Exploratory Data Analysis (EDA) on Multi-Site Factory Carbon Capture as a Service (CCaaS) Implementation

REPORT PRESENTED BY

SACHITH M. GUNAWARDANE

2024

TABLE OF CONTENTS

1 Introduction

Carbon Capture as a Service (CCaaS) is a burgeoning business model designed to facilitate the adoption of carbon capture technologies, providing a scalable, cost-effective solution for industries with CO2 emissions. This model helps overcome the high initial costs, operational complexities, and technical risks often associated with carbon capture systems [1].

CCaaS providers deliver comprehensive services that include financing, construction, and operation of carbon capture facilities, enabling client businesses to concentrate on their primary operations while managing CO2 emissions externally [2]. This approach helps businesses minimise their carbon footprint, achieve sustainability targets and ensure compliance with strict environmental regulations [3].

Especially beneficial for smaller emitters and industries where independent carbon capture initiatives are cost-prohibitive, the CCaaS model offers variable pricing, such as payment per ton of captured CO2. It also opens revenue opportunities through carbon utilisation and trading in carbon credits [4].

In the global pursuit of reducing greenhouse gas emissions and combating climate change, CCaaS is set to significantly influence the widespread adoption of carbon capture technologies in various sectors. By lowering the barriers to entry and fostering technological innovation, CCaaS could substantially contribute to worldwide climate mitigation efforts [5].

Key Features of CCaaS:

No Upfront Capital Expenditure: CCaaS providers typically finance, build, and operate carbon capture equipment, eliminating the need for client companies to make significant upfront investments in the technology.

Operational Expertise: The service provider brings specialised knowledge and experience in carbon capture technology, ensuring efficient and effective operation of the capture systems.

Risk Mitigation: By outsourcing carbon capture operations to a third party, companies can mitigate these systems' technical and operational risks.

Flexible Pricing Models: CCaaS agreements may include various pricing models, such as pay-per-ton of CO_2 captured, which aligns the costs directly with the amount of carbon mitigated.

Regulatory Compliance: For companies facing strict environmental regulations regarding their CO_2 emissions, CCaaS offers a pathway to compliance without developing in-house solutions.

Revenue from Carbon Utilization: Some CCaaS models also explore opportunities for monetising the captured CO_2, whether through sale to other industries (e.g., for use in beverage carbonation, enhanced oil recovery, or as a feedstock for synthetic fuels) or by claiming carbon credits in regulatory or voluntary carbon markets.

Benefits of CCaaS:

Accessibility: Makes carbon capture technology accessible to smaller emitters and those in sectors where it might not have been economically feasible to deploy such solutions independently.

Cost-Effectiveness: Spreads out the cost of carbon capture over time, based on actual usage or capture rates, making it a more cost-effective solution for many businesses.

Innovation and Improvement: Encourages continuous improvement and innovation in carbon capture technologies as service providers seek to enhance efficiency and reduce costs to stay competitive.

Sustainability Goals: Helps companies achieve their sustainability and environmental goals by reducing their carbon footprint in a measurable and verifiable way.

Market Expansion: Opens up new markets for carbon capture solutions by lowering entry barriers, potentially accelerating the global adoption of carbon capture technologies.

1.1 Objectives of the EDA

The primary objective of conducting an Exploratory Data Analysis (EDA) on the Carbon Capture as a Service (CCaaS) dataset is to gain valuable insights into the operational efficiency, financial viability, and environmental impact of implementing CCaaS across multiple factory sites. The EDA aims to uncover data patterns, trends, and anomalies, providing a foundation for informed decision-making and strategic recommendations. The specific objectives of the EDA are as follows:

Assess Operational Efficiency:

Analyse the CO_2 capture volumes, energy consumption, and plant uptime percentage across different sites to evaluate operational efficiency. Identify sites with higher CO_2 capture rates and lower energy consumption, indicating better operational performance. Investigate any variations or anomalies in operational metrics across sites and periods to identify potential areas for improvement.

Evaluate Financial Viability:

Examine the revenue generated from CO_2 sales and carbon credits across different sites to assess the financial benefits of CCaaS. Analyse the operational costs associated with CCaaS implementation and compare them against the revenue streams to determine profitability. Identify sites with higher revenue generation and lower operational costs, indicating better financial performance. Explore the relationship between CO_2 capture volumes, energy consumption, and financial metrics to understand the factors influencing financial viability.

Quantify Environmental Impact:

Assess the total CO_2 emissions mitigated through CCaaS implementation across different sites and periods. Analyse the trends in CO_2 capture rates over time to evaluate the effectiveness of CCaaS in reducing carbon footprint. Identify sites with higher CO_2 capture volumes and calculate the corresponding environmental benefits regarding reduced greenhouse gas emissions.

Identify Patterns and Trends:

Explore temporal patterns and trends in CO_2 capture rates, operational costs, revenue, and other relevant metrics across different sites and periods. Identify any seasonal variations or long-term trends in the data that may influence CCaaS performance. Investigate the relationship between different variables to uncover potential correlations and dependencies.

Detect Anomalies and Outliers:

Identify any unusual or unexpected values in the dataset that may indicate anomalies or outliers. Investigate the reasons behind these anomalies and assess their impact on the overall CCaaS performance. Determine if the anomalies require further investigation or if genuine data points provide valuable insights.

Compare Site Performance:

Conduct a comparative analysis of CCaaS performance across different factory sites. Identify top-performing sites based on operational efficiency, financial viability, and environmental impact metrics. Explore the factors contributing to the success of high-performing sites and identify best practices that can be replicated across other sites.

Provide Strategic Recommendations:

Based on the insights gained from the EDA, provide strategic recommendations to optimise CCaaS implementation and performance. Identify potential areas for improvement regarding operational efficiency, financial viability, and environmental impact. Suggest data-driven strategies to enhance CCaaS adoption, maximise CO2 capture rates, minimise costs, and increase revenue generation.

1.2 Dataset Description

In this exploratory data analysis (EDA), we used a synthetic GPT dataset. The dataset encompasses records from five distinct manufacturing processes across five organisational sites, ensuring a diverse range of CO2 capture rates, operational costs, and energy consumption. The selected manufacturing processes include:

- **Chemical Production**: Known for substantial CO2 emissions due to chemical reactions and high-energy equipment, this process likely incurs high operational costs and substantial energy use but offers significant CO2 capture opportunities.
- **Cement Manufacturing**: As a major emitter of industrial CO2, primarily from limestone calcination, this process likely records high CO2 capture and potential revenue from carbon credits.
- **Steel Manufacturing**: Characterised by high energy consumption and CO2 emissions from the smelting process, steel plants may demonstrate varied operational efficiencies and significant CO2 capture potential.
- **Food Processing**: Generally lower in CO2 emissions than heavy industries, food processing varies widely in energy usage and operational costs based on the type of food processed.

- **Paper and Pulp Manufacturing**: This industry sees considerable energy use and CO_2 emissions from wood processing and the operation of energy-intensive machinery.

The dataset covers these five sites over the past two years, with data recorded monthly, resulting in 120 records. The fields in the dataset include:

Date: The date of data recording.

Site_ID: A unique identifier for each factory site.

CO2_Captured_Tons: The amount of CO_2 captured is measured in tons.

Operational_Costs: Operational costs associated with the CCaaS are in USD.

Energy_Consumption: Energy consumed by the carbon capture process in kWh.

Revenue_CO2_Sales: Revenue generated from selling captured CO_2 in USD.

Revenue_Carbon_Credits: Revenue obtained from carbon credits in USD.

Plant_Uptime_Percentage: The percentage of time the CCaaS plant was operational.

Customer_Satisfaction_Score: The factory's satisfaction with the CCaaS provider on a scale of 1 to 10.

This dataset provides a foundation for analysing carbon capture implementations' efficiency and financial viability across different industrial processes.

2 Visualisation of Captured Data

The following line graphs visualise various performance metrics from five different sites over a period of two years.

2.1 CO2 Capture Tons

The CO_2 capture tons show significant variability across sites, with some sites like Site 2 and Site 5 displaying high peaks and troughs, indicating fluctuating capture efficiencies or operational challenges. Site 1 and Site 3 show more stability in their performance, suggesting more consistent operations in terms of CO_2 capture.

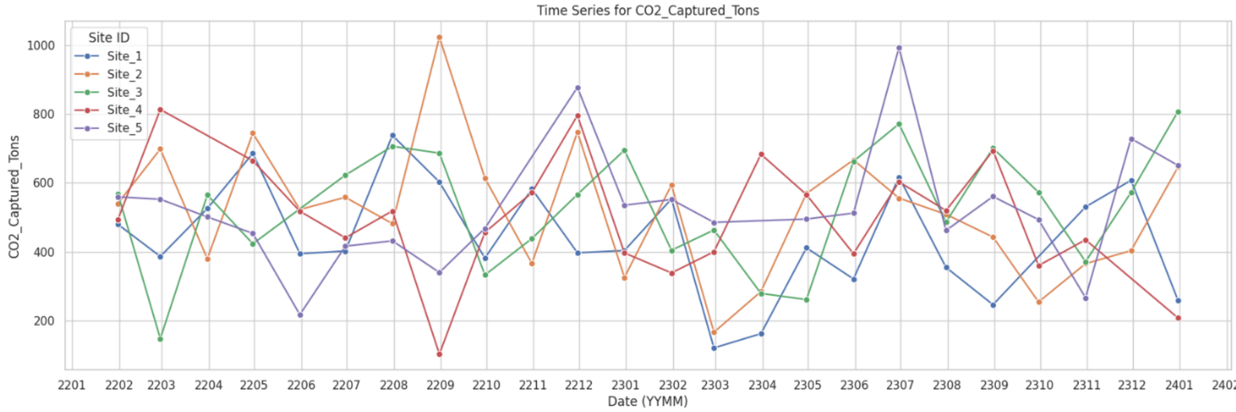


Figure 1: CO2 Captured Tons

2.2 Operational Costs

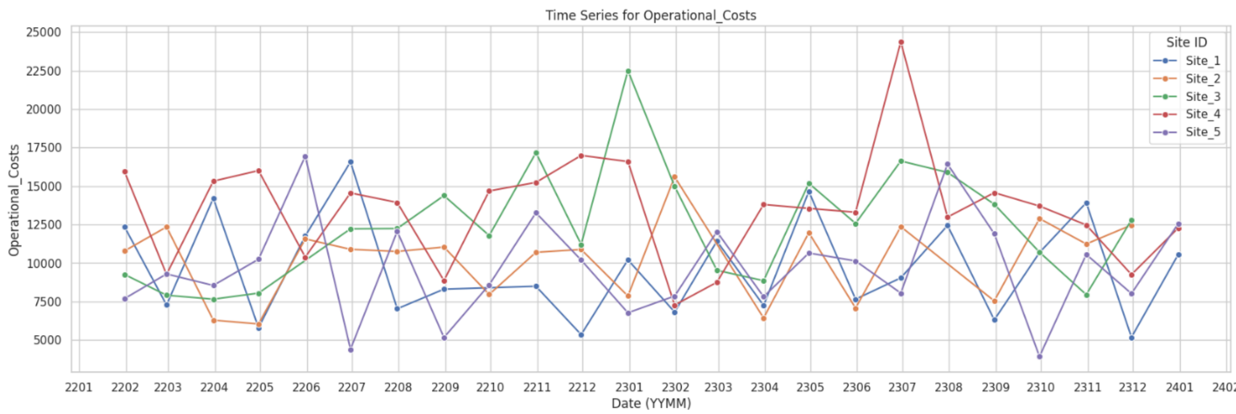


Figure 2: Operational Costs

Operational costs fluctuate across all sites but display a relatively correlated trend with CO_2 capture tons, suggesting that higher operational activities (or issues) directly impact costs. Site 4 and Site 3 often have higher operational costs than others, which could reflect higher activity levels or less efficient operations.

2.3 Energy Consumption

Energy consumption closely mirrors the trends seen in CO2 capture, particularly at sites with high operational variability. Site 5, while efficiently capturing CO2, seems to consume more energy, pointing to potential inefficiencies that could be targeted for improvement.

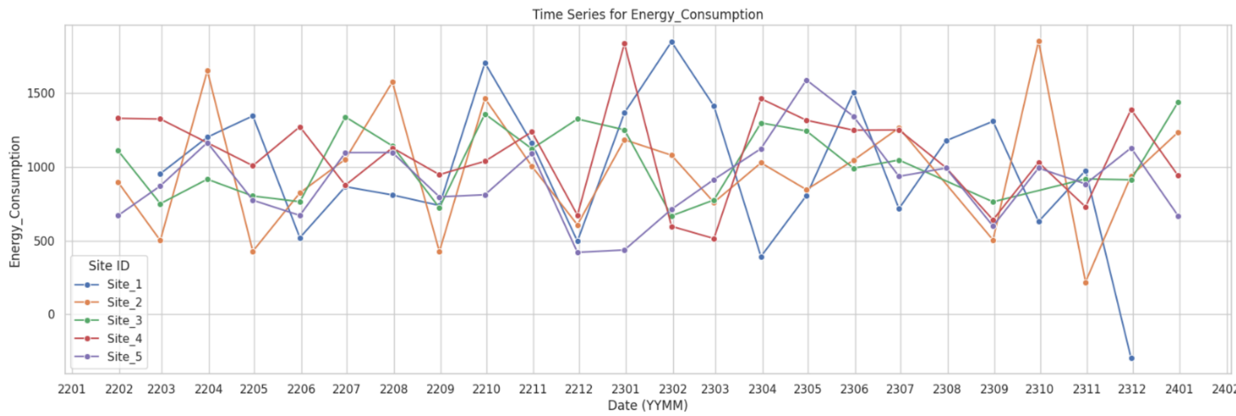


Figure 3: Energy Consumption.

2.4 Revenue from CO2 Sales

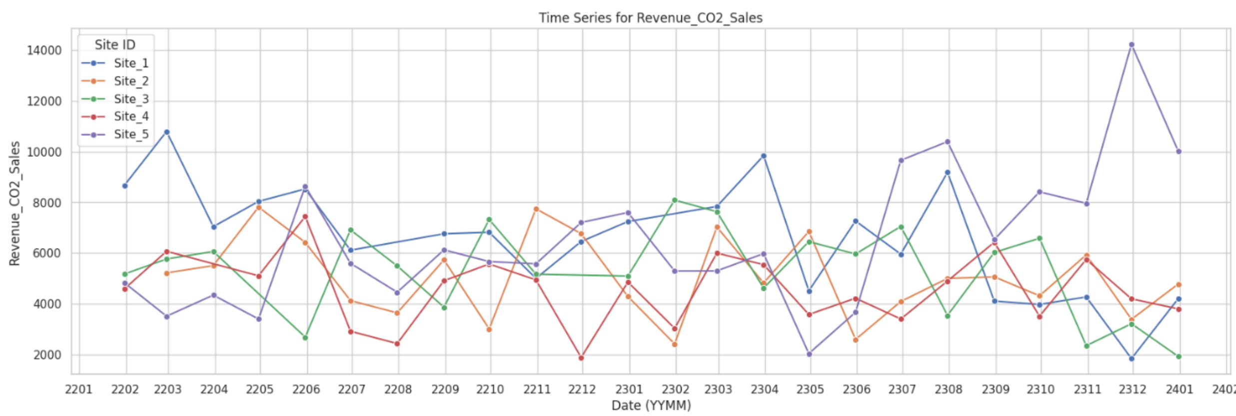


Figure 4: Revenue CO2 Sales

Revenue trends are not as tightly aligned with CO2 capture tons as expected, indicating that pricing or market factors also play significant roles. Site 5, despite its high energy consumption, does not always lead in revenue, suggesting that its efficiency gains are not directly translating into financial performance.

2.5 Revenue from Carbon Credits

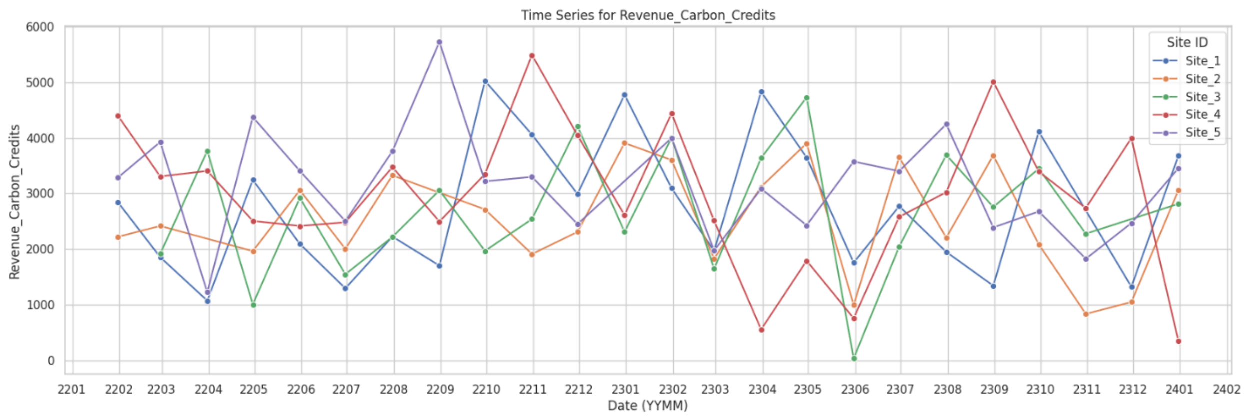


Figure 5: Revenue Carbon Credits

This revenue stream shows less correlation with CO2 captured tons, which might be influenced by external market conditions or regulatory changes affecting carbon credit values. Sites 2 and 3 occasionally spike in revenue from carbon credits, which could indicate strategic management of carbon credit assets or external influences affecting these sites differently.

