

**ANL312 July 2023 Final Report**

|  |  |
| --- | --- |
| **Course Code** | ANL488 |
| **Project Title** | Prediction of potential CO2 storage locations |
| **SUSS PI No** | N1981115 |
| **Name** | Low Kai Boon Calvin |
| **Submission Date** | 27 October 2022 |

**Table of Contents**

[1. Introduction 2](#_Toc144671967)

[2. Literature Review 3](#_Toc144671968)

[3. Problem and Data Understanding 5](#_Toc144671969)

[3.1. Problem Understanding 5](#_Toc144671970)

[3.2. Data Understanding 5](#_Toc144671971)

[3.2.1. The CSS Dataset 6](#_Toc144671972)

[3.2.2. The CSRC Dataset 6](#_Toc144671973)

[4. Data Preparation & Cleaning 7](#_Toc144671974)

[4.1. Cleaning the CCS Dataset 7](#_Toc144671975)

[4.2. The CSRC Database 10](#_Toc144671976)

[4.3. Combining Datasets & Final Cleaning 11](#_Toc144671977)

[5. Proposed Modelling Techniques & Evaluation 15](#_Toc144671978)

[6. Timeline of Project And Proposed Schedule 16](#_Toc144671979)

[7. References 17](#_Toc144671980)

# Introduction

Every year, global temperatures are increasing at a rapid rate due to climate change, which is caused by the greenhouse and the El Nino effect. One method that can mitigate the effects caused by climate change is using Carbon Capture and Storage (CCS) technologies, where carbon dioxide is captured and stored during the pre-combustion or post-combustion stage of an industrial process, thus preventing the carbon dioxide from entering the atmosphere.

However, despite CCS being a promising technology that aims to stop the increase of greenhouse gases in the atmosphere, the logistics and costs with constructing and using new CCS facilities are often complex and expensive. Thus, there are not many CCS facilities existing around the world. Therefore, we will predict the feasibility of potential CCS storage sites by utilising information obtained from CCS databases to determine how the various factors will affect the feasibility of various CCS projects.

To aid in our study, we are given two datasets. The first dataset contains information about the existing and future CCS facilities around the world, whereas the second dataset contains information regarding geological features around the world that can be used to store carbon dioxide. Based on the non-categorical nature of the two datasets, we will only be using decision trees and Artificial Neural Network (ANN) to construct our predictive model. After our predictive model has been constructed, we will be using a confusion matrix to evaluate the performance of our predictive model.

# Literature Review

According to the World Meteorological Organisation (2023), global temperatures from the year 2023 to 2027 is predicted to have an estimated increase of 1.1℃ to 1.8℃ higher than the average temperature increase recorded from the year 1850 to 1900. This radical increase in temperatures is caused by the greenhouse effect and El Nino Effect. The greenhouse effect is caused by the trapping of heat by greenhouse gases, including carbon dioxide, whereas the El Nino Effect is caused by warmer waters releasing heat into the atmosphere, thus contributing to the increase of temperatures around the globe.

With global temperatures increasing at a drastic rate, there will be adverse effects felt in population centres around the world. According to Haines & Hatz (2004), the adverse effects of climate change include thermal stress, increased occurrence of floods, droughts, malnutrition, and contracting of diseases. As the climate becomes warmer, heat waves will become more prominent, resulting in wildfires, causing air pollution and causing more heat-related injuries or deaths. A warmer climate also means that weather patterns will become more erratic, resulting in an increased number of floods and droughts, destroying crops, and causing malnutrition. Lastly, a warmer climate can cause disease-carrying carriers to thrive in tropical areas, leading to an increased number of people contracting animal-borne illnesses.

One way to mitigate the effects of global warming is to reduce the amount of greenhouse gases in the atmosphere, such as carbon dioxide. Such technologies are known as Carbon Capture and Storage (CCS), where carbon dioxide is captured during industrial processes, such as the burning of fossil fuels. Using various CCS technologies, the captured carbon dioxide gas is then stored a storage facility or pressurised and injected into a stable geological feature (Boot-Handford et al, 2014). This will trap carbon dioxide for millennia and prevent it from being released into the atmosphere.

According to Boot-Handford et al (2014), there are several methods in which CCS can be done. These methods include but are not limited to solvent scrubbing and oxyfuel combustion. Solvent scrubbing is the process of letting the carbon dioxide react in a chemical reaction with a solvent that creates a by-product that contains carbon and oxygen particles, the by-product can then be stored in a storage facility. Other CCS methods, such as oxyfuel combustion, involve the use of a medium to reduce the flue gas generated from the combustion process to just containing carbon dioxide and water, making it easier to capture the carbon dioxide before it reaches the atmosphere.

