**Academic Partner**

**Industry Partner**

A logo with text on it

Description automatically generated

|  |
| --- |
| Person planting a flower |
| Ecosoil Insights AKL  Interactive Dashboard for Auckland Soil Quality Monitoring |
| |  |  |  | | --- | --- | --- | | Presented by: Li Tao & Surangi Gunaratna |  | Capstone Project GDDA 7124C | |

A white text on a black background

Description automatically generated

**School of Tech**

**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:** | Li Tao  Surangi Gunaratna | **Student ID:** | 850001413  850000826 |
| **Assessment No & Type:** | Assessment 2, Research Project Report | **Cohort:** | GDDA7124 |
| **Due Date:** | 31/01/2025 | **Date**  **Submitted:** | 30/01/2025 |
| **Tutor’s Name:** | Dr. Sara Zandi | | |
| **Assessment**  **Weighting** | 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.

A black background with lines

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

**Student signature:**



**Date: 30/01/2025**

|  |  |  |  |
| --- | --- | --- | --- |
| **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** | |

Table of Contents

[Abstract 6](#_Toc188793223)

[Keywords 7](#_Toc188793224)

[Introduction 7](#_Toc188793225)

[Project Questions 8](#_Toc188793226)

[Project Design 9](#_Toc188793227)

[Data sources 9](#_Toc188793228)

[Data ingestion and storage 9](#_Toc188793229)

[Data processing and analytics 10](#_Toc188793230)

[Machine learning models 12](#_Toc188793231)

[AI components 13](#_Toc188793232)

[Output and visualization 14](#_Toc188793233)

[Feedback loop 15](#_Toc188793234)

[Integration points 16](#_Toc188793235)

[Security and Compliance 16](#_Toc188793236)

[Project development 17](#_Toc188793237)

[Data Cleaning Application development and setup 17](#_Toc188793238)

[Dashboard Code Development and Setup 23](#_Toc188793239)

[Discussion 32](#_Toc188793240)

[Conclusion and Future Directions 33](#_Toc188793241)

[References 35](#_Toc188793242)

[Appendix 1 – Python code for data cleaning 36](#_Toc188793243)

[Appendix 2: Capstone Project Checklist 42](#_Toc188793244)

[Appendix 3: Capstone Project MOU 43](#_Toc188793245)

[Appendix 4: Confidentiality 46](#_Toc188793246)

[Appendix 5: Checkpoint Checklist # 47](#_Toc188793247)

[Appendix 6: Capstone Project Weekly Work Log – Surangi Gunaratna 48](#_Toc188793248)

[Appendix 7: Capstone Project Weekly Work Log-Li Tao 49](#_Toc188793249)

# Table of Figures

[Figure 1- Comparison of KS Statistics between Miss Forest & Iterative Imputation 12](#_Toc188792965)

[Figure 2 -Comparison of p-values between Miss Forest and Iterative Imputer 13](#_Toc188792966)

[Figure 3- Filtering the land use by 'Dairy' and site number '8' 14](#_Toc188792967)

[Figure 4 - Filtering land use by 'Forestry' 14](#_Toc188792968)

[Figure 5 - Filtering the site number by '7' 15](#_Toc188792969)

[Figure 6 -User Login Page 16](#_Toc188792970)

[Figure 7 - Integrating streamlit and python 17](#_Toc188792971)

[Figure 8 - Initial user interface 17](#_Toc188792972)

[Figure 9 - Visualizing original dataset in data cleaning app 18](#_Toc188792973)

[Figure 10 Viewing raw data set information 18](#_Toc188792974)

[Figure 11 - Data type conversion and visualizing afterwards 18](#_Toc188792975)

[Figure 12- Validating critical columns 19](#_Toc188792976)

[Figure 13 - Visualizing duplicate value removal 19](#_Toc188792977)

[Figure 14 - Visualizing '<' value replacement 19](#_Toc188792978)

[Figure 15- Applying Iterative Imputation model 20](#_Toc188792979)

[Figure 16 - Visualizing validation checks 20](#_Toc188792980)

[Figure 17- Visualizing sample count extraction 21](#_Toc188792981)

[Figure 18- Dataset after period label assigning 21](#_Toc188792982)

[Figure 19- Dataset after retaining latest sample 21](#_Toc188792983)

[Figure 20- Visualizing dataset after calculating Contamination Index 22](#_Toc188792984)

[Figure 21- Final dataset validating information 22](#_Toc188792985)

[Figure 22 - Cleaned data set downloading option 22](#_Toc188792986)

[Figure 23- Python script for dashboard 23](#_Toc188792987)

[Figure 24 - Python script for dashboard 24](#_Toc188792988)

[Figure 25- Python script for dashboard 25](#_Toc188792989)

[Figure 26 - Python script for dashboard 25](#_Toc188792990)

[Figure 27- Requirement document 26](#_Toc188792991)

[Figure 28- GitHub streamlit integration 26](#_Toc188792992)

[Figure 29 - File upload interface 27](#_Toc188792993)

[Figure 30 - Dynamic Filtering 27](#_Toc188792994)

[Figure 31 - Soil Quality Indicators 28](#_Toc188792995)

[Figure 32- Interactive visualizations 28](#_Toc188792996)

[Figure 33- Trend Visualization 29](#_Toc188792997)

[Figure 34 - Contamination level 30](#_Toc188792998)

# Abstract

Project Eco Soil Insights focused on enhancing communication and providing dynamic insights using complex, periodic soil quality data gathered by Auckland council through developing an interactive, dynamic dashboard. Aim of the project is to develop a user-friendly dynamic dashboard using a dataset spanning multiple time periods which contains data related to site meta data, soil quality metrics and trace elements levels, offering actionable insights to different stakeholders. The final product is usable as a complementing tool for Auckland Council’s State of the Environment (SoE) report related to Soil quality insights in the Auckland Region.

The project encompassed two main phases as data preprocessing and dashboard building. Preprocessing and data cleaning was performed using python and advanced machine learning techniques such as iterative imputation were used to tackle missing values. The whole data cleaning and preprocessing steps were integrated to Streamlit application, enabling both technical and non-technical internal staff to interactively observe and understand data transformation journey and its impact on raw data. The cleaned data was integrated into another Streamlit application developed using python coding to create an interactive and dynamic dashboard to visualize trends, contamination levels and key soil quality indicators, creating a user-friendly interface for both technical and non-technical users.

The interactive dashboard promotes data driven decision making by allowing users to filter and view each visual based on different metrics, providing tailored insights to address diverse stakeholder needs. The Eco soil Insights dashboard serves as an efficient tool by addressing the timely need for interactive, dynamic and user-friendly analytic platform, to support the region's soil quality analysis by promoting informed data driven decision making for sustainable land management, compliance with environment health standards and mitigating land degradation risk. Simply this tool elevates traditional static soil analysis reports by transforming them to an interactive platform, thus amplifying Auckland Council’s effort to timely communicate soil quality data and support informed decision making for sustainable land use.

