A white text on a black background

Description automatically generated

**School of Tech**

**Graduate Diploma in Data Analytics (Level 7)**

Automating PM10 Source Apportionment Data Analysis Using Python

**Submitted By: Priyank Dangwal**

Student ID: 764707660

**Submitted to:**

External Supervisor: Dr. Louis K Boamponsem

Internal Supervisor: Dr. Sara Zandi

Contents

[Abstract 4](#_Toc172814388)

[Keywords 4](#_Toc172814389)

[Introduction 4](#_Toc172814390)

[Project Questions 5](#_Toc172814391)

[Project Design 5](#_Toc172814392)

[Components of schematic diagram 6](#_Toc172814393)

[Sources of Data: 6](#_Toc172814394)

[Analytics and Data Processing: 6](#_Toc172814395)

[Outputs and visualisations: 7](#_Toc172814396)

[User Interface: 7](#_Toc172814397)

[Integration Points: 7](#_Toc172814398)

[Security and Compliance: 7](#_Toc172814399)

[In Summary: 8](#_Toc172814400)

[Project Development 8](#_Toc172814401)

[Preprocessing Part 8](#_Toc172814402)

[Automation Part 9](#_Toc172814403)

[Visualization Part 10](#_Toc172814404)

[Discussion 12](#_Toc172814405)

[Limitations 12](#_Toc172814406)

[Use of Kaupapa Māori research Methodology 13](#_Toc172814407)

[Conclusion with future directions 13](#_Toc172814408)

[References: 14](#_Toc172814409)

[Appendix 1: Capstone Project Checklist 14](#_Toc172814410)

[Appendix 2: Capstone Project MOU 16](#_Toc172814411)

[Appendix 3: Confidentiality 20](#_Toc172814412)

[Appendix 4: Checkpoint Checklist #\_\_ 20](#_Toc172814413)

[Appendix 5: Capstone Project Weekly Work Log 0](#_Toc172814414)

[Appendix 7: Student Capstone Project Evaluation 0](#_Toc172814415)

A white text on a black background

Description automatically generated

**Graduate Diploma in Data Analytics (Level 7)**

**Cover Sheet and Student Declaration**

This sheet must be signed by the student and attached to the submitted assessment.

|  |  |  |  |
| --- | --- | --- | --- |
| **Course Title:** | **Capstone Project (DA)** | **Course code:** | **GDDA713** |
| **Student Name:** | Priyank Dangwal | **Student ID:** |  |
| **Assessment No** | Assessment 2, Research Project Report | **Cohort:** |  |
| **Due Date:** | 26/07/2024 | **Date** **Submitted:** |  |
| **Tutor’s Name:** | Mohammad Norouzifard | | |
| **Assessment Weight** | 65% | | |
| **Total Marks** | 100 | | |

**Student Declaration:**

I declare that:

• I have read the New Zealand School of Education Ltd policies and regulations on assessments and understand what plagiarism is.

• I am aware of the penalties for cheating and plagiarism as laid down by the New Zealand School of

Education Ltd.

• This is an original assessment and is entirely my own work.

• Where I have quoted or made use of the ideas of other writers, I have acknowledged the source.

• This assessment has been prepared exclusively for this course and has not been or will not be submitted as assessed work in any other course.

• It has been explained to me that this assessment may be used by NZSE Ltd, for internal and/or external moderation.

**Student signature: Priyank Dangwal**

**Date: 26/07/2024**

|  |  |  |  |
| --- | --- | --- | --- |
| **Assessor only to complete:** | | | |
| **Assessment results:** | **All Tasks except formatting /95 marks** | | Report formatting **/5 marks** |
| **Total Marks: /100** |  | |
| **LO1 Requirements** | **Met                            ​☐​**  **Not Met                    ​☐​** | **Assessor Signature** | |
| **LO2 Requirements** | **Met                            ​☐​**  **Not Met                    ​☐​** | **Assessor Signature** | |
| **LO3 Requirements** | **Met                            ​☐​**  **Not Met                    ​☐​** | **Assessor Signature** | |

## Github link

[Priyank Dangwal Portfolio Template - Index (dangwal06.github.io)](https://dangwal06.github.io/Portfolio-1/index.html#projects)

[Capstone\_Project\_Final/Capstone 2nd.ipynb at main · dangwal06/Capstone\_Project\_Final (github.com)](https://github.com/dangwal06/Capstone_Project_Final/blob/main/Capstone%202nd.ipynb)

## Abstract

This project addresses the intricate business requirements for air quality management and evidence-based policymaking by developing an automated system for PMF-generated PM10 source apportionment data analysis using Python. We developed a system that automates data pretreatment, analysis, and visualisation by utilising our extensive programming abilities and technical expertise. This ensures consistency and scalability while drastically minimising human labour and mistakes. To preserve data quality, the system incorporates strong data preparation methods as validation, normalisation, and cleaning. The comprehension of pollution sources is improved by interactive visualisations that were made with Matplotlib and Plotly. These visualisations offer insightful information about the contributions of PM10 sources. Both technical and non-technical stakeholders may simply submit data, choose visualisation kinds, and interact with the data thanks to the system's user-friendly graphical user interface (GUI), which was created with `Ipywidgets` and Dash. The project's all-encompassing methodology showcased Python's capacity to manage intricate data operations and facilitate sophisticated analytical procedures. Notwithstanding the difficulties in attaining flawless visual quality in Plotly and guaranteeing data consistency across several sources, the system efficiently facilitates corporate decision-making by offering precise, prompt, and useful insights. Upcoming advancements will focus on improving the accuracy of visualisation, using machine learning to get more profound understanding, and extending the capacity to handle data in real-time. This project demonstrates how programming knowledge, data analytics proficiency, and project management abilities may be combined to create creative solutions that enhance environmental sustainability and public health. Our automated approach helps to improve air quality management methods by promoting evidence-based policymaking, demonstrating the revolutionary power of technology in tackling environmental issues.

