**Applied machine learning to estimate CO2 adsorption in different materials**

*By*

*Daniela Mariano Barreto -* 2023278

*Rachel Mae Lademora – 2023396*

*Higher Diploma in Science in Data Analytics for Business*

*Strategic Thinking*

*James Garza*

*CCT College*

*Dublin, Ireland*

CCT College Dublin

**Assessment Cover Page**

|  |  |
| --- | --- |
| **Module Title:** | Strategic Thinking |
| **Assessment Title:** | CA 1 – Capstone Project Proposal |
| **Lecturer Name:** | James Garza |
| **Student Full Name:** | Daniela Mariano Barreto  Rachel Mae Lademora |
| **Student Number:** | 2023278  2023396 |
|  |  |
|  |  |
|  |  |
|  |  |
| **Assessment Due Date:** | 17th December 2023 |
| **Date of Submission:** |  |

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

Table of Contents

[Introduction 4](#_Toc149304071)

[Objectives 5](#_Toc149304072)

[Problem Definition 5](#_Toc149304073)

[Scope 5](#_Toc149304074)

[Data Source 6](#_Toc149304075)

[Ethical Considerations 6](#_Toc149304076)

[Appendix 1 7](#_Toc149304077)

[Appendix 2 8](#_Toc149304078)

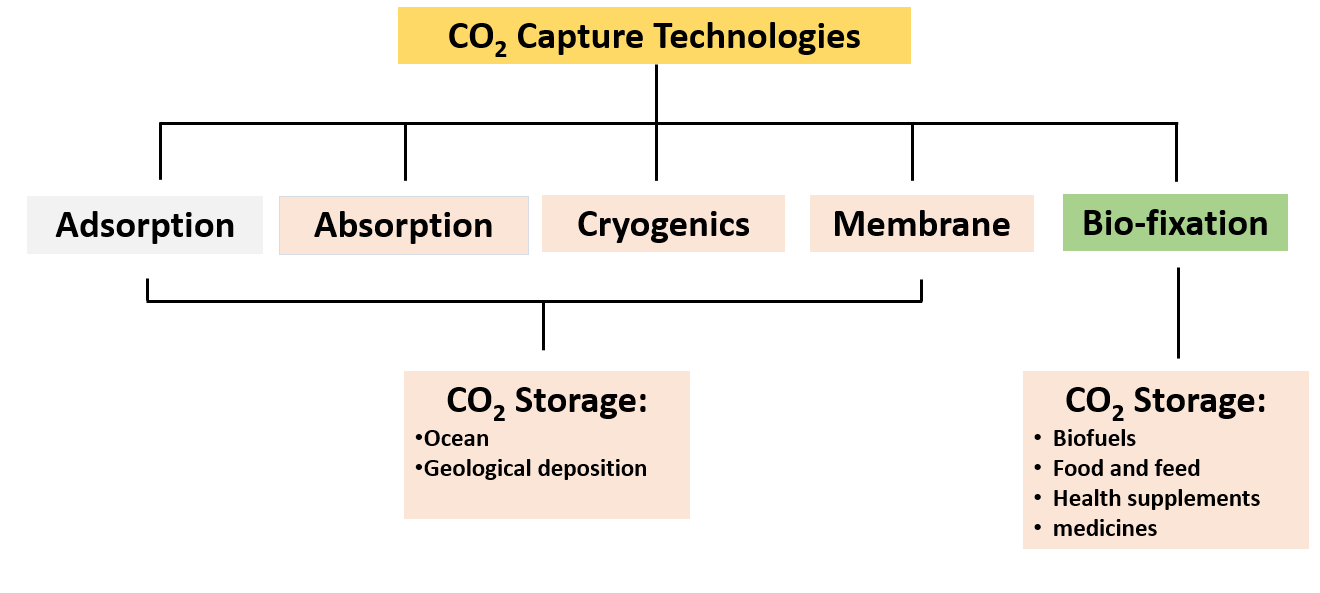
[References 10](#_Toc149304079)

# Introduction

Carbon capture and storage have been of extreme importance, as the concentration of this gas has increased, and data has shown that electricity generation will increase by 69% in 2040, with coal-based generation rising by 25% until 2040, achieving high levels of carbon dioxide (CO2) emissions. (Huetteman, Bowman and Slater-Thompson, 2016, p.81; Ren and Liu, 2023, p.1)

Studies agree that CO2 emissions have caused massive global damage because of the greenhouse effect, which is the potential explanation for all the climate changes. Those changes may be decrease in biodiversity, impairment in human lives with tsunamis, earthquakes, and others; consequently, they drastically diminish the productivity of ecosystems, which causes economic disparities. (Daneshvar *et al*., 2022, p.1; Ren and Liu, 2023, p.1)

The mitigation of CO2 remotion from the atmosphere to reduce the greenhouse effects has become the focus of universities, government, and private institutions; furthermore, developing technologies and products to capture CO2 has gained more power. (Ren and Liu, 2023, p.1). Asif *et al*. (2018, p.4) and Daneshvar *et al*. (2022, p.6) have shown the general panorama of all techniques that we have so far (Figure 1.)