# Keywords

1. Soil Quality
2. Trace Elements
3. Contamination Index
4. Machine Learning Imputation
5. Environmental sustainability
6. Dashboard

# Introduction

Soil quality is fundamental to environmental sustainability, agricultural productivity, and community well-being. In Auckland, diverse land uses, including agricultural use, forestry and conservation, and urban and mixed-use, place significant pressure on soil health, requiring effective monitoring and management to prevent degradation. The Auckland Council’s Soil Quality Monitoring Program gathers extensive data on critical metrics, such as pH, Total Carbon (TC%), and trace elements like arsenic (As), and cadmium (Cd). However, the data is presented in a static report, which limits accessibility, engagement, and usability for diverse stakeholders, including policymakers, landowners, and environmental scientists.

This project addresses the challenges of transforming static soil reports into a dynamic and interactive web-based dashboard. The minimum viable product (MVP) aims to enhance decision-making for sustainable land management by integrating machine learning and advanced data visualisation techniques through Streamlit. Specifically, the dashboard will simplify complex datasets, identify contamination levels, and provide period insights, enabling stakeholders to make informed decisions. This initiative not only aligns with Auckland Council’s sustainability goals but also addresses the accessibility for both technical and non-technical users.

Key obstacles observed and solved in this project include handling incomplete dataset, ensuring compliance with privacy regulations, and designing an intuitive and accessible interface. Advanced imputation methods, such as Iterative Imputer, are employed to fill missing data accurately. Python and Streamlit form the backbone of the platform, enabling efficient data processing and visualization. The dashboard provides a comprehensive, scalable, and dynamic tool for monitoring soil quality and empowering stakeholders with actionable insights.

The structure of this report begins with the project’s scope and questions, followed by the detailed project design blueprint and framework for the solution. The development process, including implementation, testing, and evaluation, is then outlined. Finally, the report concludes with a discussion of the challenges encountered, limitations identified, and future directions for improving the MVP and its contribution to soil quality management.

# Project Questions

#### 1. How can transforming the existing static soil quality report into an interactive dashboard provide added value to diverse stakeholders and meet Auckland Council’s environmental sustainability goals?

The aim is to create a dynamic, user-friendly dashboard that enhances the experience for both technical and non-technical users by making complex soil quality data accessible and actionable. This will allow landowners, policymakers, and agricultural practitioners to easily interpret the soil health of the land, determine suitable agrarian activities, and make effective management actions. The design will ensure that key soil metrics, contamination trends, and visual alerts are easily identifiable, empowering users to make informed decisions about land use and take timely corrective actions to mitigate environmental risks. By transforming static soil reports into interactive insights, stakeholders can understand land suitability, optimize fertilizer use, and adapt land practices for sustainable development.

#### 2. What measures are required to ensure the accuracy, completeness, and security of soil quality data while meeting Auckland Council’s data privacy standards?

Given the presence of missing values and incomplete datasets, a significant part of this project involves determining appropriate imputation techniques, such as machine learning models, to enhance the completeness and accuracy of the soil data. Moreover, the project must address the challenges related to data privacy and security when sharing and presenting soil data, ensuring compliance with ethical standards, as well as maintaining long-term usability and trust through the dashboard.

# Project Design

The design phase of the Project *Eco Soil Insights* expanded from 14th October to 9th November, establishing a detailed blueprint for building dynamic and interactive soil quality dashboard, integrating number of data preprocessing, machine learning, and visualization tools. However numerous changes were adopted to initial design as the project ongoing, due to practical implementation issues arose.

## Data sources

The primary, raw data set for the project was provided by Auckland Council, in the format of a Excel file, which contained total of 24 columns and 490 rows. Columns encompassed data related to site metadata, soil quality metrics and Trace element concentration levels spanning over multiple periods from 1995 – 2023. Columns related to each core category are listed as below.

* **Site Metadata**: Site Number, Year, Soil Series, Soil Texture, Soil Type, Soil Order, NZ Soil Classification, Land use
* **Soil Quality Metrics**: PH, Total Carbon %, Total Nitrogen %, Olsen P, Ammonia, BD, MP5, MP10
* **Trace Element levels:** As, Cd, Cr, Cu, Ni, Pb, Zn

## Data ingestion and storage

The raw data set was received in the Excel format. The data ingestion process began with loading an Excel file to Jupiter’s note book and carrying out initial exploration and preprocessing using Python. A data cleaning script was developed using Python, and a Streamlit-based application was used to facilitate the data cleaning process.

The cleaned data set was stored in a cloud-based GitHub repository for easy accessibility among teammates and version control. Cleaned data was retrieved from GitHub and again used in another Streamlit application which was developed using Python script to create dynamic, interactive dashboard visualizations.

## Data processing and analytics

Data cleaning and preprocessing workflow was executed using following key steps.

#### 1. Data cleaning

* **Validating critical columns**

During literature review, it was identified that soil quality metrics (pH, TC %, TN %, Olsen P, AMN, BD) are essential indicators of soil health and recognized as foundational parameters in environmental and agricultural soil analysis. As the first step of data cleaning, data presence of critical columns was verified, and missing rows were removed to maintain data integrity.

* **Handling duplicate rows**
* **Handling Missing Values**

Remaining missing values in numerical columns were imputed using Machine learning based Iterative imputation model.

#### 2. Data Transformation

* **Assigning Period labels**

Auckland Council’s soil quality program was conducted not every year but periodically such as 1995-2000, 2008-2012, 2013-2017, 2018-2023. Based on the year column relevant period labels were assigned to derive meaningful visualizations when filtered based on period.

* **Retaining latest samples from each site from each period**

Raw data set contained multiple rows for each period from the same site as soil samples were tested repeatedly within the same period. However as per the guidelines received from industry supervisor (client) and mentioned in the SoE report, only the latest sample values for each unique site-period combination were retained to avoid duplication.

* **Handling values with “<” sign**

Trace elements for certain sites were below the detection limit of the used methods; therefore, they had been indicated as “less than minimal value” using “<” sign. To ensure numerical value presence of the column, in aligned with the industry standards and SoE Report, < values were replaced with half the detection limit (Curran-Cournane et al., 2020).

#### 3. Feature Engineering

In aligned with the Auckland Council soil related SoE report, contamination indexes for heavy metals (As, Cd, Cr, Cu, Ni, Pb, Zn) were calculated using the following formula.

**Contamination Index = Observed Value/ Native Mean value**

Native mean values were obtained using the threshold provided in SoE Report.

Then an Integrated Contamination Index was obtained for each site by calculating the mean contamination level and classified contamination levels based on below thresholds.

* Low Contamination (ICI ≤ 1)
* Moderate Contamination (1 < ICI ≤ 3)
* High Contamination (ICI > 3)

#### 4. Performing validation checks

Automated Python scripts with error handling were used to secure data integrity and validation. Furthermore, distribution comparison using histogram visualizations and Kolmogorov Simrnov Test validation was performed before and after imputation of missing values to ensure that imputation did not significantly deviate from the original distribution. KS Test statistics and p-values were documented and demonstrated to the user using a streamlet interface to showcase the accuracy and transparency of the process.