## Keywords

PM10 Source Apportionment, Data Analysis, Data Automation, Interactive Visualization, Python, Air Quality

Introduction

Particulate matter (PM) is one of the most dangerous pollutants for the environment and human health, and air pollution is a major worldwide problem. Particulate matter of a diameter of 10 micrometres or smaller, or PM10, can cause a variety of respiratory and cardiovascular conditions by deeply penetrating the lungs and even into the bloodstream. For the purpose of creating methods for managing air quality, it is essential to comprehend the sources of PM10. This capstone project uses Python to automate the study of PM10 source apportionment data with the goal of giving academics and decision-makers a powerful tool for effectively locating and managing pollution sources.

This project's main context is found in the fields of environmental research and public health, where rapid and precise data analysis is crucial to make well-informed decisions. Conventional techniques for allocating PM10 sources are frequently labour-intensive, slow, and prone to human error. This project aims to increase productivity, repeatability, and scalability by automating the analysis process and utilising Python's robust data processing and visualisation features. This is in line with the Minimum Viable Product (MVP) development methodology, which aims to provide a working product that can be refined repeatedly in response to user input and practical use.

The project's primary concerns centre on whether automated PM10 source apportionment is feasible and efficient. More specifically, can the preparation, analysis, and visualisation of data be made more efficient with Python? To what extent is the automated system able to reproduce the outcomes of the human analysis? What difficulties exist in combining several data sources and guaranteeing the accuracy and consistency of the data? It is vital to tackle these enquiries to design a solution that satisfies the intricate operational requirements of public health organisations and environmental monitoring authorities.   
  
The PM10 data produced by Positive Matrix Factorization (PMF) in Queen Street, Auckland CBD, is the focus of this study. The project intends to use Python's extensive ecosystem of modules and tools to automate the processing, analysis, and visualisation of PM10 source apportionment data obtained from the PMF model. This will be accomplished by automating the analysis of the data. This will entail preparing the raw data for analysis, running statistical tests, and producing summaries of the findings in the form of reports or visualisations. Ensuring more precise and repeatable results through automation would not only expedite the analytical process but also enable more informed decision-making for air quality management.   
  
This research is important in ways that go beyond data analysis technicalities. The project is to promote evidence-based policymaking and aid in better air quality management by offering a dependable and effective approach for PM10 source apportionment. As a result, the project's results are relevant to a wide range of stakeholders in relation to public health, environmental sustainability, and regulatory compliance.

## Project Questions

The industrial initiative focuses on the intricate business requirements of environmental monitoring and public health management, with the goal of addressing several important issues. To preserve consistency in scope and objectives, the major questions are constructed with clarity and alignment with the original project proposal in mind.

1. Can Python effectively automate PMF-generated PM10 source apportionment data preprocessing, analysis, and visualisation?

* This topic examines if Python can streamline workflow, reduce human labour and mistakes, and ensure repeatability and scalability of the analytical process.

1. To what extent is the automated system able to replicate and even improve the outcomes of the manual PM10 source apportionment analysis?

* This enquiry centres on how accurate and dependable the automated system contrasts with more conventional techniques. It looks at whether automation can improve or preserve the precision of locating the sources of pollution.

1. What challenges arise when integrating several data sources and guaranteeing consistency and quality of data in the automated workflow?

* To provide solid and trustworthy analytical results, this question explores the operational and technological challenges of managing a variety of data sets. It covers topics such data cleansing, normalisation, and validation.

1. In what ways may the automated system aid in evidence-based policymaking and enhance methods for managing air quality?

* This question connects project technicalities to commercial and social implications. It examines how automated analysis may improve policy decisions, regulatory compliance, improve public health and environmental sustainability.

1. How can a user-friendly GUI and interactive graphs improve the automated system's usability and accessibility?

* This topic analyses if the automated system needs a GUI and interactive charts to be more accessible. A well-designed GUI with interactive plots simplifies data upload, visualisation selection, and data interaction using intuitive controls.

## Project Design

The design stage helps data analysts understand the architecture and parts of the system. In this research, elements of big data analytics, artificial intelligence (AI), and machine learning (ML) are integrated to automate the study of PM10 source apportionment data. The schematic diagram for our project makes it easier to comprehend the architecture of the system and how different modules work together.

A diagram of data flow

Description automatically generated

Figure 1 : Schematic Diagram of proposed Automating PM10 Source Apportionment Data Analysis.

### Components of schematic diagram

Sources of Data:  
The foundation of our system is made up of the data sources. The primary data source is the Auckland Council, which provides PM10 model findings generated by the Positive Matrix Factorization (PMF) approach. Because they offer the initial dataset for additional research, their findings are essential. The dataset, which is provided in CSV format, is crucial for comprehending the many factors influencing PM10 pollution levels.  
  
Ingestion and Storage of Data:  
After the data is acquired, it moves into the phase of data intake and storage. This stage involves several necessary procedures:  
Data Ingestion: To ensure that the data is easily available for additional processing, the raw CSV files are imported into the system using Pandas, a potent Python data processing tools.   
Preprocessing: This involves cleaning the data to remove errors or variabilities and normalizing it to make sure it is formatted consistently for precise analysis.

Analytics and Data Processing:  
This stage requires several important adjustments and evaluations.   
Data Transformation: Ensuring the dataset is full and prepared for analysis by filling in any missing values and modifying data types as necessary.   
Data Aggregation: Putting the data into appropriate groups and doing a statistical summary of the results to identify more general trends and patterns.   
Automation: By automating data processing pipelines, system dependability and overall efficiency are maximised through effective, repeatable processing with no need for human involvement.