2.6 Plant Uptime Percentage

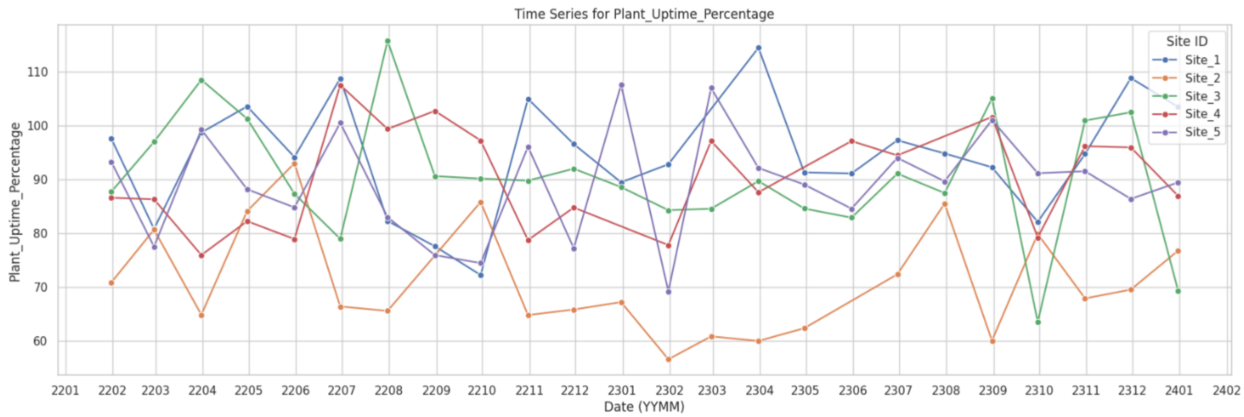


Figure 6: Plant Uptime Percentage

Plant uptime is critical for operational efficiency. The charts show that uptime does not always correlate with high CO2 capture, suggesting that simply running more often does not necessarily mean running better. Site 2 shows lower uptime but does not always result in lower CO2 capture;

it shoots higher than other sites, hinting at possible efficiencies when operational or high maintenance quality when the plant is running.

2.7 Customer Satisfaction Score

Figure 7: Customer Satisfaction Score

Customer satisfaction does not directly follow the patterns of operational metrics, indicating that other factors (such as customer service or pricing strategies) also significantly impact satisfaction. Site 3 and Site 5 generally maintain higher satisfaction scores, which might be linked to their operational approaches or customer management practices.

Overall, there is a need to balance energy consumption with CO2 capture effectiveness. Sites with high fluctuations in operational metrics need to investigate the causes and possibly stabilise their operations. While revenue from CO2 sales is crucial, managing operational costs and maximising revenue from carbon credits are equally important. Financial strategies should consider market conditions and regulatory changes. Consistent high performance in CO2 capture combined with efficient operations and energy use is essential. Sites showing lower energy efficiency despite high CO2 capture volumes may need targeted improvements.

3 Data Cleaning and Preparation

3.1 Handling Missing Values

The initial analysis identifies missing values in several key variables, each having six missing entries. This suggests a pattern or issue in data collection or recording at specific intervals.

```
Handling Missing Values:
Number of missing values in each column:
Date                          0
Site_ID                       0
CO2_Captured_Tons             6
Operational_Costs             6
Energy_Consumption            6
Revenue_CO2_Sales             6
Revenue_Carbon_Credits        6
Plant_Uptime_Percentage       6
Customer_Satisfaction_Score   6
dtype: int64
```

Figure 8: Missing Values

When applying imputation techniques, understanding the data distribution is crucial. Figure 9 presents the skewness values for each variable. A skewness value close to 0 signifies a symmetrical distribution. Conversely, values greater than 1 indicate a right-skewed distribution and values less than -1 suggest a left-skewed distribution, highlighting significant asymmetry.

```
Skewness of columns with missing values:
CO2_Captured_Tons: 0.1940862619246729
Operational_Costs: 0.590003844104742
Energy_Consumption: -0.1346702244048381
Revenue_CO2_Sales: 0.7909146525450073
Revenue_Carbon_Credits: 0.039551264103468306
Plant_Uptime_Percentage: -0.27252346939474204
Customer_Satisfaction_Score: 0.11625610999413542
```

Figure 9: Skewness Analysis

Imputation Techniques Applied:

Given the nature of the data and its distribution (as indicated by skewness values), different imputation techniques were appropriately chosen:

Mean imputation for moderately skewed continuous variables (*CO2_Captured_Tons*).

Median imputation for more skewed continuous variables (*Operational_Costs*) is beneficial for reducing the influence of outliers.

Mode imputation for categorical data (*Customer Satisfaction Score*), ensuring the most frequent category is used for filling gaps.

Forward and Backward Fill for variables like *Energy Consumption* and *Revenue CO2 Sales,* respectively, which can help maintain trends in time-series data without introducing artificial changes.

Linear and Polynomial Interpolation for *Revenue_Carbon_Credits* and *Plant Uptime Percentage*, respectively, help make the data more coherent for time series analysis by using existing data points to estimate missing values.

Skewness values were crucial in determining the most appropriate imputation method. Variables with mild skewness used mean or median strategies, depending on whether the skew was positive or negative, whereas more neutral skewness levels allowed for more straightforward methods like linear interpolation.

3.1.1 Visual Analysis of Imputation Impact

The comparison of Revenue_CO2_Sales before and after imputation (seen in the attached graph) illustrates the effectiveness of the chosen imputation techniques. The imputed values blend well with the original data, maintaining existing trends and variability without introducing visible biases or anomalies, which is crucial for accurate subsequent analyses.

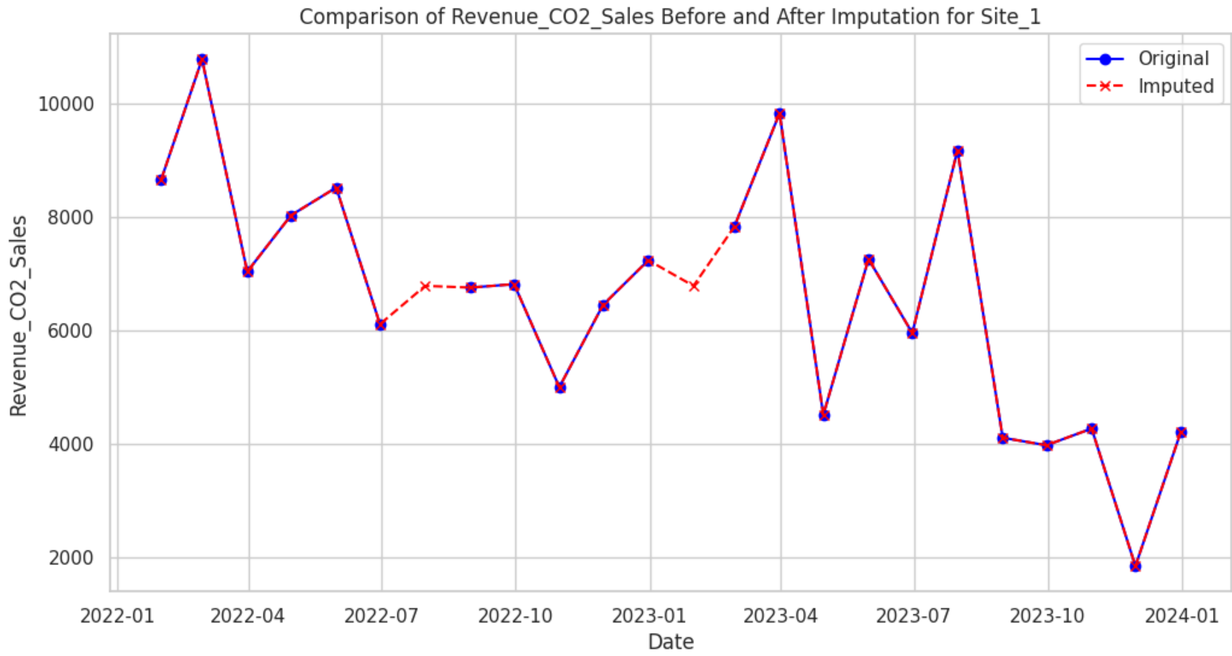


Figure 10: Revenue for CO2 Sales after Imputation

Site-Wise Imputation Consideration

The method acknowledges the possibility of site-specific patterns and conditions affecting the data by conducting imputation site-wise. This approach helps preserve the intrinsic data characteristics at each site, leading to more tailored and accurate data restoration.

3.1.2 Overall Impact on Analysis

Post-imputation, the dataset is likely more robust and reliable for further analysis. The imputation helps prevent potential biases in the analysis phases arising from missing data, such as in predictive modelling or trend analysis. It ensures that the insights drawn from the data represent the actual operational and financial conditions at the sites.

3.2 Outlier Detection and Treatment

The box plots were generated for each variable across five sites to help us visualise the distribution and presence of outliers.

CO2 Captured Tons: Sites 2 and 5 show multiple outliers on the upper end, indicating some instances of exceptionally high CO2 capture volumes. Site 5 displays a narrower interquartile range (IQR) and several outliers, suggesting variability in the CO2 capture process or data entry errors.

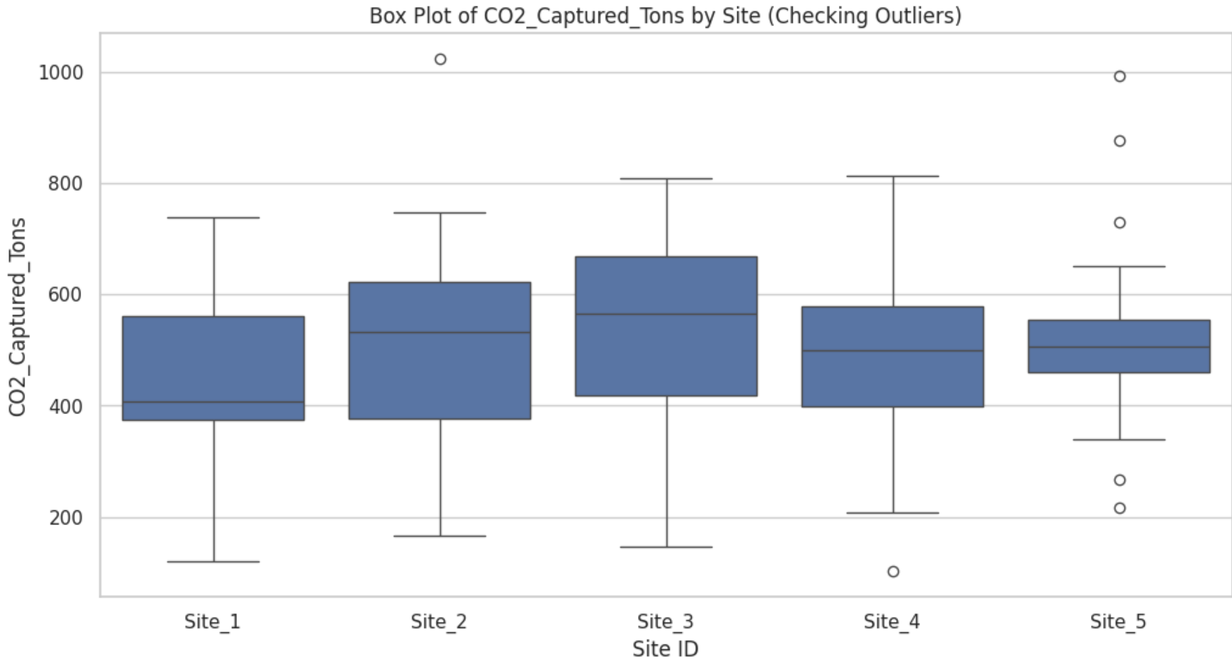


Figure 11: Box plot for CO2 Captured Tons

Operational Costs: The box plot analysis of operational costs across five sites reveals distinct patterns that suggest varying levels of cost management efficiency. Sites 2 and 5 are notable for their controlled operational costs, evidenced by compact distributions and fewer outliers, indicating effective and stable cost management practices. Conversely, Sites 3 and 4 exhibit higher medians and a more comprehensive range of costs with several upper outliers, pointing to potential inefficiencies that may involve occasional spikes in spending due to operational challenges or inefficiencies.

Site 1 presents a unique case with a broader range of operational costs and outliers on the lower end, suggesting periods of significant cost efficiency, potentially from effective management strategies or reduced operational scales. This variability highlights opportunities for cross-site learning to enhance cost efficiency.

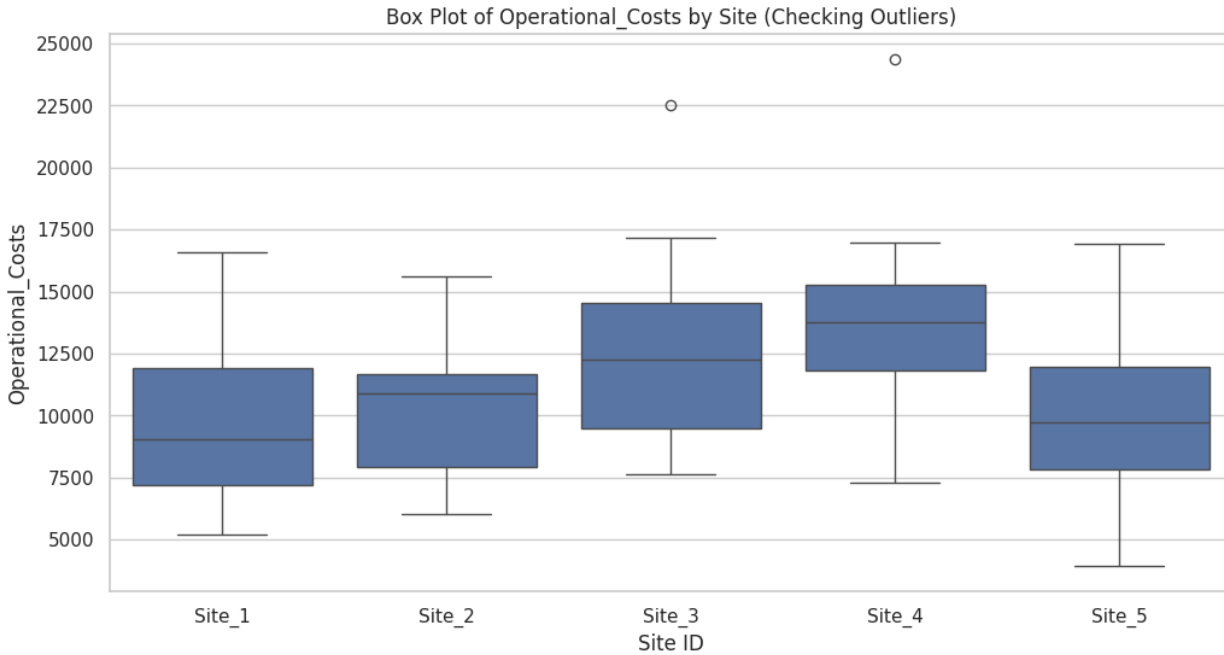
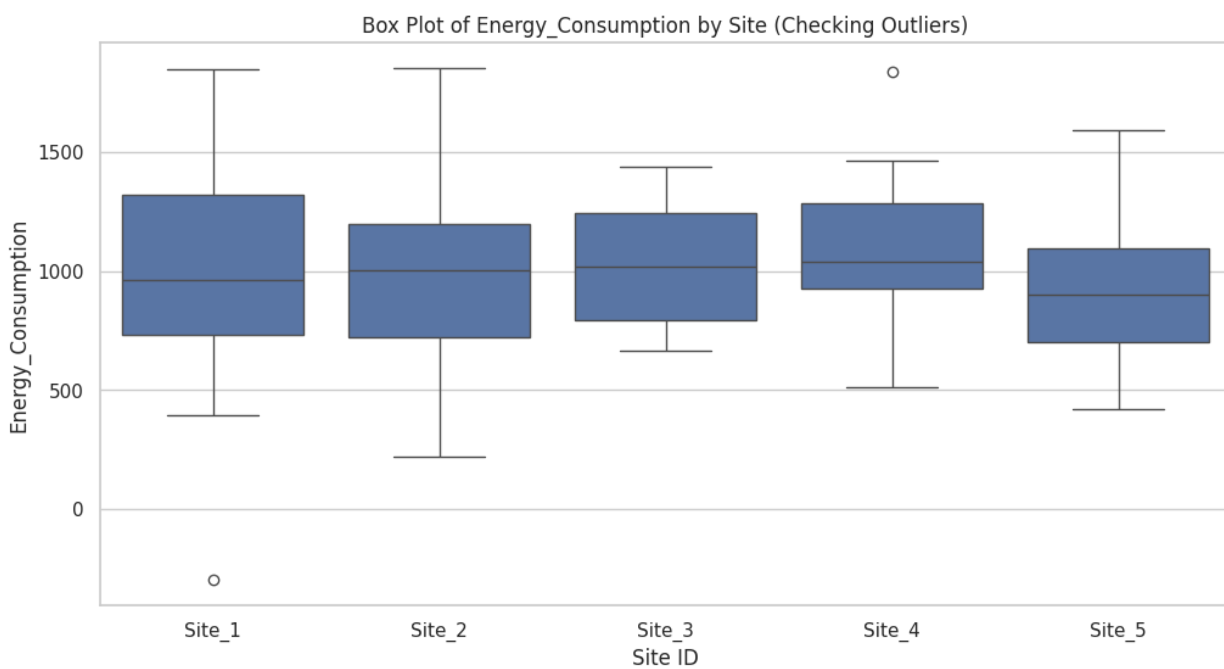
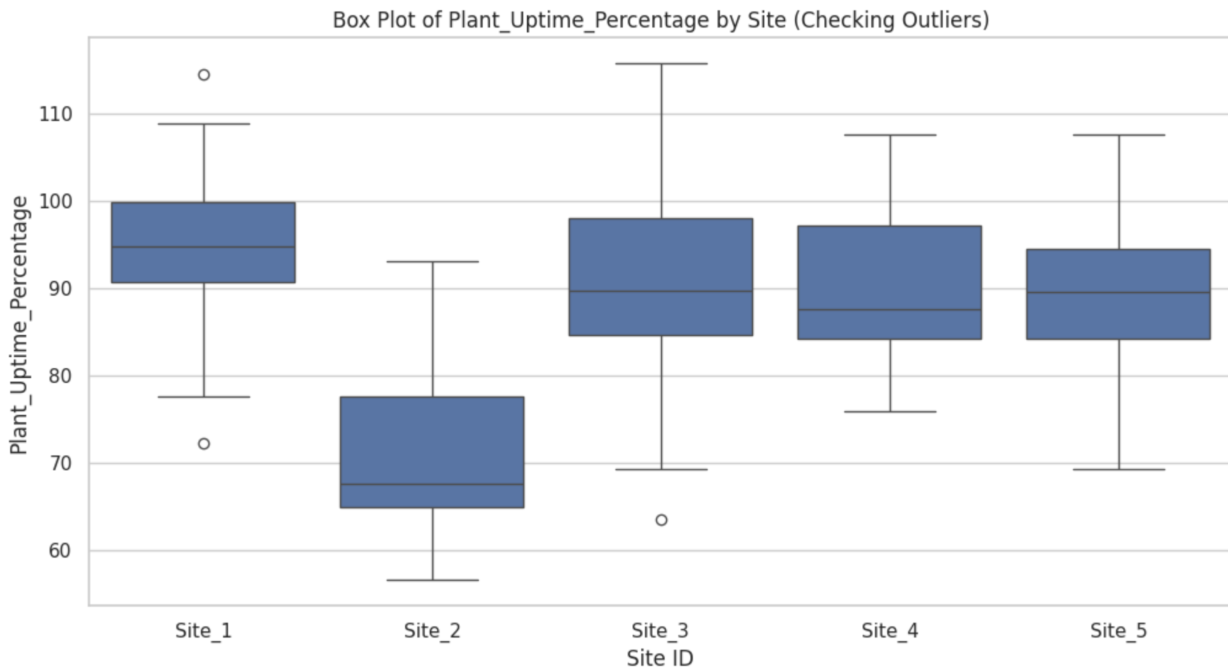


Figure 12: Box Plot for Operational Costs

Energy Consumption: Sites 1 outliers on the lower end, which could indicate periods of lower operational activity or improved energy efficiency measures inconsistent across the dataset.