## Machine learning models

In this project Machine learning models were used to impute missing data. Initially Miss Forest Imputation method was used based on the literature review conducted based on reliable sources. Unlike simple methods like mean or median imputation, MissForest leveraged inter-feature relationships to provide more accurate and data-driven estimates, making it ideal for datasets where features are correlated.

However, MissForest Dependency created challenges when deploying the code in environments such as Streamlit application which is a a key framework used in this project. Further, not being a part of the core Python ecosystem, Miss Fores was less supported in terms of updates and maintenance.  
  
Alternatively, a similar but updated machine learning method called “Iterative Imputer” was used for imputation process due to its：

* **Technical Environment Compatibility**

Iterative Imputer integrates seamlessly into Python pipelines, especially those built with widely used machine-learning libraries like scikit-learn. This makes it easier to manage data preprocessing steps in a unified environment.

* **Consistency and Maintainability**

Iterative Imputer is actively maintained as part of scikit-learn library which ensures regular updates and support, in terms of reliability and long-term maintainability.

Performance comparison of both Miss Forest and Iterative Imputation was conducted using Kolmogorov-Smirnov (KS) test values and p-values and comparison drafts are as per below.

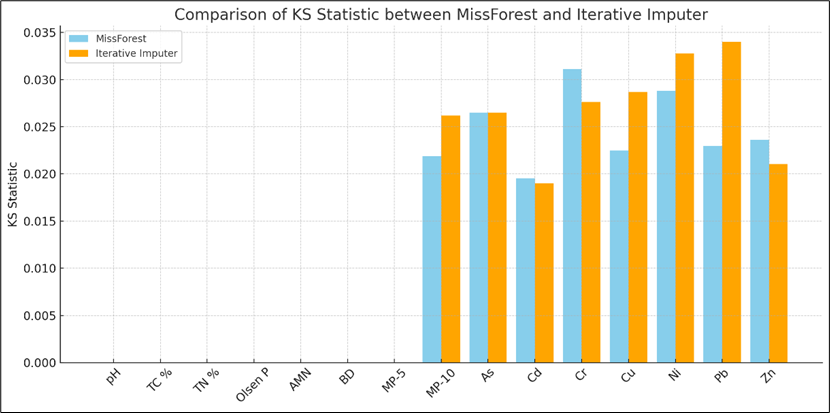


Figure 1- Comparison of KS Statistics between Miss Forest & Iterative Imputation

A graph with blue and orange lines

Description automatically generated

Figure 2 -Comparison of p-values between Miss Forest and Iterative Imputer

The results indicate that both methods produce similar imputation results. KS test values (statistic and p-value results) show minimal difference, indicating both have been capable of sufficiently preserving the original data distribution.

Iterative Imputer showed marginally higher KS Statistic for some variables like Ni and Pb while MissForest yielded slightly higher p-values for a few variables, but the difference is negligible. Therefore, it’s clear that Iterative imputation is suitable alternative Machine Learning model to impute missing data in this scenario.

## AI components

At this level, based on the project requirement, it was only Machine Learning models were integrated as an AI component. However, AI integrated components like Natural in Integrating AI to aid in analysis and enhance data-driven insights for soil quality monitoring.

## Output and visualization

The results of data analytics in the EcoSoil Insights Dashboard are visualized through an interactive and user-friendly interface powered by Streamlit. This platform allows stakeholders to explore soil quality metrics, contamination levels, and patterns via visualizations.

#### Example 1

A screenshot of a computer

Description automatically generated

Figure 3- Filtering the land use by 'Dairy' and site number '8'

#### Example 2

A screenshot of a computer

Description automatically generated

Figure 4 - Filtering land use by 'Forestry'

Example 3

A screenshot of a computer

Description automatically generated

Figure 5 - Filtering the site number by '7'

## Feedback loop

This design incorporates robust feedback loop to ensure evolving data needs and stakeholder requirements. Using python scripts for data pipelines enables versatile and amendable environment for continuous Improvement. It enables

* Continuous improvement with updated data points
* Integrating stakeholder feedback
* Adopting to industry standards and upgrades

## Integration points

The EcoSoil insights dashboard is designed to seamlessly interact with multiple systems and platforms, ensuring efficient data flow, accessibility, and scalability. The following integration points demonstrate how the solution interacts with other systems.

* **Github for version control and deployment**

The Python code and related files are hosted on a Github repository, providing centralized version control and collaboration. Github is integrated with Streamlit Cloud, ensuring that any updates in the repository are automatically deployed to the live dashboard.

* **Streamlit Cloud for hosting and scalability**

The dashboard is deployed on Streamlit Cloud, allowing stakeholders to access the application via a secure web link. The integration ensures high scalability, with the ability to handle increased user traffic without compromising performance.

* **Cloud-based for data storage**

Cleaned and processed datasets are stored in a cloud-based repository. This setup enables secure and centralized data access, ensuring that stakeholders can retrieve the most up-to-date information.

## Security and Compliance

Ensuring that data privacy, security protocols, and compliance standards are upheld throughout the data handling process, including storage, analysis, and visualization. All data was securely and ethically handled in adherence with Auckland Council’s data privacy guidelines. Both data cleaning app and dashboard adheres to privacy and security protocols by integrating authentication username and passwords to prevent unauthorized access.

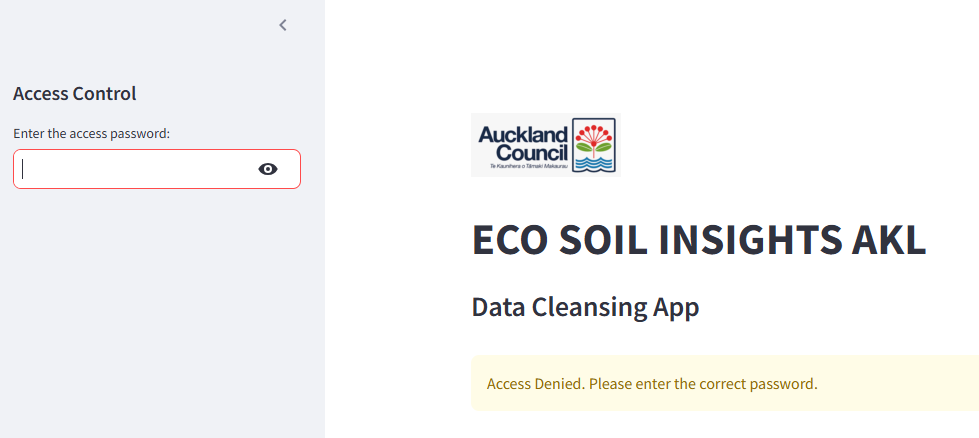


Figure 6 -User Login Page

# Project development

## Data Cleaning Application development and setup

Data cleaning application was developed using python-based scrip with integration of GitHub and streamlit application.