Outputs and visualisations:  
Visualisation: Different graphs and plots are made to graphically depict the data using programmes like Matplotlib and Plotly, which are essential for results reporting and help with intuitive comprehension.   
Interactive Plots: Plots that are made using software such as Plotly enable users to interactively examine the data, offering a more interesting and perceptive means of interacting with it.   
Charts: Python modules such as Dash or Plotly may be used to construct dashboards that offer a centralised platform for visualising significant indicators and trends. These dashboards are helpful for both continuous monitoring and real-time analysis.

User Interface:  
A graphical user interface (GUI) is created to improve accessibility and usability.   
GUI Creation: Libraries such as Ipywidgets, matplotlib. pyplot, Ipython.display etc. are used to build user interfaces, which facilitate user interaction with the system. The GUI makes the system more intuitive and user-friendly by facilitating a variety of tasks, from data intake to visualisation.

Integration Points:  
For the system to function smoothly, several components must be integrated:  
Dashboard Integration: Python libraries such as Dash and Plotly are two technologies that are used to generate dashboards that are integrated with the system. This interface guarantees easy accessibility to the processed data and visualisations, allowing for ongoing monitoring.

#### Security and Compliance:

When processing data, security and compliance are crucial, especially when sensitive information is involved:  
Data ethical Approval: An agreement with the Auckland Council established the ethical permissions, and the data utilised in this study is publicly accessible. This agreement makes sure that rules regarding privacy are followed.

Data Protection and Compliance: Auckland Council enforces data laws. The agreement acknowledged appropriate data processing and data protection rules. Compliance includes data security, client instructions, and data storage best practices.   
  
Feedback Loop:

Constant improvement and consistency to stakeholder expectations are guaranteed by the feedback loop.   
Weekly Meetings: Consistent get-togethers with the supervisor to go over updates, talk about difficulties, and fine-tune the strategy.   
Client Interactions: Regular meetings with the Auckland Council to discuss progress, solicit comments, and make sure the project satisfies their needs.   
Progressive Refinement: Project components were improved to increase accuracy and relevance based on input from these sessions, guaranteeing that the final products closely matched the client's expectations.

While not explicitly used in this project, artificial intelligence (AI) components and machine learning models are acknowledged for their potential use.

AI Components: AI-driven analysis and advanced data mining techniques may improve insights by spotting complex patterns or forecasting future trends.   
Models for Machine Learning: To support proactive environmental management, predictive models based on past data might be created to estimate PM10 levels.

Justification for Exclusion: Project Range- The main goal was to use the PMF model for source apportionment to automate the examination of data supplied by the Auckland Council. Beyond the project's timeframe and scope, further AI or machine learning models were to be included.   
Future Improvements: To improve capabilities and offer predictive insights, the incorporation of AI and machine learning models may be taken into consideration for further versions.

In Summary:  
Our system integrates various components and techniques to automate PM10 source apportionment data analysis. From data input and preprocessing to automation and visualisation, every step is designed for usability, accuracy, and efficiency. A GUI and live graphs make the system easy to use and effective for identifying PM10 pollution sources. Data processing, system integrity, and dependability are ensured by security and compliance. Frequent feedback loops ensure project progress and goals. AI components and machine learning models were not included of this research, although their potential usage suggest future improvements. This architecture is ideal for Python-based PM10 source apportionment and data analysis automation.

## Project Development

### Preprocessing Part

We developed a thorough data preparation strategy as part of our project development for automating PM10 source apportionment data analysis. First, we verified that our dataset was there by checking the file location and then putting the data into a Data Frame. We looked over the first few rows, found any missing values, and produced a statistical summary to comprehend the dataset.

To continue cleaning the data, we eliminated duplicate rows and items that lacked dates. We substituted the appropriate column means for any missing values in other columns. We computed the Interquartile Range (IQR) for the PM10 measurements and eliminated data points that were outside of the permitted range since we understood how important it is to handle outliers.

We next turned our attention to feature selection, figuring out which columns—including certain aspects like Sc, Co, Ga, etc., were less relevant and eliminating them. To guarantee consistent data scales, we used Min-Max Scaling to all columns except for the "Date" column.

To aid in temporal analysis, we additionally extracted date-related variables and added year, month, and day columns. Finally, to facilitate more informative analysis and visualisation of patterns over time, we aggregated the data by month and year and calculated the mean values for these periods. This thorough preparation created a solid framework for our ensuing analytical work.