There are several advantages of using CCS (Bui et al, 2018). One advantage that CCS provides is the easy integration with existing industrial systems as CCS can easily retrofitted without causing much change to existing systems. Another advantage that CCS provides is the decarbonisation of emission-intensive industries. As such industries are notorious for producing carbon dioxide emissions, using CCS to capture these emissions will lower the industries’ carbon footprint. Lastly, CCS can be used in conjunction with carbon-neutral bioenergy (BECCS) to produce negative emissions as CCS will be able to remove as much carbon dioxide from the atmosphere as the amount of carbon dioxide produced by BECCS.

However, according to Haszeldine (2009), the main disadvantage is that CCS will be expensive to commercialise at the initial stages as CCS requires 25% to 40% of a power plant’s energy needs. Despite the costs needed to commercial CCS, it is useful to note that towards the path of commercialisation, CCS requires larger equipment to be constructed for prototyping or demonstration purposes. Hence, with the lessons learned from the CCS prototypes or demonstrations, it can then be used to improve the reliability of these CCS technologies and bring the costs down. Therefore, this will make CCS a viable option for reducing carbon dioxide emissions or towards the path of decarbonisation.

# Problem and Data Understanding

## Problem Understanding

Because of climate change, global temperatures are gradually increasing year on year, with the World Meteorological Organisation predicting in 2023 that there is a 66% chance that at least a year between 2023 and 2027 will have temperatures that surpass 1.5 above the average temperatures recorded during the pre-industrial era (World Meteorological Organisation, 2023). Furthermore, the World Meteorological Organisation (2023) also predicts a 98% chance that at least a year between 2023 and 2027, the temperature will surpass the highest temperature recorded in 2016. The World Meteorological Organisation’s statistics show that global temperatures will increase uncontrollably over the next few years without any mitigation measures.

Out of the several mitigation measures proposed to lessen the effects of climate change, one viable mitigation measure that can be undertaken will be using CCS technologies to capture and store carbon dioxide before it reaches the atmosphere. However, as constructing and using CCS facilities is often expensive and several factors need to be considered, including the project cost and the design capture rate of carbon dioxide, we are going to predict the availability of potential CCS locations in this case study.

## Data Understanding

For the datasets, we are given two datasets to work with, the CCS Dataset and CSRC Dataset. The CCS Dataset contains information about man-made facilitates that deals with capture or storage of carbon dioxide or both. On the other hand, the CSRC Dataset is done by the CO2 Storage and Resource Catalogue (CSRC) to record natural-occurring geological features that can be used to store carbon dioxide for millennia.

### The CSS Dataset

An overview of what the dataset contains shown in Table 3.2.1.

|  |  |  |
| --- | --- | --- |
| **S/N** | **Header Name** | **Description** |
| 1 | Project Name | The name of the project |
| 2 | Company | The company who oversees the project |
| 3 | Plant Name | The name of the CCS plant |
| 4 | Storge and/or Capture | Whether the CCS plant in involved with Capture, Storage or Capture and Storage |
| 5 | Overall Status | Whether the project is active, completed, hold, potential or terminated |
| 6 | Plant Status | Whether the plant is cancelled, decommissioned, existing, in development or planned |
| 7 | Country Location | The country in which the CCS plant is located. |
| 8 | State Location | The country’s state in which the CCS plant is located. |
| 9 | Specific Site Location | The location in which the CCS plant is located. |
| 10 | Plant Sie or Capture Amount | The size of the CCS plant or the capture amount intended for the CCS plant. |
| 11 | Combustion/Separation | The types of combustion/separation technologies used by the CCS plant |
| 12 | Capture Technology | The types of CCS technologies used by the plant |
| 13 | Amount of C02 Captured/Stored | The amount of carbon dioxide stored by the plant |
| 14 | Captured/Stored Unit | The measurements used to determine the volume of carbon dioxide stored |
| 15 | Project Summary | A short description of the project |
| 16 | Project Start Date | The starting date of the project |
| 17 | Project Cost | The cost of the project |
| 18 | Currency | The currency which the company pays for the project |
| 19 | Project Information Webpage | The hyperlink that directs to the information about the project |

*Table 3.2.1. CCS Dataset Fields and Descriptions*

### The CSRC Dataset

The CO2 Storage and Resource Catalogue (CSRC) is a multi-year project held over six annual cycles commissioned by the Oil and Gas Climate Initiative (OGCI) and is spearheaded by Global CCS Institute (GCCSI), starting with cycle 1 in 2020 and ending with cycle 6 in 2025. The OGCI aims to obtain a global perspective of the commercial readiness in using CCS.