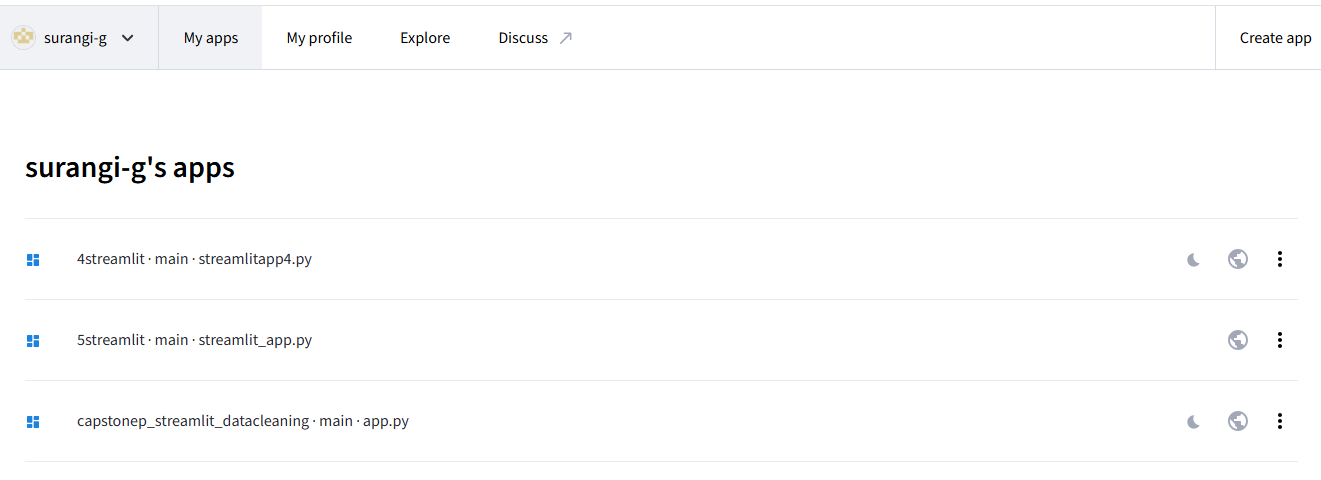


Figure 7 - Integrating streamlit and python

Each step undertaken was clearly visualized using before and after data set demonstration, along with message indication on which action performed and data validation actions. This ensures the transparency of whole data cleaning and preprocessing stage. Steps involved:

#### 1. Uploading dataset

Initial interface of the data cleaning app provides simple and clear guidelines including a welcoming message, short description of app’s purpose and instructions on how to upload the data; making it handle even by a non-technical user.

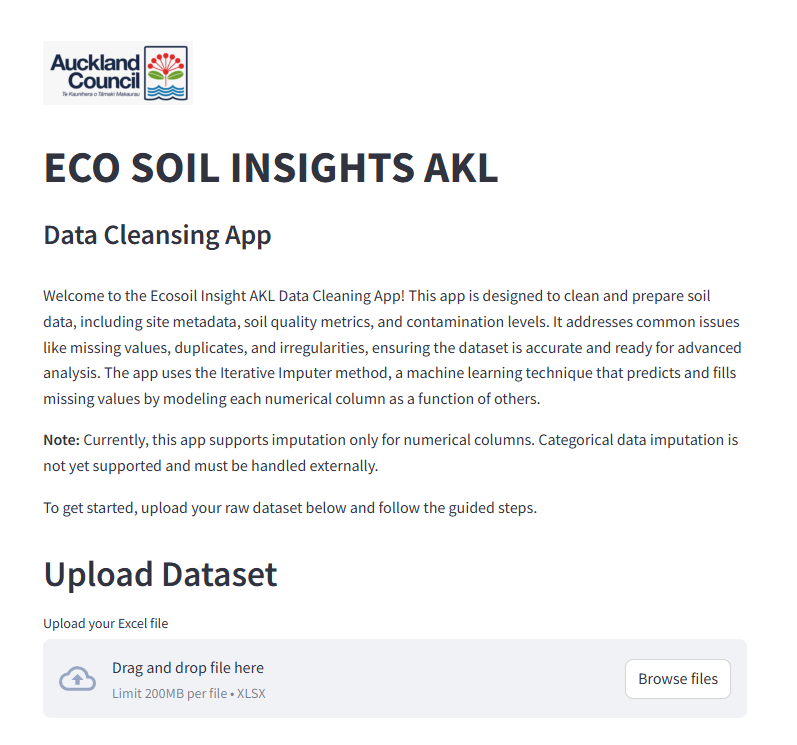


Figure 8 - Initial user interface

#### 2. Exploratory Data Analysis (EDA)

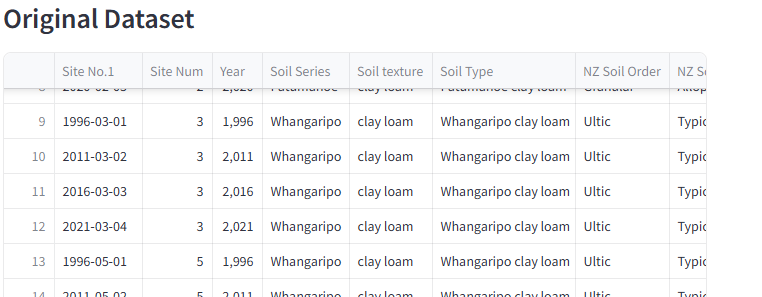


Figure 9 - Visualizing original dataset in data cleaning app

A screenshot of a data information

AI-generated content may be incorrect.

Figure 10 Viewing raw data set information

#### 3. Data preprocessing

* **Data type conversion and visualizing data set information afterwards**

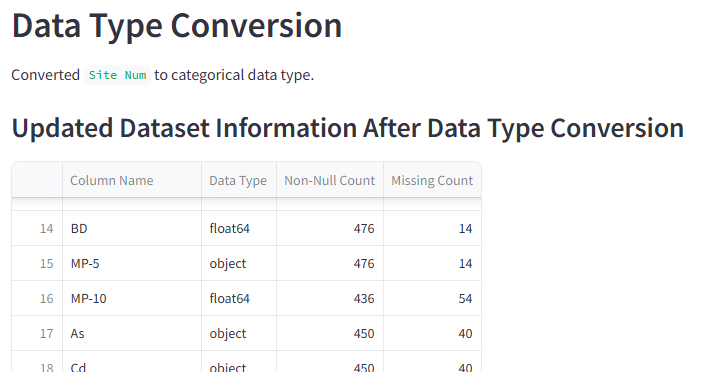


Figure 11 - Data type conversion and visualizing afterwards

* **Validating critical columns**

Ensuring data integrity by validating the presence of critical columns necessary for soil quality analysis. If any critical values are missing, the app generates an alert and then handle by dropping rows with missing values.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 12- Validating critical columns

* **Detecting and removing duplicate values**



Figure 13 - Visualizing duplicate value removal

* **Replacing values with "<" sign**

To ensure numerical value presence of the column, in aligned with the industry standards and SoE Report, < values (values below the detection limit) were replaced with half the detection limit.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 14 - Visualizing '<' value replacement

* **Imputing missing values using a machine learning algorithm**

Remaining missing values in numerical columns were imputed using Machine learning based Iterative imputation model.