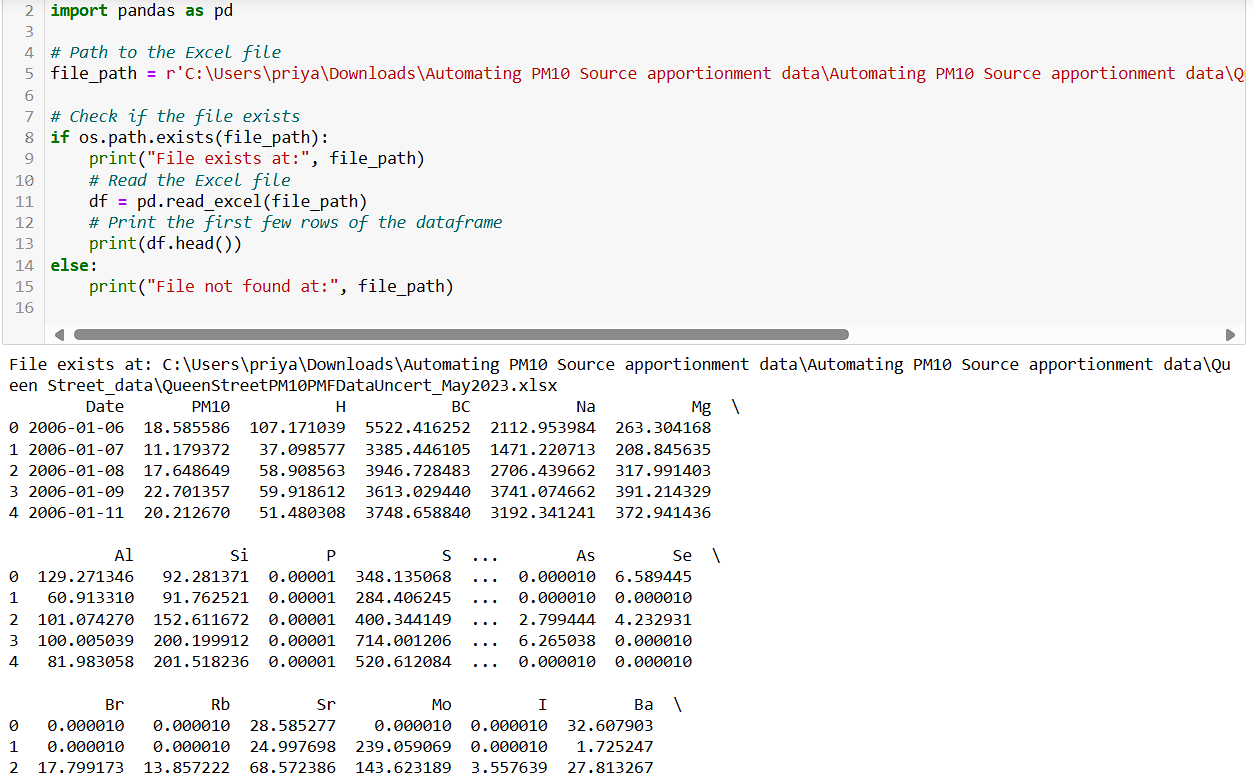


Figure 2: Screenshot of Data Loading and Reading from CSV File using Pandas.

### Automation Part

To advance the data extraction, computation, and visualisation procedures, we created several automation scripts as part of the project development. First, we imported the dataset and took out columns pertaining to different sources and their dispersion statistics. We generated a new Data Frame with computed values for each source by normalising these values against base values.   
We processed this Data Frame and stored it so that our data could be accessed for additional research. To further improve the comprehensiveness of our dataset, we also included new columns for the explained variance percentages for each source.  
  
Using bar and scatter plots to visualize the data was our primary goal. For each source—soil, sulphate, diesel, marine aerosol, biomass, petrol, and construction—we created subplots to show the concentrations, asymmetry, errors, and dispersion. A twin y-axis was included in each subplot to show the concentration and provide an explanation of the fluctuation percentages. To improve visibility of fluctuations, we changed the y-axis scales to log scale, added gridlines for clarity, and labelled the plots for simple comprehension. We saved the final plots as high-resolution photos and were able to communicate the essential findings from the data with effectiveness.

A graph of different types of fuel

Description automatically generated with medium confidence

Figure 3: Screenshot of Bar Plot Illustrating the Aggregated PM10 Data by Source, Automating Python Script.

An automated Python script was used to create above figure, which shows the species that are linked to PM10 sources such as diesel vehicles, sulphate/marine diesel, and soil/road dust. Species (e.g., H, Al, Si) are listed on the x-axis, and their concentrations are displayed on the y-axis (logarithmic scale). Species concentrations are represented by bars, while error margins are shown by black dots and red circles. The display supports PM10 source apportionment by making it easier to comprehend how each species contributes to various sources. The appendix contains comprehensive data for in-depth study, including explained variation and concentration.

A pie chart with numbers and text with Crust in the background

Description automatically generated

Figure 4: Screenshot of Pie Chart Displaying the Distribution of PM10 Sources.

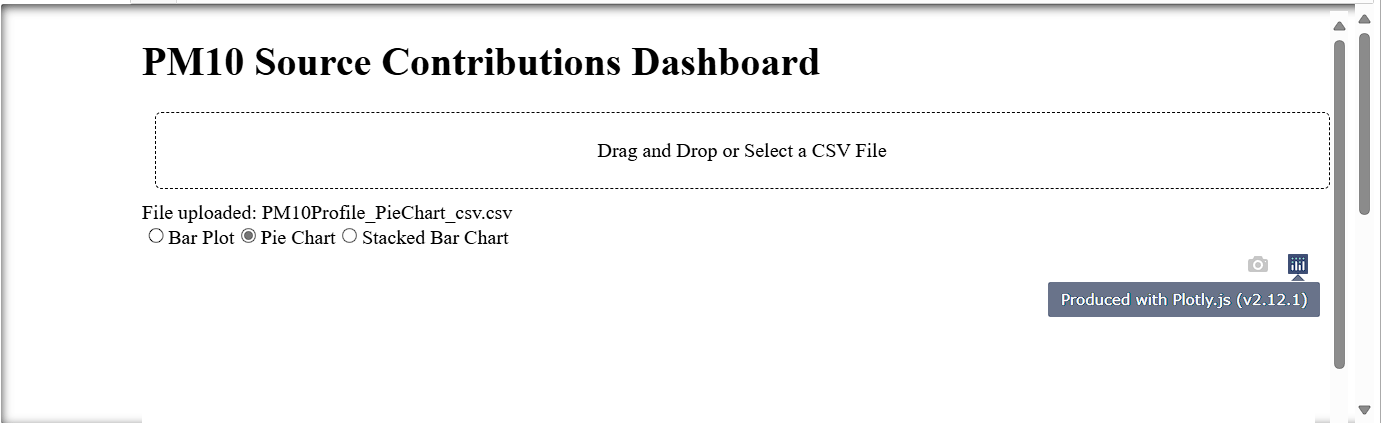
To provide a clear picture of the percentage contributions from each source, we have built a pie chart to depict PM10 source contributions. The method of automation and visualisation greatly enhanced our capacity to examine and display data related to PM10 source apportionment.

### Visualization Part

We carried out several data visualisation, GUI, and interactive plotting activities as part of the development. To plot the PM10 source contributions over time, we first saved the processed Data Frame to a CSV file and added titles, labels, and legends to make the chart more readable.   
  