The CSRC database is thus set up to record information regarding the CO2 storage sites around the world, including saline aquifers and oil fields, after evaluation using the Storage Resource Management System (SRMS) to determine the storage site’s maturity level. Furthermore, the CSRC database contains information including the storage site’s location and its storage capabilities. The CSRC Database is updated yearly, with newly discovered CO2 storage sites added in during each cycle. However, as there are numerous columns contained in the dataset, the overview of what the dataset contains shown in Table 3.2.2 will only the show the information that .

|  |  |  |
| --- | --- | --- |
| **S/N** | **Header Name** | **Description** |
| 1 | Type | Whether the geological feature is an oil field (petroleum) or a saline aquifer (saline). |
| 2 | Discovery Status | Whether there is enough information or regulations to justify the geological feature’s discovery. |
| 3 | Aggregated SR (Mt) - Mid | The total average amount of carbon dioxide the geological feature can store. |

# Data Preparation & Cleaning

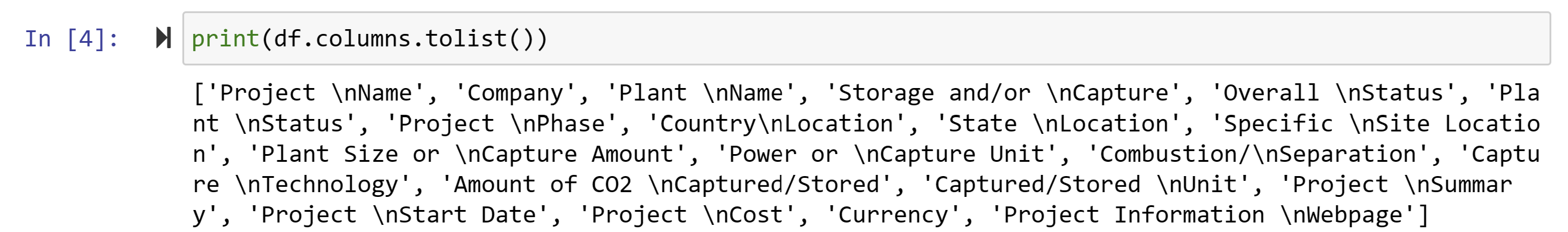
## Cleaning the CCS Dataset

Before we need start our data preparation process, we will need to read the CSS dataset into a Pandas Dataframes, as shown in figure 3.3.1.



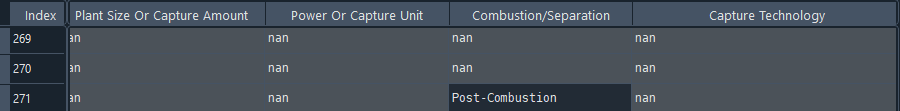
*Figure 3.3.1. Reading CCS Data into Pandas Dataframe*

According to Figures 3.3.2 and 3.3.3, there are a few issues which prevents the Pandas library in python to read the headers, therefore we will use Python to clean it up.



*Figure 3.3.2. Pandas Dataframe Columns Headers*

As shown in figure 3.3.2, one issue facing the dataset is the naming of the column header as the column headers included line breaks, as evident in the “\n” newline characters. The second issue the dataset faces are missing values, as shown in figure 3.3.3 below. The missing values will need to be removed as the null values will affect the accuracy of the predictive models’ performance. Thus, the code shown in figure 3.3.4 is aimed at resolving the issue.



*Figure 3.3.3. Missing values in Pandas Database*



*Figure 3.3.4. Codes to clean the “CCS-Database.xlsx”*

According to Figure 3.3.4, we rename each of the column headers in a for loop by capitalising each word and replacing the line break character “\n” with an empty string. As the dataset contains additional project information that is redundant to the prediction column, we use the drop function to remove these columns. We also use the dropna function to drop any missing values in the “Plant Size Or Capture Amount” and “Project Cost” columns by using the subset parameter to specify these columns we want to drop the missing columns, setting the how parameter to ‘any’ to delete the columns where the missing values are encountered and set the inplace parameter to ‘True’ to commit the changes to the existing dataframe.