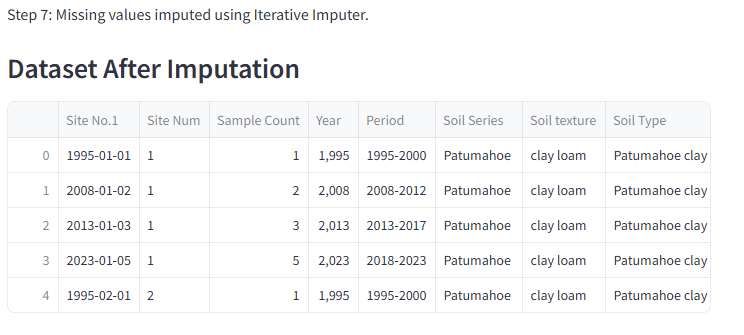


Figure 15- Applying Iterative Imputation model

* **Performing validation checks after imputation**

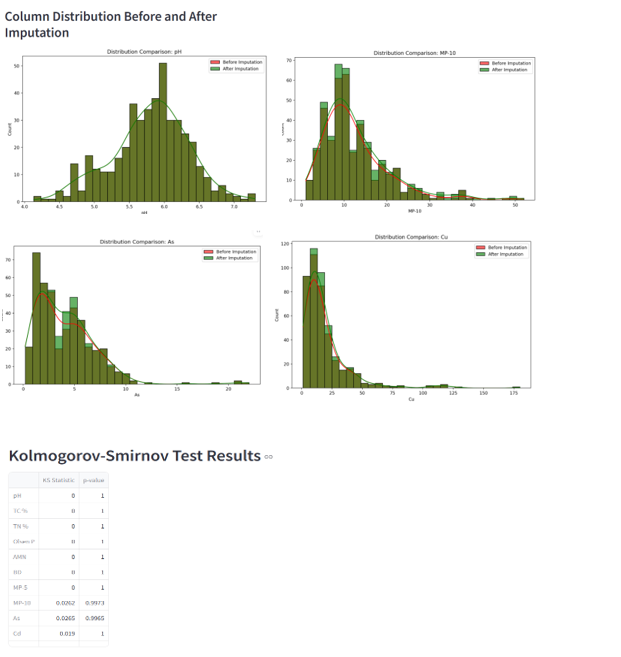


Figure 16 - Visualizing validation checks

#### 4. Data Transformation

* Sample count extraction using regex function

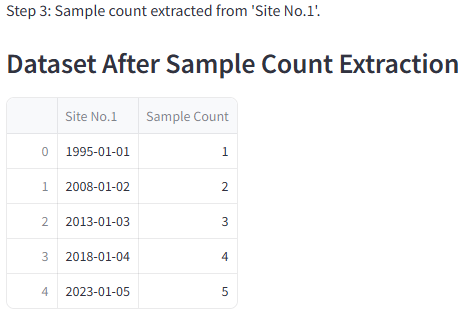


Figure 17- Visualizing sample count extraction

* Assigning period labels to categorize based on sample collection year

A screenshot of a computer

AI-generated content may be incorrect.

Figure 18- Dataset after period label assigning

* Retaining latest samples from each site from each period

A screenshot of a computer

AI-generated content may be incorrect.

Figure 19- Dataset after retaining latest sample

#### 5.Feature Engineering

* Calculating contamination index in aligned with Auckland Council soil related SoE report Formula.

A table with numbers and letters

AI-generated content may be incorrect.

Figure 20- Visualizing dataset after calculating Contamination Index

6. Validating final data set

#### A close-up of a white background AI-generated content may be incorrect.

Figure 21- Final dataset validating information

#### 7. Exporting final Output

A close-up of a white background

AI-generated content may be incorrect.

Figure 22 - Cleaned data set downloading option

**\*\*\*Python script used for data cleaning is attached as appendix and saved in GitHub Repository.**

## Dashboard Code Development and Setup

The EcoSoil insights dashboard was implemented using a combination of Python, Streamlit, and Github to ensure seamless development, deployment, and accessibility.

#### 1. Code development and repository setup

The Python code and a ‘requirements.txt’ file were uploaded to a dedicated GitHub repository, serving as the central hub for version control and collaboration.

The soil quality dashboard coding is shown below:

A screenshot of a computer program

Description automatically generated

Figure 23- Python script for dashboard

A screenshot of a computer code

Description automatically generated

Figure 24 - Python script for dashboard

A screenshot of a computer code

Description automatically generated

Figure 25- Python script for dashboard

A screenshot of a computer program

Description automatically generated

Figure 26 - Python script for dashboard

The requirements document is shown below:

|  |
| --- |
| A screenshot of a computer  Description automatically generated  Figure 27- Requirement document |

#### 2. Streamlit integration

The GitHub repository was linked to Streamlit Cloud, enabling the deployment of Python applications as an interactive web-based dashboard.

A screenshot of a computer

Description automatically generated

Figure 28- GitHub streamlit integration

#### 3. Testing and optimisation

Testing was conducted to ensure seamless integration and functionality.

The integration between GitHub and Streamlit was optimised to ensure that updates made in the repository were automatically reflected in the live dashboard.

#### 4. Maintenance and scalability

Hosting the dashboard on Streamlit Cloud ensures high scalability and availability, with GitHub facilitating continuous development and deployment.

#### 5. The Dashboard platform output

The EcoSoil insights dashboard is a dynamic platform designed to transform static soil quality reports into an interactive, user-friendly interface. It supports stakeholders in sustainable land management by visualizing complex soil quality data, contamination levels, and actionable insights. The dashboard utilizes Streamlit for dynamic and interactive visualizations, offering the following features:

* File upload and data integration:

Allows users to upload cleaned datasets in real-time; the cleaned dataset from the data preprocessing Streamlit app is directly utilized in the dashboard. This will supports both CSV and Excel formats, ensuring flexibility for different users.

A screenshot of a dashboard

Description automatically generated

Figure 29 - File upload interface

* Dynamic filtering:

Sidebar filters allow users to refine data views based on ‘Land use’, ‘Period’, and ‘Site number’. These filters provide the flexibility to customize analyses based on specific regions or timeframes.

A screenshot of a screen

Description automatically generated

Figure 30 - Dynamic Filtering

* Soil quality indicator:

Soil quality indicators offer users insights into understanding the soil health and overall ecosystem condition by filtering the sidebar. Indicator are dynamically updated based on selected filters, making it easy to explore variations across different land type or monitoring sites.

A chart of soil quality data

Description automatically generated

Figure 31 - Soil Quality Indicators

* Interactive visualizations:

Bar charts display the average value of soil quality metrics grouped by land use, giving a high-level overview of soil performance.