We created an interactive interface with `Ipywidgets` to improve user involvement, where users could upload files, choose different plots, and save plots as JPEG photos. Plotting, generating visualisations, and handling file uploads were all defined as functions. Three distinct data presentation options are available to users: pie charts, stacked bar charts, and bar plots.

The reason of choosing Ipywidgets over Tkinter alongside other libraries Dash, and Plotly for creating GUI was their smooth integration within Jupyter notebooks, facilitating the creation of dynamic and user-friendly web-based graphical user interfaces. This combination offers effective data visualisation and manipulation, resulting in a notable decrease in development time and enhanced performance as compared to standalone desktop apps such as Tkinter. Tkinter was taking longer than usual time to run more than an hour which was affecting the development phase of GUI.   
  
We created a Dash application to provide more sophisticated interactive visualisation. Plot type selection, file upload, and a graph display area are all included in the app layout. The application parses the content and creates the chosen plot type after receiving a CSV file. We created subplots for bar graphs using Plotly, which showed each factor's concentrations, asymmetry, dispersion, and explained variation. For a thorough understanding of PM10 source contributions, Plotly was also used to construct pie charts and stacked bar charts.

Using a strong and intuitive framework for analysing and visualising PM10 source apportionment data, this method allowed for improved data interpretation and deeper insights.



A colorful square with a green and blue background

Description automatically generated with medium confidence

A screenshot of a graph

Description automatically generated

Figure 5: Stacked Bar Chart and Pie Chart Displaying Distributions of Various PM10 sources in Interactive GUI.

A thorough understanding of PM10 source contributions is offered via the interactive GUI's visualisations. By dividing each source into smaller groups, the stacked bar chart facilitates in-depth study and a more complex understanding of PM10 composition. In contrast, the pie chart provides an instantaneous and unambiguous visual assessment of the primary sources of PM10 pollution, facilitating the identification of the most prominent sources.  
When taken as a whole, these charts improve knowledge of the origins of PM10 pollution and help inform data-driven environmental management and policy development decisions.

To keep the report simple, the appendix includes the interactive GUI's stacked bar chart and pie chart. This keeps the report's main body simple while giving in-depth data analysis for those who want more. The appendix provides a detailed examination of PM10 source distributions and visual insights.

## Discussion

There were several difficulties in the planning and execution of our automating PM10 source apportionment data analysis project. Making sure the scripts for data automation and visualisation could adequately manage the volume and complexity of the data posed a major problem. It was imperative to have strong data cleaning, normalisation, and aggregation procedures in place since there are many different sources of PM10 and thorough analysis was required.  
We encountered technical difficulties in producing interactive and user-friendly visualisations throughout the development stage. Plotly's interactive graphs, as opposed to Matplotlib's static plots, made disparities in data manipulation and plot arrangements visible. These discrepancies included inconsistent marker style and subplot configurations, mismatched data extraction from the Data Frame, and problems with axis scaling (such as using a log scale). For instance, subplot titles and legends could be precisely controlled using Matplotlib but reproducing similar features in Plotly necessitated further tweaks and occasionally resulted in visual disparities.

The goal of the coding phase's practical application was to create interactive graphical user interface (GUI) utilising Python packages such as Dash and Ipywidgets. Users could choose from a variety of plot styles, including pie charts, stacked bar charts, and bar plots, and save visualisations as JPEG pictures using the interface. They could also input CSV data. Making the data analysis accessible and easy to utilise.  
  
The Dash application was made to offer a smooth user experience. It had components for file uploads, a dynamic graph display area, and radio buttons for choosing the plot type. The Dash app was effective in enabling users to interactively visualise PM10 source contributions, despite the early rendering issues with Plotly. This allowed users to get insights into concentrations, asymmetry, dispersion, and explained variation.

To carry out this job successfully, several knowledge gaps had to be filled. It was crucial to comprehend the degrees of Plotly and Matplotlib to produce accurate visualisations. Furthermore, to guarantee a seamless user experience while developing dynamic web apps, prowess of Dash was required. To effectively interpret and visualise the data, the project also required a thorough grasp of PM10 source apportionment data and statistical analysis methodologies.  
  
In summary, the study offered significant understandings of PM10 source allocation and sophisticated data visualisation methods however, it also highlighted the necessity of ongoing education and adjustment to prevail technical obstacles and enhance the accuracy and usefulness of the visual aids.

## Limitations

Despite the best efforts, the initiative had its limitations. Compared to Matplotlib, Plotly's executing issues with bar plots showed how difficult it is to maintain perfect visual fidelity. Disparities in data treatment and plot design often led to visual disparities, such as mismatched axis scales and inconsistent marker styles. Furthermore, producing an understandable bar plot with confidence intervals and concentration data was difficult, even if interactive visualisations like the pie chart and stacked bar chart performed admirably. The intricacy of the widgets and layouts involved, blocking procedures, and restrictions in the local computer hardware might all be to blame for these problems. The challenge of creating an interactive and completely customisable graphical user interface exposed gaps in our current knowledge of complex Python visualisation packages.

## Use of Kaupapa Māori research Methodology

Incorporating the principles of Kaupapa Māori research methodology, our project was designed and executed with a focus on respecting Māori perspectives and values. This approach was manifested in several ways throughout our project. Firstly, we maintained open and continuous communication with the Auckland Council, ensuring that the data provided, and the findings derived were aligned with their expectations and needs. Weekly meetings and feedback sessions with both, the supervisor and the client from Auckland Council facilitated a collaborative environment, embodying the collective benefit and shared knowledge aspects of Kaupapa Māori. Moreover, we ensured that our data handling practices adhered to ethical standards, respecting the integrity and confidentiality of the data, thus aligning with the principles of guardianship and stewardship. The iterative refinement process, where feedback was actively sought and incorporated, highlights the respect for community input and collective decision-making, central to Kaupapa Māori methodology. This methodology not only guided our interactions but also shaped the ethical and respectful handling of environmental data, emphasizing sustainability and community well-being in addressing PM10 pollution sources.

## Conclusion with future directions

Our project effectively created an automated Python system for source apportionment data for PMF-generated PM10 pollution. By minimising human error and labour, this technology assures data analysis that is repeatable and scalable. We preserved data consistency and quality by incorporating sophisticated data preparation procedures, such as data cleansing, normalisation, and validation.  
  
The project made use of the robust Matplotlib and Plotly libraries for Python to provide interactive visualisations that offered insightful information on PM10 source contributions. Key information was effectively communicated via the visualisations, despite Plotly's difficulties in producing perfect visual accuracy. A user-friendly GUI built using Ipywidgets and Dash was another component of the system that improved accessibility for stakeholders.  
  
Important discoveries showed that Python could automate intricate processes while preserving the accuracy and dependability of human analysis. The system efficiently handled the issues associated with preprocessing and data handling by integrating numerous data sources.  
  
Our study has important applications in real life. The automated system offers a scalable and reproducible analytical method that improves air quality control and promotes evidence-based decisions. Non-technical people may interact with the data through the user-friendly interface and interactive charts, which helps with well-informed decision-making.  
  
Plotly's visualisation accuracy should be improved in the future to better match Matplotlib. While increasing the system to accommodate real-time data can yield immediate insights and predictive analytics, integrating powerful machine learning algorithms can offer deeper insights on PM10 source apportionment. Robustness may be enhanced by adding external data sources, such as meteorological data, and addressing issues with data consistency and quality. The system's applicability and efficacy will be validated by a wider deployment across other geographic locations, and future improvements will be driven by cooperation between environmental agencies and lawmakers.  
  
Our study shows how Python can be used to automate complicated data operations for PM10 source apportionment, making a unique contribution to data analytics and environmental science. The way in which the effect of technology on environmental management methods is highlighted is through the creative combination of interactive visualisations, a user-friendly GUI, and rigors data preparation. By promoting evidence-based policymaking, this method improves environmental sustainability and public health.

## References:

Boamponsem, L. K., De Freitas, C. R., & Williams, D. E. (2017). Source apportionment of air pollutants in the Greater Auckland Region of New Zealand using receptor models and elemental levels in the lichen, Parmotrema reticulatum. *Atmospheric Pollution Research*, *8*(1), 101–113. <https://doi.org/10.1016/j.apr.2016.07.012>

Wang, J., Hailin, G., An, L., Huang, C., Zhang, T., & Zhang, B. (2021). Modeling for the source apportionments of PM10 during sand and dust storms over East Asia in 2020. *Atmospheric Environment*, *267*, 118768. <https://doi.org/10.1016/j.atmosenv.2021.118768>

Farahmandkia, Z., Moattar, F., Zayeri, F., Sekhavatjou, M. S., & Mansouri, N. (2017). Contribution of point and small-scaled sources to the PM10 emission using positive matrix factorization model. *Journal of Environmental Health Science & Engineering*, *15*(1). https://doi.org/10.1186/s40201-016-0265-8

Manousakas, M. I., Furger, M., Daellenbach, K. R., Canonaco, F., Chen, G., Tobler, A., Rai, P., Lu, Q., Tremper, A. H., Green, D., Hueglin, C., Slowik, J. G., Haddad, I. E., & Prévôt, A. S. H. (2022). Source identification of the elemental fraction of particulate matter using size segregated, highly time-resolved data and an optimized source apportionment approach. *Atmospheric Environment. X*, *14*, 100165. <https://doi.org/10.1016/j.aeaoa.2022.100165>

## 

## Appendix 1: Capstone Project Checklist

##### Purpose

This checklist ensures that a selected workplace is adequate and appropriate for the purposes of the Capstone Project. The Checklist will be completed by the NZSE Capstone Project Leader in conjunction with the student and the Capstone Project Capstone Project Leader.

|  |  |  |
| --- | --- | --- |
| **Criteria** | **Yes/No** | **Initials** |
| Is the selected workplace suitable for the learning objectives of the Capstone Project? | Yes | Priyank |
| Have the student and Capstone Project Supervisors received a copy of the Capstone Project Handbook? | Yes | Priyank |
| Have the student and the Capstone Project Supervisors been fully briefed on the information in the Capstone Project Handbook and the learning and assessment requirements of the Capstone Project?   * roles and responsibilities of all parties. * requirements for liaison and contact with the NZSE Capstone Project Leader. * problem-solving process to be used in the event of any difficulty; and * learning and assessment requirements of the Capstone Project. | Yes | Priyank |
| Have the students and the Capstone Project Supervisors discussed and agreed on the project description and deliverables? | Yes | Priyank |
| Have the students and the Capstone Project Supervisors read and signed the Internship contract? | Yes | Priyank |
| Have the roles and responsibilities of all parties in the Capstone Project been explained to the student? | Yes | Priyank |
| Have the students read and signed the confidentiality agreement (**Appendix 3**)? | Yes | Priyank |
| Have the students been inducted on the Health and Safety policies and procedure of the organisation (if applicable)? | Yes | Priyank |
| Have the student and the Capstone Project Supervisors been advised of the problem-solving process to be used in the event of any difficulty? | Yes | Priyank |
| Have the requirements for liaison and contact with the New Zealand School of Education Capstone Project Leader been discussed and acknowledged by the  student? | Yes | Priyank |