Plant Uptime Percentage: The Plant Uptime Percentage box plot across the five sites displays a general consistency in operational uptime, with most values clustering around the 80% to 90% range. Notably, Site 1 and Site 3 present a couple of outliers indicating moments of reduced uptime, possibly due to maintenance or operational disruptions. On the other hand, the spread of data in Sites 4 and 5 is tighter around the median, suggesting more stable operations. This visualisation underscores potential areas for operational improvements, especially for sites showing lower uptime outliers. Further, notably, some values exceed 100%, which is impractical and indicates data entry errors.



3.2.1 Outlier Detection Function:

The function defined to flag outliers operates by calculating the first and third quartiles (Q1, Q3) and the interquartile range (IQR) for each site-specific subset of the data. It then defines outliers as any values that fall below Q1 - 1.5IQR or above Q3 + 1.5IQR. This method effectively flags data points statistically distant from the central tendency, which are marked for further analysis or exclusion.

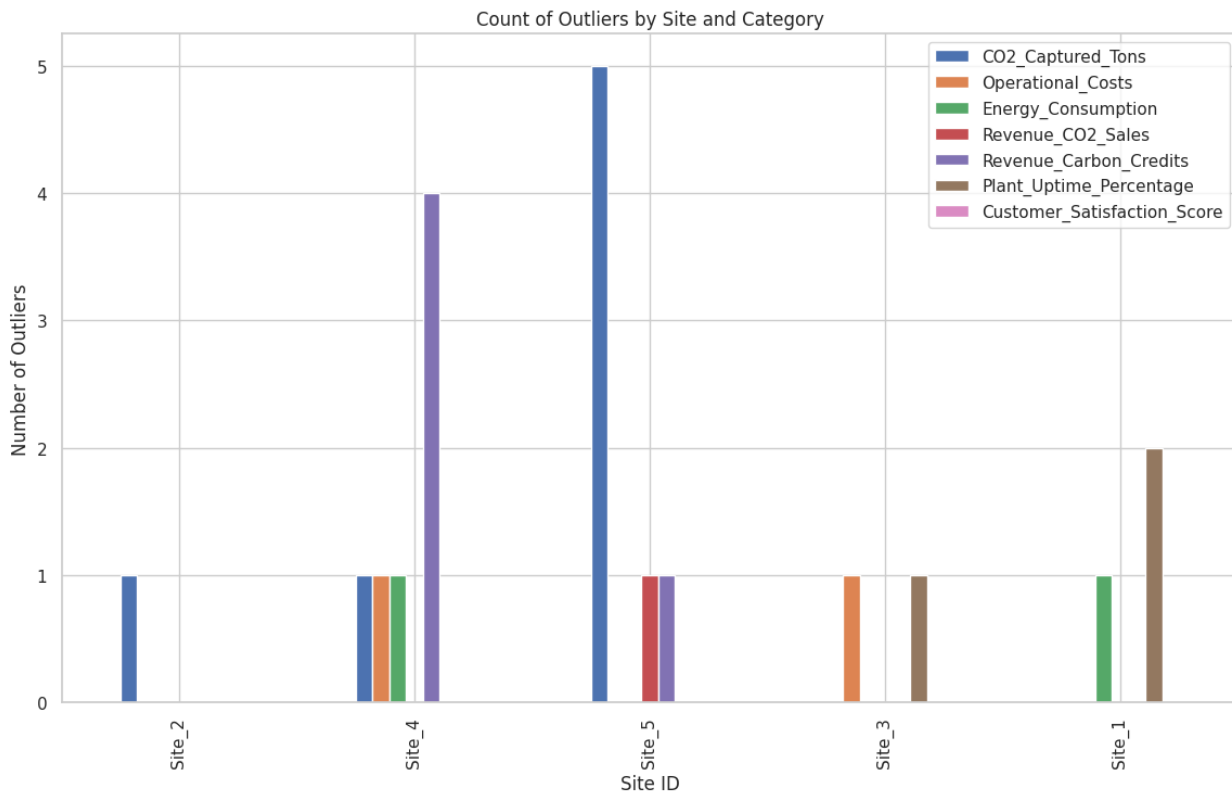


Figure 13: Count of Outliers

The above chart provides a site-wise summary of outlier detection across various parameters. It shows that Site 5 exhibits outliers in nearly every category, particularly in CO2 captured tons and customer satisfaction score, suggesting variability in operational efficiency or data recording practices. Site 4 also shows a significant presence of outliers across operational costs, revenue from CO2 sales, and energy consumption, indicating possible fluctuations in operational management or external factors affecting these metrics. In contrast, Sites 2 and 3 show fewer outliers.

3.2.2 Treatment of Identified Outliers:

Outliers are treated using a cap and floor approach, where values are set to the 1st and 99th percentiles to reduce the impact of extreme data points on the analysis. This method preserves the integrity of the data while minimising the influence of anomalous values. Additionally, for the ‘Plant_Uptime_Percentage’, values above 100% are clipped to 100%, correcting any impossible values in the dataset.

3.3 Data Type Consistency

For data type consistency in the dataset, two key parameters were adjusted:

Date: The 'Date' parameter was converted into a datetime data type. This adjustment is crucial for any operations involving time series analysis, where chronological ordering and time-based filtering are essential. Using the datetime data type allows for more intuitive and accurate handling of dates and times, enabling straightforward computation of intervals, durations, and time-based aggregation.

Site_ID: The 'Site_ID' parameter was transformed into a categorical data type. This change is beneficial when dealing with non-numeric, qualitative data representing discrete categories or groups, such as site identifiers. Converting to a categorical data type reduces memory usage and can speed up operations like sorting and grouping. It also aids in statistical analyses where distinguishing between groups is necessary, as categorical data are handled more efficiently than plain text data.

Both adjustments ensure the data is stored in the most appropriate formats for analysis, enhancing performance and ensuring compatibility with various data manipulation and analysis techniques.

3.4 Feature Engineering

Feature engineering plays a crucial role in enhancing the effectiveness of data analysis by creating new variables that capture essential aspects of the dataset.

3.4.1 CO2 Capture Efficiency

CO2_Capture_Efficiency = CO2_Captured_Tons / Energy_Consumption

This feature calculates the CO2 captured per unit of energy consumed (kWh). It quantifies the energy efficiency of the carbon capture process, providing a direct measure of how effectively energy is utilised. This metric is crucial for assessing operational efficiency. Higher values indicate more efficient CO2 capture per unit of energy, highlighting effective energy use and potentially lower environmental impact.

3.4.2 Revenue per Ton of CO2

Revenue_per_Ton_CO2 = Revenue_CO2_Sales / CO2_Captured_Tons

This feature measures the revenue generated from each ton of CO2 sold. It's calculated by dividing the total revenue from CO2 sales by the total tons of CO2 captured. Understanding revenue per ton is essential for evaluating the financial viability of carbon capture operations. It helps in assessing whether the pricing of captured CO2 aligns with market rates and operational costs.

3.4.3 Profit

Profit = (Revenue_CO2_Sales + Revenue_Carbon_Credits) - Operational_Costs

This feature computes the overall profit by subtracting total operational costs from the combined CO2 sales and carbon credits revenues. It consolidates the primary financial flows into a single profitability metric. Profit is a critical indicator of financial health. Positive values suggest a financially viable operation, whereas negative values could indicate areas needing cost reduction or revenue enhancement.

3.4.4 Cost per Ton

Cost_per_Ton = Operational_Costs / CO2_Captured_Tons

This feature calculates the operational cost per ton of CO2 captured. Cost per ton is vital for understanding the economic efficiency of the capture process. Lower values indicate more cost-effective operations, which are critical for long-term sustainability and competitive pricing.

3.4.5 Rolling Averages

RollingAvg_CO2_Captured = Rolling average of CO2 captured over the past three months.

RollingAvg_Operational_Costs = Rolling average of operational costs over the past three months.

These features use a rolling window to calculate the average CO2 captured and operational costs for each site over specified periods, helping smooth out short-term fluctuations and reveal longer-term trends. Rolling averages provide a clearer view of the data by mitigating the effect of outliers and short-term anomalies. They help in strategic planning and performance assessment over time.

3.4.6 Lagged CO2 Captured

Lagged_CO2_Captured = Previous month's CO2 captured amount.