A screenshot of a graph

Description automatically generated

Figure 32- Interactive visualizations

* Trend analysis of selected metric

Trend visualisation tool the line graphs helps users evaluate trends in the selected soil metric over time, helping in identify patterns and long-term changes. This feature is particularly valuable for monitoring the effectiveness of land management practices.

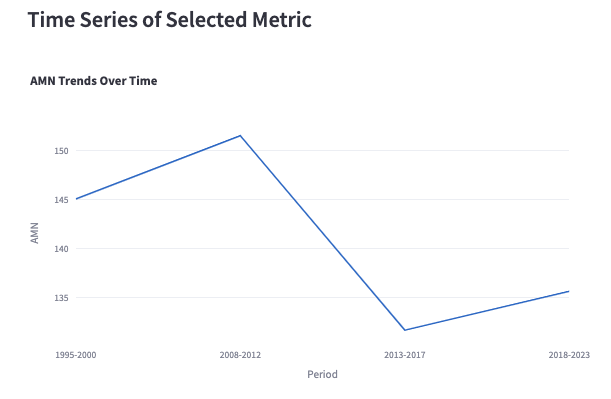


Figure 33- Trend Visualization

* Contamination analysis

Gauge graphs represent the integrated contamination index (ICI) with the threshold for low, moderate, and high contamination levels. This tool helps users quickly assess potential risks and prioritize remediation efforts.

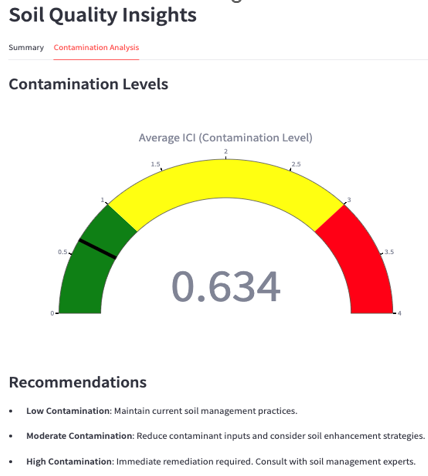


Figure 34 - Contamination level

* Downloadable data

Filtered datasets can be downloaded in CSV files, allowing stakeholders to utilise the insights offline, ensuring flexibility for reporting or further research purpose.

A screenshot of a computer

Description automatically generated

Figure 35- Downloading Filtered data

Summary of Pipeline Flow:

Data Ingestion: Raw data

Data Processing: Cleaning and exploration in Python with Streamlit library, Iterative imputer imputation.

Storage: Cleaned data to Cloud-Based SharePoint.

Integration: Cleaned data upload to Streamlit dashboard app

Dashboard: Visual insights accessible to stakeholders with the Streamlit app link.

Feedback Loop: Insights and Pipeline Improvement

Security Compliance: Throughout the process.

# 

# Discussion

**Challenges**

* Lack of domain knowledge and technical background related to soil science

Domain knowledge was a big challenge as the data set and provided SoE report contained lots of technical jargon related to soil science. Extensive literature review had to be done in order to get clear understanding of variable meanings, interrelationships and user cases.

* Missing or incomplete dataset

Handling missing values was a significant challenge. Initially, we considered using the traditional MissForest imputation method. However, this approach presented several issues, including compatibility problems with tools like Google Colab and Streamlit Cloud, and limitations in scalability for future use. After discussion with the stakeholder and tutor, we adopted Iterative Imputation, a more modern and widely accepted method. This approach integrates seamlessly with Streamlit and provides a strong foundation for future cases.

**Limitations**

* Lack of predictive analytics

While the dashboard effectively visualises trends and patterns, it does not currently include predictive analysis due to the nature of the data and its limitations. The dataset lacks the scale and richness needed to train a predictive model. Each monitoring site is unique, with significant differences in land use and periods. These variations make it difficult to generalise patterns or develop machine-learning models that apply across sites. For now, the dashboard focuses on time series analysis using line graphs to visualize trends over time. While this provides valuable insights into past, predictive analysis will require more extensive and diverse datasets in the future.

* Geographical data unavailability

Due to privacy considerations, the datasets provided did not include geographical information. The absence of location-based references limited our ability to integrate map-based visualisations into the dashboard. With geographical data, we could enrich the dashboard with features such as heat maps or site-specific soil analysis, offering greater visual context and actionable insights.

**Achievements**

* Transforming static reports

The dashboard successfully converts static sheets into an interactive, user-friendly platform, providing stakeholders with actionable insights for sustainable land management. Engaging with stakeholders throughout the development process ensured the dashboard addressed their real-world needs and making it a practical tool for industrial use.

* Scalability and Accessibility

The adoption of Iterative Imputation improved data processing capabilities, setting a foundation for future integrations and extensions. Deployment through Streamlit Cloud ensures the solution is highly accessible and scalable, accommodating a diverse range of users.

# Conclusion and Future Directions

The Eco soil Insights Dashboard project is a significant step forward taken from traditional soil quality reporting format practiced by Auckland Council over the period. Leveraging with python based Streamlit interface, this new dynamic, interactive and user-friendly platform not only compliment the traditional report-based communication but go beyond mere reporting by empowering stakeholders with data based analytical insights.

The project followed a structured workflow which includes literature review, project design, data gathering and exploration, data cleaning, project development and user testing and finetuning. Consistent communication with both industry and academic supervisors was carried out at each stage in order to integrate their valuable knowledge and feedback to craft a more refined product.

key achievements of project remarked at imputing missing data using a machine learning model which scored very high accuracy level, implementing transparent data cleaning workflow that both technical and nontechnical users can witness, developing contamination index by leveraging the same methods that SoE report presented and ultimately presenting a structured, user-friendly, dynamic dashboard.

However, we have identified several areas related to dashboard’s functionality and impact which could enhance future developments and directions.

**Future Directions**

#### 1. User based customized dashboard

Not every stakeholder has the same requirements or priorities when using the dashboard. For instance, landowners may be severely concerned about their land’s contamination or soil fertility level while policy makers may look for a bigger picture on regional trends. By enabling user profile based customized features will provide more tailored insights for diverse stakeholder needs.

#### 2. Predictive Analytics

This dashboard currently provides only descriptive analytics based on past data as there are not enough periodic data points to support reliable future predictions. If Council focus on expanding the dataset by collecting more periodic data, it’s possible to incorporate machine learning models and forecast future soil quality trends along with potential contamination risks to support more proactive interventions and more effective resource allocation.

#### 3. AI-Based Recommendations

With incorporating AI based machine learning algorithms that can train and analyze historic patterns, there’s a potential ability to provide customized recommendations for users by suggesting remediations such as optimal land management practices, crop selection and

#### 4. IoT-Enabled Real-Time Data Updates

In future if council place Integrating IoT devices, such as soil sensors to collect data, this model could facilitate real time data cleaning and ingestion to dashboard which provide up-to-date insights on changing soil conditions.

The current choice of Python based dashboard and analysis models provides robust foundation and versatile space for such future expansion and scalability, where the Eco Soil Insights could leverage remarkable experience for users and provide proactive, meaningful insights to improved soil quality monitoring and sustainable environmental practices in the future.