**Figure 1:** Techniques for CO2 capture.

For this capstone project proposal, we will focus on the adsorption technique. This promising method has gained notable attention due to its low operational cost, lower energydemand, ease of handling, and general reliability. (Daneshvar *et al*., 2022, p.4)

Regarding the increase in the technologies to fix CO2, we propose implementing machine learning model to evaluate which one can efficiently estimate CO2 capture. After we obtain our results, it might be possible to understand which material is more efficient and, consequently, which one should be focused on to reduce production costs and become more economically affordable.

# Objectives

Principal objective:

We propose assessing some machine learning models to predict CO2 adsorption in different materials, such as rice rusk, activated carbons, and carbon nanotubes, considering data from academic papers.

Secondary objectives:

* Identify the principal materials used in the adsorption technique.
* Use the Exploratory Data Analysis (EDA) to understand the data.
* Apply machine learning models to choose one with a high accuracy score.
* Implement the chosen ML model to predict CO2 adsorption in these materials.

# Problem Definition

Considering that the emission of CO2 is a global problem, government and private institutions are investing in research to find solutions in this field. Consequently, we expect to have a lot of data, however, we might face difficulties finding articles with comparable experimental conditions. The second problem might be in choosing the most suitable machine learning model.

Therefore, the result of this project tends to be important as it aims to provide information on which material has the best potential to be invested in and improved by public and private institutions for the removal of CO2.

# Problem Definition

Considering that the emission of CO2 is a global problem, government and private institutions are investing in research to find solutions in this field. Consequently, we expect to have a lot of data, however, we might face difficulties finding articles with comparable experimental conditions. The second problem might be in choosing the most suitable machine learning model.

Therefore, the result of this project tends to be important as it aims to provide information on which material has the best potential to be invested in and improved by public and private institutions for the removal of CO2.

# Scope

This project aims to assess some machine learning models to predict CO2 adsorption in different materials. We will give an actual situation of the research in this field, with a compilation of articles analyzed by descriptive statistics. Secondly, we will implement a ML model, such as a decision tree, k-NN, and regression. Probably we will exclude models that would not be interesting to our objective.

The first boundary would be using data from exclusively adsorption technique, not others. This project will be limited by data with a maximum of five years. The third frontier is to use academic papers from reliable sources to maintain data credibility. Lastly, we will gather data with the same variables to compare the materials fairly.

At the end of this project, we expect to deliver an in-depth analysis of this technique with descriptive statistics and implement a ML model to predict CO2 adsorption in different materials. To achieve our goals, we break down the project into phases and delegate tasks as proposed in the Cross Industry Standard Process for Data Mining (CRISP-DM) method. In Appendix 1, we are presenting the timeline for this project.

Our action plan is to search for data with comparable experimental conditions, such as pressure and temperature. We will combine this data into an Excel spreadsheet. Later, we will work with Python using Jupyter Notebook to clean and organize the data and create graphs to have an overview of our data.

After we have a well-developed EDA, we will test some ML models in Python to choose the best one. Subsequently, we hope to have the answer as to which model presented the best estimate of results and which has the potential to be used in industries.

# Data Source

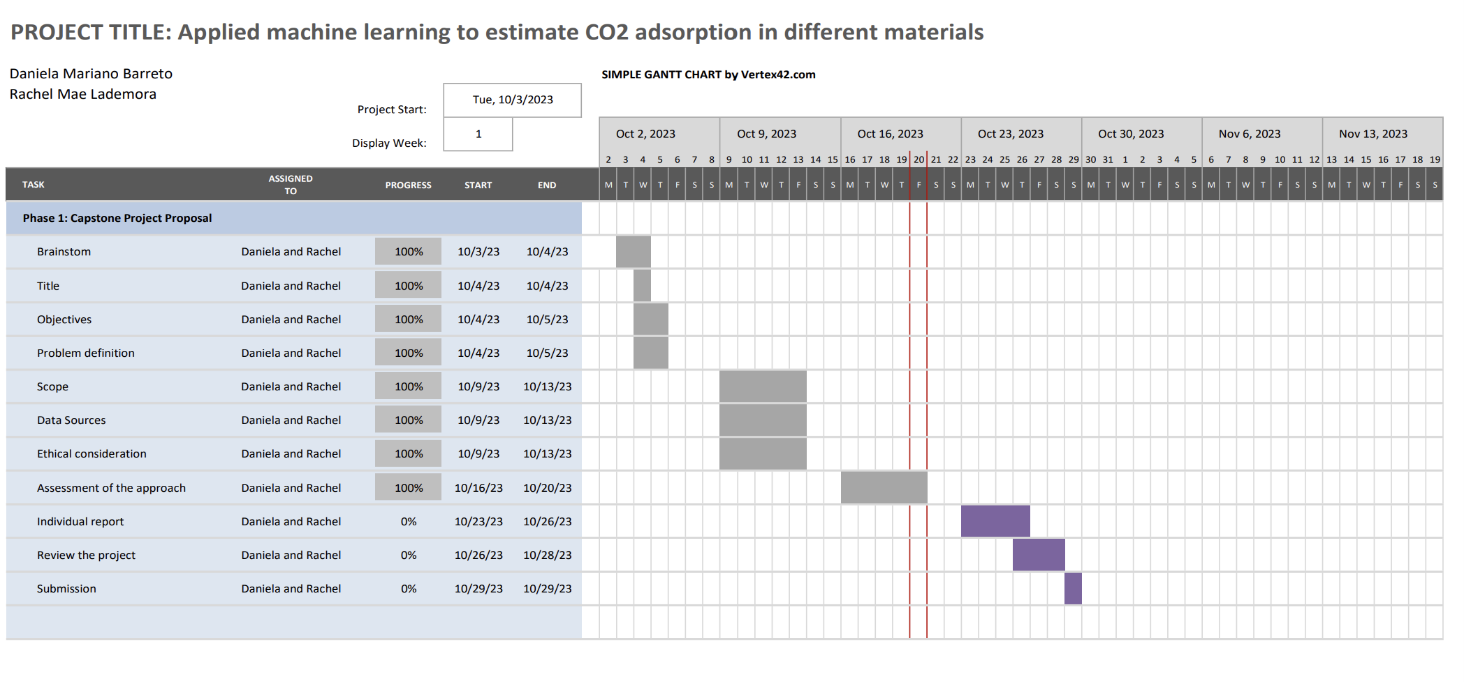
We will obtain data from academic papers in compiling at least 50 records for materials used in the adsorption CO2 capture technique. If necessary, we will also use the WebPlotDigitizer software to extract data from graphs in the articles.