##### Comments

**None**

**A close-up of a logo

Description automatically generatedMohammad Norouzifard 24/7/2024**

NZSE Capstone Project Leader Name & Signature Date

## Appendix 2: Capstone Project MOU

##### Student Details

Name Priyank Dangwal

Student ID 764707660 Cohort GDDA7123C

Mobile Number 0221003855

Email priyank07\_dangwal@yahoo.com

##### Industry Partner Details

Organization Name Auckland Council

Physical Address Victoria Street West Auckland 1142

Website

##### Capstone Project Supervisor/s

Name Dr Loius Boamponsem

Mobile Number

Email louis.boamponsem@aucklandcouncil.govt.nz

##### Capstone Project Supervisor/s

Name Dr Sara Zandi

Mobile Number

Email saraz@nzse.ac.nz

##### NZSE Capstone Project Leader

Name Dr Mohammad Norouzifard

Mobile Number

Email mohammad@nzse.ac.nz

##### Capstone Project Description

Capstone Project Dates from 16/04/2024 to 02/08/2024

Weekly hours of work 15

List below the description of the tasks, roles and responsibilities of the student undertaking the for the Capstone Project. Else, attach the description provided by the company. (These may be added to or amended during the period of the Capstone Project. If this is the case a new contract should be signed by all parties – see section 4 of this document under **Inadequate, inappropriate or insufficient work)**

##### Capstone Project tasks (milestones), roles and responsibilities:

Project 3: Automating PM10 source apportionment data analysis using python.

This project focuses on automating the analysis of PM10 source apportionment data using Python programming language. The project will utilize Python's rich ecosystem of libraries and tools to develop automated workflows for processing, analyzing, and visualizing PM10 source apportionment data that have been generated from the positive matrix factorization model. This will involve preprocessing raw data, performing statistical analyses, and generating reports or visualizations summarizing the results. Key components of the project will include Data Preprocessing: Developing scripts to preprocess raw PM10 source apportionment data, including data cleaning, formatting, and transformation tasks to ensure data consistency and quality. Automation: Designing automated workflows to streamline the analysis process, from data ingestion to result generation, reducing manual effort and increasing efficiency. Visualization and Reporting: Utilizing Python libraries such as Matplotlib, Seaborn, or Plotly to create informative visualizations, such as pie charts, bar plots, or stacked area plots, to illustrate the contributions of various emission sources. Additionally, generating automated reports summarizing the analysis results for stakeholders' review. By automating PM10 source apportionment data analysis with Python, the project aims to provide researchers with a powerful tool for understanding the sources of PM10 pollution and formulating evidence-based air quality management strategies. This automation will improve efficiency, reproducibility, and scalability of the analysis process, ultimately contributing to better-informed decision-making and improved air quality outcomes.

This is a joint agreement between the student, the organization and the NZSE.

##### The student agrees to:

* take responsibility for their own learning
* act ethically and responsibly at all times
* abide by organisational policy
* students read and signed the confidentiality agreement **Appendix 3**
* attend and participate in meetings with the Capstone Project Supervisors and/or NZSE Capstone Project Leader
* speak with their Capstone Project Supervisors should they have any concerns or if the Capstone Project is not meeting the learning needs of the student. If this is not an option, the student is to speak with their NZSE Capstone Project Leader; and
* take 30 minutes per day to spend time on reflective practice and completion of practice requirements.

##### The organisation agrees to:

* provide students with opportunities to meet the learning outcomes as specified in the course descriptor
* familiarise students with organisational policy, structure, accountability systems, codes of conduct and reporting systems at the commencement of the placement
* ensure the student has regular weekly professional supervision
* ensure the student is not left unsupervised or alone in a dangerous situation
* capstone Project Supervisors conduct mid-placement and closure interviews at the end of placement.
* communicate to the NZSE Capstone Project Leader, any sign of difficulty regarding the student’s performance, attendance or other aspects relevant to the success of their Capstone Project/ Capstone Project provide learning opportunities which acknowledge the student’s status as a beginning practitioner; and provide a healthy and safe workplace for the student.

##### NZSE agrees to:

* ensure the organisation is fully informed of everyone’s roles and responsibilities
* keep the NZSE Capstone Project Leader up to date with factors affecting Capstone Project arrangements meet (physically if applicable) with the Capstone Project Supervisors at least three times during the Capstone Project and use the checkpoint checklist in **Appendix 4**; and
* complete the final assessment of the student’s work and award a grade.

##### Student

Signature Priyank Dangwal Date 09/04/2024

##### Capstone Project Supervisors (Industrial supervisor)

Signature: A black line drawing of a person shaking hands

Description automatically generated Date: 10/5/2024

##### Capstone Project Supervisors

Signature   
 Date 15/05/2024

##### NZSE Capstone Project Leader

Signature **A close-up of a logo

Description automatically generated** Date 24/7/2024

This is a joint agreement between the student, the business organisation, and the New Zealand School of Education.

## Appendix 3: Confidentiality

As a student you must always be aware of the confidentiality of information gained during the course of your duties. It is expected that you understand the importance of treating information in a discreet and confidential manner and your attention is drawn to the following:

1. Written records and correspondence must be kept securely at all times, including when not in use
2. Business organisational documentation must not be submitted as appendices to the Final Assessment Report unless prior agreement from the organisation is received
3. Information regarding the business must not be disclosed either orally or in writing to unauthorised persons. It is particularly important that the authenticity of phone, e-mail and text enquirers should be checked
4. Conversation relating to confidential matters should not take place in situations where they may be heard by passers-by i.e., in corridors, reception areas etc.
5. The same confidentiality must also be preserved in dealing with matters relating to departmental personnel.