This feature shifts the CO2 captured data by one month, creating a lagged variable that represents the previous month's capture amount for each site. Lagged features can help identify dependencies and forecast future trends based on previous performance. They allow analysis of how changes in operations affect CO2 capture over time.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 120 entries, 0 to 119
Data columns (total 16 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   Date                          120 non-null    datetime64[ns]
 1   Site_ID                       120 non-null    category
 2   CO2_Captured_Tons             120 non-null    float64
 3   Operational_Costs             120 non-null    float64
 4   Energy_Consumption            120 non-null    float64
 5   Revenue_CO2_Sales             120 non-null    float64
 6   Revenue_Carbon_Credits        120 non-null    float64
 7   Plant_Uptime_Percentage       120 non-null    float64
 8   Customer_Satisfaction_Score   120 non-null    float64
 9   CO2_Capture_Efficiency        120 non-null    float64
 10  Revenue_per_Ton_CO2           120 non-null    float64
 11  Profit                        120 non-null    float64
 12  Cost_per_Ton                  120 non-null    float64
 13  RollingAvg_CO2_Captured       120 non-null    float64
 14  RollingAvg_Operational_Costs  120 non-null    float64
 15  Lagged_CO2_Captured           115 non-null    float64
dtypes: category(1), datetime64[ns](1), float64(14)
memory usage: 14.5 KB
```

Figure 14: Full list of Features

4 Descriptive Statistics

4.1 Summary Statistics

4.1.1 CO2 Captured Tons

Descriptive - Summary Statistivs for: CO2_Captured_Tons

Site_ID	count	mean	std	min	25%	50%	75%	max	mode
Site_1	24.0	441.772106	157.020283	125.532603	374.636181	407.224113	560.384554	737.805448	125.532603
Site_2	24.0	516.666010	183.246882	166.060343	375.845913	531.290393	621.496997	970.720286	166.060343
Site_3	24.0	526.093710	170.563493	147.391969	418.278946	565.534740	668.442879	807.607313	147.391969
Site_4	24.0	499.414069	163.918340	125.532603	398.822933	498.453803	579.230257	812.928731	498.453803
Site_5	24.0	523.875304	163.851186	216.925852	459.416499	506.534367	553.964305	970.720286	524.788976

Figure 15: Summary Statistics for CO2 Captured Tons

Site 2 and Site 3 have higher mean CO2 capture (516.67 and 526.09 tons, respectively) than other sites, indicating potentially higher productivity or efficiency in their operations. Site 1 has the lowest standard deviation (157.02), suggesting more consistent CO2 capture operations than other sites with more comprehensive operation ranges, as indicated by their standard deviations. The maximum values at Site 2 and Site 5 are high, suggesting possible outlier days of very high productivity or reporting errors.

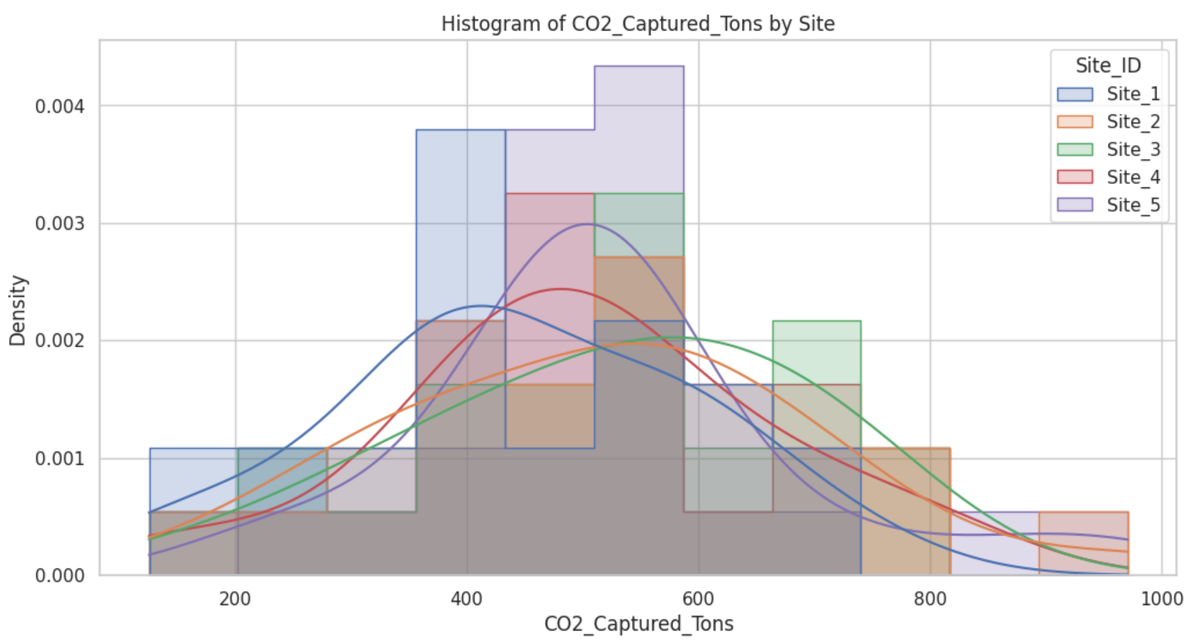


Figure 16: Distribution Analysis for CO2 Captured Tons

4.1.2 Operational Costs

Descriptive - Summary Statistivs for: Operational_Costs

Site_ID	count	mean	std	min	25%	50%	75%	max	mode
Site_1	24.0	9678.289884	3201.503141	5177.550296	7188.393229	9033.815451	11913.702421	16571.366877	9033.815451
Site_2	24.0	10303.911943	2420.168987	6038.630161	7934.180813	10879.217420	11666.783729	15597.323533	10879.217420
Site_3	24.0	12363.757394	3444.695104	7650.215699	9461.879226	12232.174096	14529.524610	21489.976621	12232.174096
Site_4	24.0	13383.924716	3227.176506	7269.968219	11803.592702	13752.381286	15258.875056	21489.976621	7269.968219
Site_5	24.0	9742.517971	3221.241640	4543.297535	7842.791680	9713.129850	11936.016112	16943.975700	4543.297535

Figure 17: Summary Statistics for Operational Cost

Site 3 and Site 4 exhibit higher mean operational costs (12,363.76 and 13,383.92 USD, respectively), indicating larger-scale operations or less efficiency in managing costs. Site 3 shows a high standard deviation, suggesting significant fluctuations in operational costs, possibly due to variable production levels or prices of inputs.

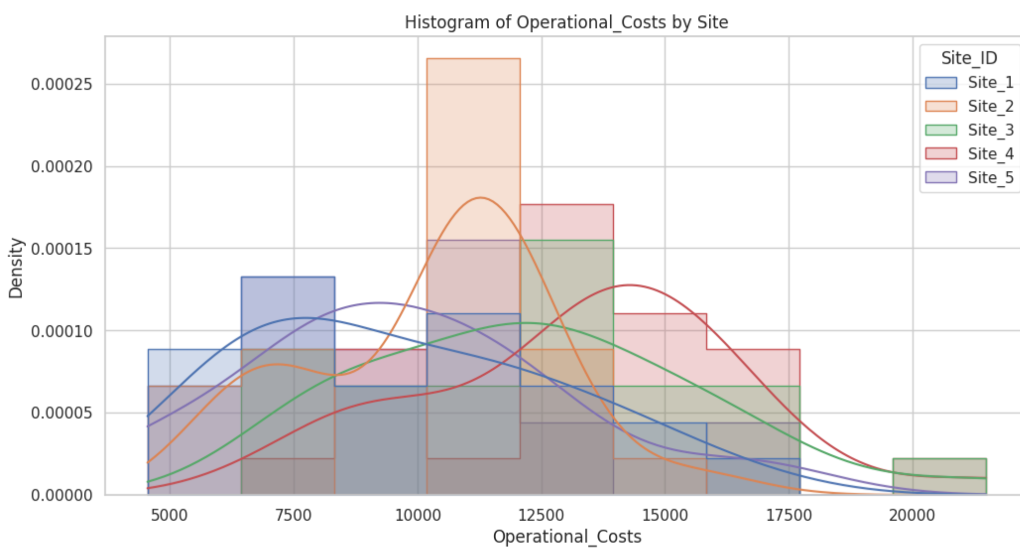


Figure 18: Distribution Analysis for Operational Cost

4.1.3 Energy Consumption

Descriptive - Summary Statistivs for: Energy_Consumption

Site_ID	count	mean	std	min	25%	50%	75%	max	mode
Site_1	24.0	1005.213587	411.128232	252.051436	733.234284	965.147640	1318.965735	1846.600041	965.147640
Site_2	24.0	976.109586	408.661297	252.051436	721.784072	1004.093224	1197.462763	1846.600041	1004.093224
Site_3	24.0	1029.659563	233.762301	668.401996	796.081616	1019.285158	1246.073964	1439.510741	1019.285158
Site_4	24.0	1076.469864	312.643292	514.324555	924.856110	1038.798386	1283.752161	1836.954910	1038.798386
Site_5	24.0	908.102356	275.189554	420.766263	702.244694	899.578842	1097.602821	1590.142487	420.766263

Figure 19: Summary Statistics for Energy Consumption

Lower mean and tightly grouped standard deviations in energy consumption at Sites 2 and 5 suggest more consistent energy use, which could indicate better energy management practices.

4.1.4 Revenue from CO2 Sales

Descriptive - Summary Statistivs for: Revenue_CO2_Sales

Site_ID	count	mean	std	min	25%	50%	75%	max	mode
Site_1	24.0	6588.329449	2090.985599	1906.876678	4886.588075	6794.657134	7888.617982	10713.893480	6794.657134
Site_2	24.0	5071.706218	1518.335089	2430.298206	4126.217754	5007.865550	6057.115113	7814.256259	5007.865550
Site_3	24.0	5350.366425	1675.956451	1941.151733	4439.232882	5635.064635	6489.926502	8091.121887	5635.064635
Site_4	24.0	4589.334804	1337.184965	1906.876678	3572.987973	4866.319941	5556.447570	7442.923318	4866.319941
Site_5	24.0	6376.549294	2379.682496	2051.613335	4746.011554	5827.682339	8075.657940	10713.893480	2051.613335

Figure 20: Summary Statistics for Revenue from CO2 Sales

Site 1 shows the highest mean revenue from CO2 sales, potentially indicating higher prices obtained or more effective sales strategies. Site 4 shows the lowest standard deviation in revenue, which could imply a stable pricing or sales volume environment.

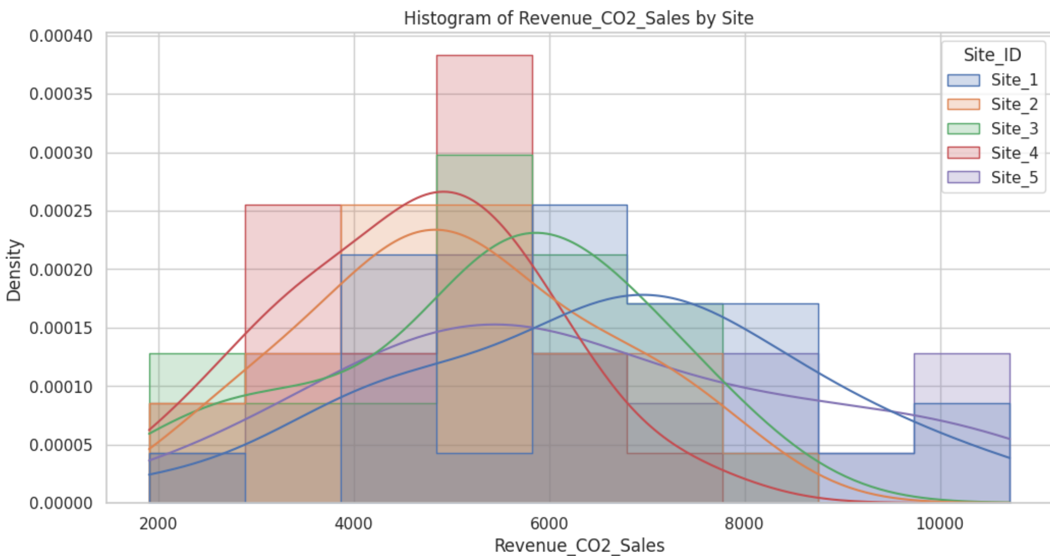


Figure 21: Distribution Analysis for Revenue from CO2 Sales

4.1.5 Revenue from Carbon Credits

Descriptive - Summary Statistivs for: Revenue_Carbon_Credits

Site_ID	count	mean	std	min	25%	50%	75%	max	mode
Site_1	24.0	2766.997139	1198.940261	1072.264416	1828.839757	2771.593586	3653.909810	5024.569962	2771.593586
Site_2	24.0	2523.144677	897.034527	834.141479	1992.947664	2361.661971	3170.536123	3906.469471	2361.661971
Site_3	24.0	2673.788158	1035.649357	386.084086	2024.670260	2647.439442	3492.972539	4729.882345	2647.439442
Site_4	24.0	2960.809329	1280.791584	386.084086	2491.346511	2879.590710	3603.828602	5401.957869	386.084086
Site_5	24.0	3151.415240	920.063867	1225.696505	2462.327796	3282.737470	3620.916403	5401.957869	3282.737470

Figure 22: Summary Statistics for Revenue from Carbon Credits

Site 5 shows the highest mean revenue from carbon credits, possibly due to more effective participation in carbon credit markets or more favourable regulatory conditions.

4.1.6 Plant Uptime Percentage

Descriptive - Summary Statistivs for: Plant_Uptime_Percentage

Site_ID	count	mean	std	min	25%	50%	75%	max	mode
Site_1	24.0	94.298231	10.228699	72.212798	90.681499	94.714684	99.885595	113.386438	94.714684
Site_2	24.0	70.782861	9.414343	59.969587	64.813187	67.521133	77.521952	92.988981	59.969587
Site_3	24.0	90.444126	11.392315	63.490302	84.557167	89.715608	97.950315	113.386438	63.490302
Site_4	24.0	89.866717	8.841509	75.924883	84.121556	87.596746	97.114696	107.522704	87.596746
Site_5	24.0	89.268358	9.979581	69.180706	84.152248	89.527098	94.508546	107.547942	69.180706

Figure 23: Summary Statistics for Plant Uptime Percentage

Site 1 shows the highest mean uptime percentage, possibly better maintenance practices or fewer technical issues. Site 2 has a significantly lower average uptime, which might highlight operational challenges or maintenance issues.

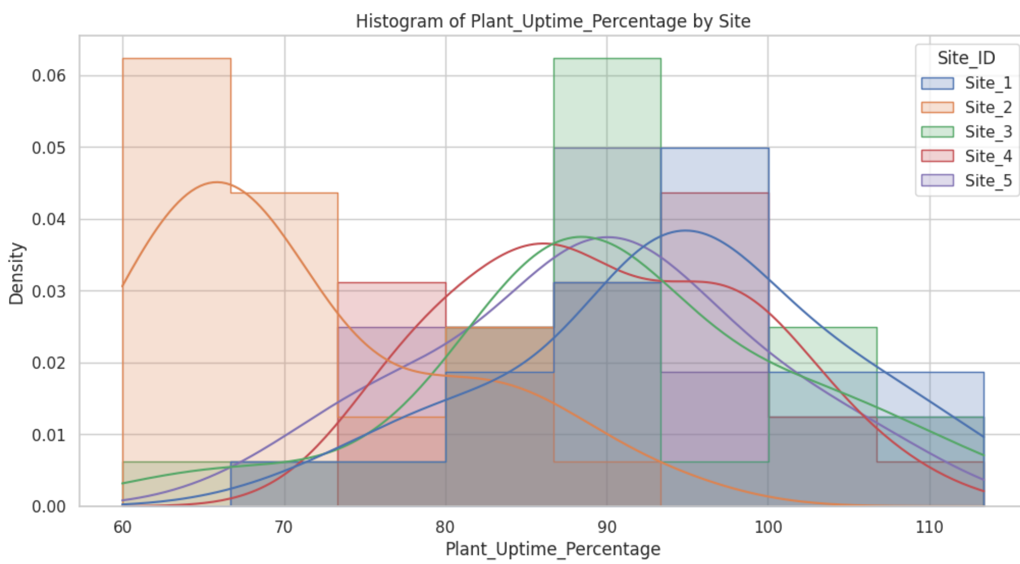


Figure 24: Distribution Analysis for Plant Uptime Percentage

4.1.7 Profit

Descriptive - Summary Statistivs for: Profit

Site_ID	count	mean	std	min	25%	50%	75%	max	mode
Site_1	24.0	-322.963296	4115.361563	-9158.879270	-2108.368344	-574.512720	2196.653934	7425.968394	-9158.879270
Site_2	24.0	-2709.061049	3074.750567	-9565.924702	-4504.515831	-2987.444306	-1168.116192	3738.837174	-9565.924702
Site_3	24.0	-4339.602811	3741.493172	-14080.969143	-7059.087062	-3875.696893	-1394.096646	2183.517569	-14080.969143
Site_4	24.0	-5833.780582	3975.128010	-15500.613085	-8205.496166	-6876.650785	-2707.340859	214.784237	-15500.613085
Site_5	24.0	-214.553437	3769.738124	-6170.951370	-2969.654172	-653.425799	1972.612117	6547.616441	-6170.951370

Figure 25: Summary Statistics for Profit

Negative mean profits at all sites suggest operational challenges or high costs relative to revenues—site 4 exhibits exceptionally high losses on average, which could be a significant concern.

5 Temporal Analysis

Verifying whether the data is stationary is fundamental in analysing time series data. Stationarity refers to the property of a time series to have constant statistical characteristics over time, such as mean, variance, and autocorrelation. This is crucial because many predictive modelling techniques and statistical tests assume stationarity as a prerequisite. Non-stationary data can lead to misleading trends and results in time series analysis, making it challenging to model and forecast accurately. Therefore, checking for stationarity helps choose the appropriate analytical and forecasting tools and ensures reliable predictions and insights.

One of the most commonly used statistical tests for checking stationarity is the **Augmented Dickey-Fuller** (ADF) test, available in the *statmodels.tsa.stattools* module as *adfuller*. This test is designed to determine the presence of a unit root in a series, which is a telltale sign of non-stationarity. The ADF test operates under the null hypothesis that the time series has a unit root, thus indicating it is non-stationary.

The ADF test statistic is compared against critical values for different confidence levels. If the test statistic is less than the critical value, the non-stationary time series null hypothesis is rejected. This implies that the series does not have a unit root, suggesting that it is stationary. On the other hand, if the test statistic is greater than the critical value, we fail to reject the null hypothesis, indicating that the time series is likely non-stationary.

5.1 Site 1 Analysis

For Site 1, the CO_2 capture rates and operational costs do not exhibit stationary properties as indicated by their respective Dickey-Fuller test statistics (-0.927825 and -1.974626) and high p-values (0.778671 and 0.297808), suggesting the presence of a unit root. The energy consumption, however, shows stationarity with a test statistic of -4.671094 and a p-value significantly below

0.05, indicating no unit root. Revenue from CO2 sales is marginally stationary (p-value of 0.028491), while the revenue from carbon credits is not stationary, with a p-value of 0.069725.

```
Dickey -Fuller Test: Site_1 parameter: CO2_Captured_Tons
Results of Dickey-Fuller Test for CO2_Captured_Tons at Site_1:
Test Statistic                 -0.927825
p-value                         0.778671
#Lags Used                      8.000000
Number of Observations Used    15.000000
Critical Value (1%)            -3.964443
Critical Value (5%)            -3.084908
Critical Value (10%)           -2.681814
dtype: float64
```

```
Dickey -Fuller Test: Site_1 parameter: Operational_Costs
Results of Dickey-Fuller Test for Operational_Costs at Site_1:
Test Statistic                 -1.974626
p-value                         0.297808
#Lags Used                      4.000000
Number of Observations Used    19.000000
Critical Value (1%)            -3.832603
Critical Value (5%)            -3.031227
Critical Value (10%)           -2.655520
dtype: float64
```

```
Dickey -Fuller Test: Site_1 parameter: Energy_Consumption
Results of Dickey-Fuller Test for Energy_Consumption at Site_1:
Test Statistic                 -4.671094
p-value                         0.000095
#Lags Used                      1.000000
Number of Observations Used    22.000000
Critical Value (1%)            -3.769733
Critical Value (5%)            -3.005426
Critical Value (10%)           -2.642501
dtype: float64
```

```
Dickey -Fuller Test: Site_1 parameter: Revenue_CO2_Sales
Results of Dickey-Fuller Test for Revenue_CO2_Sales at Site_1:
Test Statistic                 -3.074601
p-value                         0.028491
#Lags Used                      0.000000
Number of Observations Used    23.000000
Critical Value (1%)            -3.752928
Critical Value (5%)            -2.998500
Critical Value (10%)           -2.638967
dtype: float64
```

```
Dickey -Fuller Test: Site_1 parameter: Revenue_Carbon_Credits
Results of Dickey-Fuller Test for Revenue_Carbon_Credits at Site_1:
Test Statistic                 -2.725522
p-value                         0.069725
#Lags Used                      9.000000
Number of Observations Used    14.000000
Critical Value (1%)            -4.012034
Critical Value (5%)            -3.104184
Critical Value (10%)           -2.690987
dtype: float64
```

Figure 26: Site 1 - Dickey-Fuller test results

5.2 Site 2 Analysis

At Site 2, all tested parameters—CO2 captured tons, operational costs, energy consumption, and revenues from CO2 sales and carbon credits—show stationarity, with very low p-values (all below 0.05) and test statistics more negative than the critical values. This suggests that these time series do not have unit roots and are suitable for traditional time series forecasting and analyses without the need for differencing.

```
Dickey -Fuller Test: Site_2 parameter: CO2_Captured_Tons
Results of Dickey-Fuller Test for CO2_Captured_Tons at Site_2:
Test Statistic                 -4.618445
p-value                         0.000119
#Lags Used                      0.000000
Number of Observations Used    23.000000
Critical Value (1%)            -3.752928
Critical Value (5%)            -2.998500
Critical Value (10%)           -2.638967
dtype: float64
```

Figure 27: Site 2 - Dickey-Fuller test results for CO2 Capture Tons

5.3 Site 3 Analysis

For Site 3, the CO2 capture rates are close to being stationary but still above the significance level with a p-value of 0.078516. Operational costs and energy consumption exhibit stationary characteristics with p-values of 0.007843 and 0.001556, respectively. However, CO2 sales and carbon credit revenues are not stationary, with p-values considerably above 0.05, indicating potential underlying trends or seasonal components in these financial metrics.

```
Dickey -Fuller Test: Site_3 parameter: CO2_Captured_Tons
Results of Dickey-Fuller Test for CO2_Captured_Tons at Site_3
Test Statistic                 -2.674721
p-value                         0.078516
#Lags Used                      9.000000
Number of Observations Used    14.000000
Critical Value (1%)            -4.012034
Critical Value (5%)            -3.104184
Critical Value (10%)           -2.690987
dtype: float64
```

```
Dickey -Fuller Test: Site_3 parameter: Operational_Costs
Results of Dickey-Fuller Test for Operational_Costs at Site_3:
Test Statistic                 -3.505647
p-value                         0.007843
#Lags Used                      0.000000
Number of Observations Used    23.000000
Critical Value (1%)            -3.752928
Critical Value (5%)            -2.998500
Critical Value (10%)           -2.638967
dtype: float64
```

```
Dickey -Fuller Test: Site_3 parameter: Energy_Consumption
Results of Dickey-Fuller Test for Energy_Consumption at Site_3:
Test Statistic                 -3.973477
p-value                         0.001556
#Lags Used                      1.000000
Number of Observations Used    22.000000
Critical Value (1%)            -3.769733
Critical Value (5%)            -3.005426
Critical Value (10%)           -2.642501
dtype: float64
```

```
Dickey -Fuller Test: Site_3 parameter: Revenue_CO2_Sales
Results of Dickey-Fuller Test for Revenue_CO2_Sales at Site_3:
Test Statistic                 -1.809565
p-value                         0.375724
#Lags Used                      9.000000
Number of Observations Used    14.000000
Critical Value (1%)            -4.012034
Critical Value (5%)            -3.104184
Critical Value (10%)           -2.690987
dtype: float64
```