We will maintain the PDF versions of the articles from which we took data as proof of permission to access the data, and all will be appropriately referenced.

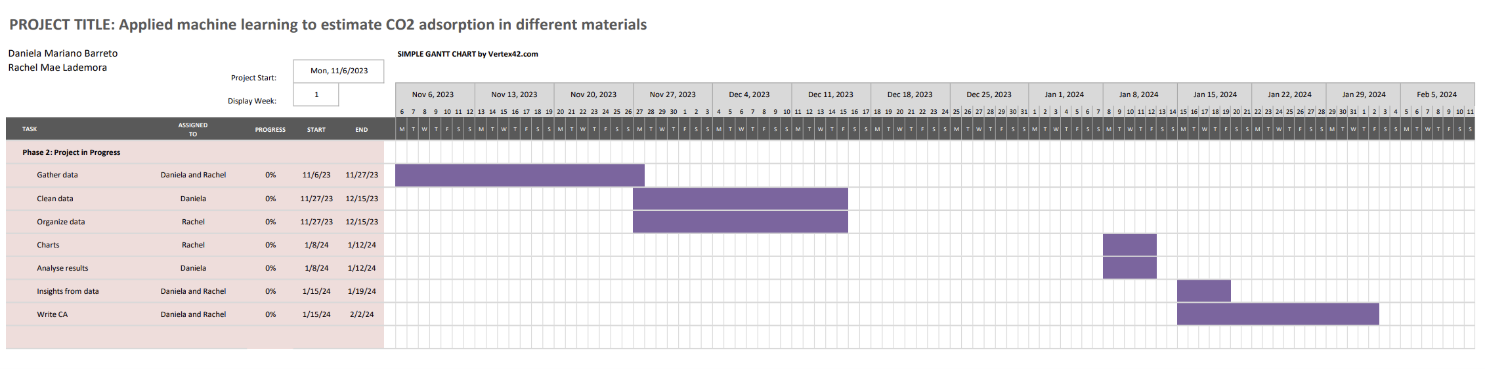
# Ethical Considerations

We will not intend to work with data that involves sensitive data, user privacy or potential social impacts; our data will be essential from laboratory research. In Appendix 2, we are presenting our Ethics Form signed.

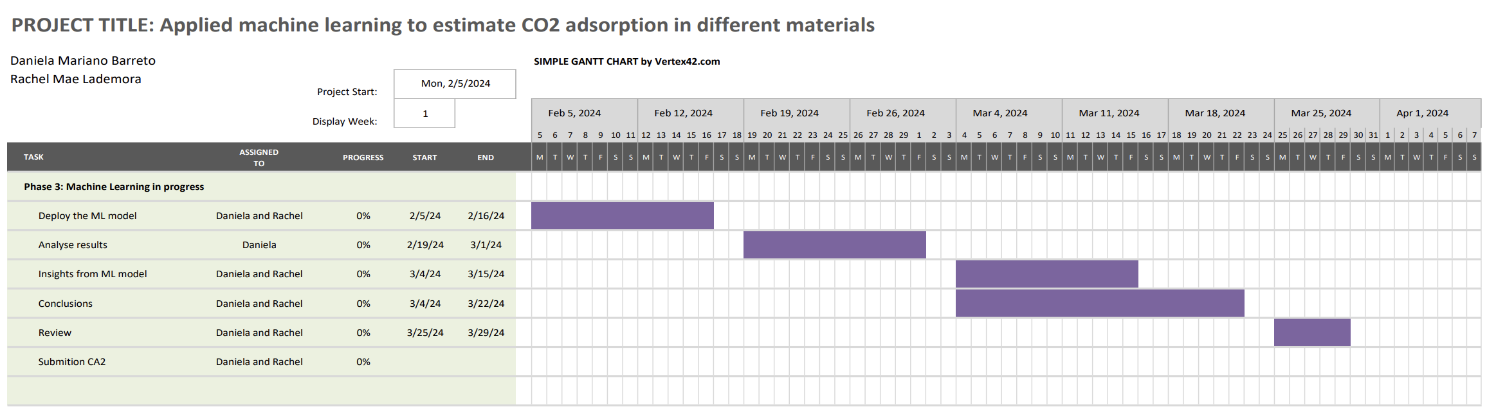
# Appendix 1

Phase 1:

Phase 2:



Phase 3:



# 

A close-up of a paper

Description automatically generated

# References

Ehsan Daneshvar *et al*. (2022). Biologically-mediated carbon capture and utilization by microalgae towards sustainable CO2 biofixation and biomass valorization – A review. *Chemical Engineering Journal*, *427*. <https://doi.org/10.1016/j.cej.2021.130884>

Furao Ren and Weijun Liu (2023). Review of CO2 Adsorption Materials and Utilization Technology. *Catalysts*, *13*(8), 1176. <https://doi.org/10.3390/catal13081176>

Muhammad Asif *et al.* (2018). Post-combustion CO2 capture with chemical absorption and hybrid system: current status and challenges. In *Greenhouse Gases: Science and Technology* (Vol. 8, Issue 6, pp. 998–1031). John Wiley and Sons Inc. https://doi.org/10.1002/ghg.1823

Thaddeus Huetteman, Michelle Bowman and Nancy Slater-Thompson (2016). Electricity. In *International Energy Outlook 2016 With Projections to 2040* (pp. 81–100). Energy Information Administration (EIA).

***CA2 Strategic Thinking***

***Predicting CO2 Adsorption materials using***

***one machine Learning Algorithms***

*By*

*Rachel Mae Lademora - 2023396*

*Higher Diploma in Science in Data Analytics for Business*

*Strategic Thinking*

*James Garza*

*CCT College*

*Dublin, Ireland*

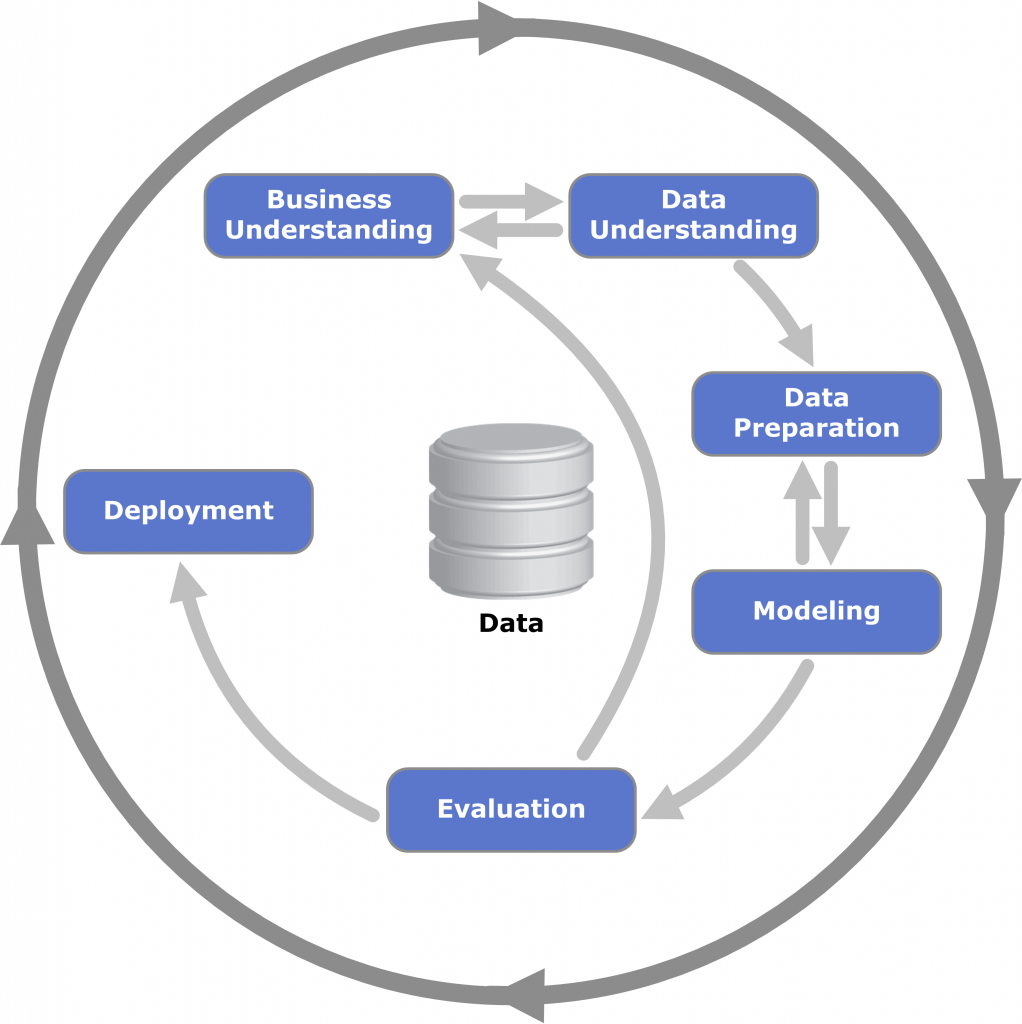
CCT College Dublin

**Assessment Cover Page**

|  |  |
| --- | --- |
| **Module Title:** | Strategic Thinking |
| **Assessment Title:** | CA 1 – Capstone Project Proposal |