##### I have read and accept the terms of the above on confidentiality:

Signed: Priyank Dangwal Date: 15/04/2024

Printed Name: Priyank Dangwal ID#: 764707660

## Appendix 4: Checkpoint Checklist #\_\_

##### Purpose

This checklist ensures that the selected Capstone Project is progressing as intended for the student and business organization. The Checklist will be completed by the NZSE Capstone Project Leader in conjunction with the student and the Capstone Project Supervisors.

|  |  |  |
| --- | --- | --- |
| **Criteria** | **Yes/No** | **Initials** |
| Are the student and Capstone Project Supervisors free from any performance or other issues? If no, please comment below. | Yes | Priyank |
| Have the Capstone Project Supervisors and student developed work- based expectations/goals for the Capstone Project period? | Yes | Priyank |
| Does the student’s physical workplace remain suitable for the Capstone Project placement? | Yes | Priyank |
| Does the student have a clear understanding of expectations required to complete the Capstone Project? | Yes | Priyank |
| Has the student received input from their Capstone Project Supervisors on how to complete any required reports/assignments? | Yes | Priyank |
| Have you asked for feedback on how to improve the placement? Please comment below. | Yes | Priyank |
| Do you recommend continuing with the Capstone Project assignment? | Yes | Priyank |

##### Feedback from Student

##### Feedback from Capstone Project Supervisors

**Priyank Dangwal 31/07/2024**

Student Signature Date

Capstone Project Supervisor(s) Signature Date

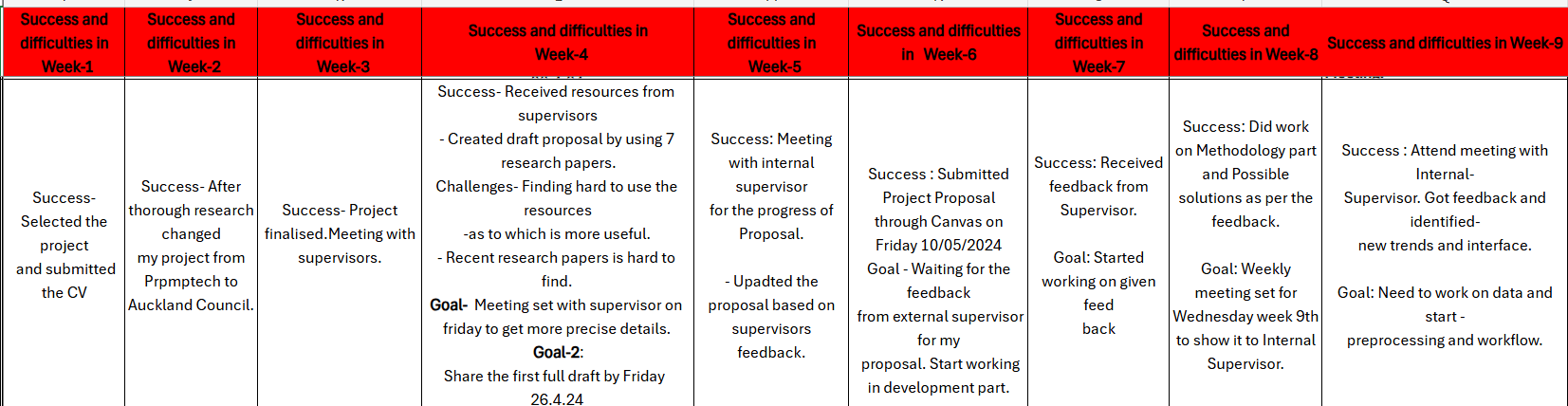
NZSE Capstone Project Leader Signature Date

## Appendix 5: Capstone Project Weekly Work Log

Between: Priyank Dangwal (Student Name)

And Louis K Boamponsem (Auckland Council)

Sara Zandi (NZSE) (Name of Company, Supervisors/Capstone Project Leader Details, Address of company, contact details of Supervisors) **Hours Log**

****

**A screenshot of a computer

Description automatically generated**

##### Total Hours of Capstone Project = 15-20

Student Signature: Priyank Dangwal Date: 31/07/2024

Capstone Project Supervisor(s) Signature: Date:

## Appendix 7: Student Capstone Project Evaluation

In order to provide quality and innovative work experiences, on-going evaluation is important. Your opinion, recommendations and comments help us monitor placements and provide feedback to the business organisation, Capstone Project Supervisors and NZSE Capstone Project Leader.

##### Cohort: GDDA7123C Capstone Project commencement date: 10/05/2024

##### Course: Capstone Project Business organisation: NZSE

##### NZSE Capstone Project Leader: Mohammad Norouzifard

##### Please rate the placement using the following criteria:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Organisation** | **Strongly Agree** | **Agree** | **Disagree** | **Strongly Disagree** | **Uncertain N/A** |
| Was expecting me. | YES |  |  |  |  |
| Orientation was well organised and included Health and Safety briefing. | YES |  |  |  |  |
| Provided guidelines around their expectations. | YES |  |  |  |  |
| Gave me the type of work specified in my project proposal (MOU). | YES |  |  |  |  |
| Provided access to resources (research resources, dataset, etc.) | YES |  |  |  |  |
| **Capstone Project Supervisors** | **Strongly Agree** | **Agree** | **Disagree** | **Strongly Disagree** | **Uncertain N/A** |
| Is willing to guide | YES |  |  |  |  |
| Provides opportunities to ask questions. | YES |  |  |  |  |
| Is actively involved in helping me reach the course objectives & LOs. | YES |  |  |  |  |
| Is friendly and approachable. | YES |  |  |  |  |
| Is sensitive to my needs. | YES |  |  |  |  |
| Has a thorough knowledge of their area of practice/expertise. | YES |  |  |  |  |
| **NZSE Capstone Project Leader** | **Strongly Agree** | **Agree** | **Disagree** | **Strongly Disagree** | **Uncertain N/A** |
| Provided guidance on the assessment requirements. | YES |  |  |  |  |
| Actively involved in helping me reach my objectives & Los in class. | YES |  |  |  |  |
| Facilitated learning experiences. | YES |  |  |  |  |
| Maintained a good relationship with the business organisation. | YES |  |  |  |  |
| Provided timely and constructive feedback. | YES |  |  |  |  |