```
Dickey -Fuller Test: Site_3 parameter: Revenue_Carbon_Credits
Results of Dickey-Fuller Test for Revenue_Carbon_Credits at Site_3:
Test Statistic                 -2.430330
p-value                         0.133358
#Lags Used                      8.000000
Number of Observations Used    15.000000
Critical Value (1%)            -3.964443
Critical Value (5%)            -3.084908
Critical Value (10%)           -2.681814
dtype: float64
```

Figure 28: Site 3 - Dickey-Fuller test results

5.4 Site 4 Analysis

Site 4 presents a mixed result where CO2 capture rates are stationary (p-value of 0.021149), and the operational costs and energy consumption also display stationarity (p-values of 0.000101 and less than 0.00001, respectively). However, the revenue from carbon credits fails to reject the null hypothesis of non-stationarity with a p-value of 0.991868, suggesting it may require differencing or transformation for further time series analysis.

5.5 Site 5 Analysis

At Site 5, CO2 capture rates and operational costs show strong evidence of stationarity with p-values significantly lower than 0.05, indicating the absence of a unit root. However, energy consumption and revenue from CO2 sales do not show stationarity with p-values of 0.273125 and 0.101534, respectively. Interestingly, revenue from carbon credits at this site is stationary, with a p-value near zero, contrasting with the revenue from CO2 sales.

These analyses highlight the importance of treating each site's data individually, recognising the unique characteristics of each operational metric. Stationary series are more straightforward to model and predict, whereas non-stationary series might require transformations or differencing to stabilise the mean and variance over time.

5.6 Introduction to Seasonal Decomposition

Another essential tool in time series analysis is *seasonal_decompose* from the *statsmodels.tsa.seasonal module*. This function decomposes a time series into its seasonal, trend, and residual components, making it easier to understand and model. Seasonal decomposition is particularly useful in identifying and quantifying the seasonal patterns and trends within the data, which are critical for accurate forecasting and strategic decision-making.

While *seasonal_decompose* can technically be applied to both stationary and non-stationary data, the interpretation of the decomposed components can vary significantly between the two. The decomposition can reliably isolate the trend and seasonal effects for stationary data since the mean and variance are constant over time. However, applying seasonal decomposition to non-stationary data can lead to misleading results because the changing mean and variance can distort the seasonal and trend components. Therefore, it is often recommended to first stabilise the series (make it stationary) before applying seasonal decomposition to ensure that the resulting components reflect true underlying patterns rather than artefacts of non-stationarity.

The seasonal decomposition graphs for Site 1 across various parameters provide a detailed look into how each metric behaves over time, split into trend, seasonal, and residual components.

Analysis of Seasonal Trends

CO2 Captured Tons

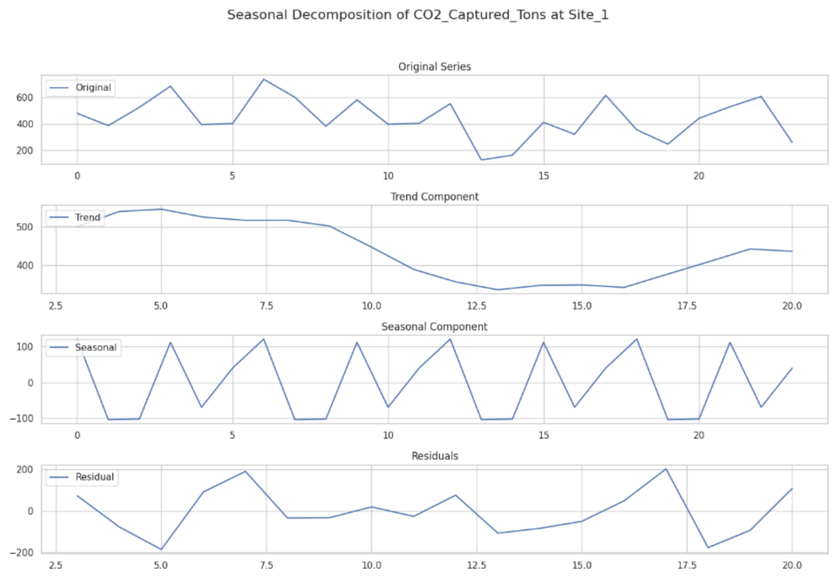


Figure 29: Site 1 - Seasonal Decomposition - CO2 Captured Tons

Exhibits a clear seasonal pattern, repeating roughly every six months. Peaks are consistent and suggest higher CO2 capture at specific times of the year, possibly related to production cycles or operational enhancements during certain months.

Operational Costs

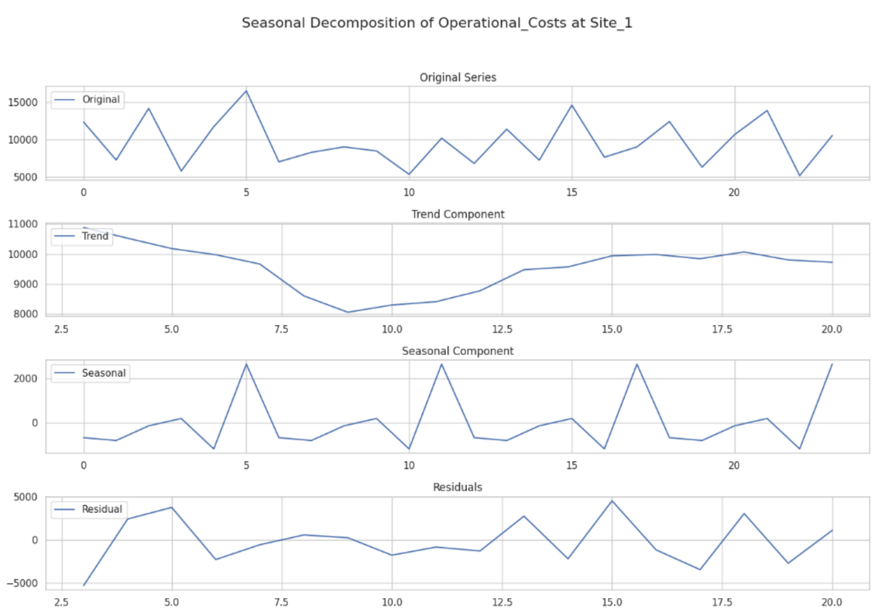


Figure 30: Site 1 - Seasonal Decomposition - Operational Costs

Like CO_2 captured tons, the operational costs show a repeating seasonal pattern every six months. Peaks in operational costs often correspond with peaks in CO_2 captured, which may indicate higher costs incurred during increased production or capture activity periods.

Energy Consumption

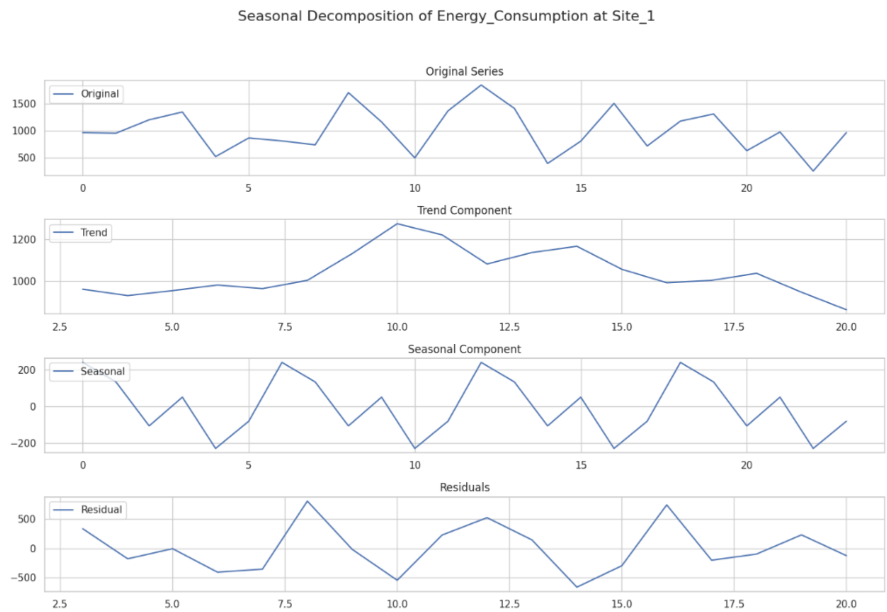


Figure 31: Site 1 - Seasonal Decomposition - Energy Consumption

The seasonal component also repeats every six months but with less amplitude than CO_2 capture and operational costs, indicating some degree of decoupling between energy consumption and CO_2 capture volumes. The trend component is increasing, suggesting a gradual rise in energy consumption over time, potentially due to scaling up operations or decreasing energy efficiency.

Revenue from CO2 Sales

Exhibits clear seasonal fluctuations with a similar frequency to the other parameters, indicating that revenue peaks follow the same seasonal pattern. The trend is downwards, suggesting that the overall revenue from CO_2 sales is decreasing despite seasonal peaks.

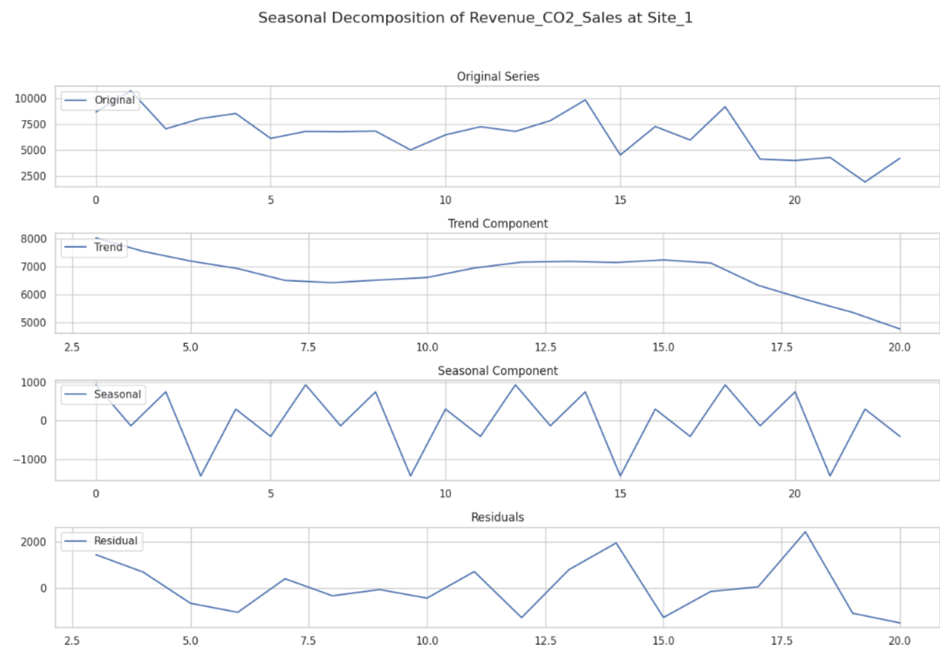


Figure 32: Site 1 - Seasonal Decomposition - Revenue CO2 Sales

Revenue from Carbon Credits

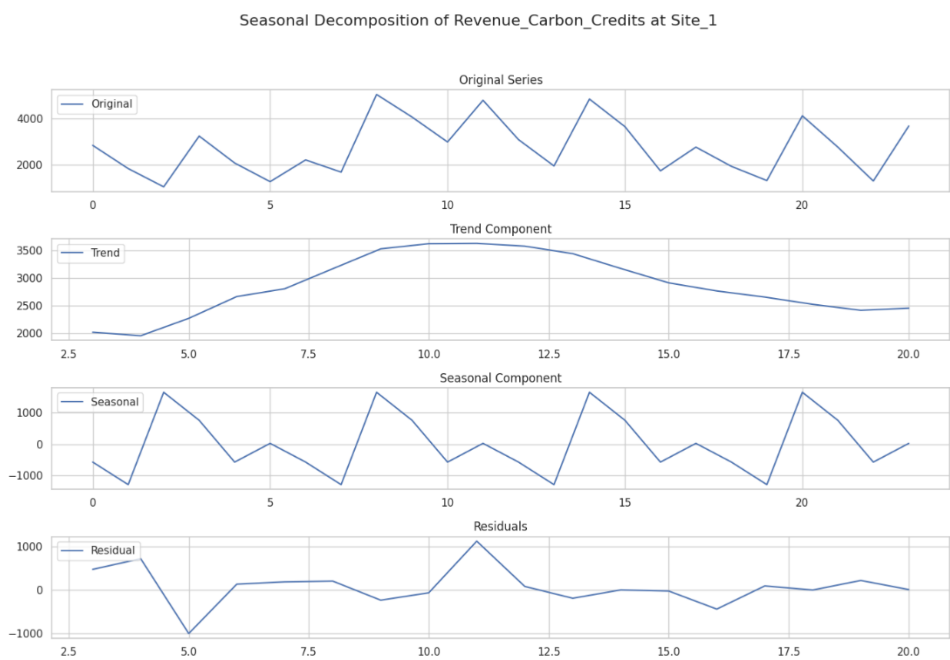


Figure 33: Site 1 - Seasonal Decomposition - Revenue from Carbon Credits

This parameter also shows a clear seasonal pattern that aligns with the others in frequency. Unlike CO2 sales, the trend for carbon credit revenue is upward, potentially indicating an increasing reliance or efficiency in generating revenue from carbon credits instead of direct CO2 sales.

All parameters display a standard seasonal frequency of about every six months. This consistency across different metrics suggests that the underlying business cycles or operational factors influencing one similarly impact the others. Strong seasonal components in all parameters suggest that any predictive modelling or forecasting should account for these seasonal effects to ensure accuracy. While all parameters share similar seasonal frequencies, their trends differ significantly. Operational costs and CO_2 sales revenues are decreasing, whereas energy consumption and carbon credit revenues are increasing. This trend divergence could be key to understanding the site's financial health and operational efficiency. Considering these differences when planning operational improvements or financial strategies is crucial, particularly in balancing operational costs against the revenues from CO_2 sales and carbon credits. Further analysis might delve deeper into what specific months or periods the peaks occur and how these align with external factors like market conditions or internal operational changes.

6 Correlation Analysis

The detailed site-wise correlation matrices highlight relationships between key operational metrics across different sites. Below is an analysis of the relationships indicated by the correlation coefficients for each site, which provides insights into how different factors may be interacting within specific contexts:

6.1 Site 1 Analysis:

CO_2 Captured and Operational Costs have a negative correlation (-0.171670), suggesting that higher CO_2 capture does not necessarily lead to increased operational costs at this site, which could indicate efficient capture processes. Customer Satisfaction is positively correlated with CO_2 Captured Tons (0.238925) and Operational Costs (0.264730), suggesting that higher satisfaction may be associated with higher operational activity and possibly better customer-perceived performance. Revenue from CO_2 Sales shows a negative correlation with CO_2 Captured Tons (-0.224393), indicating that increased capture does not directly translate to increased revenue from sales, which market prices or strategies could influence.

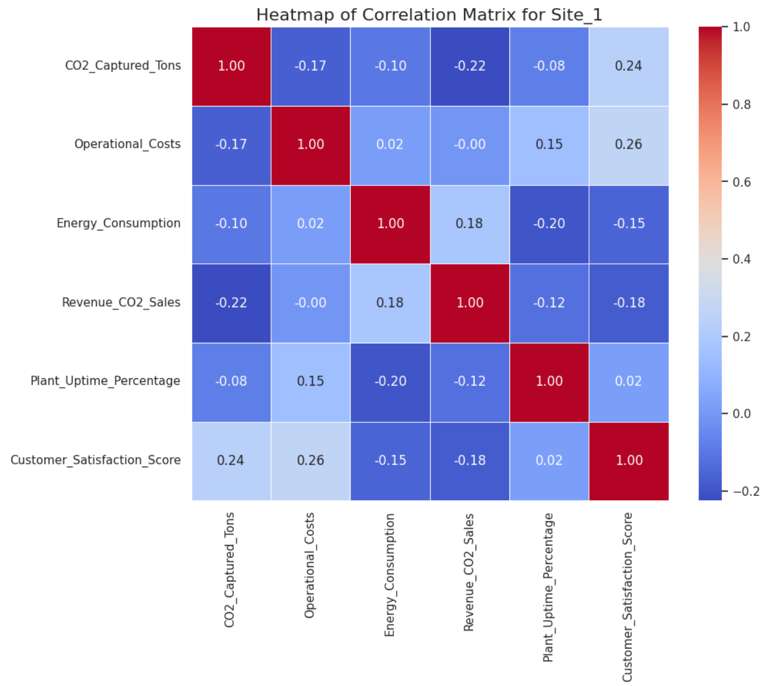


Figure 34: Correlation Heatmap for Site 1

6.2 Site 2 Analysis:

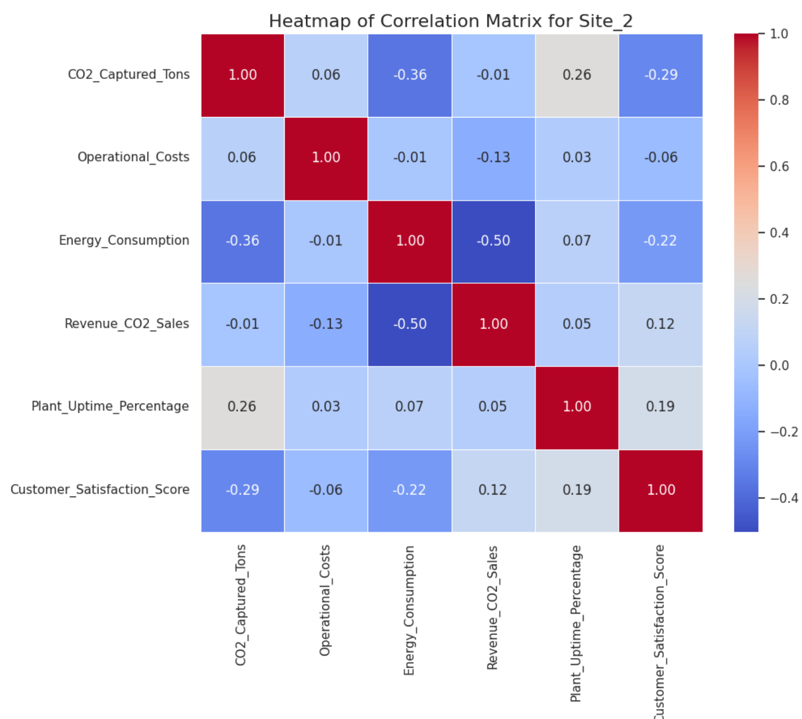


Figure 35: Correlation Heatmap for Site 2

Energy Consumption has a strong negative correlation with Revenue CO2 Sales (-0.502219), and CO2 Captured Tons (-0.358025), indicating that higher energy consumption at this site does not correspond with higher revenue or CO2 capture, suggesting inefficiencies. Customer Satisfaction is negatively correlated with CO2 Captured Tons (-0.292514), which could indicate issues with the process or other factors negatively impacting customer perceptions as CO2 capture increases.

6.3 Site 3 Analysis:

Customer satisfaction has a very strong positive correlation with CO2 Captured Tons (0.845508), highlighting that customer satisfaction is highly dependent on the volume of CO2 captured, potentially reflecting direct benefits to customers from higher capture rates. Operational Costs and CO2 Captured got a positive correlation (0.373822), suggesting that increases in CO2 captured are associated with higher costs, possibly due to scaling challenges or increased resource usage.