# References

Curran-Cournane, F., Research and Evaluation Unit (RIMU), & Auckland Council. (2020b). Differences in Soil Quality and Trace Elements Across Land Uses in Auckland and Changes in Soil Parameters from 1995-2017 (Peer Review Panel, E. McLaren, & M. Carbines, Eds.; Technical Report No. 2020/001). Auckland Council. <https://knowledgeauckland.org.nz/media/1815/tr2020-001-differences-in-soil-quality-and-trace-elements-auckland.pdf>

Differences in soil quality and trace elements across land uses in Auckland and changes in soil parameters from 1995-2017 - Knowledge Auckland. (n.d.-b). <https://knowledgeauckland.org.nz/publications/differences-in-soil-quality-and-trace-elements-across-land-uses-in-auckland-and-changes-in-soil-parameters-from-1995-2017/>

Guinto, D. F. (2022). Changes in soil quality under different land uses in the Manukau Harbour catchment area, 1995-2017. In Adaptive Strategies for Future Farming (Ed.), Occasional Report No. 34 (p. 11 pages). Farmed Landscapes Research Centre, Massey University. <https://www.massey.ac.nz/~flrc/workshops/22/Manuscripts/Guinto_Poster.pdf>

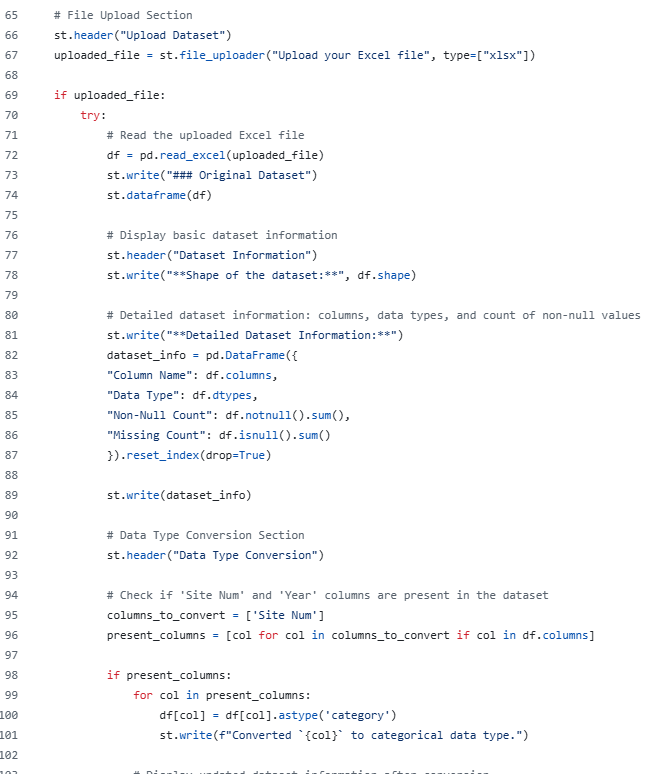
# Appendix 1 – Links to Project Portfolio

## GitHub Link

## e-portfolio Link

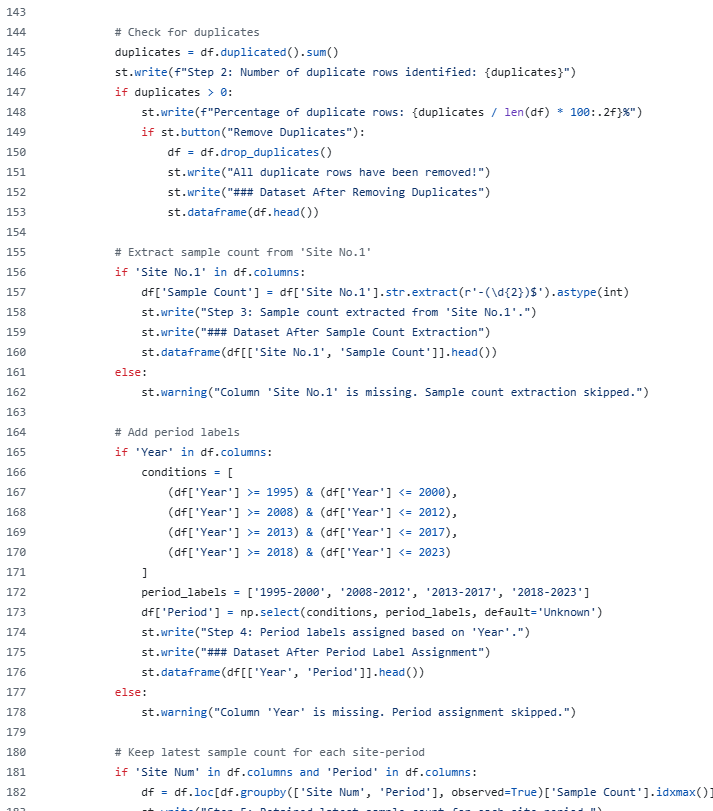
# 

# Appendix 2 – Python code for data cleaning



A screenshot of a computer code

AI-generated content may be incorrect.



A screenshot of a computer code

AI-generated content may be incorrect.

A screenshot of a computer code

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

# Appendix 3: 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 |  |
| Have the student and Capstone Project Supervisors received a copy of the Capstone Project Handbook? | Yes |  |
| 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 |  |
| Have the students and the Capstone Project Supervisors discussed and agreed on the project description and deliverables? | Yes |  |
| Have the students and the Capstone Project Supervisors read and signed the Internship contract? | Yes |  |
| Have the roles and responsibilities of all parties in the Capstone Project been explained to the student? | Yes |  |
| Have the students read and signed the confidentiality agreement (**Appendix 3**)? | Yes |  |
| Have the students been inducted on the Health and Safety policies and procedure of the organisation (if applicable)? | Yes |  |
| 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 |  |
| 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 |  |

##### Comments

NZSE Capstone Project Leader Name & Signature Date

# Appendix 4: Capstone Project MOU

**Student Details**

Name Li Tao Student ID 850001413 Cohort GDDA7124C Mobile Number 0221249729 Email [850001413@nzse.ac.nz](mailto:850001413@nzse.ac.nz)

Name Surangi Gunaratna Student ID 850000826 Cohort GDDA7124C Mobile Number 02102680952 Email [850000826@nzse.ac.nz](mailto:850001413@nzse.ac.nz)

**Industry Partner Details**

Organisation Name Auckland Council Physical Address: Level 16, 135 Albert Street, Auckland 1010 Website: https://www.aucklandcouncil.govt.nz/

**Capstone Project Supervisor/s**

Name Dani Guinto Mobile Number: 027 283 4089 Email [dani.guinto@aucklandcouncil.govt.nz](mailto:saraz@nzse.ac.nz)

**NZSE Capstone Project Leader**

Name Sara Zandi Mobile Number Email saraz@nzse.ac.nz

**Capstone Project Description**

Capstone Project Dates from 16/09/2024 to 10/02/2025 Weekly hours of work 10 hours by each team member