| **Lecturer Name:** | James Garza |
| **Student Full Name:** | Rachel Mae Lademora |
| **Student Number:** | 2023396 |
|  |  |
|  |  |
|  |  |
|  |  |
| **Assessment Due Date:** | 17th December 2023 |
| **Date of Submission:** |  |

**Declaration**

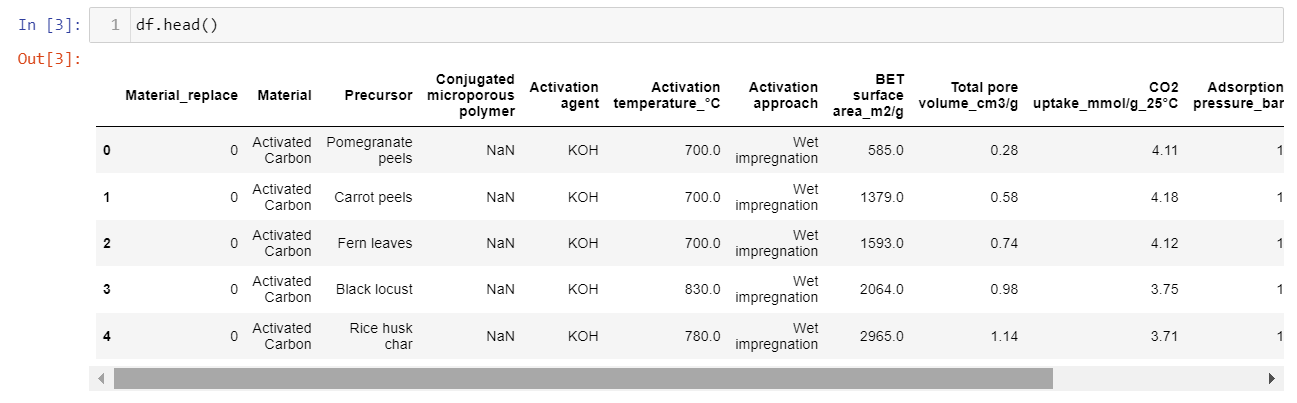
By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.



Business Understanding

As we mention in our CA 1 enable for us to achieve our goal of this capstone project we will use the CRISP-DM process as a guid for the step-by-step process of data analysis of the CO2 adsorption dataset. As I read academic papers that we use in our CA1. I discover that there are natural materials that can be used as alternative to help lessen the emission of CO2 adsorption. It is vital because the materials are not so expensive and help the business sector to implement it to be used, the business company can sell the materials to other companies in affordable price. I see that these materials would have a huge help for business companies, since they can offer it to the customers in affordable price. It has a good benefit both in environment and in the business industry.

Data Understanding

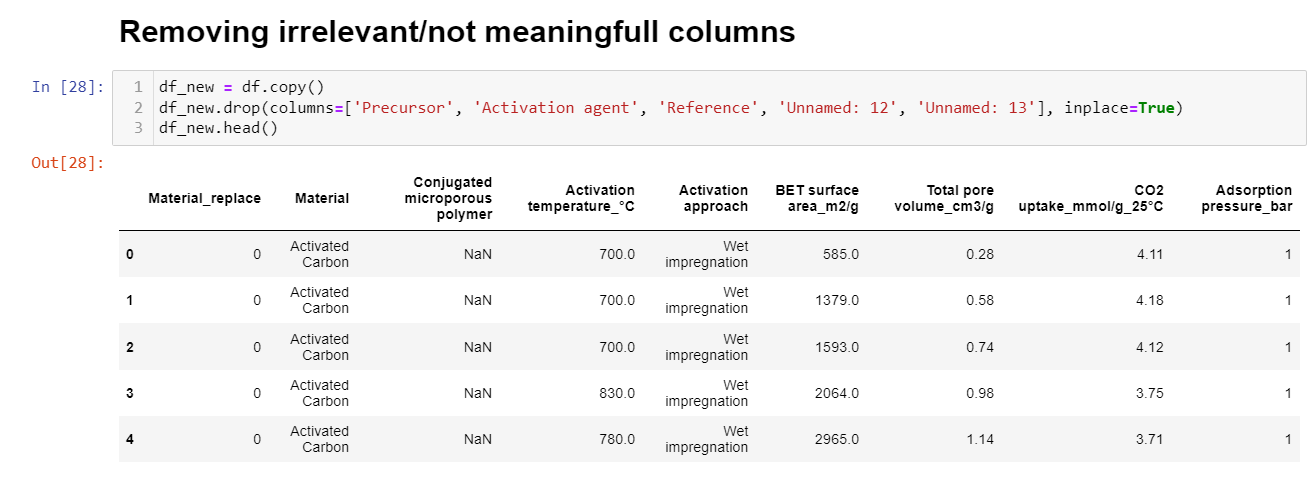
The dataset consists natural materials which is very useful to have a lower effect of adsorption to lessen the greenhouse effect. To better understand the dataset this are the dataframe head and data dictionary with definitions and the data type that the CO2 adsorption dataset. 

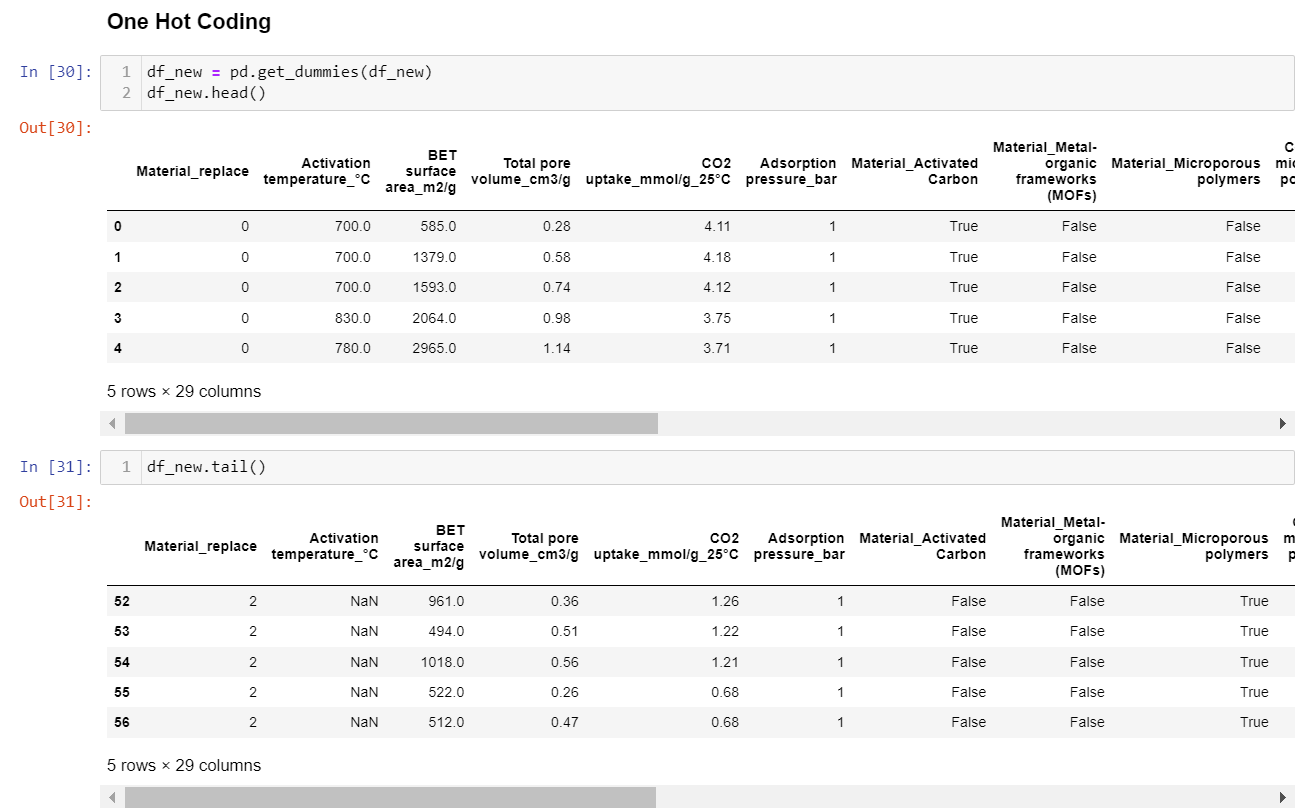
Data Dictionary

|  |  |  |
| --- | --- | --- |
| Columns name | Definition | Data Types |
| Material Replace | Code for the class material. | Int64 |
| Material | Material used in the CO2 adsorption process. | object |
| Precursor | Is the material used in the initial adsorption process. | object |
| Conjugated microporous polymer | Chemical used in polymers material in the adsorption process. | object |
| Activation Agent | Chemical compound used in the carbon- based material. | object |
| Activation\_Temperature\_oC | Temperature used to agent activation in carbon-based material. | float64 |
| Activation approach | Technique used in the activation in carbon-based material. | object |
| BET surface area\_m2/g | Surface area of the material used in the CO2 adsorption. | float64 |
| Total pore volume\_cm3/g | Volume of the pore material used in the CO2 adsorption in 25oC. | float64 |
| CO2 uptake mmol/g\_25\_ oC | CO2 adsorption in 25oC. | float64 |
| Adsorption pressure\_bar | Pressure of CO2 adsorption. | int64 |
| Reference | Reference for academic papers results (Dziejarski et al., 2023, p.69-74). | object |

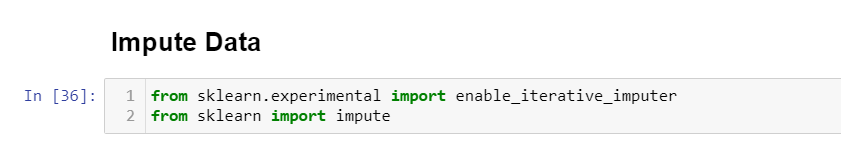
**Data Preparation**

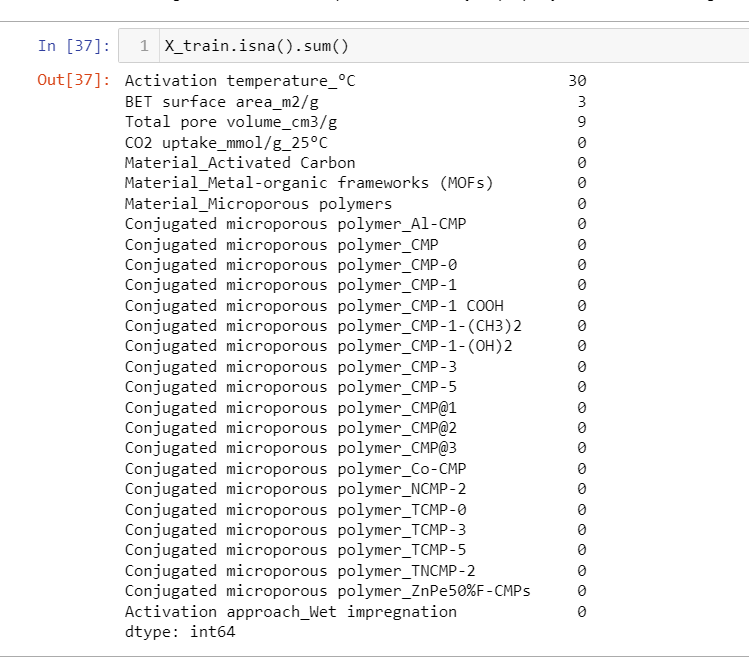
The data preparation is very important because the dataset need to be clean first before performing the machine learning. I drop some columns that I see that they are not necessary to be use in executing my data analysis. There are two unnamed columns that it will not be useful since it that columns have a NAN. Also, I drop the precursor as I see it is only categorical variables and it has a NAN values, activation agent is also categorical and has NAN values, and the reference has only a numerical variable to tells the numbers of reference in the academic paper.

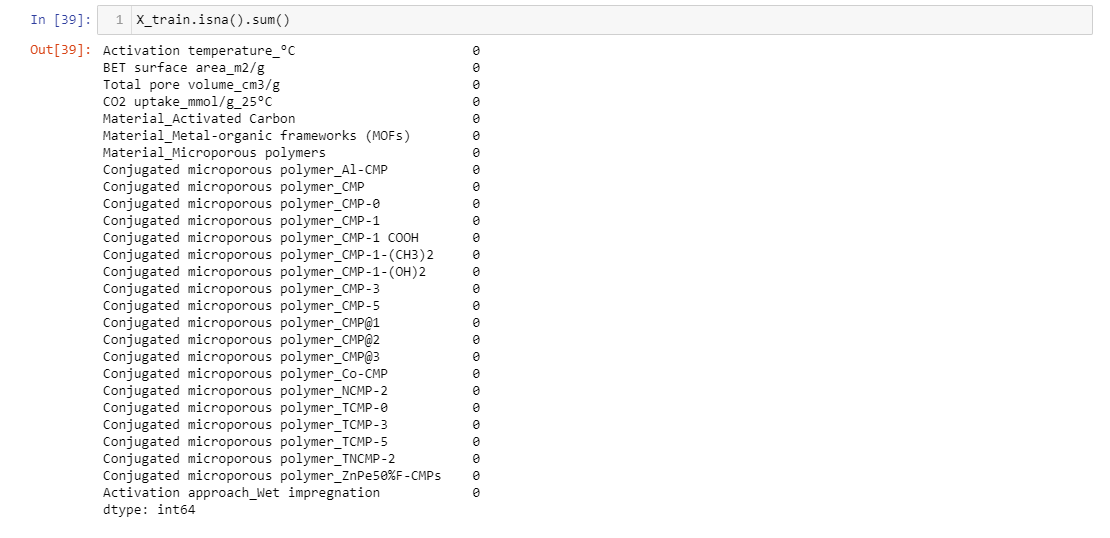




I used one hot coding to transform the categorical to numerical variables. But it shows that is not successfully transform to numerical variables. Because the material activated carbon and the rest of the materials are only transformed by true or false. That is why I find another library which I can use to convert the categorical columns.

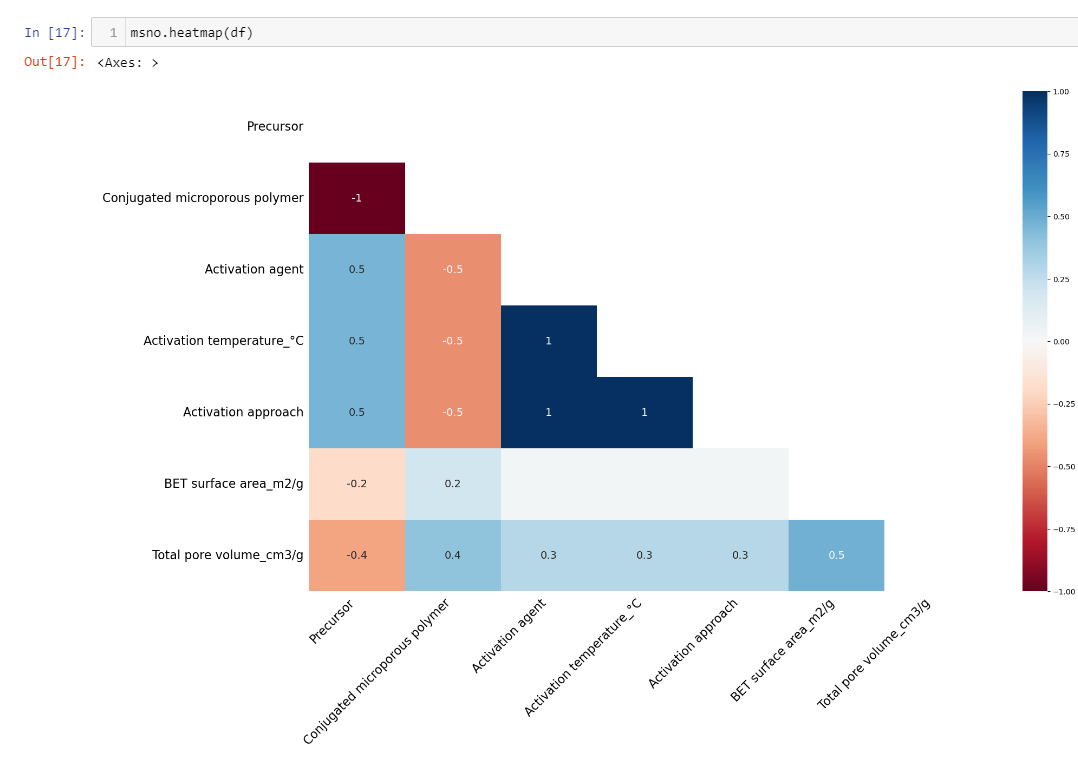


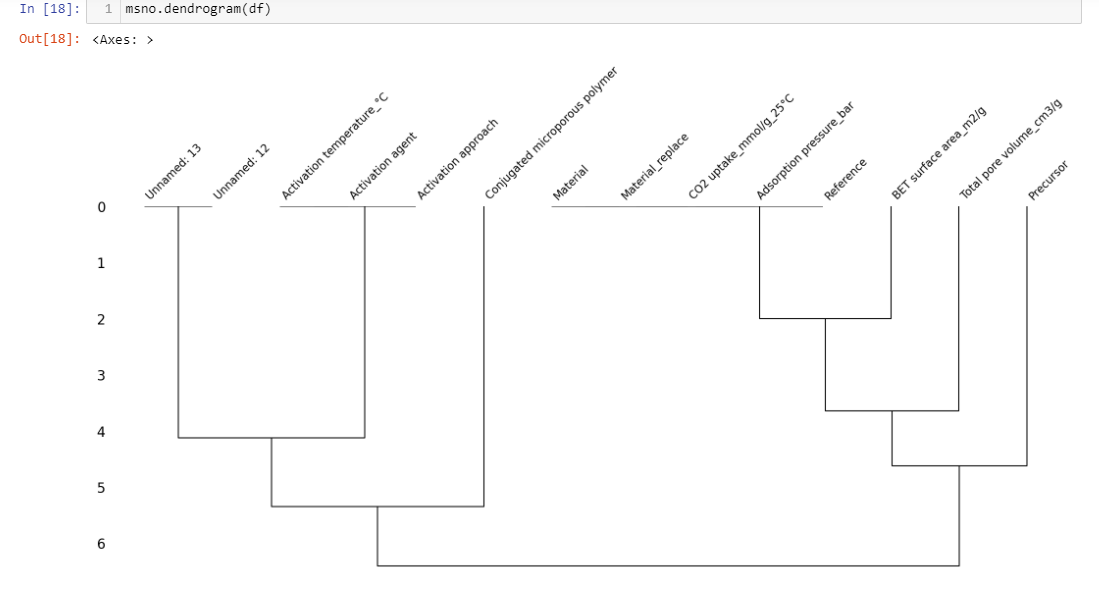




I decide to use to impute the data to transform the categorical to numerical variables before I will perform the machine learning model, because machine learning only read numbers. It is a very important to do the data cleaning first so that I will have a good test accuracy score as a result of training the model. By using the impute data I transform it gives me 0 values.

Modeling





Modeling is the fundamental method to execute the visualization of the missing data and the correlation of the columns upon using the heatmap to determine which columns has a high result of missing values. The conjugated microporous polymer has -1 values it means that it has a lot of missing values on it. On the dendrogram It is clearly stated