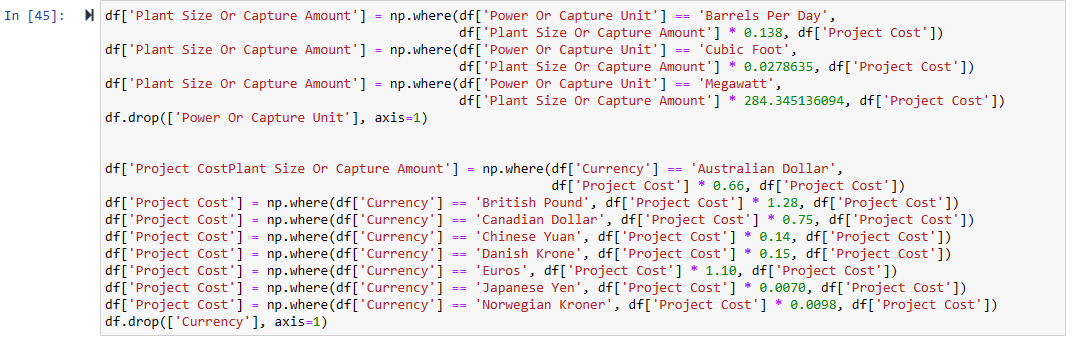
According to the measurements shown in the CCS Dataframe, there are different units of measurement used in the collection of the data, including the capture unit and the currency. Thus, we will need to convert all amounts to a common unit, metric ton for capture amounts and United States Dollar (USD) for currency, with the conversion rates as shown in Table 3.3.1 and Table 3.3.2 respectively.

|  |  |
| --- | --- |
| **Original Capture Unit** | **Converted Capture Unit In Metric Ton** |
| Barrels Per Day | 0.138 Metric Ton |
| Cubic Foot | 0.0278635 Metric Ton |
| Megawatt | 284.345136094 Metric Ton |

*Table 3.3.1. Conversion to Metric Ton*

|  |  |
| --- | --- |
| **Original Currency** | **US Currency** |
| Australian Dollar | 0.66 USD |
| British Pound | 1.28 USD |
| Canadian Dollar | 0.75 USD |
| Chinese Yuan | 0.14 USD |
| Danish Krone | 0.15 USD |
| Euro | 1.10 USD |
| Japanese Yen | 0.0070 USD |
| Norwegian Kroner | 0.0098 USD |

*Table 3.3.2. Conversion to United States Dollar (USD)*



*Figure 3.3.5. Codes to convert measurements to a common unit*

As shown in Figure 3.3.5, we use the np.where function to aid in the conversion of the measurements. The np.where function works similarly to an if-else statement, where the condition is specified in the first parameter, the value we want to change if the condition is met is specified in the second parameter and the value we want to change if the condition is not met is specified in the third parameter. In this case, we want to keep the original value when the condition is not met.

After we are done with the conversion, we can drop the columns that contain the capture units or currency from the dataframe.

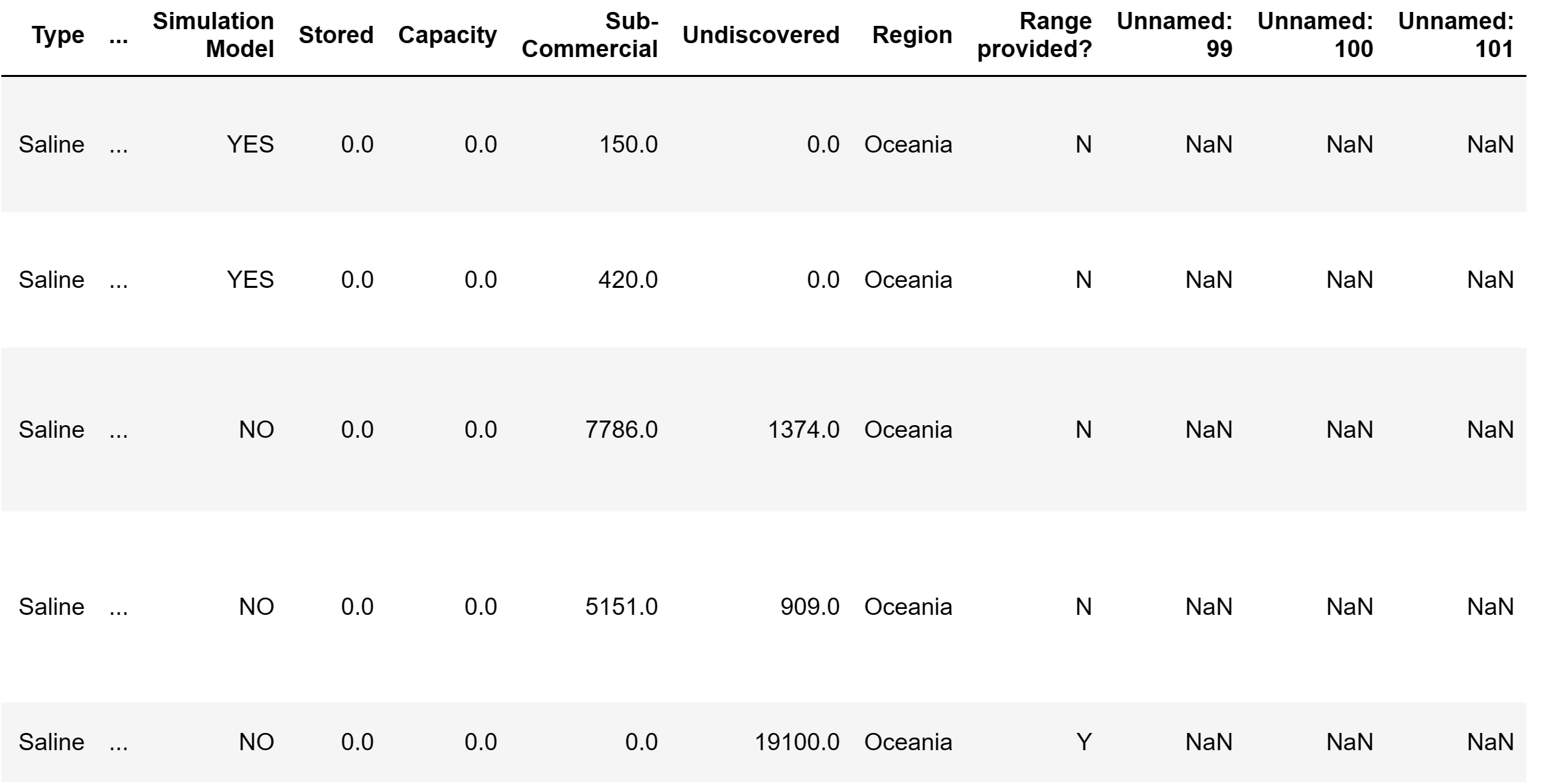
## The CSRC Database

To open the “CSRC\_Database\_Cycle.xlsx” dataset in Python, there are some special considerations to take note of. As the “CSRC\_Database\_Cycle.xlsx” dataset’s headers start from the second row onwards, we need to include the skiprows parameter and set it to one so that Pandas will skip the first row and read the data from the second row onwards, as shown in figure 4.2.1.



*Figure 4.2.1. Reading CSRC Data into Pandas Dataframe*

As there are unnamed columns included in the dataset, as shown in figure 4.2.2 below, we will remove these unnamed columns that contains the string “Unnamed”, as shown in figure 4.2.3 below.



*Figure 4.2.2. Unnamed Columns in Dataframe*



*Figure 4.2.3. Removing Unnamed Columns From Dataframe*

After we remove the unnamed columns, we can choose what columns we want in our dataframe by including only the columns we needed in a new dataframe, as shown in figure 4.2.4.