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:**

**Roles**

**Data Analyst: Li Tao**

**Data Analyst: Surangi Gunaratna**

**Phase 1: Project Scoping and Proposal Development**

**Key Milestone** – Project proposal submission

**Phase 2: Data Collection, Cleaning and Analysis**

**Key Milestone** - Cleaned, processed and complete data set

**Phase 3: Dashboard Design and Development**

**Key Milestone** - Initial dashboard with visualizations and interactive filters

**Phase 4: Final Review, Delivery and Presentation**

**Key Milestone** - Final interactive, QA tested dashboard

**Key Milestone** - Final report and presentation

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

**The student agrees to:**

* take responsibility for their own learning
* act ethically and responsibly at all times
* abide by organizational 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 organization agrees to:**

* provide students with opportunities to meet the learning outcomes as specified in the course descriptor
* familiarize students with organizational 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:**

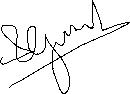
* ensure the organisation is fully informed of everyone’s roles and responsibilities
* keep the NZSE Capstone Project Leader up to date with factors affecting Capstone Project arrangements meet (physically if applicable) with the Capstone Project Supervisors at least three times during the Capstone Project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

**A black background with lines

AI-generated content may be incorrect.Student 1**

Signature Date 7/10/2024

**Student 2**



Signature Date 7/10/2024

**A signature on a white background

Description automatically generatedCapstone Project Supervisors**

Signature Date 21/10/2024 25/10/2024

**NZSE Capstone Project Leader**

Signature Date 25/10/2024

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

# Appendix 5: 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 organizational 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.

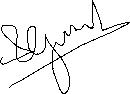
**I have read and accept the terms of the above on confidentiality:**

A black background with lines

Description automatically generated

Signed: Date: 7/10/2024

Printed Name: Li Tao ID#: 850001413



Signed: Date: 7/10/2024

Printed Name: Surangi Gunaratna ID#: 850000826

# Appendix 6: 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. |  |  |
| Have the Capstone Project Supervisors and student developed work- based expectations/goals for the Capstone Project period? |  |  |
| Does the student’s physical workplace remain suitable for the Capstone Project placement? |  |  |
| Does the student have a clear understanding of expectations required to complete the Capstone Project? |  |  |
| Has the student received input from their Capstone Project Supervisors on how to complete any required reports/assignments? |  |  |
| Have you asked for feedback on how to improve the placement? Please comment below. |  |  |
| Do you recommend continuing with the Capstone Project assignment? |  |  |

**Feedback from Student**

**Feedback from Capstone Project Supervisors**

A black background with lines

AI-generated content may be incorrect.

**7/11/2024**



Student Signature Date

Capstone Project Supervisor(s) Signature Date

NZSE Capstone Project Leader Signature Date

# Appendix 7: Capstone Project Weekly Work Log – Surangi Gunaratna

Between: Surangi Gunaratna (Student Name)

And NZSE / Auckland Council

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

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Hours worked** | **Work done (itemized)** | **Student Initials** |
| 14/09/2024 | 2 | Clarifying business need and defining project scope |  |
| 20/09/2024 | 4 | Literature review and exploring sample data set |  |
| 25/09/2024 | 5 | Literature review and exploring sample data set |  |
| 05/10/2024 | 3 | Developing project plan/Gantt chart |  |
| 11/10/2024 | 4 | Drafting project proposal |  |
| 20/10/2024 | 3 | Reviewing and discussing project proposal |  |
| 02/11/2024 | 5 | Report writing and final review |  |
| 05/11/2024 | 4 | Report writing and creating project proposal presentation |  |
| 11/11/2024 | 3 | EDA and preliminary data analysis with sample data |  |
| 16/11/2024 | 5 | Further literature review on data cleaning processes used in environmental analysis |  |
| 20/11/2024 | 2 | Data cleaning |  |
| 25/11/2024 | 4 | Implementing ML model for missing data imputation |  |
| 28/11/2024 | 4 | Developing data cleaning app |  |
| 01/12/2024 | 4 | Developing data cleaning app |  |
| 02/12/2024 | 1 | Testing data cleaning app |  |
| 05/12/2024 | 4 | Documenting data cleaning process |  |
| 06/12/2024 | 3 | Trying dashboard development with different applications Tableau/PowerBI |  |
| 12/12/2024 | 1 | Conducting QA Testing for data cleaning app & dashboard |  |
| 03/01/2025 | 3 | Summarizing project findings |  |
| 08/01/2025 | 2 | Exploring future directions and discussion on final report |  |
| 15/01/2025-  27/01/2025 | 8 | Finalizing project report |  |
| 28//1/2025 | 3 | Preparing power point presentation and proofreading final report |  |

##### Total Hours of Capstone Project = 76 hours

Student Signature: Date: 29/01/2025



Capstone Project Supervisor(s) Signature: Date:

# Appendix 8: Capstone Project Weekly Work Log-Li Tao

Between: Li Tao (Student Name)

And NZSE / Auckland Council

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

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Hours worked** | **Work done (itemized)** | **Student Initials** |
| 15/09/2024 | 3 | Understand the business needs and the project process |  |
| 21/09/2024 | 3 | Review the dataset and research for industry information |  |
| 26/09/2024 | 4 | Literature review and terminology familiarity |  |
| 03/10/2024 | 3 | Find methods for data cleaning and transformations |  |
| 10/10/2024 | 4 | Handle missing value with initial machine learning method |  |
| 18/10/2024 | 5 | Research dashboard design and use Tableau for visualization |  |
| 01/11/2024 | 5 | Work on the proposal report |  |
| 07/11/2024 | 2 | Proofread and submit the proposal report |  |
| 12/11/2024 | 3 | Design dashboard layout with using different methods such as Tableau, PowerBI, and Streamlit |  |
| 17/11/2024 | 3 | Develop visualizations for dashboard by using Python |  |
| 22/11/2024 | 2 | Push Python code to Github and integrate with Streamlit cloud to build the app |  |
| 24/11/2024 | 2 | Improve the visual design of dashboard visuals |  |
| 26/11/2024 | 3 | Draft documentation for project pipeline and designation |  |
| 03/12/2024 | 3 | Fill out report documentation with project questions and scope |  |
| 06/12/2024 | 3 | Fine tune dashboard with adjustments based on team meeting |  |
| 11/12/2024 | 2 | Review dashboard to ensure requirements are met |  |
| 14/12/2024 | 3 | Finalize dashboard design and documentation |  |
| 17/12/2024 | 4 | Conduct quality check for accuracy and performance |  |
| 07/01/2025 | 5 | Work on the project report |  |
| 11/01/2025 | 4 | Summarize the challenges, limitations, and achievements for the final report |  |
| 18/01/2025 | 2 | Prepare the presentation slides and summarize findings |  |
| 26/1/2025 | 3 | Work on the project presentation |  |

##### Total Hours of Capstone Project = 71 hours

Student Signature: Date: 29/01/2025



Capstone Project Supervisor(s) Signature: Date: