Business Intelligence Project Report

Carbon Footprint and Emissions Analysis of Apple Products

Elaborated By Terrabythians :

Nada Latrach
Mohamed Aziz B'CHIR
Mohamed Aziz Allaya
Syrine Laabidi
Nour EL Houda Ouni
Fatma Naoui

2024/2025

Table of contents

Table of Figures

General Introduction

As part of our fourth year at ESPRIT, we, the Terrabythians group, will work on an extensive Business Intelligence project focused on environmental responsibility and sustainability as part of our fourth year at ESPRIT. The academic year 2024–2025 will feature our study, "Carbon Footprint and Emissions Analysis of Apple Products."

In the contemporary corporate landscape, sustainability and environmental responsibility are growing in importance for companies globally. As a pioneer in technology, Apple Inc. has always placed a high priority on lowering greenhouse gas emissions and its carbon footprint in order to lessen its impact on the environment. The objective of this project is to enhance our comprehension of business intelligence principles by exposing us to practical applications and obstacles.

To guide our analysis, we will employ the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, which consists of the following phases:

- <u>Business Understanding:</u> Define project objectives and requirements from a business perspective.
- <u>Data Understanding:</u> Explore the data relevant to the project to gain insights into the problem at hand.
- <u>Data Preparation:</u> Prepare the final dataset that will be used for modeling, including cleaning and transforming the data.
- <u>Modeling:</u> Select and apply various modeling techniques to analyze the data and generate insights.
- <u>Evaluation:</u> Assess the model results to ensure they meet business objectives and requirements.
- <u>Deployment:</u> Implement the model in a production environment to deliver results and insights effectively.

By following this structured approach, we aim to contribute valuable insights into Apple Inc.'s sustainability practices while deepening our understanding of business intelligence concepts.

Chapter 1: Business Understanding

1- *Project Overview :*

The initial step in our project is to gain a comprehensive understanding of the business elements and challenges that Business Intelligence seeks to address or enhance. In this analysis, our primary focus will be on assessing the carbon footprint and greenhouse gas emissions associated with Apple products, alongside a review of the company's annual profit. Specifically, we aim to answer several critical questions:

1. What is the carbon footprint of Apple products (PCF)?
2. What are Apple's total greenhouse gas emissions (GHG)?
3. What are the implications of these emissions for Apple as a corporation?
4. What strategies can be implemented to mitigate this environmental impact?

Regarding the first two questions, the answers have already been provided earlier in this report: the carbon footprint of Apple products (PCF) and Apple's total greenhouse gas emissions (GHG). These metrics are crucial for understanding the environmental impact of Apple's operations. However, the implications of these emissions for Apple as a corporation add a layer of complexity to this analysis, given the growing environmental concerns and regulatory pressures.

As for the last question, answering it leads us to begin the next step of this study: identifying and recommending strategies that can be implemented to mitigate Apple's environmental impact.

2- *Business Goals and Data Science Objectives :*

Business Goal 1: Reduce the carbon footprint of Apple products year-over-year (Decision Makers: Sustainability teams, Investors)

Focus: Sustainability teams will use these insights to design products with a lower environmental impact, and investors will see the long-term value in sustainable product designs that contribute to Apple's reputation and market positioning.

Data Science Objective 1: Build a predictive model to estimate the carbon footprint for future Apple products

The model will help sustainability teams predict emissions based on past trends, allowing them to design more eco-friendly products. Investors will value the foresight into Apple's sustainability improvements, which can positively impact the company's brand and long-term growth.

Business Goal 2: Optimize corporate emissions and sustainability initiatives **(Decision Makers: Sustainability teams, Investors)**

Focus: Sustainability teams need data-driven insights to optimize their emission reduction strategies, while investors want to ensure that Apple's investments in sustainability initiatives lead to measurable results.

Data Science Objective 2: Develop a model to identify the most significant contributors to corporate emissions

This model will assist sustainability teams in targeting key areas for emission reductions, enabling more efficient environmental strategies. Investors will benefit from the data showing effective reductions in corporate emissions, aligning with sustainable growth goals.

Business Goal 3: Assess the impact of Apple's sustainability efforts on financial performance **(Decision Makers: Sustainability teams, Investors)**

Focus: Investors are increasingly interested in the financial benefits of sustainability, while sustainability teams need to demonstrate how their initiatives positively influence Apple's financial health.

Data Science Objective 3: Create a time-series analysis to evaluate the relationship between sustainability efforts and financial performance

This analysis will help both teams understand how emission reductions and sustainability initiatives correlate with financial performance indicators like revenue growth and market value, giving investors confidence in Apple's commitment to sustainable practices.

Business Goal 4: Improve the efficiency of corporate sustainability initiatives **(Decision Makers: Sustainability teams, Investors)**

Focus: Sustainability teams seek to maximize the impact of their initiatives, while investors are focused on cost-effectiveness and profitability of sustainability efforts.

Data Science Objective 4: Build an optimization model to assess the cost-effectiveness of sustainability initiatives

The model will identify the most cost-efficient sustainability strategies, allowing the sustainability team to allocate resources effectively. Investors will appreciate the focus on maintaining profitability while advancing Apple's environmental goals.

Business Goal 5: Increase transparency in Apple's sustainability reporting **(Decision Makers: Sustainability teams, Investors)**

Focus: Both sustainability teams and investors require transparent reporting to track Apple's environmental progress and ensure that sustainability efforts align with business objectives.

Data Science Objective 5: Develop a dashboard to visualize corporate and product life cycle emissions data

This dashboard will enable sustainability teams and investors to easily track Apple's sustainability progress. Enhanced transparency will strengthen investor confidence in Apple's environmental initiatives and improve accountability.

Business Goals	Data Science Goals
Reduce the carbon footprint of Apple products year-over-year.	Build a predictive model to estimate the carbon footprint for future Apple products
Optimize corporate emissions and sustainability initiatives.	Develop a model to identify the most significant contributors to corporate emissions
Assess the impact of Apple's sustainability efforts on financial performance.	Create a time-series analysis to evaluate the relationship between sustainability efforts and financial performance.
Improve the efficiency of corporate sustainability initiatives	Build an optimization model to assess the cost-effectiveness of sustainability initiatives.
Increase transparency in Apple's sustainability reporting.	Develop a dashboard to visualize emissions data across all categories.

3- Conclusion :

These goals and objectives now focus solely on the needs of **sustainability teams and Investors**, ensuring that Apple's sustainability strategies align with business priorities and provide measurable financial and environmental benefits.

Chapter 2: Data Understanding

1-Introduction:

Data Understanding is a crucial phase in the CRISP-DM methodology. It involves gathering initial insights about the data, identifying data quality issues, and laying the groundwork for subsequent analysis.

2-Data Description :

This phase ensures that you are well-acquainted with the data, its characteristics, and its potential limitations. For this project, there are three datasets:

1 . **Greenhouse Gas Emissions**: Contains information on the sources and amounts of greenhouse gas emissions.

2. **Carbon Footprint by Product**: Details the carbon footprint associated with various Apple products.

3. **Normalizing Factors**: Provides financial and employee data for normalization purposes.

- ***Data Characteristics :***

Table	Filed	Description
Greenhouse Gas Emissions	Fiscal year	- Apple's fiscal calendar starts on the last Sunday of September and is 364 days long
Greenhouse Gas Emissions	Category	- Emissions are divided into two categories: corporate emissions and product life cycle emissions
Greenhouse Gas Emissions	Type	- There are two types of emissions data included: gross emissions (which add to the carbon footprint) ...

<table>
<tr><td>Greenhouse Gas Emissions</td><td>Scope</td><td>- Greenhouse gas emissions are classified into three categories,known as Scopes:
● Scope 1(Direct emissions) : covers emissions from sources that Apple owns or controls directly.

Example:Natural gas, diesel , propane ,fleet vehicles, Other (R&D processes & refrigerant leaks)….

● Scope 2(Indirect Emissions from Energy): come from The energy the company purchases and then consumes.

Example: Electricity,stream,heating,and cooling …

● Scope 3(Indirect Emissions from the value chain): emissions that are not produced by the company itself and are not the result of activities from assets owned or controlled by them, but by those that it's indirectly responsible for up and down its value chain.

Example : Product manufacturing ,product use, product transport , Business travel…..</td></tr>
<tr><td>Greenhouse Gas Emissions</td><td>Description</td><td>- The source of the greenhouse gas emissions</td></tr>
<tr><td>Greenhouse Gas Emissions</td><td>Emissions</td><td>- Greenhouse gas emissions (metric tons CO_2e)</td></tr>
<tr><td>Carbon footprint by product</td><td>Release Year</td><td>- Year the product was released</td></tr>
<tr><td>Carbon footprint by product</td><td>Product</td><td>- Product name</td></tr>
<tr><td>Carbon footprint by product</td><td>Baseline Storage</td><td>- Lowest storage option</td></tr>
<tr><td>Carbon footprint by product</td><td>Carbon Footprint</td><td>- A carbon footprint quantified in tonnes</td></tr>
</table>

		of greenhouse gas emissions from the product lige cycle(kg CO2e)
Normalizing factors	Fiscal Year	- Figures are as of the end of the fiscal year
Normalizing factors	Revenue	- Net sales (in millions, US$)
Normalizing factors	Market Capitalization	- Value of the company (in billions, US$)
Normalizing factors	Employees	- Number of full-time equivalent employees

3-Data Exploration(EDA) :

For the EDA, we will use machine learning techniques and tools, specifically Jupyter Notebook. This interactive environment will allow you to perform various analyses, visualize data, and draw insights from your datasets.

- **Statistical and Descriptive Study :**

In this step, we retrieved the required 3 CSV files by using the Pandas library and displayed all The information that was stored in the DataFrames.

Display of greenhouse_gas_emissions DataFrame :

```
greenhouse_gas_emissions=pd.read_csv(r'D:\4DS1\Sem1\BI\Proj\apple_emissions\greenhouse_gas_emissions.csv')
greenhouse_gas_emissions
```

	Fiscal Year	Category	Type	Scope	Description	Emissions
0	2022	Corporate emissions	Gross emissions	Scope 1	Natural gas, diesel, propane	39700.0
1	2022	Corporate emissions	Gross emissions	Scope 1	Fleet vehicles	12600.0
2	2022	Corporate emissions	Gross emissions	Scope 1	Other (R&D processes & refrigerant leaks)	2900.0
3	2022	Corporate emissions	Gross emissions	Scope 2 (market-based)	Electricity	0.0
4	2022	Corporate emissions	Gross emissions	Scope 2 (market-based)	Steam, heating, and cooling	3000.0
...	...	...	...	...	...	...
131	2015	Product life cycle emissions	Gross emissions	Scope 3	Manufacturing (purchased goods and services)	29600000.0
132	2015	Product life cycle emissions	Gross emissions	Scope 3	Product transportation (upstream and downstream)	1300000.0
133	2015	Product life cycle emissions	Gross emissions	Scope 3	Product use (use of sold products)	6600000.0
134	2015	Product life cycle emissions	Gross emissions	Scope 3	End-of-life processing	500000.0
135	2015	Product life cycle emissions	Carbon removals	NaN	Product carbon offsets	NaN

136 rows × 6 columns

figure 1 : greenhouse gas emissions dataframe

Display of carbon footprint DataFrame :

```
carbon_footprint_by_product=pd.read_csv(r'D:\4DS1\Sem1\BI\Proj\apple_emissions\carbon_footprint_by_product.csv')
carbon_footprint_by_product
```

	Release Year	Product	Baseline Storage	Carbon Footprint
0	2023	iPhone 15	128	56
1	2022	iPhone 14	128	61
2	2021	iPhone 13	128	64
3	2020	iPhone 12	64	70
4	2019	iPhone 11	64	72
5	2018	iPhone Xs	64	70
6	2017	iPhone X	64	79
7	2017	iPhone 8	64	57
8	2016	iPhone 7	32	56
9	2015	iPhone 6s	32	54

figure 2: carbon footprint dataframe

Display of normalizing factors DataFrame :

```
In [139]: normalizing_factors=pd.read_csv(r'D:\4DS1\Sem1\BI\Proj\apple_emissions\normalizing_factors.csv')
          normalizing_factors
```

Out[139]:

	Fiscal Year	Revenue	Market Capitalization	Employees
0	2022	394328	2490	164000
1	2021	365817	2450	154000
2	2020	274515	1720	147000
3	2019	260174	1090	137000
4	2018	265595	830	132000
5	2017	229234	740	123000
6	2016	215639	600	116000
7	2015	233715	580	110000

figure 3: normalizing factors dataframe

Overview of DataFrames' Structure:

```
carbon_footprint_by_product.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 4 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Release Year      10 non-null     int64
 1   Product           10 non-null     object
 2   Baseline Storage  10 non-null     int64
 3   Carbon Footprint  10 non-null     int64
dtypes: int64(3), object(1)
memory usage: 452.0+ bytes
0
```

- The **Carbon Footprint by Product** DataFrame consists of four columns. Among these columns, one is a categorical variable['Product'], and the other three are numerical variables['Release Year', 'Baseline Storage', 'Carbon Footprint'].

figure 4: carbon footprint structure

- The **Greenhouse Gas Emissions** DataFrame consists of six columns. Among these columns, four are categorical variables['Category', 'Type', 'Scope', 'Description'], and two of them are numerical variables['Fiscal Year', 'Emissions']

```
greenhouse_gas_emissions.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 136 entries, 0 to 135
Data columns (total 6 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   Fiscal Year  136 non-null    int64
 1   Category     136 non-null    object
 2   Type         136 non-null    object
 3   Scope        120 non-null    object
 4   Description  136 non-null    object
 5   Emissions    109 non-null    float64
dtypes: float64(1), int64(1), object(4)
memory usage: 6.5+ KB
```

figure 5: greenhouse gas emissions structure

```
normalizing_factors.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8 entries, 0 to 7
Data columns (total 4 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Fiscal Year           8 non-null      int64
 1   Revenue               8 non-null      int64
 2   Market Capitalization 8 non-null      int64
 3   Employees             8 non-null      int64
dtypes: int64(4)
memory usage: 388.0 bytes
```

figure 6: normalizing factors structure

- The factorizing factors DataFrame consists of four columns. And they are all numerical variables.

- **Visualization :**

Carbon Footprint by product :

As we can see in the data, the iPhone X, iPhone 11, and iPhone 12 exhibit the highest carbon footprints compared to other models, indicating that these products have significant environmental impacts.

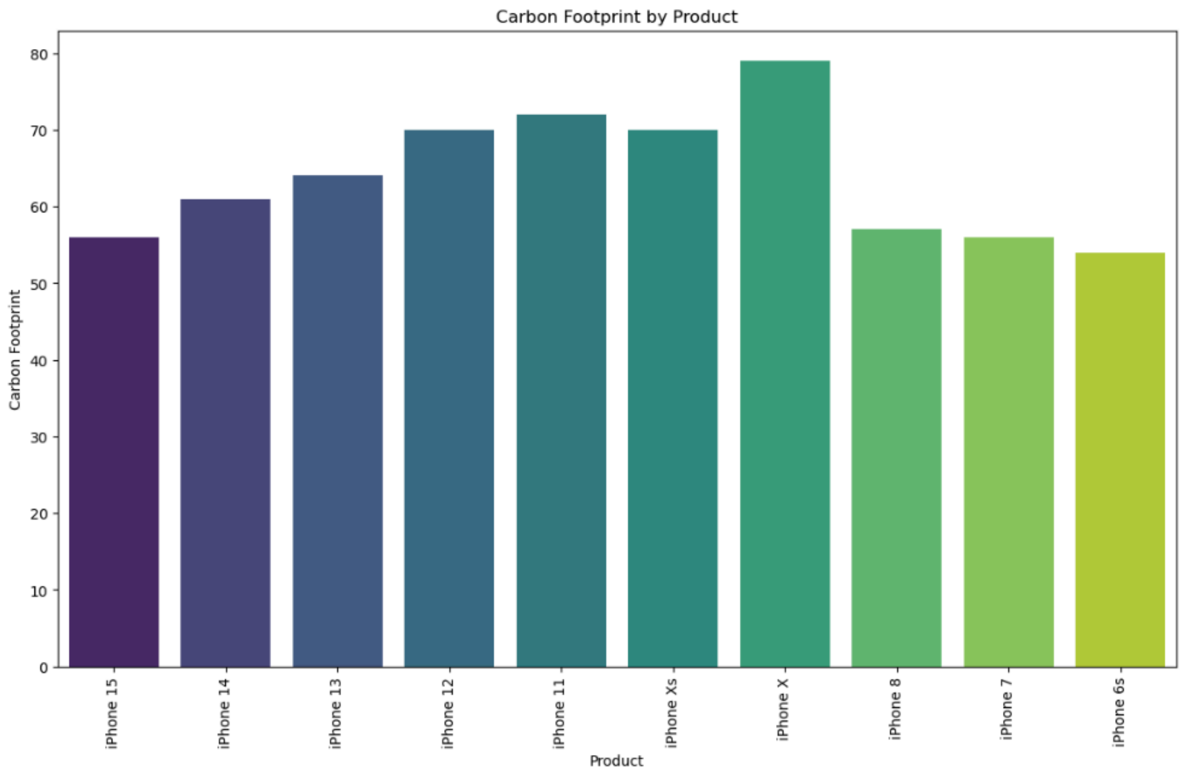


figure 7: Bar plot for carbon footprint by product

Emissions by Fiscal year and Scope:

- From 2015 to 2022, the highest emissions came from Scope 3, which includes indirect emissions from the supply chain.

- In 2015 and 2016, the second highest emissions were from Scope 2, reflecting indirect emissions from purchased electricity, steam, heating, and cooling.
- From 2017 to 2022, the second highest emissions shifted to Scope 1, which includes direct emissions from owned or controlled sources.

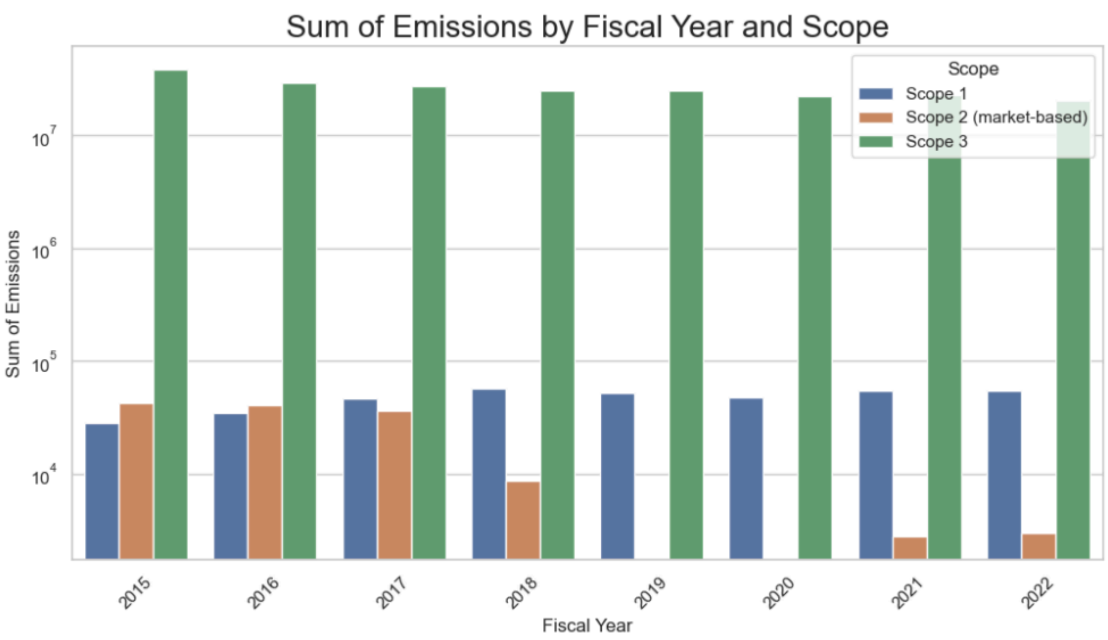


figure 8: Bar plot of emissions by Fiscal year and Scope

Emissions by Fiscal year:

This pie chart showcases that the highest emissions were in 2015 (18.2%)

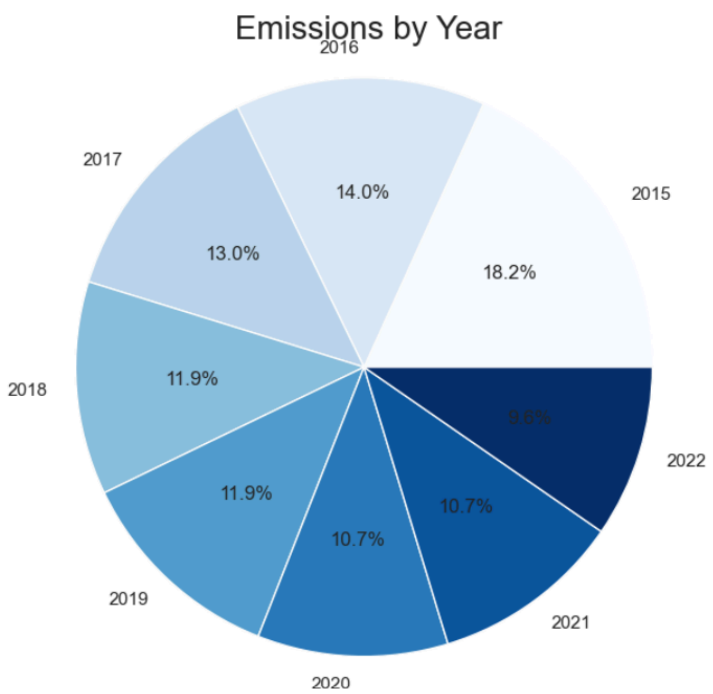


figure 9: Pie chart of emissions by Fiscal year

Market Capitalization and Revenue By year:

The total market value known as “Market Capitalization” has been steadily increasing year by year, along with revenue growth. Notably, 2022 marked the best performance to date, showcasing significant advancements in both market value and revenue generation.

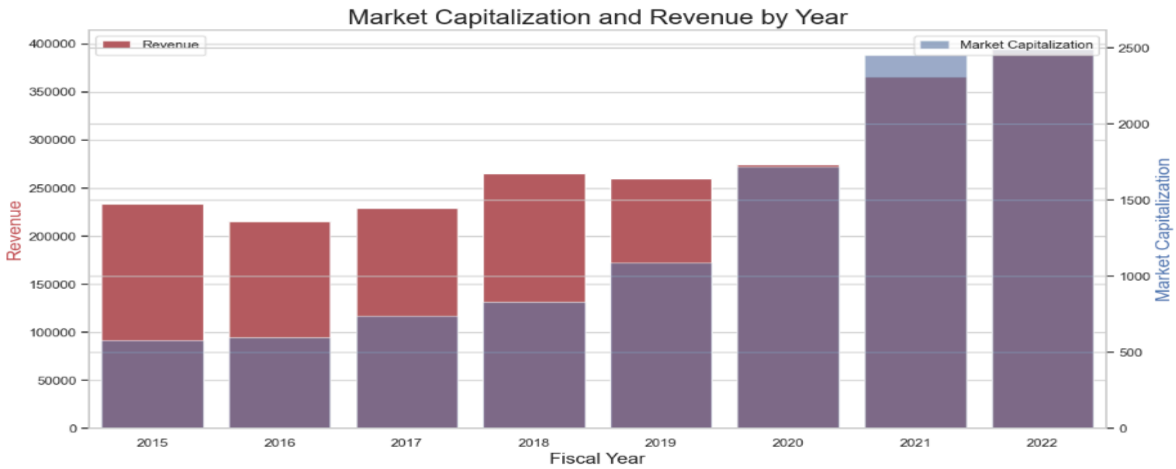


figure 10: Bar plot of Market Capitalization and Revenue By year

Carbon footprint and greenhouse gas emissions over years :

While the carbon footprint, measured by CO2 emissions, has been increasing each year, overall greenhouse gas emissions, including methane and nitrous oxide (NO_2), have been declining annually. This positive trend helps mitigate climate change and its adverse environmental effects. Apple Inc. has been working tirelessly to achieve these results through innovative practices and sustainability initiatives.

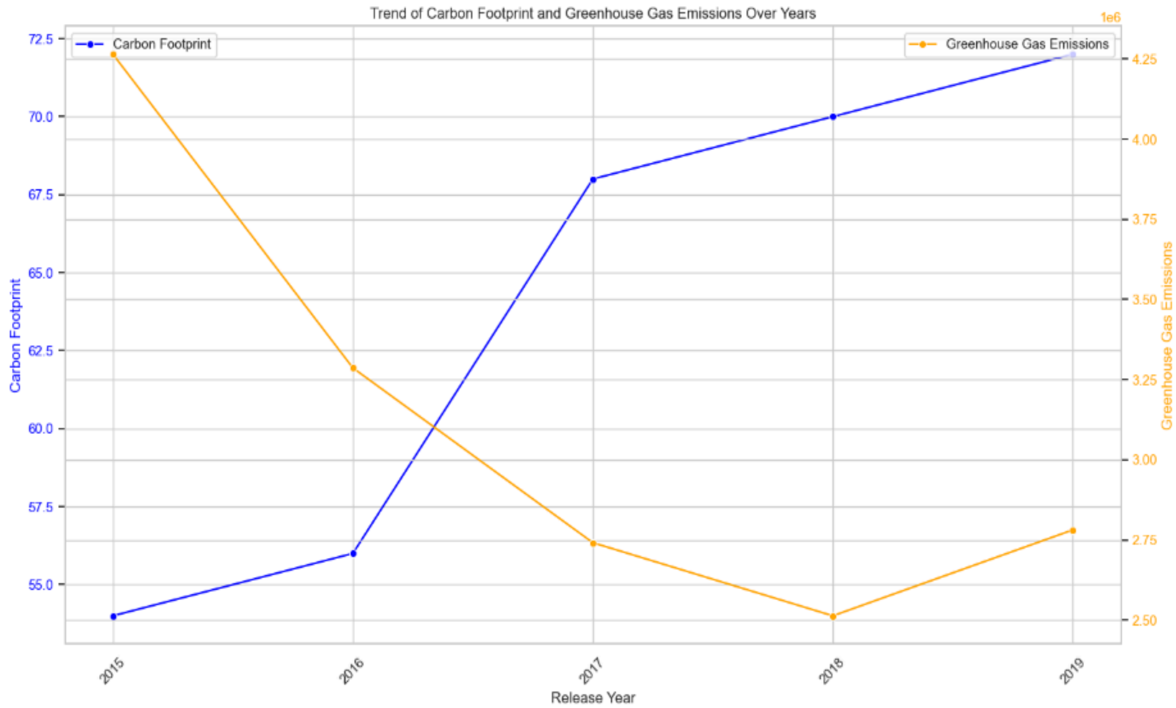


figure 11: Line plot of Carbon Footprint and Greenhouse Gas emissions over years

Correlations in Greenhouse Gas Emissions DataFrame:

There is a clear correlation between the **Fiscal Year** and **emissions**, as emissions have decreased each year. This trend suggests that the company has been successful in implementing measures to reduce its environmental impact over time, indicating a commitment to sustainability and improved operational efficiency.

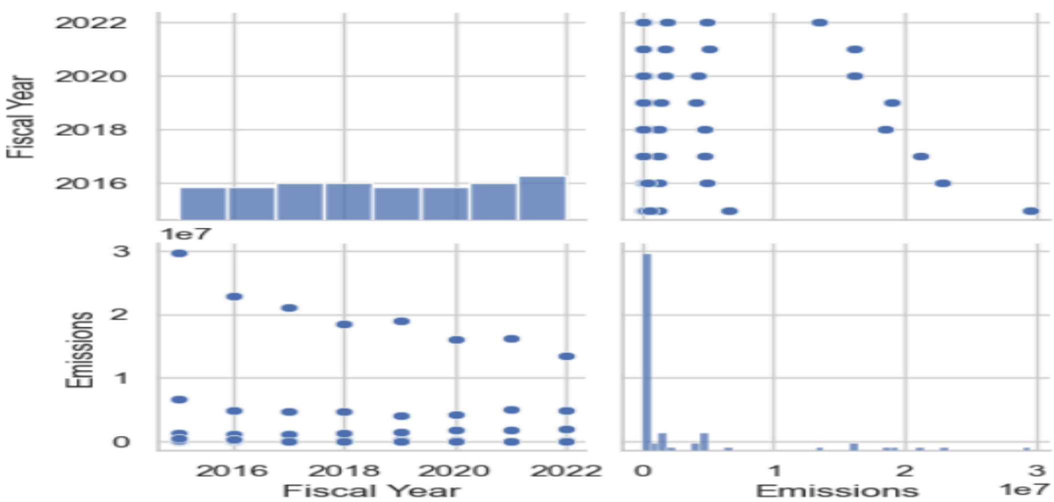


figure 12: Pairplot of Greenhouse Gas emissions

Correlations of Carbon Footprint By product DataFrame:

There is a clear correlation between the **Release Year** and **baseline storage**, as **baseline storage has** decreased each year.

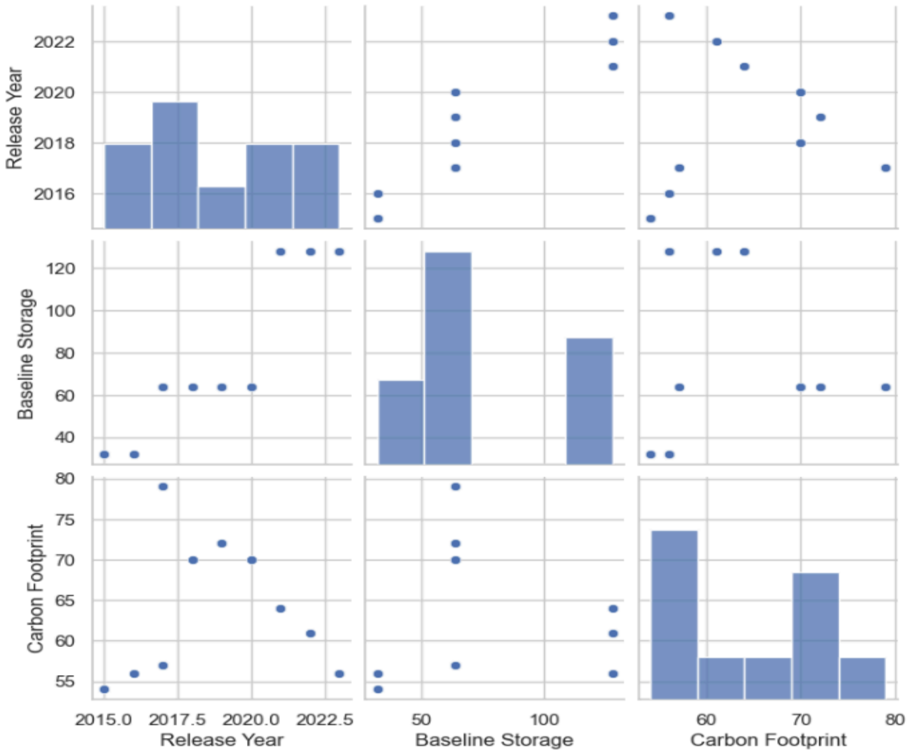


figure 13: Pairplot of Greenhouse Gas emissions

Correlations in Normalizing Factors DataFrame:

The pairplot reveals high correlations among the **Fiscal Year**, **Revenue**, **Market Capitalization**, and **Employees**, indicating that the company has consistently grown over time. A strong positive relationship between **Revenue** and **Market Capitalization** suggests that higher revenues contribute to increased market value. Additionally, the correlation between **Revenue** and **Employees** indicates that a larger workforce is associated with higher productivity and income.

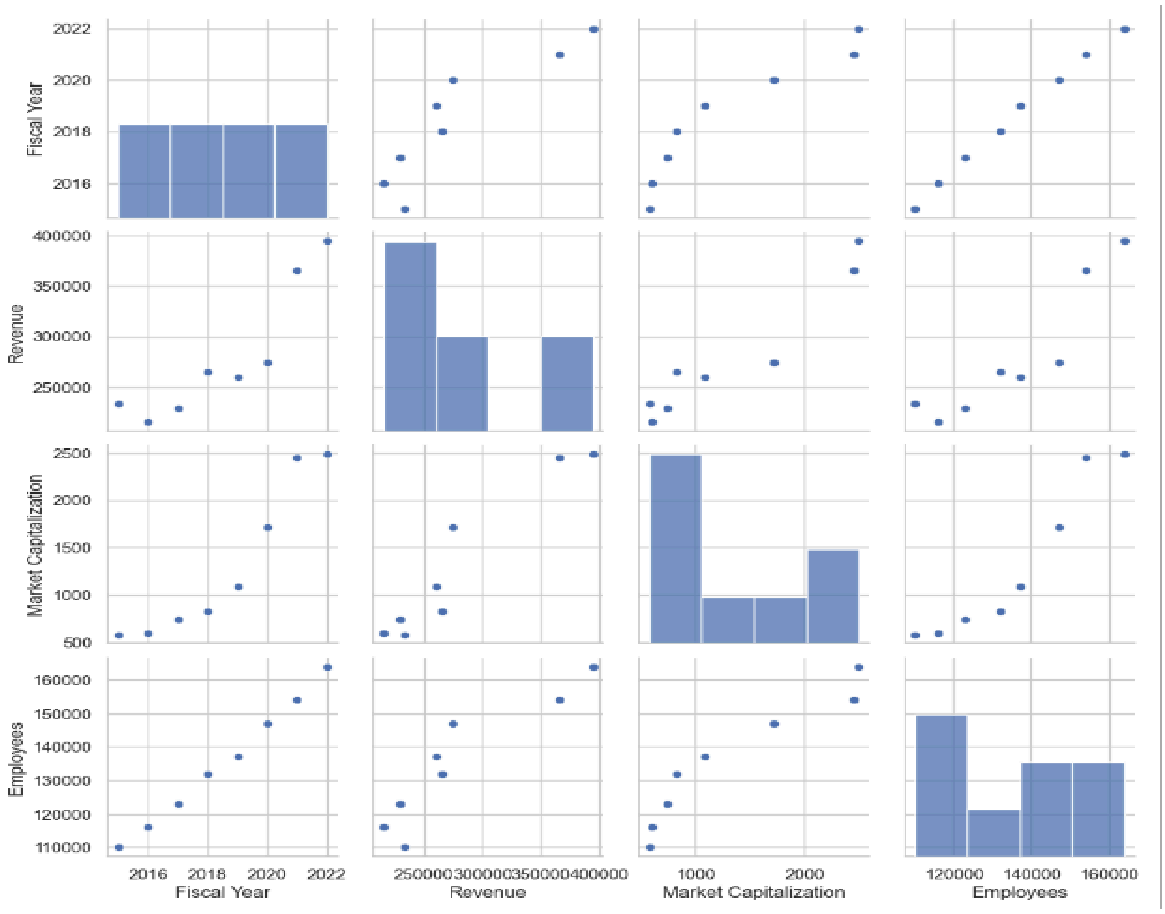


figure 14: Pairplot of Normalizing Factors

Chapter 3: Data Preparation

1-Introduction :

Data preprocessing is a mandatory step in data analysis that involves transforming raw data into a clean and usable format. It includes data cleaning (handling missing values and correcting outliers), data integration (combining data from different sources), data transformation (normalizing and converting data types), data reduction (simplifying data complexity), and data discretization (converting continuous data into discrete bins). Effective preprocessing ensures data quality and consistency, enhancing the performance of analytical models and the reliability of insights derived from the data.

Machine Learning Data Preparation:

2-Handling missing values :

Sum of missing values in carbon footprint DataFrame:

```
print(carbon_footprint_by_product.isna().sum().sum())# sum of all missing values
0
```

figure 15: Missing values in carbon footprint dataframe

Sum of missing values in greenhouse_gas emissions DataFrame :

```
print(greenhouse_gas_emissions.isna().sum().sum())# sum of all missing values
43
```

figure 16: Missing values in greenhouse gas emissions dataframe

We removed with missing values using KNNImputer and the figure shown below demonstrates this.

```
Before Imputing Data:
Fiscal Year      0
Category         0
Type             0
Scope            0
Description      0
Emissions       27
dtype: int64
After Imputing Data:

Fiscal Year     0
Category        0
Type            0
Scope           0
Description     0
Emissions       0
dtype: int64
```

figure 17: Removed Missing values in greenhouse gas emissions dataframe

Sum of missing values in normalizing factors DataFrame :

```
print(normalizing_factors.isna().sum().sum())# sum of all missing values in the whole dataset
0
```

figure 18: Missing values in normalizing factors dataframe

3-Handling Duplicates Values :

Duplicates in the three DataFrames :

No duplicates values found in the three dataframes

```
# Check for duplicate rows
duplicates = greenhouse_gas_emissions.duplicated()
print("Duplicate rows:")
print(greenhouse_gas_emissions[duplicates])# no duplicates

Duplicate rows:
Empty DataFrame
Columns: [Fiscal Year, Category, Type, Scope, Description, Emissions]
Index: []

# Check for duplicate rows
duplicates = carbon_footprint_by_product.duplicated()
print("Duplicate rows:")
print(carbon_footprint_by_product[duplicates])

Duplicate rows:
Empty DataFrame
Columns: [Release Year, Product, Baseline Storage, Carbon Footprint]
Index: []

# Check for duplicate rows
duplicates = normalizing_factors.duplicated()
print("Duplicate rows:")
print(normalizing_factors[duplicates])

Duplicate rows:
Empty DataFrame
Columns: [Fiscal Year, Revenue, Market Capitalization, Employees]
Index: []
```

figure 19: Duplicates in the three dataframes

4-Dealing with outliers :

Outliers in carbon footprint by product DataFrame:

There are no outliers in the carbon footprint by product dataframe for the numerical variables as showing in these boxplots .

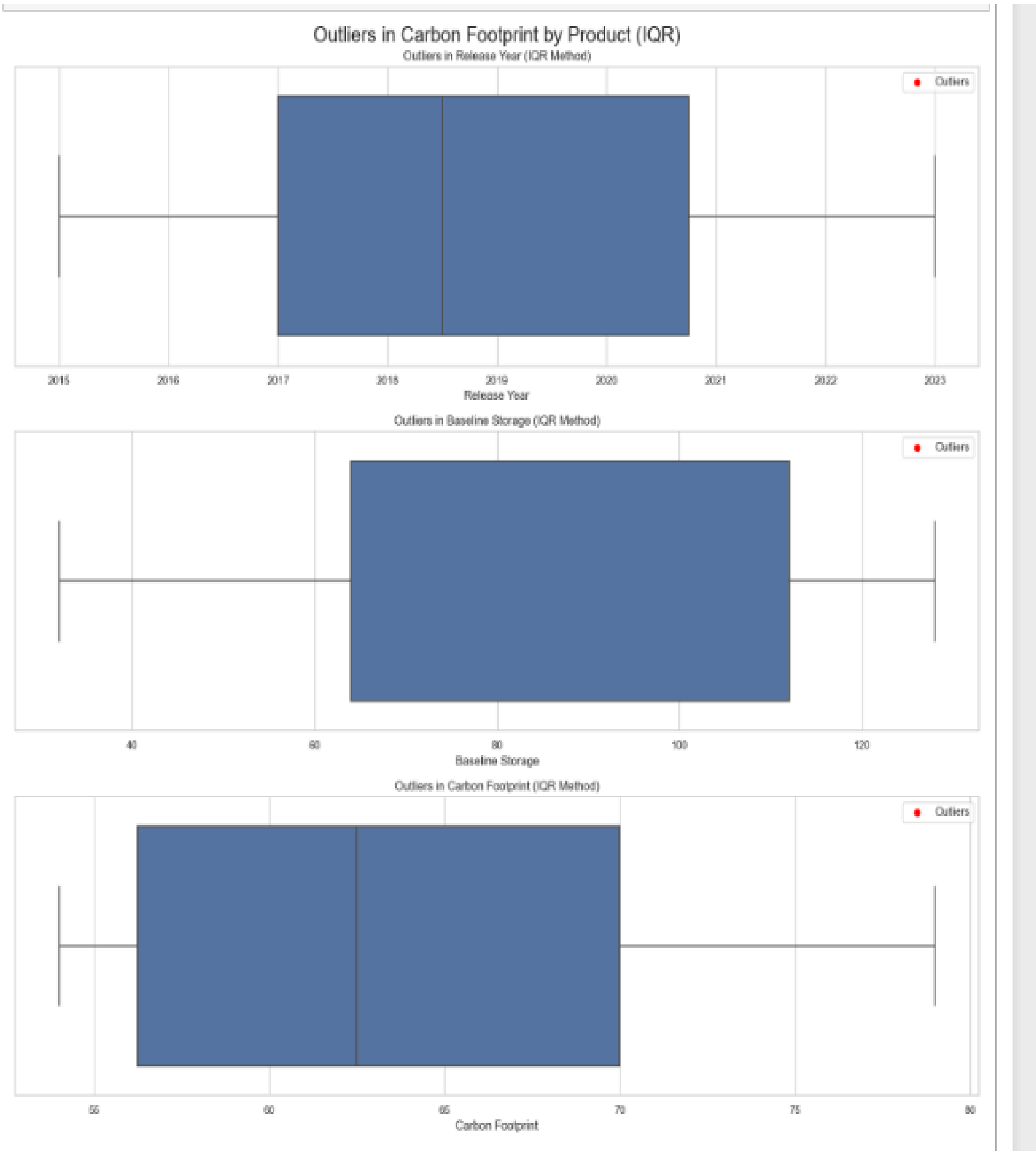


figure 20: Outliers in carbon footprint by product dataframe

Outliers in greenhouse gas emissions DataFrame:

There are outliers in the greenhouse gas emissions dataframe and more specifically in Emissions variable as it contains multiple zeros which is impossible .

We can see that most outliers are in upper border so we won't be removing any outlier .

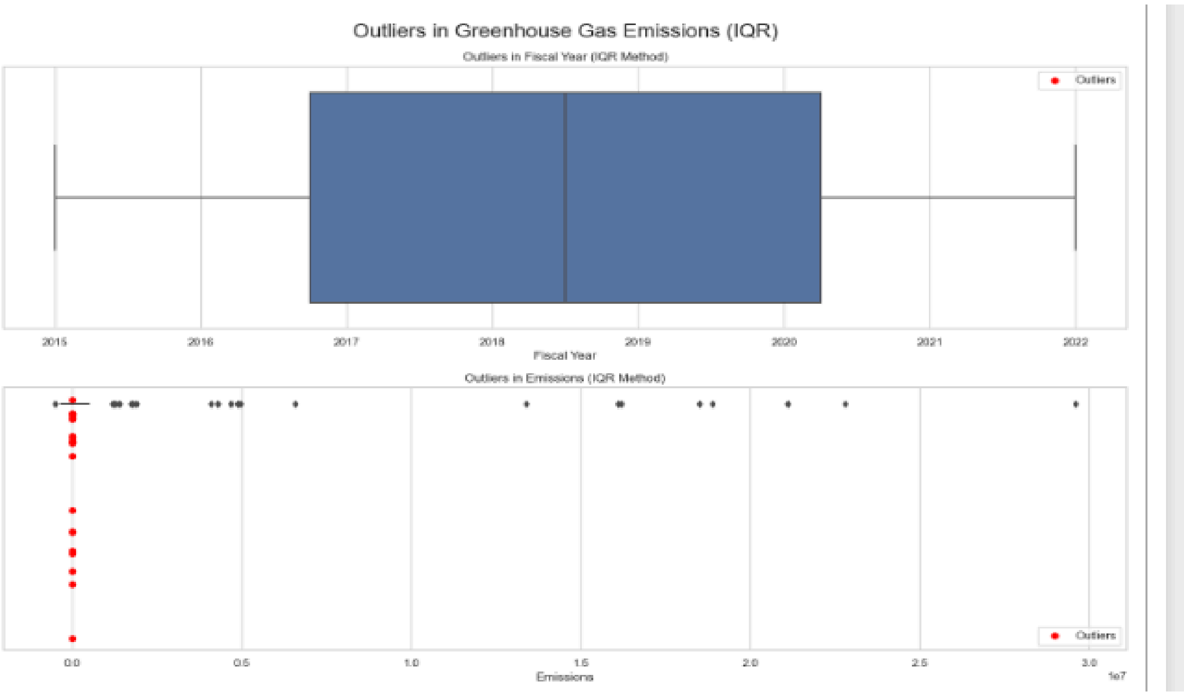


figure 21: Outliers in greenhouse gas emissions dataframe

Outliers in normalizing factors DataFrame:

There are no outliers in the normalizing factors dataframe for the numerical variables as showing in these boxplots .

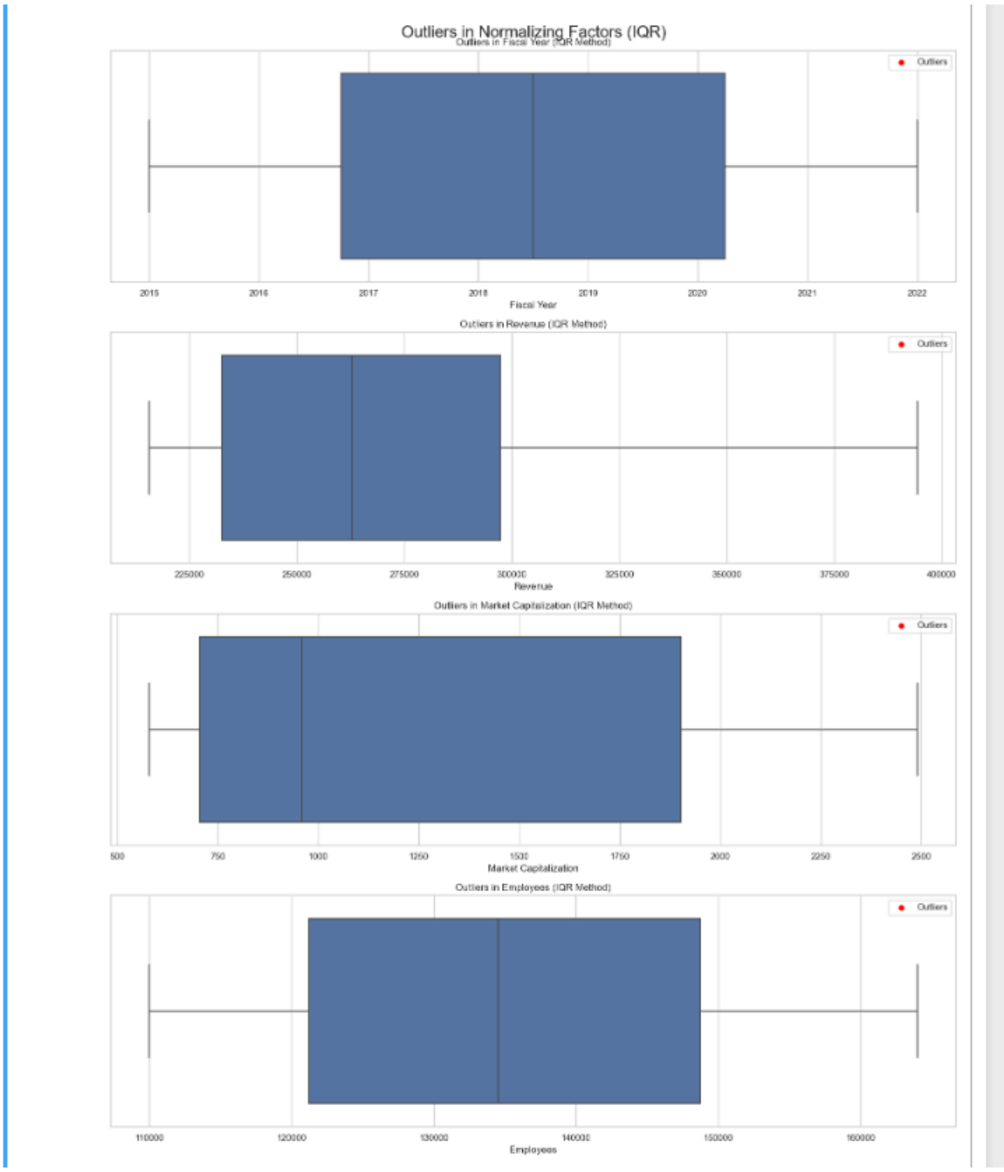


figure 22: Outliers in normalizing factors dataframe

5-Data Encoding :

We encoded the categorical variable ['Product'] in the carbon_footprint_by_product dataframe using the One Hot encoding to a numerical variable

```
# One Hot Encoding
categorical_columns = ['Product']
carbon_footprint_by_product = pd.get_dummies(carbon_footprint_by_product, columns=categorical_columns)
```

```
#now our data is encoded
plt.figure(figsize=(30,30))
sns.heatmap(carbon_footprint_by_product.corr(), annot= True,cmap='mako')
```

figure 23: OneHot encoding

Correlation Matrix (Carbon footprint by product DataFrame):

We can deduce from the correlation matrix displayed below that there is a correlation between the release year and the baseline storage .

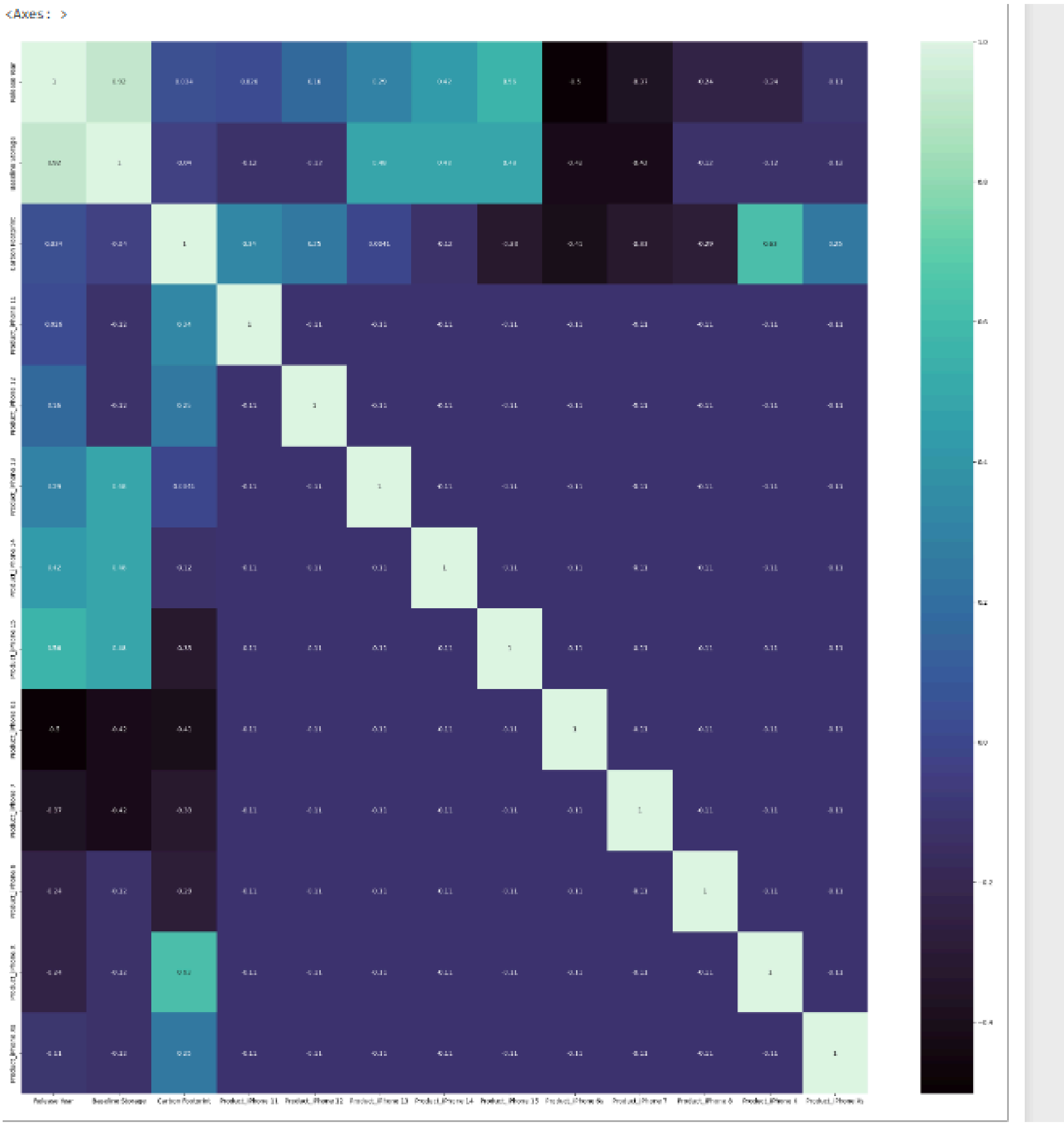


figure 24: Correlation matrix carbon footprint by product dataframe

Correlation Matrix (Greenhouse Gas Emissions DataFrame):

We encoded the categorical variables ['Scope', 'Category', 'Type', 'Description'] in the greenhouse_gas_emissions dataframe using the label encoding to a numerical variable . We can deduce from the correlation matrix displayed below that there is a correlation between the release year and the baseline storage.

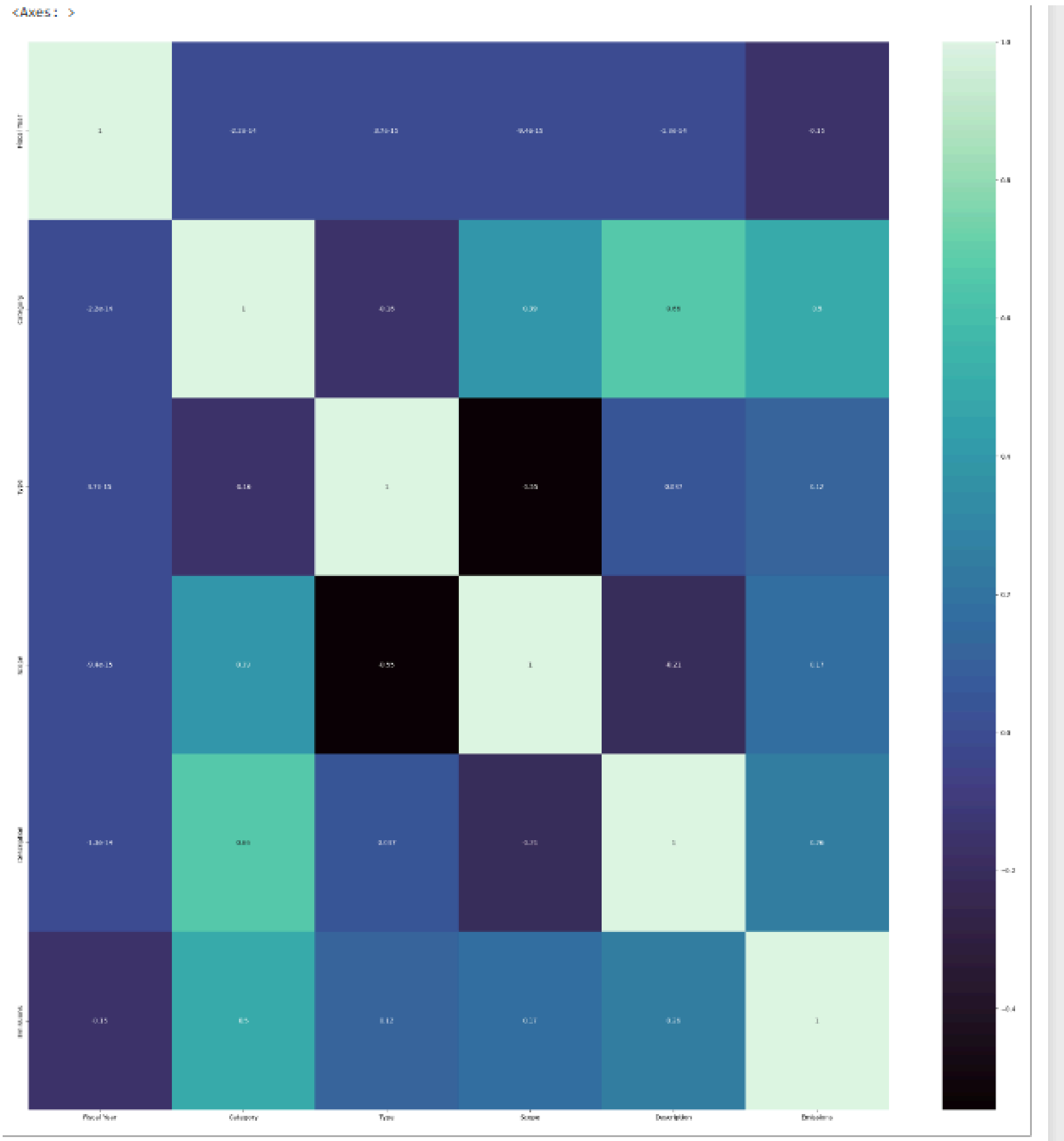


figure 25: Correlation matrix greenhouse gas emissions dataframe

6-Data Transformation(normalization):

Appling Min-Max Scaling on all Dataframes which is a normalization technique used to scale features to a fixed range, usually [0, 1].

```
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
greenhouse_gas_emissions_normalized = scaler.fit_transform(greenhouse_gas_emissions_imputed)
```

figure 26: Normalizing greenhouse gas emissions dataframe

```python
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
carbon_footprint_by_product_normalized = scaler.fit_transform(carbon_footprint_by_product)
```

figure 27: Normalizing carbon footprint by product dataframe

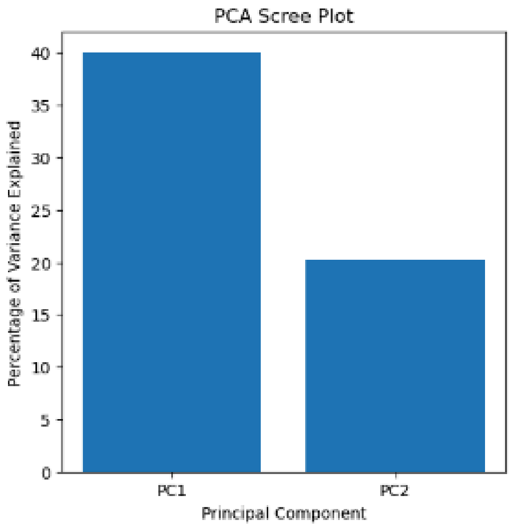
```python
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
normalizing_factors_normalized = scaler.fit_transform(normalizing_factors)
```

figure 28: Normalizing the normalizing factors dataframe

7-Dimensionality Reduction (PCA) :

PCA is a powerful tool for dimensionality reduction, helping to simplify datasets while retaining essential information.

```
Percentage of variance explained by each component: [40.03 20.25]
```




```
C:\Users\nadal\AppData\Local\Temp\ipykernel_20020\1792141154.py:27: UserWarning: No data for colormapping provided via 'c'. Par
ameters 'cmap' will be ignored
  plt.scatter(principal_df['PC1'], principal_df['PC2'], c='blue', cmap='viridis')
```

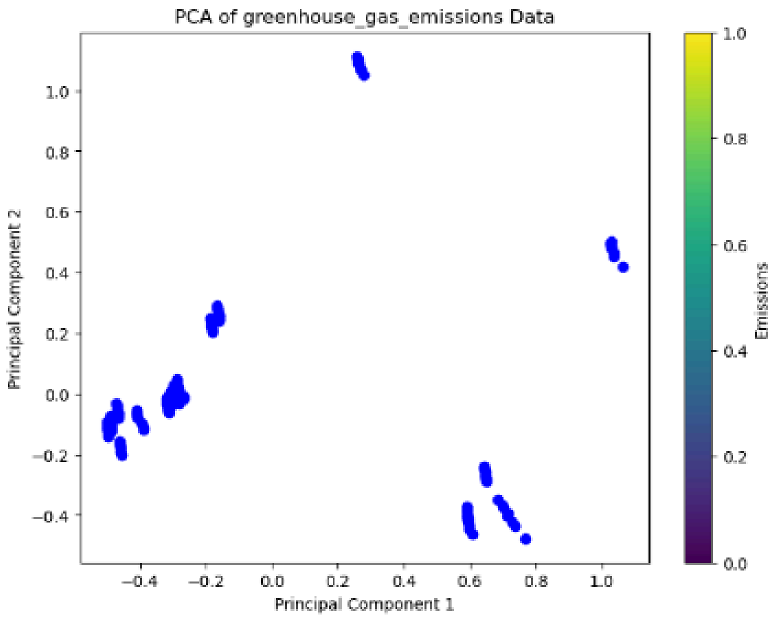
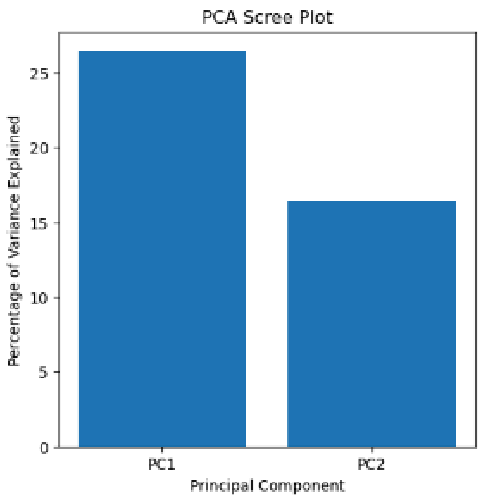


figure 29: PCA greenhouse gas emissions dataframe




```
C:\Users\nadal\AppData\Local\Temp\ipykernel_20020\3381151671.py:27: UserWarning: No data for colormapping provided via 'c'. Parameters 'cmap' will be ignored
  plt.scatter(principal_df['PC1'], principal_df['PC2'], c='blue', cmap='viridis')
```