*Figure 4.2.4. Subletting CSRC Dataframe*

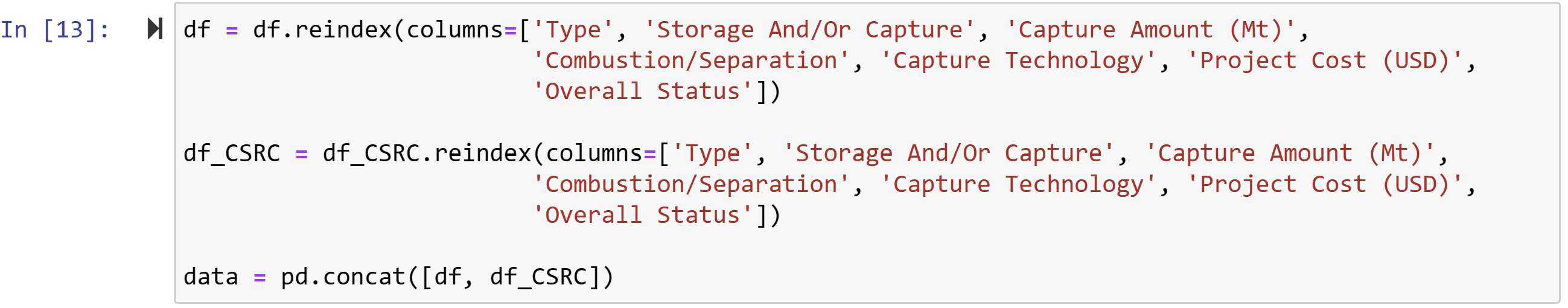
## Combining Datasets & Final Cleaning



*Figure 4.3.1. Subletting CSRC Dataframe*

According to Figure 4.3.1, to lengthen our dataframe, we will add additional information to match what is contained in both the CSS and CSRC dataframes. First, we add a 'Combustion/Separation' column in the CSRC Dataframe and set the value to 'Post-Combustion' as the carbon dioxide is capture after an industrial process. Secondly, we add a 'Capture Technology' column in the CSRC Dataframe and set the value to 'Injection' as the captured carbon dioxide will be injected into the geological feature to be stored indefintely. Thirdly, we add a 'Project Cost' column in the CSRC Dataframe and set the value to 0 as the geological features are naturally occurring, there is no need to construct new CCS storage sites. Thus, the project cost will be negligible. Lastly, we add a 'Type' column in the CSS dataframe and set the value to 'Plant' as the CSS facilities are man-made.

After ensuring that we have the same number of columns in both dataframes, we will rename the dataframe columns in both dataframes to match each other, as shown in figure 4.3.2. After we have the same columns header in both dataframe, we can join the CSRC dataframe and the CSS dataframe together to form a new “data” dataframe using the pd.concat function.



*Figure 4.3.2. Combining CSS and CSRC Dataframe Together*

Lastly, as predictive modelling algorithms only allows numerical values, we will need to convert the non-numerical values, as shown in tables 4.3.1, 4.3.2, 4.3.3, 4.3.4 and 4.3.5 below.

|  |  |
| --- | --- |
| **Converted Numerical Value** | **Original Text Value** |
| 1 | Plant |
| 2 | Petroleum |
| 3 | Saline |

*Table 4.3.1. Conversion Values For “Type” Column*

|  |  |
| --- | --- |
| **Converted Numerical Value** | **Original Text Value** |
| 1 | Capture |
| 2 | Storage |
| 3 | Capture and Storage |

*Table 4.3.2. Conversion Values For “Storage And/Or Capture” Column*

|  |  |
| --- | --- |
| **Converted Numerical Value** | **Original Text Value** |
| 0 | Not Specified |
| 1 | Pre-Combustion |
| 2 | Post-Combustion |
| 3 | Oxy-Combustion |
| 4 | Industrial |
| 5 | Separation |

*Table 4.3.3. Conversion Values For “Combustion/Separation” Column*

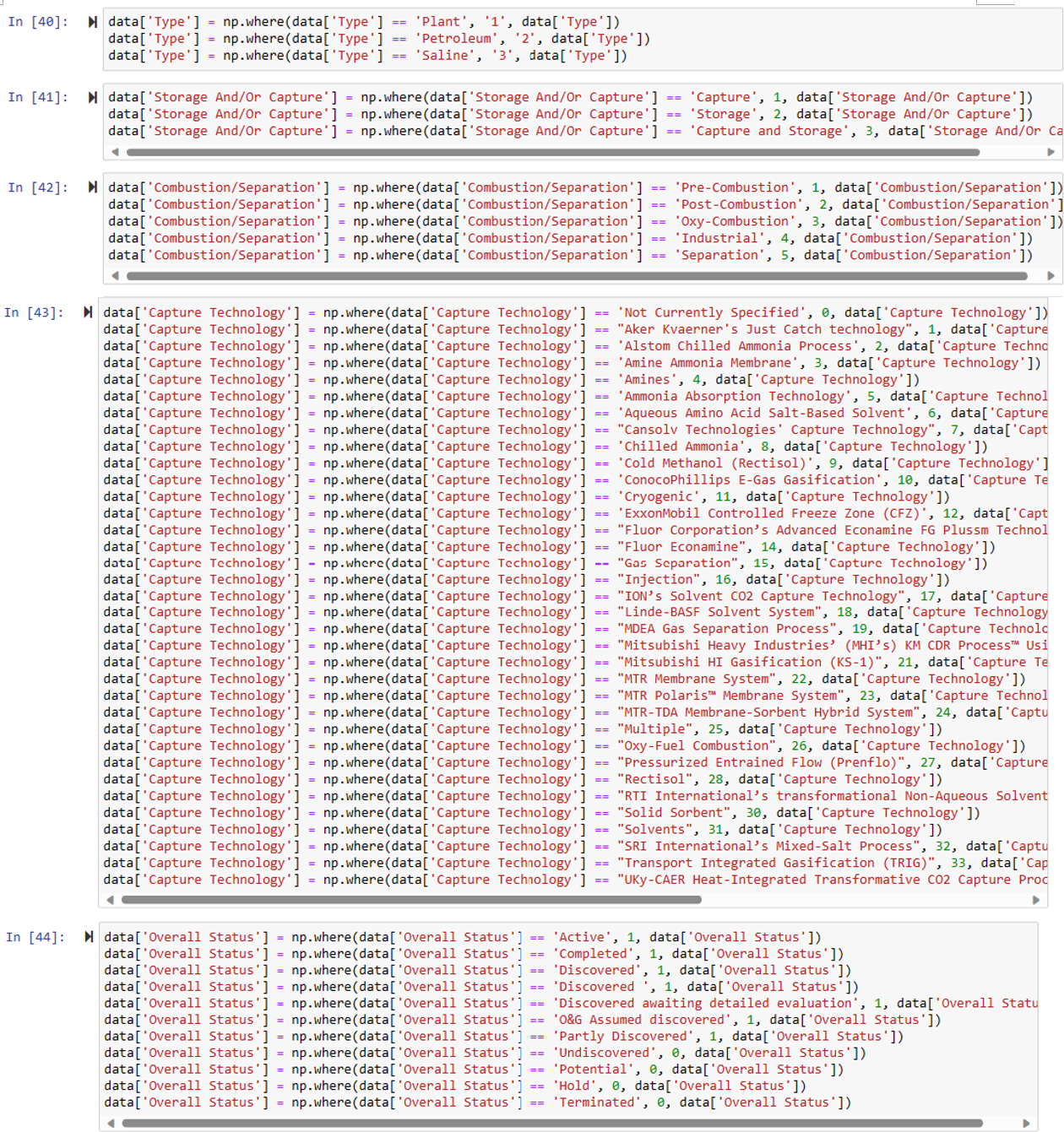
|  |  |
| --- | --- |
| **Converted Numerical Value** | **Original Text Value** |
| 0 | Not specified |
| 1 | Aker Kvaerner's Just Catch technology |
| 2 | Alstom Chilled Ammonia Process |
| 3 | Amine Ammonia Membrane |
| 4 | Amines |
| 5 | Ammonia Absorption Technology |
| 6 | Aqueous Amino Acid Salt-Based Solvent |
| 7 | Cansolv Technologies' Capture Technology |
| 8 | Chilled Ammonia |
| 9 | Cold Methanol (Rectisol) |
| 10 | ConocoPhillips E-Gas Gasification |
| 11 | Cryogenic |
| 12 | ExxonMobil Controlled Freeze Zone (CFZ) |
| 13 | Fluor Corporation’s Advanced Econamine FG Plussm Technology |
| 14 | Fluor Econamine |
| 15 | Gas Separation |
| 16 | Injection |
| 17 | ION’s Solvent CO2 Capture Technology |
| 18 | Linde-BASF Solvent System |
| 19 | MDEA Gas Separation Process |
| 20 | Mitsubishi Heavy Industries’ (MHI’s) KM CDR Process™ Using the KS-1™ Solvent |
| 21 | Mitsubishi HI Gasification (KS-1) |
| 22 | MTR Membrane System |
| 23 | MTR Polaris™ Membrane System |
| 24 | MTR-TDA Membrane-Sorbent Hybrid System |
| 25 | Multiple |
| 26 | Oxy-Fuel Combustion |
| 27 | Pressurized Entrained Flow (Prenflo) |
| 28 | Rectisol |
| 29 | RTI International’s transformational Non-Aqueous Solvent (NAS)-Based CO2 Capture Technology |
| 30 | Solid Sorbent |
| 31 | Solvents |
| 32 | SRI International’s Mixed-Salt Process |
| 33 | Transport Integrated Gasification (TRIG) |
| 34 | UKy-CAER Heat-Integrated Transformative CO2 Capture Process |

*Table 4.3.4. Conversion Values For “Capture Technology” Column*

|  |  |
| --- | --- |
| **Converted Numerical Value** | **Original Text Value** |
| 0 | Not Feasible (Overall Status: Undiscovered, Potential, Hold, Terminated) |
| 1 | Feasible (Overall Status: Active, Completed, Discovered, Discovered awaiting detailed evaluation, O&G Assumed discovered, Partly Discovered |

*Table 4.3.5. Conversion Values For “Capture Technology” Column*

To execute the conversion of values, we will the use np.where function to change all non-numerical values to numerical values, as shown in figure Figure 4.3.3.



*Figure 4.3.3. Codes to Convert All Non-Numeric Values to Numeric Values*

# Proposed Modelling Techniques & Evaluation

Among the numerous different prediction models, a few prominent prediction models include logistic regression, decision trees, Artificial Neural Networks (ANN) and Support Vector Machines (SVM). Thus, however, due to the non-categorical nature of both datasets, we only use the decision tress and ANN. Furthermore, as there are ensemble models, including boosting, bagging and random forests, to increase the performance of the predictive models, we will be applying these models to our initial models. We will then determine which of the two prediction models produces the most accurate results.

# Timeline of Project And Proposed Schedule

The timeline for the project is as shown in table 6.1 below.

|  |  |
| --- | --- |
| **Dates** | **Milestone** |
| 9 May 2023 | start of ANL488 intention survey table |
| 12 June 2023 | Allocation of supervisor |
| 16 June 2023 | 1st Meeting with supervisor for project allocation |
| 12 August 2023 | 1st seminar for ANL488 |
| 21 August 2023 | 2nd meeting with supervisor for proposal discussion |
| 8 September 2023 | Submission of business proposal |
| 18 September 2023 | 3rd meeting with supervisor for presentation discussion |
| 2 to 6 October 2023 | Oral presentation of project |
| 30 October 2023 | 4th meeting with Supervisor for report discussion |
| 6 November 2023 | Submission of final report |

*Table 6.1. Timeline of Project*

Based on the timeline of the project, the proposed project schedule is as shown in table 6.2 below.

|  |  |  |  |
| --- | --- | --- | --- |
| **From** | **To** | **No. of Days** | **Milestone** |
| 9 May 2023 | 9 June 2023 | 31 | Filling up the ANL488 intention survey table |
| 12 June 2023 | 12 June 2023 | 1 | Allocation of supervisor |
| 16 June 2023 | 16 June 2023 | 1 | 1st Meeting with supervisor for project allocation |
| 12 August 2023 | 12 August 2023 | 1 | 1st seminar for ANL488 |
| 13 August 2023 | 20 August 2023 | 7 | drafting the business proposal |
| 21 August 2023 | 21 August 2023 | 1 | 2nd meeting with supervisor for proposal discussion |
| 22 August 2023 | 7 September 2023 | 17 | Finalise business proposal based on supervisor’s comments |
| 8 September 2023 | 8 September 2023 | 1 | Submission of business proposal |
| 9 September 2023 | 17 September 2023 | 8 | Initial modelling and prepare presentation |
| 18 September 2023 | 18 September 2023 | 1 | 3rd meeting with supervisor for presentation discussion |
| 19 September | 1 October 2023 | 12 | Improve modelling and presentation based on the supervisor’s comments |
| 2 – 6 October 2023 | 2 – 6 October 2023 | 1 | Oral Presentation of Project |
| 7 October 2023 | 29 October 2023 | 22 | Improve modelling and write the final report |
| 30 October 2023 | 30 October 2023 | 1 | 4th Meeting with Supervisor for Report Discussion |
| 1 November | 5 November 2023 | 4 | Improve the final report |
| 6 November 2023 | 6 November 2023 | 1 | Submission of Final Report |

# References

Global Annual to Decadal Climate Update. (2023, May 17). World Meteorological Organisation. Retrieved June 29, 2023, from <https://library.wmo.int/doc_num.php?explnum_id=11629>

Haines, A., & Patz, J. A. (2004). *Health effects of climate change*. Jama, 291(1), 99-103.

Wirth, R., & Hipp, J. (2000, April). *CRISP-DM: Towards a standard process model for data mining.* In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining (Vol. 1, pp. 29-39).

Boot-Handford, M. E., Abanades, J. C., Anthony, E. J., Blunt, M. J., Brandani, S., Mac Dowell, N., ... & Fennell, P. S. (2014). *Carbon capture and storage update*. Energy & Environmental Science, 7(1), 130-189.

Bui, M., Adjiman, C. S., Bardow, A., Anthony, E. J., Boston, A., Brown, S., ... & Mac Dowell, N. (2018). *Carbon capture and storage (CCS): the way forward*. Energy & Environmental Science, 11(5), 1062-1176.

Haszeldine, R. S. (2009). *Carbon capture and storage: how green can black be?.* Science, 325(5948), 1647-1652.