6.4 Site 4 Analysis:

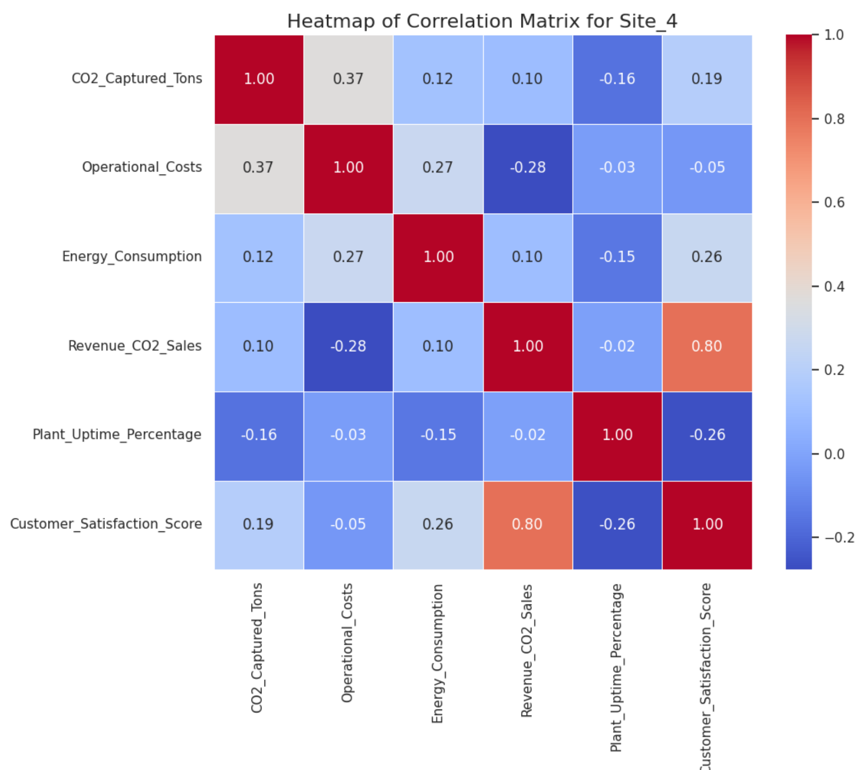


Figure 36: Correlation Heatmap for Site 4

Revenue from CO_2 Sales and Customer Satisfaction has a high positive correlation (0.796830), indicating that revenue from CO_2 sales substantially impacts customer satisfaction at this site, possibly due to direct benefits perceived by the customers from sales activities. Energy Consumption and Operational Costs have a positive correlation (0.267480), implying that energy usage and costs go hand in hand, which is typical but could be a point for efficiency improvements.

6.5 Site 5 Analysis:

Revenue from CO_2 Sales and CO_2 Captured have a positive correlation (0.261649), which indicates that more CO_2 captured leads to higher revenue, unlike some other sites. This could be due to more effective sales strategies or better market conditions. Customer Satisfaction and Plant Uptime have a positive correlation (0.328898), which suggests that higher plant uptime, which indicates reliability and efficiency, positively influences customer satisfaction.

The relationships between variables like CO_2 captured, operational costs, and revenue from CO_2 sales vary significantly from site to site, which might reflect differences in site-specific operational practices, market conditions, or technology used. It often shows a negative correlation with revenue and CO_2 capture efficiency in several sites, highlighting a potential area for improvement across the board in energy use efficiency. Various factors influence Customer Satisfaction but show particularly strong correlations with CO_2 capture and revenue metrics in several sites, underlining the importance of these factors to customer perceptions and satisfaction.

7 Operational Efficiency Analysis

The operational efficiency of carbon capture processes at different sites reveals a complex interplay between energy consumption, CO_2 captured, and plant uptime. The line graphs and correlation matrix provided for Sites 1 through 5 offer a detailed look at these relationships, highlighting potential improvement and optimisation areas.

Process Optimisation

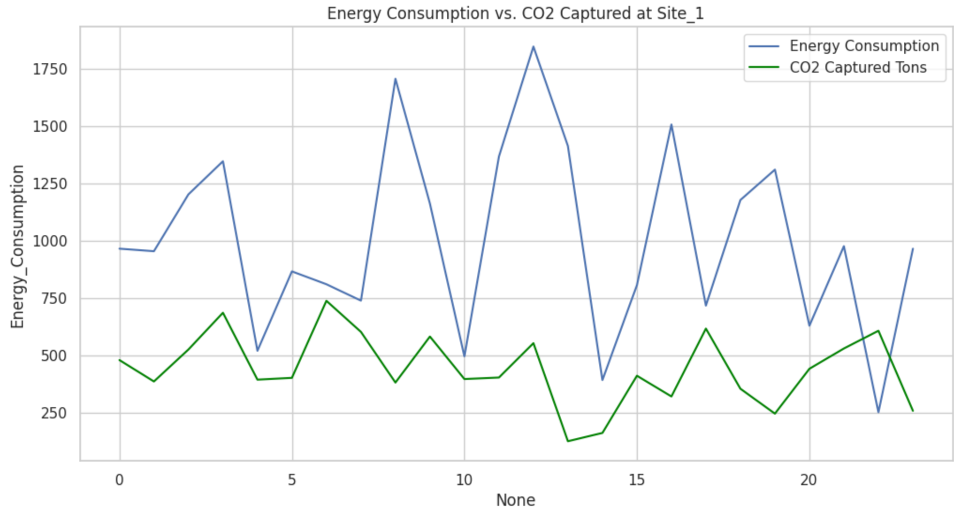


Figure 37: Site 1 - Energy Consumption Vs. CO2 Captured

Analysis of the operational data shows a fluctuating relationship between energy consumption and CO2 captured across the sites. For instance, at Site 1 and Site 4, we see significant spikes in energy consumption that do not always correspond to increases in CO2 capture, suggesting potential inefficiencies in the energy use or capture processes. This pattern calls for a detailed investigation into the specific stages of the carbon capture process where energy use is highest and least effective. Pinpointing these stages can lead to targeted interventions, such as adjusting operational parameters, updating equipment, or redesigning process flows to reduce energy wastage while maintaining or increasing CO2 capture rates.

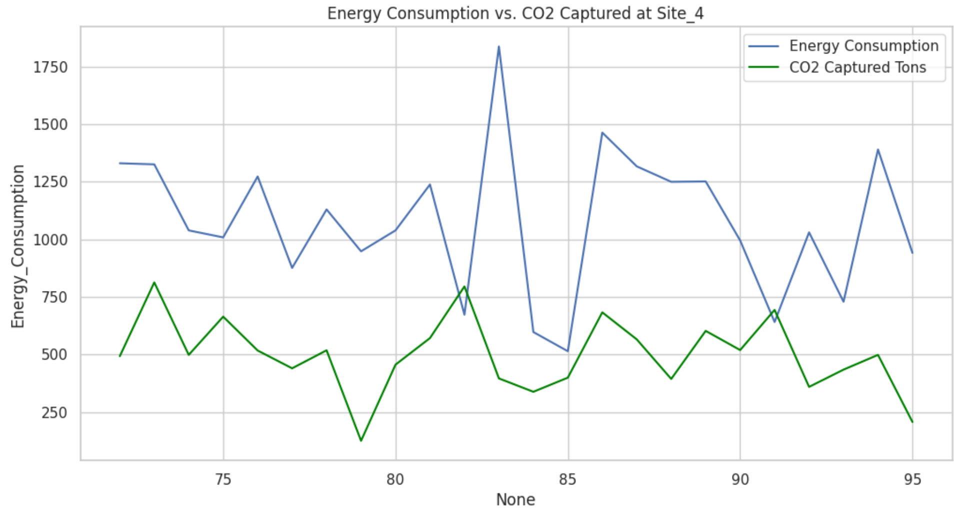


Figure 38: Site 4 - Energy Consumption Vs. CO2 Captured

Capacity Utilisation

The correlation matrices and the graphs reveal that there is generally a weak relationship between plant uptime and both CO_2 captured and energy consumption, as shown by the low correlation coefficients (around -0.07). This suggests that the plants might not be fully optimised in terms of operation times and capacity usage. For example, even when plants are operational, they may not capture CO_2 as efficiently as possible or consume more energy than necessary for the volume of CO_2 captured.

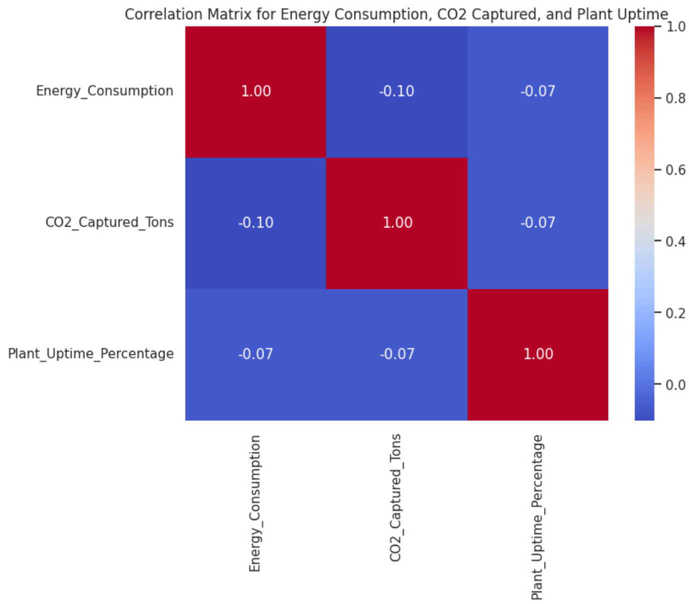


Figure 39: Correlation Matrix

Increasing capacity utilisation could involve:

Scheduling Adjustments: Aligning operation times more closely with periods of high efficiency to maximise output while minimising idle or less productive periods.

Process Automation and Control Optimisation: Implementing advanced control strategies that adjust the operation dynamically based on real-time data could help maintain optimal conditions for carbon capture, thus maximising throughput and minimising energy expenditure.

Maintenance and Upgrades: Regular maintenance and timely upgrades to the carbon capture infrastructure can ensure the equipment operates at peak efficiency, reducing downtimes and improving overall capacity utilisation.

Site-Specific Observations:

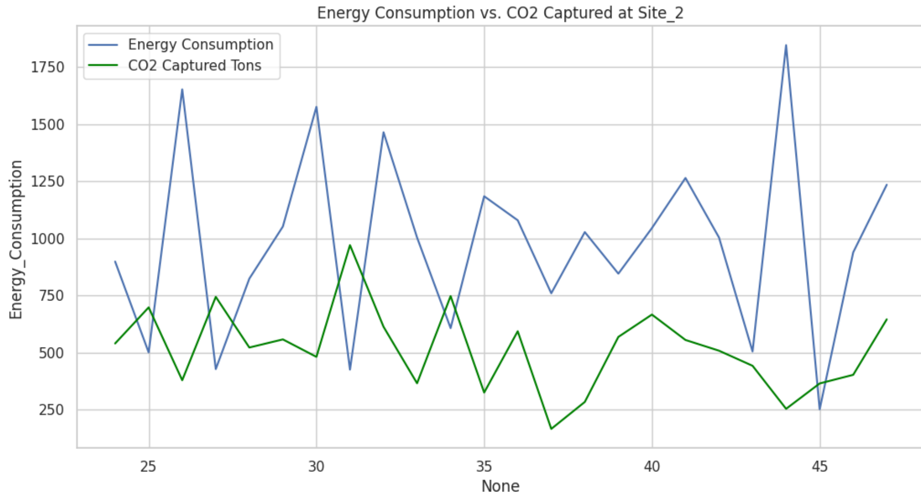


Figure 40: Site 2 - Energy Consumption Vs. CO2 Captured

Site 2 and Site 5 show relatively stable CO2 capture with less variation in energy consumption, suggesting that these sites may have more efficient or newer technologies.

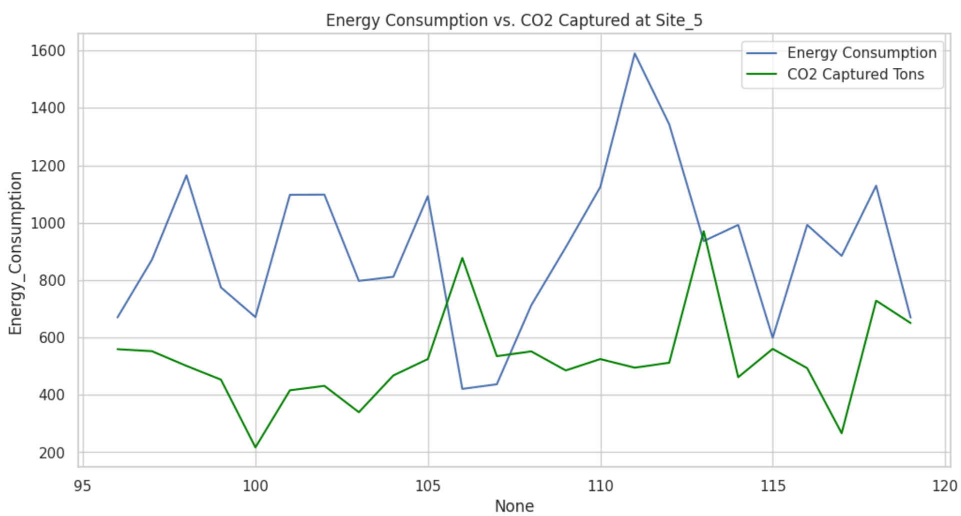


Figure 41: Site 5 - Energy Consumption Vs. CO2 Captured

Site 3, on the other hand, shows the highest fluctuations between energy use and CO2 capture, indicating possible operational inefficiencies or older, less efficient technology.

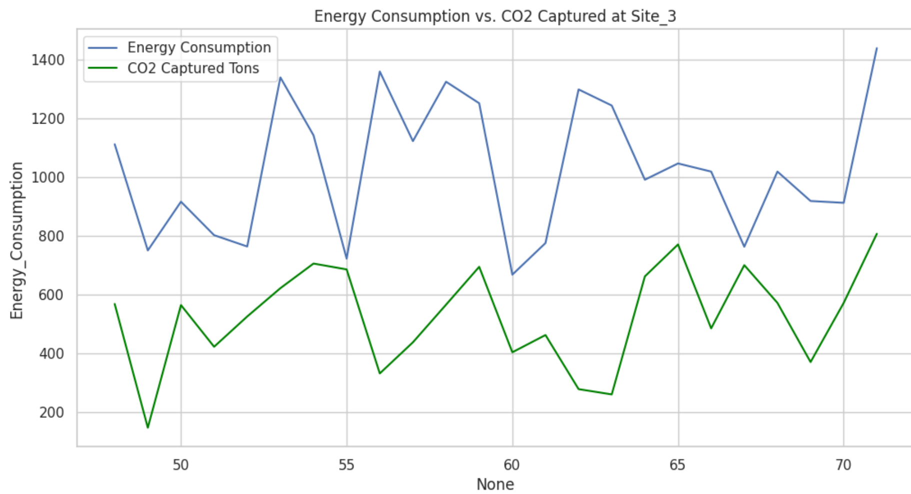


Figure 42: Site 3 - Energy Consumption Vs. CO2 Captured

In conclusion, enhancing operational efficiency in carbon capture processes across these sites will require a multifaceted approach. This includes detailed data analysis to understand energy consumption patterns, CO2 capture rates, and plant operational schedules. Implementing targeted process optimisations and capacity utilisation strategies can significantly improve efficiency, cost savings, and environmental impact reduction. These efforts will not only optimise carbon capture operations but also enhance the overall sustainability of the facilities.

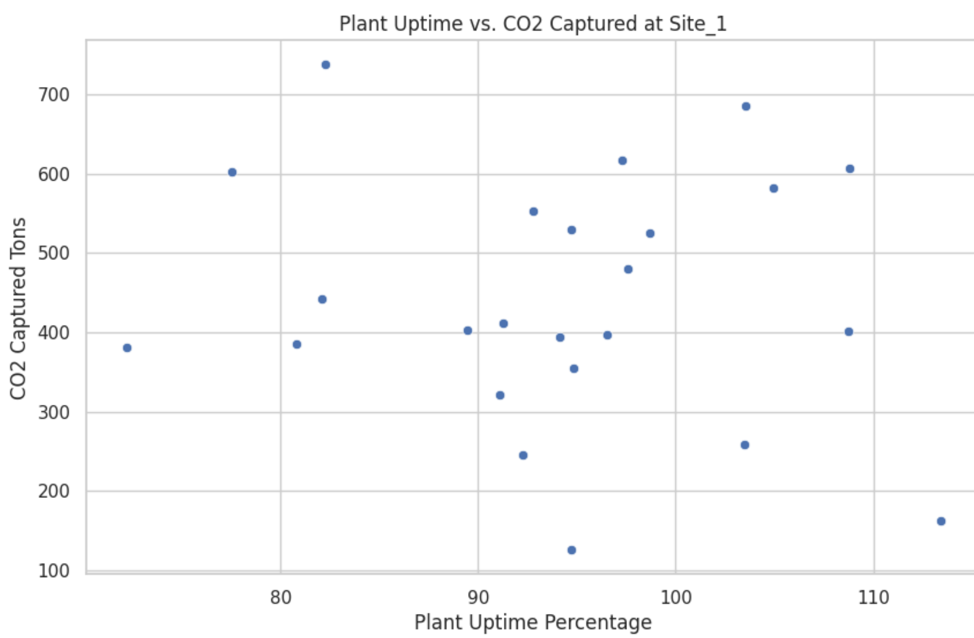


Figure 43: Site 1 - Plant Uptime Vs. CO2 Captured

8 Financial Performance Analysis

The financial performance of Carbon Capture as a Service (CCaaS) across various sites is characterised by significant variability, evident from the provided ROI and net benefits data over time. This analysis uses specific figures to better understand the financial outcomes associated with these projects.

Cost-Benefit Analysis

The "Net Benefits Over Time" graph shows fluctuations in financial performance, with net benefits ranging from as high as approximately $4,000 to as low as around -$8,000. These figures suggest a complex balance between operational and capital costs against revenues from CO2 sales and carbon credits. Periods of sharp declines in net benefits, mainly where values fall below -$6,000, may indicate when operational costs surged or revenue generation was insufficient. Conversely, peaks nearing $4,000 highlight successful phases where revenues have likely surpassed the ongoing costs significantly.

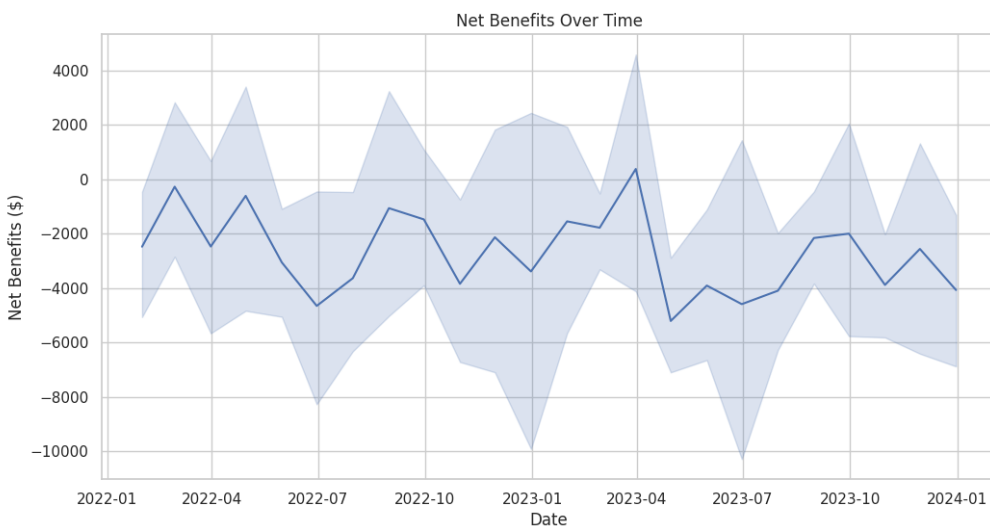


Figure 44: Net Benefit Over Time

ROI Analysis

The "ROI Over Time" graph illustrates substantial fluctuations in ROI, with values spiking above 60% and plummeting to less than -40%. This volatility indicates periods of high profitability interspersed with times of potential financial stress where the costs of running the CCaaS projects far exceeded the returns.

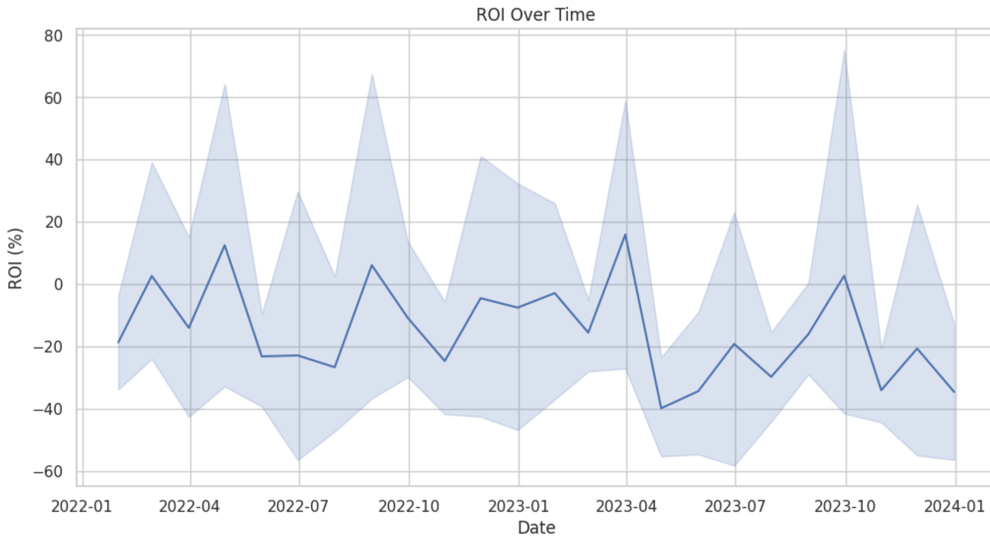


Figure 45: ROI Over Time

The "Average ROI by Site" graph shows that Site 1 has an average ROI of approximately 6.63%, suggesting modest but consistent financial gains. Site 5 outperforms other sites with a 9.91% ROI, indicating more effective cost management or higher revenue generation. In contrast, Site 2 has an ROI of -21.24%, Site 3 at -30.76%, and Site 4 with the lowest at -39.30%, all indicating financial challenges.

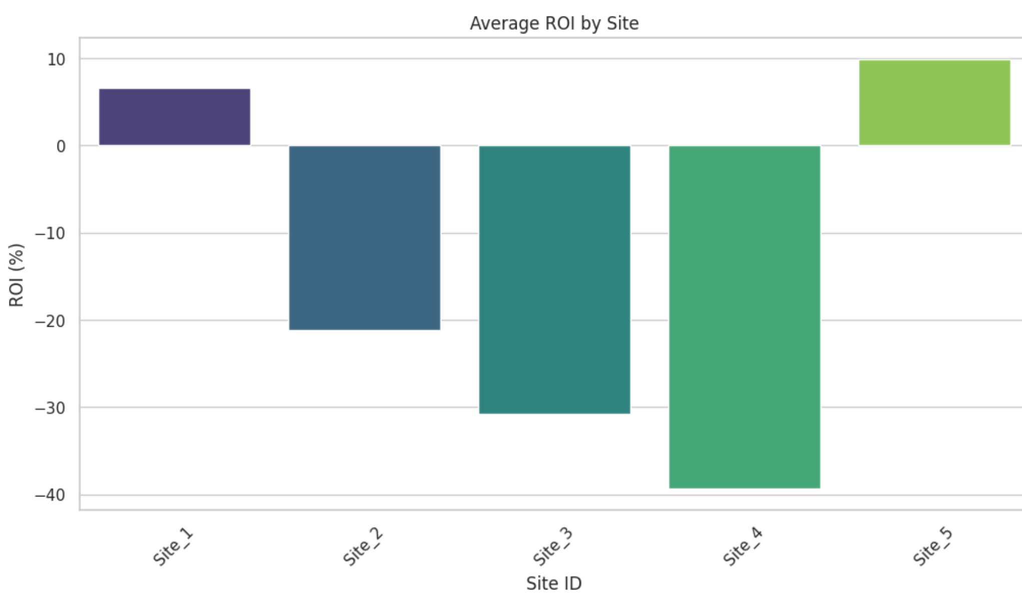


Figure 46: Average ROI

Financial Strategy Recommendations

Enhanced Cost Management: For sites like Site 4, which shows the most significant negative ROI of -39.30%, there is a critical need to scrutinise and reduce operational costs, possibly through technological upgrades or optimising operational processes for better efficiency.

Revenue Enhancement: Sites with lower ROI, such as Site 2 and Site 3, should focus on increasing revenue from CO2 sales and carbon credits. Strategies might include better market engagement, dynamic pricing strategies, or exploring new revenue streams such as partnerships with other industries.

Periodic Financial Review: Given the fluctuations seen, especially with net benefits swinging from $4,000 to -$8,000, a regular financial performance review is vital. This periodic review will enable timely strategy adjustments, ensuring financial performance is stabilised and optimised for growth.

9 Geospatial Analysis

The geographical distribution of CCaaS sites can significantly impact their operational and financial performance. By analysing the locations of these sites alongside their performance metrics, we can glean insights into how regional factors influence overall success.

Location Efficiency

The CCaaS sites are spread across various locations in the United States, each with distinct geographic and environmental characteristics that can affect operational and financial outcomes.

Site 1 (Los Angeles, CA - Latitude: 34.0522, Longitude: -118.2437): This site's moderate ROI and financial performance might be influenced by California's aggressive environmental policies and incentives for carbon capture and storage, making operations potentially more viable despite high operational costs typical for the region. Site 2 (California - Latitude: 36.7783, Longitude: -119.4179): This site's negative ROI could be impacted by local regulatory frameworks, potentially higher costs of doing business, and logistical challenges in more remote parts of California. Site 3 (New York, NY - Latitude: 40.7128, Longitude: -74.0060): The financial challenges here, indicated by a significantly negative ROI, could be attributed to the high cost of operations in an urban setting, stringent regulatory requirements, and potentially limited space for large-scale operations. Site 4 (Denver, CO - Latitude: 39.7392, Longitude: -104.9903): The negative ROI here might reflect the geographic and climatic challenges, such as colder temperatures affecting operational efficiency, and possibly the cost implications of operating in a landlocked state, which might affect logistics and supply chains. Site 5 (San Francisco, CA - Latitude: 37.7749, Longitude: -122.4194): The positive ROI observed at this site could be due to the proximity to tech hubs and innovative industries, which might offer better opportunities for selling captured CO_2 and securing carbon credits, alongside robust local and state incentives for green technologies.

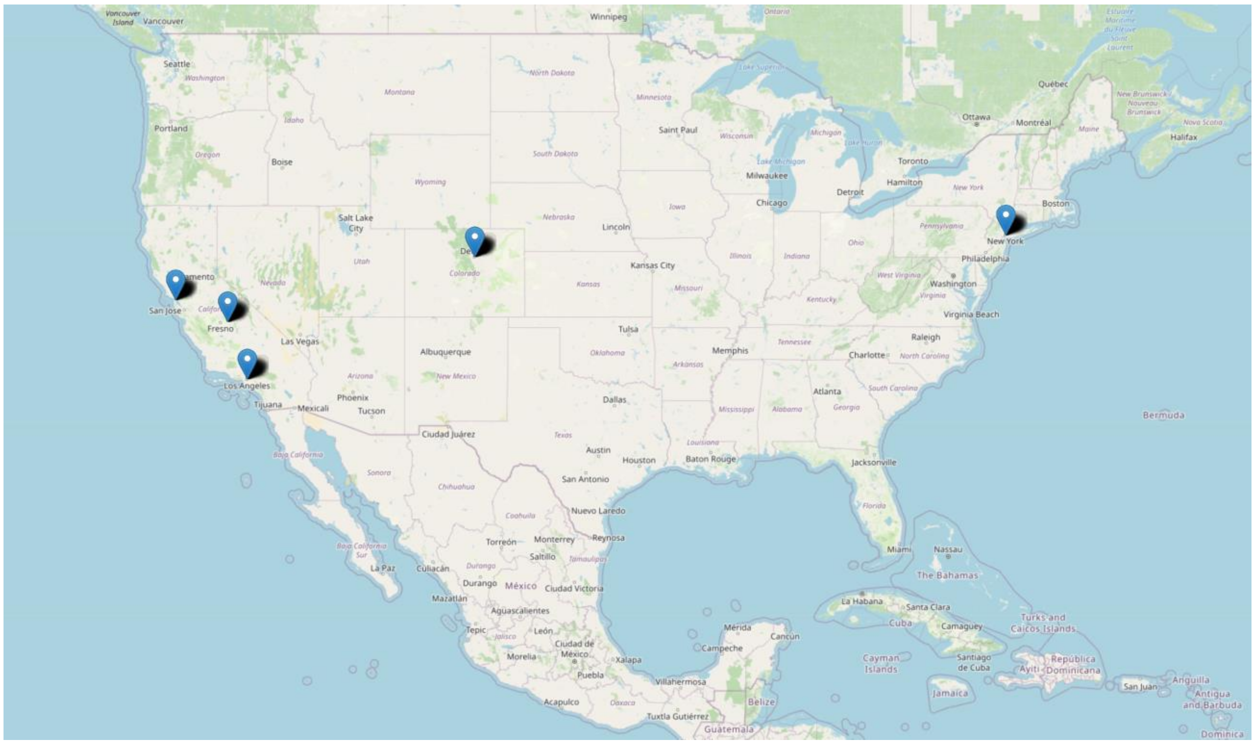


Figure 47: Site Location Map

Strategic Implications

The geospatial analysis indicates a clear link between location-specific factors and site performance. Sites in major urban centres like New York and San Francisco show differing results, likely influenced by local market conditions and regulatory environments. Similarly, sites within California show variability in performance, which may be influenced by local policies and the economic landscape specific to their exact locations.

Recommendations

Localised Strategy Development: Each site should develop strategies that align with local environmental policies, economic conditions, and market opportunities. For instance, sites in areas with incentives for green technology should capitalise on these to improve financial returns.

Operational Adjustments: Sites in challenging geographic or climatic conditions should consider technological adaptations or operational adjustments to mitigate environmental impacts on performance.

Stakeholder Engagement: Engaging with local stakeholders, including governments and businesses, can help tailor the CCaaS projects to meet local needs and capitalise on specific regional opportunities, improving ROI.

10 Environmental Impact Assessment

The environmental benefits of Carbon Capture as a Service (CCaaS) are critical to assessing the overall sustainability of this technology. The graph displays the total CO2 emissions reduced by each site, and the specific figures clearly show the impact of CCaaS implementation.

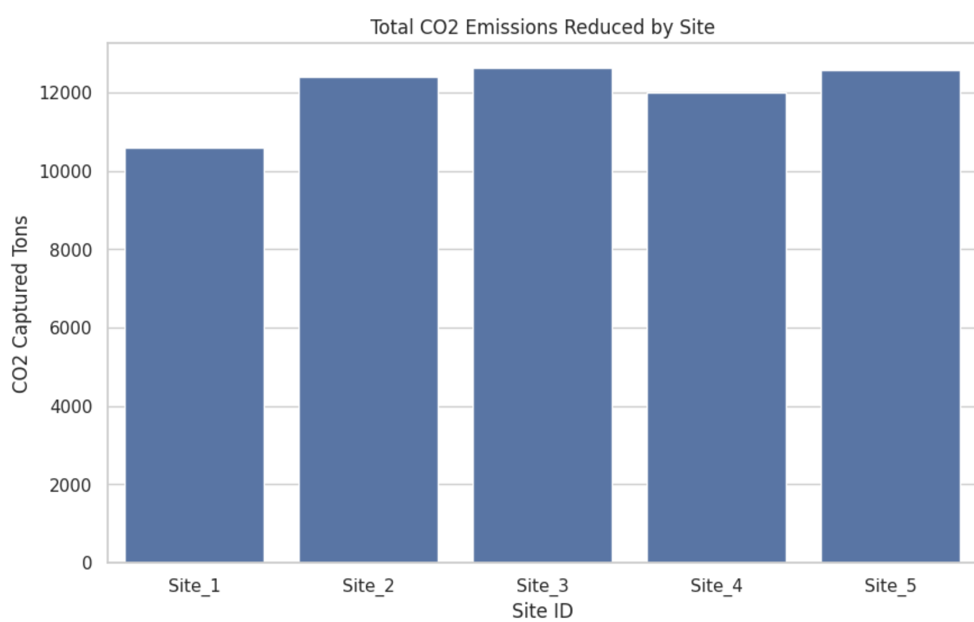


Figure 48: Total CO2 Emissions Reduction

The data indicates significant reductions in CO_2 emissions at each CCaaS site:

- Site 1: Reduced CO_2 emissions by approximately 10,602.53 tons.
- Site 2: Reduced CO_2 emissions by approximately 12,399.98 tons.
- Site 3: Achieved the highest reduction with approximately 12,626.25 tons.
- Site 4: Reduced CO_2 emissions by approximately 11,985.94 tons.
- Site 5: Reduced CO_2 emissions by approximately 12,573.01 tons.

These figures are substantial, demonstrating that each site contributes significantly to environmental sustainability by capturing large quantities of CO_2, which otherwise would contribute to atmospheric pollution and climate change.

Environmental Benefits

The reduction in CO_2 emissions at these sites not only helps in combating global warming but also contributes to improved local air quality. This can have several direct benefits, including but not limited to:

Healthier local environments: Reduced greenhouse gas emissions can lead to lower pollution levels, which in turn can decrease respiratory and cardiovascular health issues among the local population.

Compliance with environmental regulations: By capturing and storing CO_2, these sites help meet local and international environmental standards and regulations, potentially avoiding fines and facilitating smoother operations.

Enhanced biodiversity: Lower CO_2 levels can contribute to less acid rain, which can improve soil and water quality, thereby supporting healthier ecosystems.

Potential for Sustainability Improvements

While the current CO_2 capture figures are impressive, sustainability practices always have room for improvement. Each site could potentially increase its impact by:

Increasing efficiency: Improving the efficiency of the carbon capture processes can lead to higher reductions of CO_2 emissions without significantly increasing operational costs.

Utilising captured CO2: Developing methods to utilise the captured CO2, such as converting it into commercial products like concrete or biofuels, can create additional revenue streams and further offset the capture costs.

Expanding capacity: Where feasible, sites could expand their capacity to capture more CO2, particularly in regions with high industrial CO2 output.

11 Customer Satisfaction

Analysing customer satisfaction across Carbon Capture as a Service (CCaaS) sites provides insight into how operational and financial metrics impact the perception and satisfaction of customers. The box plot displaying customer satisfaction scores and the correlation matrices for each site offer a detailed view of these dynamics.

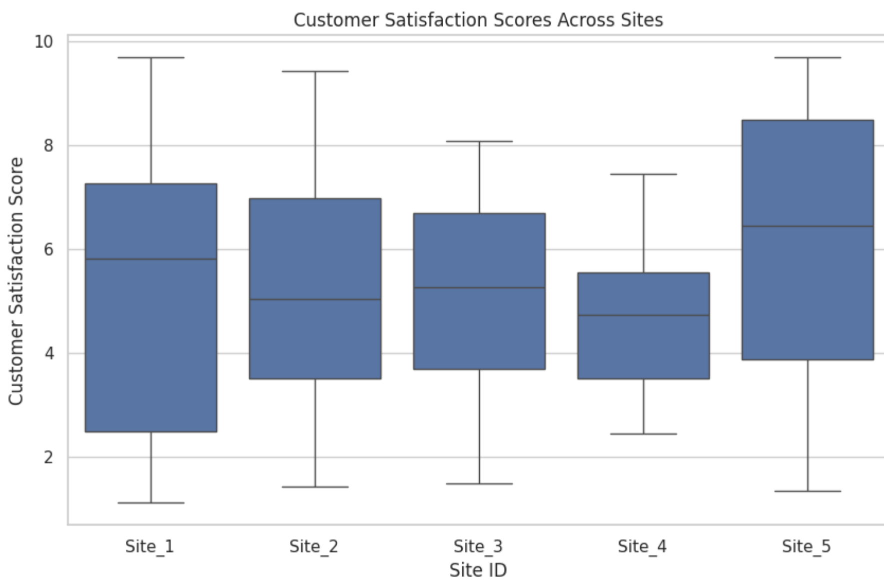


Figure 49: Customer Satisfaction Scores Across Sites

The box plot reveals varying levels of customer satisfaction across the five CCaaS sites. Site 1 and Site 5 display higher median satisfaction scores, suggesting that customers at these sites are relatively more satisfied with the services provided. Site 3 shows the lowest median customer satisfaction, indicating potential issues or unmet expectations. Site 2 and Site 4 have median scores that suggest moderate satisfaction, with Site 2 showing a broader range of scores, which could indicate inconsistent customer experiences.

Correlation Analysis

The correlation matrices highlight several exciting relationships between customer satisfaction and operational metrics:

Site 1: Customer satisfaction is negatively correlated with profit (-0.37), suggesting that higher profits at this site might be associated with cost-cutting or operational efficiencies that do not align well with customer expectations.

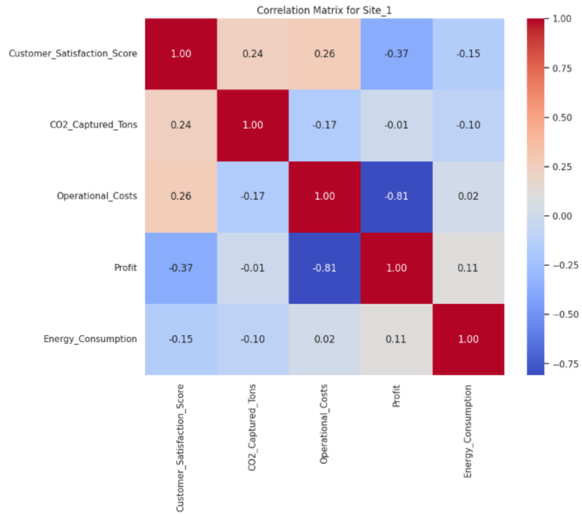


Figure 50: Site 1 - Correlation Matrix

Site 2: Shows very little correlation between customer satisfaction and CO2 captured or operational costs, indicating that other factors might be influencing satisfaction levels.

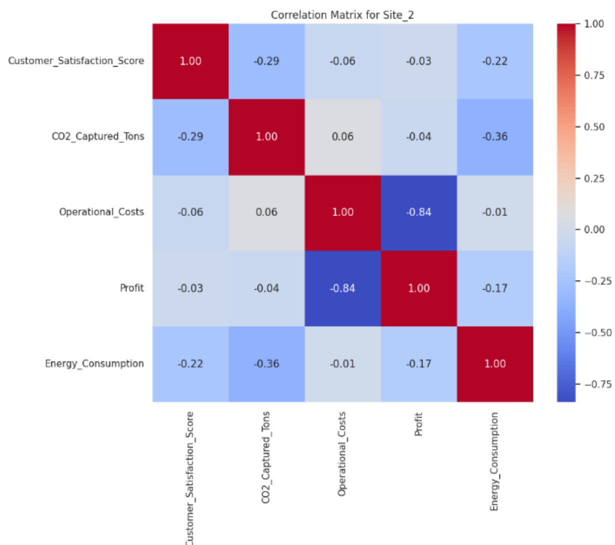


Figure 51: Site 2 - Correlation Matrix

Site 3: Displays a strong positive correlation between customer satisfaction and CO2 captured (0.85), which suggests that effective carbon capture is a significant driver of customer satisfaction at this site.

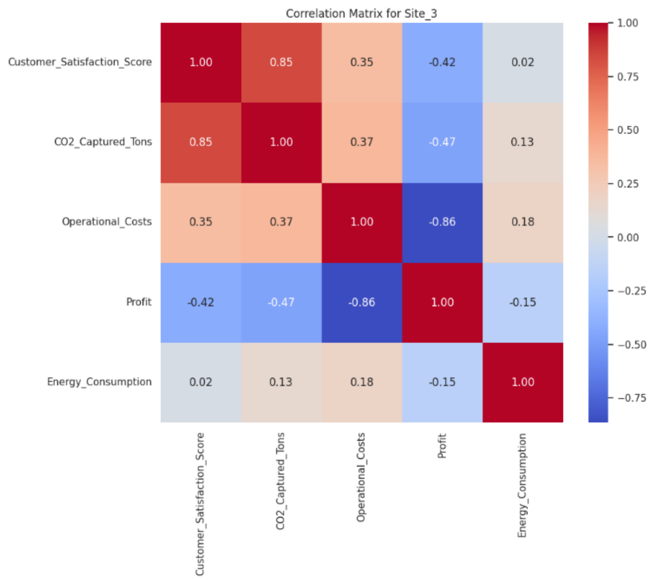


Figure 52: Site 3 - Correlation Matrix

Site 4: Profit positively correlates with customer satisfaction (0.32), indicating that financial success at this site aligns well with customer expectations, possibly due to effective service delivery that translates into customer-perceived value.

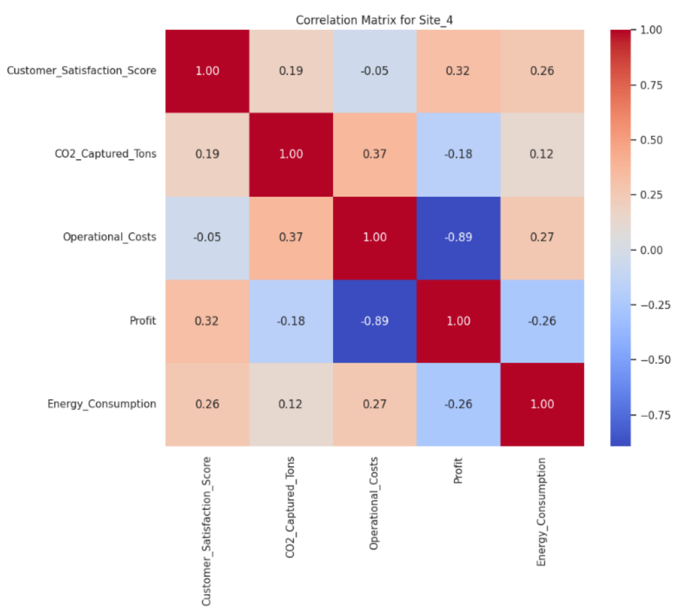


Figure 53: Site 4 - Correlation Matrix

Site 5: Shows a negative correlation between customer satisfaction and profit (-0.16), hinting that despite profitable operations, other aspects may be impacting customer satisfaction negatively.

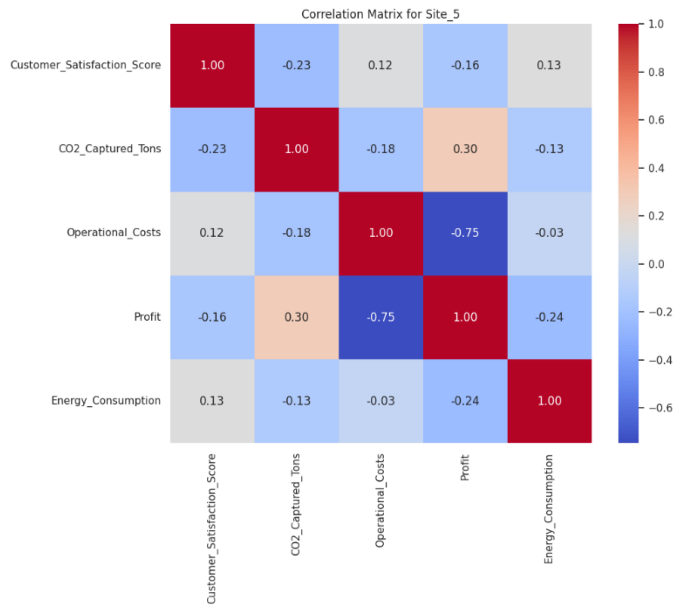


Figure 54: Site 5 - Correlation Matrix

These findings underscore the complex relationship between operational efficiency, financial performance, and customer satisfaction. For example, at Site 3, where the environmental impact (CO2 capture) strongly correlates with customer satisfaction, it suggests that customers are highly aware and appreciative of the site's effectiveness in reducing emissions. Conversely, at Site 1, where increased profitability correlates with lower satisfaction, it might be beneficial to investigate whether cost reductions or operational changes affect service quality.

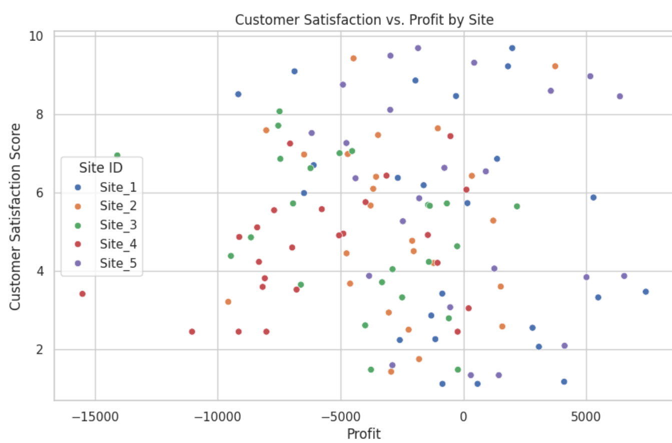


Figure 55: Customer Satisfaction Vs. Profit by Site

Strategic Recommendations

Enhance Communication and Transparency: Especially at sites like Site 1 and Site 5, improving communication about how operational decisions and profitability measures are aimed at sustaining the service quality could help align customer perceptions with the site's financial strategies.

Focus on Core Values: At sites like Site 3, maintaining high operational standards in CO2 capturing should remain a priority as it directly enhances customer satisfaction.

Customer Experience Management: For Site 2, where there is significant variability in satisfaction, standardising service delivery and focusing on consistent customer experience management could help stabilise satisfaction scores.

12Regression Analysis for Deeper Insights

The regression analysis was carried out using *statsmodels.api* Python library provides a structured approach to understanding the quantitative relationships between variables in a dataset. Specifically, this analysis aims to decipher the impact of energy consumption and plant uptime on CO2 capture amounts across different Carbon Capture as a Service (CCaaS) sites.

statsmodels.api is a powerful Python module that offers classes and functions for estimating many different statistical models, conducting statistical tests, and exploring data. Here's how it is effectively utilised in the regression analysis:

Data Preparation: Data specific to each site is filtered, and necessary predictors (Energy_Consumption and Plant_Uptime_Percentage) and the response variable (CO2_Captured_Tons) are selected.

Model Specification: The OLS (Ordinary Least Squares) model from *statsmodels.api* is used to fit the linear regression model. The process involves adding a predictor constant to account for the intercept.

Model Fitting: The model is fitted to the data, providing a full summary that includes coefficients, standard errors, p-values, and other statistical metrics.

Interpretation: The summary outputs from the model fitting process offer detailed insights into the significance and impact of each predictor.

Example Outcomes and Interpretations

The regression outcomes for each site reveal different levels of effectiveness and statistical significance:

Site 1: Shows a very low R-squared value (0.031), indicating that the model explains only 3.1% of the variability in CO_2 captured using energy consumption and plant uptime. The high p-values suggest that neither predictor is statistically significant at typical levels.

Site 2: Exhibits a slightly better R-squared value (0.207), with energy consumption showing a trend towards significance ($p = 0.074$). This suggests some predictive power, although still limited.

Other Sites: Similar patterns are observed with low R-squared values, indicating weak explanatory power of the models across the board.

Insights and Strategic Implications

The overall low R-squared values across all sites suggest that while energy consumption and plant uptime are intuitive predictors of CO_2 capture efficiency, they do not strongly predict the amount of CO_2 captured in this specific setting. This could imply that:

- Other unexamined factors may be more influential in determining CO_2 capture rates.
- A simple linear model might not capture complex interactions between variables.
- The data could have underlying issues such as multicollinearity, as suggested by the high condition numbers, or it may not meet the assumptions required for linear regression.

13 Model Selection for Forecasting CO2 Capture Volumes

Forecasting CO2 capture volumes is critical for planning and optimising operations in Carbon Capture as a Service (CCaaS). The choice of the forecasting model significantly depends on the characteristics of the data, such as trend, seasonality, and noise. Models like ARIMA, SARIMA, Random Forests, and Gradient Boosting each have unique strengths that can be leveraged based on the specific data structure and forecasting needs.

13.1 Time Series Models: ARIMA and SARIMA

ARIMA (AutoRegressive Integrated Moving Average): This model is suitable for non-seasonal data that shows autocorrelation patterns. ARIMA is effective when the data series is stationary, meaning it does not show trends or seasonality, which often requires differencing of raw data to achieve stationarity.

SARIMA (Seasonal ARIMA): An extension of ARIMA, this model is tailored for seasonal data. It addresses autocorrelation in the data and incorporates seasonal differencing, making it ideal for data with patterns that repeat over a set period. For instance, CO2 capture volumes may exhibit seasonal fluctuations due to operational cycles or external environmental factors.

13.2 SARIMA

Given the seasonal nature of CO2 capture volumes, SARIMA becomes a prime candidate for forecasting. The model configuration used:

model = SARIMAX(train_data, order=(1, 1, 1), seasonal_order=(1, 1, 1, 12))

Here, order=(1, 1, 1) specifies the ARIMA part (autoregression order 1, differencing order 1, moving average order 1), and seasonal_order=(1, 1, 1, 12) addresses the seasonality (seasonal autoregressive order 1, seasonal differencing order 1, seasonal moving average order 1, with a seasonal period of 12 months).

The Root Mean Square Error (RMSE) from this model is 153.907, indicating the average prediction error in units of CO2 captured. This metric helps evaluate the model's accuracy, with a lower RMSE indicating a better fit for the data.

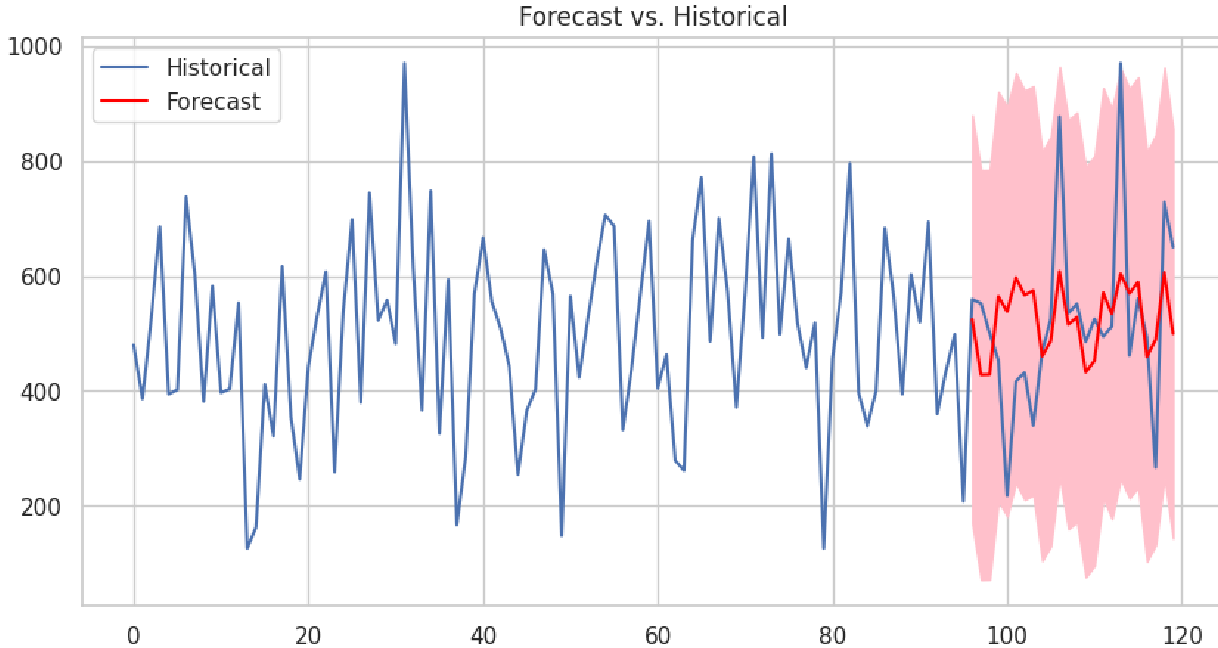


Figure 56: SARIMA - Forecast Vs. Historical

Choosing the suitable model for forecasting CO2 capture volumes involves assessing the data for trends, seasonality, and other structural patterns. SARIMA, given its capacity to model both non-seasonal and seasonal components, offers a robust approach for time series forecasting in environmental data management. As demonstrated, it can provide insights that are crucial for operational planning and environmental impact assessment in the context of CCaaS.

14 References

[1] J. Meadowcroft and O. Langhelle, "The politics and policy of carbon capture and storage," *Caching the Carbon: The Politics and Policy of Carbon Capture and Storage*, pp. 1–21, 2009, doi: 10.4337/9781849802222.00009.

[2] G. F. Nemet *et al.*, “Negative emissions—Part 3: Innovation and upscaling,” *Environmental Research Letters*, vol. 13, no. 6, p. 063003, May 2018, doi: 10.1088/1748-9326/AABFF4.

[3] Steve A. Rackley, “Carbon Capture and Storage.” Accessed: May 07, 2024. [Online]. Available: https://books.google.co.nz/books?hl=en&lr=&id=nQGqDQAAQBAJ&oi=fnd&pg=PP1&dq=Rackley,+S.+A.+(2017).+Carbon+capture+and+storage.+Butterworth-Heinemann.&ots=W0C2HNe0Ci&sig=TCOuHR6apPURPMXezLmADrv8Cfs&redir_esc=y#v=onepage&q=Rackley%2C%20S.%20A.%20(2017).%20Carbon%20capture%20and%20storage.%20Butterworth-Heinemann.&f=false

[4] E. S. Rubin, J. E. Davison, and H. J. Herzog, “The cost of CO2 capture and storage,” *International Journal of Greenhouse Gas Control*, vol. 40, pp. 378–400, Sep. 2015, doi: 10.1016/J.IJGGC.2015.05.018.

[5] Berend Smit, Jeffrey A Reimer, Curtis M Oldenburg, and Ian C Bourg, “Introduction To Carbon Capture And Sequestration.” Accessed: May 07, 2024. [Online]. Available: https://books.google.co.nz/books?hl=en&lr=&id=StS3CgAAQBAJ&oi=fnd&pg=PR5&dq=Smit,+B.,+Reimer,+J.+A.,+Oldenburg,+C.+M.,+%26+Bourg,+I.+C.+(2014).+Introduction+to+carbon+capture+and+sequestration.+Imperial+College+Press.&ots=2txaT_lghk&sig=UUSGTJBWcYB7QUfsPL5NZIvpMIA&redir_esc=y#v=onepage&q&f=false