s10661-022-10290-7

Greven, F. E., Vonk, J. M., Fischer, P., Duijm, F., Vink, N. M., & Brunekreef, B. (2019). Air pollution during New Year's fireworks and daily mortality in the Netherlands. *Scientific Reports*, *9*(1), 5735. https://doi.org/10.1038/s41598-019-42080-6

Hamad, S., Green, D. & Heo, J. Evaluation of health risk associated with fireworks activity at Central London. Air Quality, Atmosphere and Health 9, 735–741 (2016). http://dx.doi.org/10.1007/s11869-015-0384-x

Home—PDOK. (n.d.). Retrieved 28 December 2023, from https://www.pdok.nl/

Inhalable Particulate Matter and Health (PM2.5 and PM10) | *California Air Resources Board*. (n.d.). Retrieved 24 January 2024, from https://ww2.arb.ca.gov/resources/inhalable-particulate-matter-and-health

Laden F, Schwartz J, Speizer FE, Dockery DW (2006) Reduction in fine particulate air pollution and mortality extended follow-up of the Harvard six cities study. Am J Respir Crit Care Med 173:667–672 https://doi.org/10.1164/rccm.200503-443OC

Luchtmeetnet.nl. (n.d.). Luchtmeetnet.Nl. Retrieved 28 December 2023, from https://www.luchtmeetnet.nl/

Measuring Together—Data Portal. (n.d.). Retrieved 28 December 2023, from https://samenmeten.rivm.nl/dataportaal/

Plausibility of the particulate matter measurement | *Measuring air quality together*. (n.d.). Retrieved 18 January 2024, from https://www.samenmeten.nl/data/plausibiliteit-van-fijnstofmeting

Protocol to Abate Acidification, Eutrophication, and Ground-level Ozone | *UNECE*. (n.d.). Retrieved 25 December 2023, from https://unece.org/environment-policy/air/protocol-abate-acidification-eutrophication-and-ground-level-ozone

Waterstaat, M. van I. en. (2014, August 13). *Air quality standards—Air Quality—Government.nl* [Onderwerp]. Ministerie van Algemene Zaken. https://www.government.nl/topics/air-quality/air-quality-standards

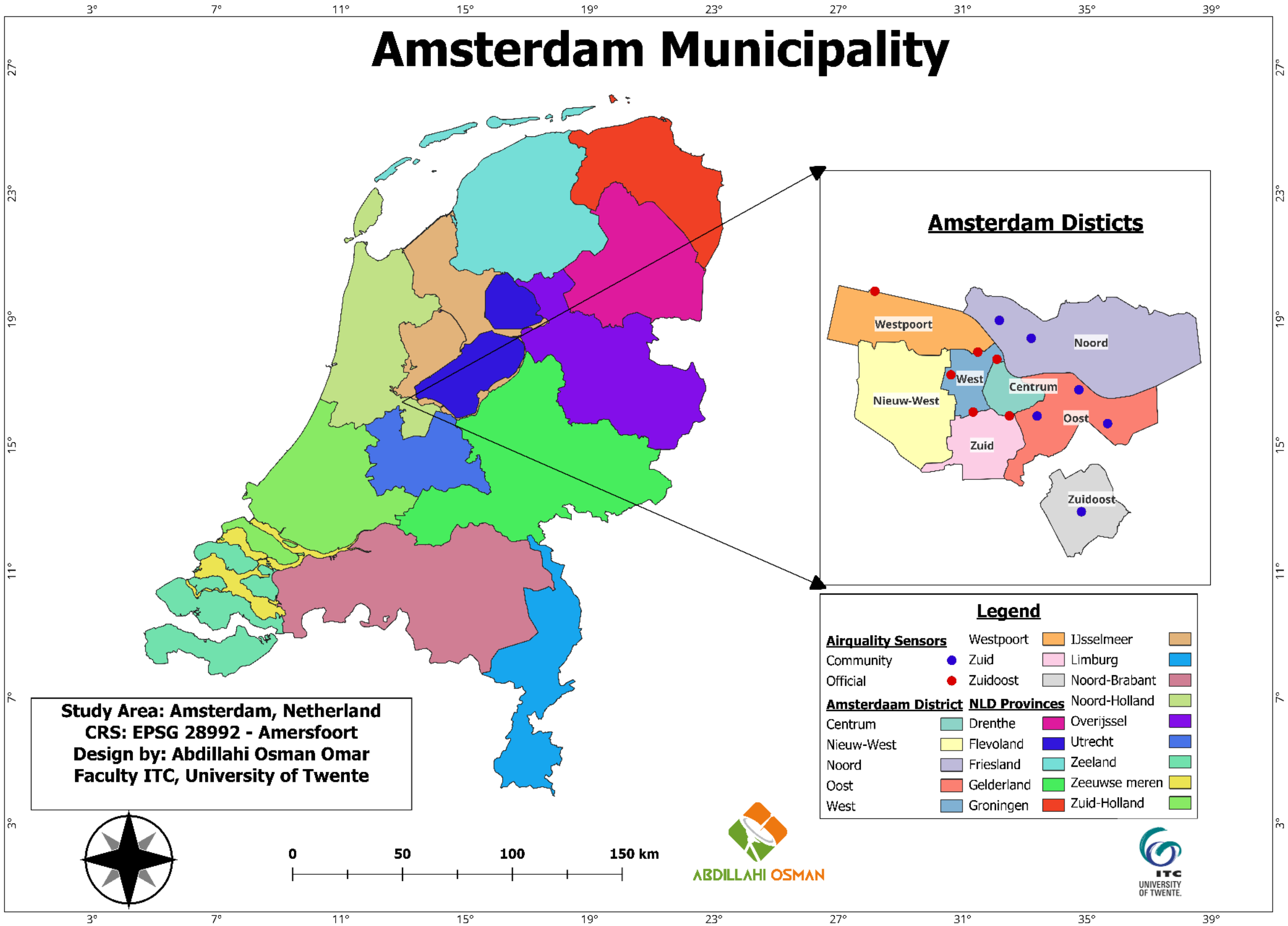
What is Data Reliability? Definition & Assessment Guide. (n.d.). Talend - A Leader in Data Integration & Data Integrity. Retrieved 25 January 2024, from https://www.talend.com/resources/what-is-data-reliability/

WHO global air quality guidelines: Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. (n.d.). Retrieved 25 December 2023, from https://www.who.int/publications-detail-redirect/9789240034228

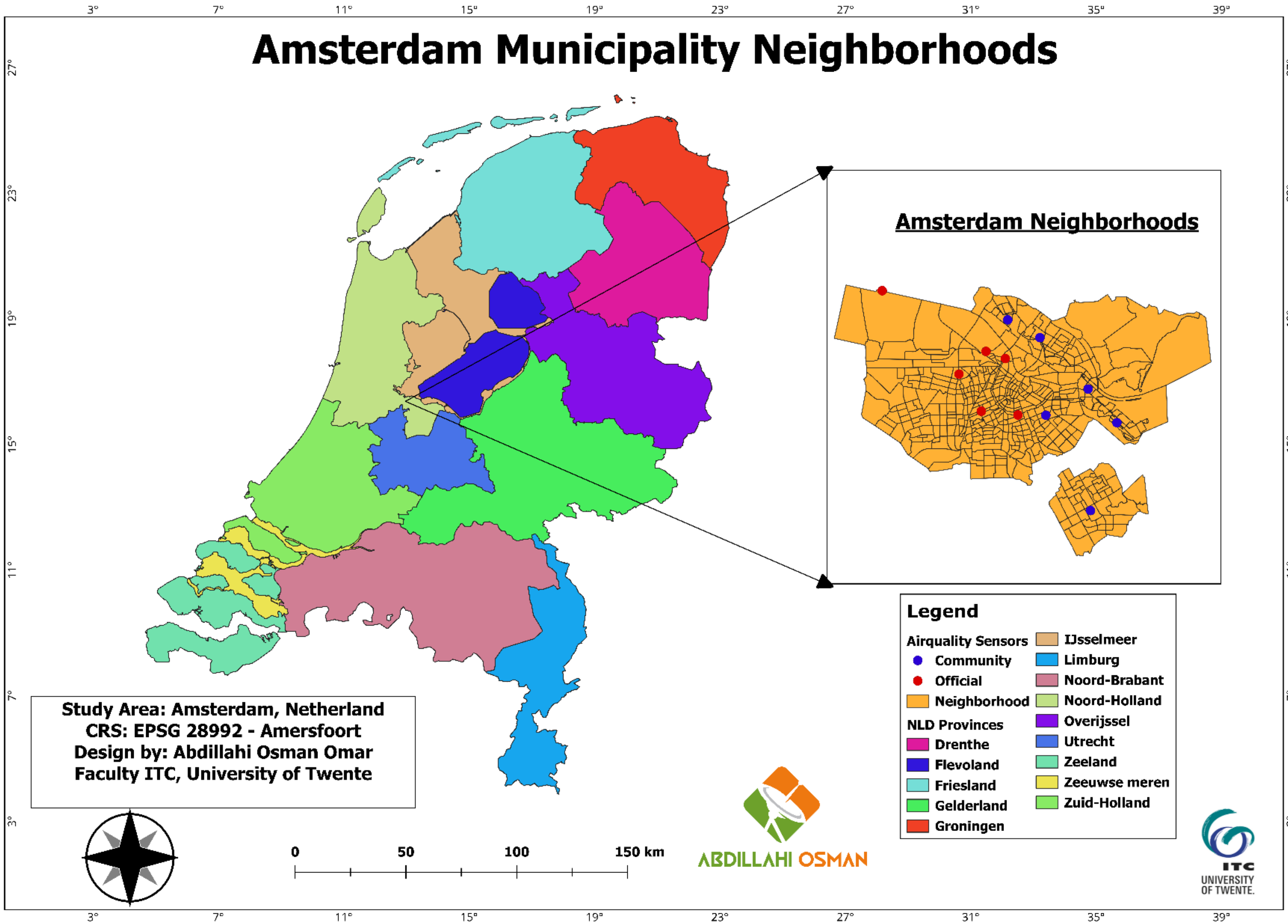
APPENDIX A: DATA MANAGEMENT PLAN

<table>
<tr><th colspan="4">Objective 1</th><th colspan="4">Objective 2</th><th>Objective 3</th></tr>
<tr><td colspan="4">Amsterdam districts shapefile</td><td colspan="4">Amsterdam neighborhoods shapefile</td><td>Amsterdam districts shapefile</td></tr>
<tr><td colspan="4">Bike Paths</td><td colspan="4">Air Quality Data CSV & Sensor Location CSV</td><td>Air Quality Data CSV & Sensor Location CSV</td></tr>
<tr><td colspan="4">Air Quality Data CSV & Sensor Location CSV</td><td>WinterPM10</td><td>SpringPM10</td><td>SummerPM10</td><td>AutumnPM10</td><td>Reliability scores for the sensors on each PM10 & PM2.5 CSV</td></tr>
<tr><td>Off-peak PM2.5 CSV</td><td>Off-peak PM10 CSV</td><td>Peak PM2.5 CSV</td><td>Peak PM10 CSV</td><td>WinterPM2.5</td><td>SpringPM2.5</td><td>SummerPM2.5</td><td>AutumPM2.5</td><td>Scores sensors CSV with sensor location & districts SHP</td></tr>
<tr><td colspan="4">Off-peak & and peak hours for PM2.5 and PM10 single CSV</td><td colspan="4">4 seasonal CSV with each PM10 and PM2.5</td><td></td></tr>
<tr><td colspan="4">Off-peak & and peak hours interpolated raster for PM2.5 and PM10</td><td colspan="4">Interpolated raster per each season</td><td></td></tr>
</table>

APPENDIX B: AMSTERDAM DISTRICTS DETAIL MAP



APPENDIX C: AMSTERDAM NEIGHBORHOODS DETAIL MAP



APPENDIX D: REGRESSION ANALYSIS: PM2.5 AND PM10 CORRELATION NL49007 & LTD_4969

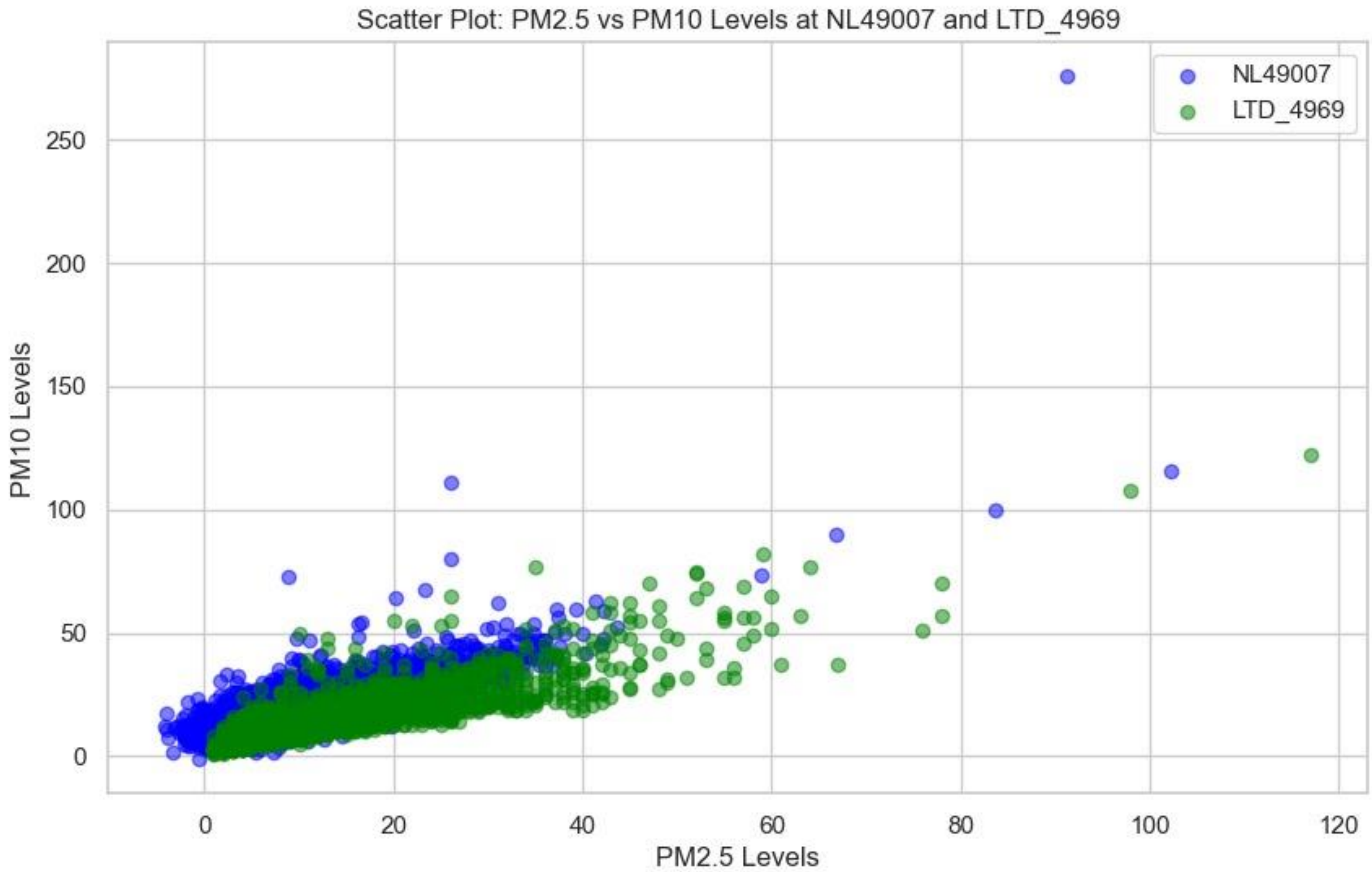
```
                            OLS Regression Results
==============================================================================
Dep. Variable:            NL49007_PM10   R-squared:                       0.635
Model:                             OLS   Adj. R-squared:                  0.635
Method:                  Least Squares   F-statistic:                     5607.
Date:                 Mon, 29 Jan 2024   Prob (F-statistic):               0.00
Time:                         13:25:51   Log-Likelihood:                -10605.
No. Observations:                 3219   AIC:                         2.121e+04
Df Residuals:                     3217   BIC:                         2.123e+04
Df Model:                            1
Covariance Type:             nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const            8.5709      0.175     48.919      0.000       8.227       8.914
NL49007_PM25     1.0635      0.014     74.881      0.000       1.036       1.091
==============================================================================
Omnibus:                     4195.868   Durbin-Watson:                   1.182
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          3056870.079
Skew:                           6.628   Prob(JB):                         0.00
Kurtosis:                     153.385   Cond. No.                         18.9
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
""",
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            LTD_4969_PM10   R-squared:                      0.749
Model:                              OLS   Adj. R-squared:                 0.749
Method:                   Least Squares   F-statistic:                    9774.
Date:                  Mon, 29 Jan 2024   Prob (F-statistic):              0.00
Time:                          13:25:51   Log-Likelihood:                -9918.7
No. Observations:                  3275   AIC:                        1.984e+04
Df Residuals:                      3273   BIC:                        1.985e+04
Df Model:                             1
Covariance Type:              nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const            4.0703      0.134     30.416      0.000       3.808       4.333
LTD_4969_PM25    0.8435      0.009     98.863      0.000       0.827       0.860
==============================================================================
Omnibus:                     1281.059   Durbin-Watson:                   0.594
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            13785.421
Skew:                           1.549   Prob(JB):                         0.00
Kurtosis:                      12.562   Cond. No.                         24.1
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
""")
```

APPENDIX E: SCATTER PLOT: PM2.5 VS PM10 LEVELS AT NL49007 AND LTD_4969



APPENDIX F: TOOLS UTILIZED IN ASSIGNMENT PREPARATION

"During the preparation of this work, the author used Chat GPT to harvest, clean, and analyze the data. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the work."

Use of AI in Education at the University of Twente

Omar, Abdillahi Osman

January 2024

...

FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION
UNIVERSITY OF TWENTE, ENSCHEDE, THE NETHERLANDS