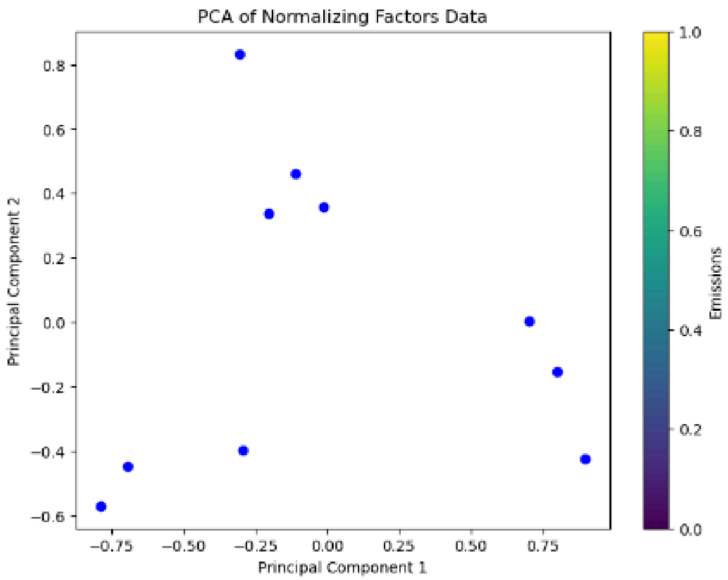
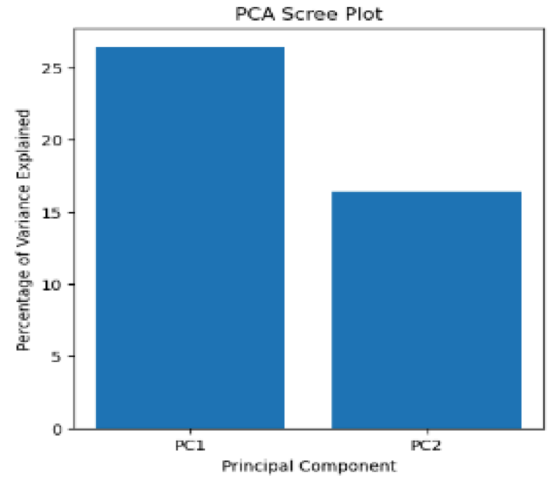


figure 30: PCA normalizing factors dataframe

```
Percentage of variance explained by each component: [26.42 16.43]
```




```
C:\Users\nadal\AppData\Local\Temp\ipykernel_20020\2857438785.py:27: UserWarning: No data for colormapping provided via 'c'. Par
ameters 'cmap' will be ignored
  plt.scatter(principal_df['PC1'], principal_df['PC2'], c='blue', cmap='viridis')
```

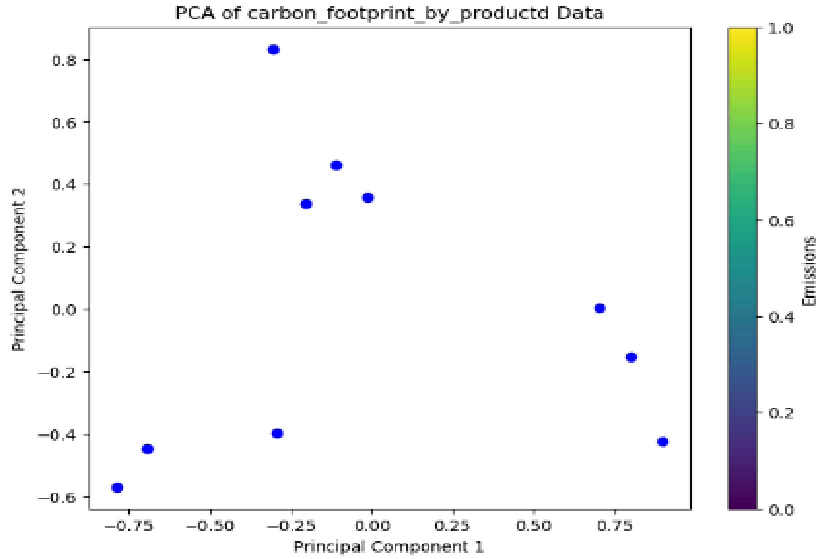
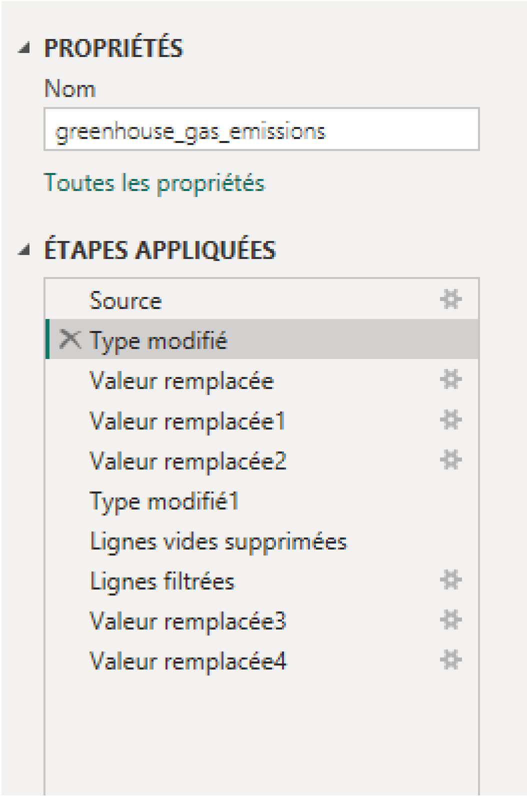


figure 31:PCA carbon footprint by product dataframe

Business Intelligence Data Preparation(ETL):

In our project, we will be leveraging the ETL process to handle and prepare the data for further analysis. This will involve extracting relevant data from multiple sources, transforming it to ensure consistency and accuracy, and finally loading it into our chosen data storage solution. By meticulously preparing our data through the ETL process, we can ensure that our analysis will be based on high-quality, reliable data.

We will delve into the specifics of this ETL process and demonstrate its implementation using the Power BI tool. Power BI will enable us to showcase how we can efficiently manage, transform, and visualize our data, providing clear and actionable insights. Through Power BI, we will highlight the effectiveness of our ETL process and its impact on our overall data analysis and business intelligence efforts.



Values Replaced: Replaced the Scope values by their number and filled the empty scope value by 'Scope 0' for rows with the category 'Carbon Removals'

Empty Rows Removed: Any rows with missing or empty values were removed in the column "Emissions" to clean the dataset.

Chapter 4: Data Warehouse Modeling

1-Introduction:

In the past ,we used to use a transactional database for our research, but we will switch to a better data warehouse modeling technique for our Online Analytical Processing (OLAP) study. The snowflake schema is perfect for our data warehouse design since it improves query efficiency and makes managing complex data relationships easier. We believe that using the snowflake schema will allow us to enable strong analytical capabilities while streamlining our data retrieval and storage procedures.

2-OLTP vs OLAP :

After conducting an in-depth comparative analysis between OLTP (Online Transactional Processing) and OLAP (Online Analytical Processing), we concluded that OLAP is the more suitable option for our needs. OLAP excels in multidimensional analysis, enabling us to perform complex queries and data analysis with greater efficiency and depth. This transition to OLAP will enhance our ability to derive meaningful insights from our data, streamline our analytical processes, and ultimately support more informed decision-making.

OLTP	OLAP
• **Short Transactions:** Optimized for fast and frequent read and write operations.	• **Multidimensional Analysis:** Allows data to be analyzed from multiple perspectives (dimensions) simultaneously.
• **High Concurrency:** Manages a large number of simultaneous users performing transactional operations.	• **Hierarchies:** Organizes data into hierarchical levels (e.g., year => quarter => month => day)
• **Data Integrity:** Uses locking	• **Aggregations:**Pre-aggregated

mechanisms and ACID transactions (Atomicity, Consistency, Isolation, Durability) to ensure data integrity.	calculations to improve the performance of analytical queries.

3-Snowflake Schema :

To implement this transition effectively, we have chosen to utilize the snowflake schema as our data model. The snowflake schema's normalized structure allows for efficient data organization and improved query performance. By leveraging this schema, we can manage complex data relationships more effectively and optimize our data retrieval and storage processes. This choice will enable us to harness the full potential of OLAP, ensuring robust and efficient analytical capabilities.

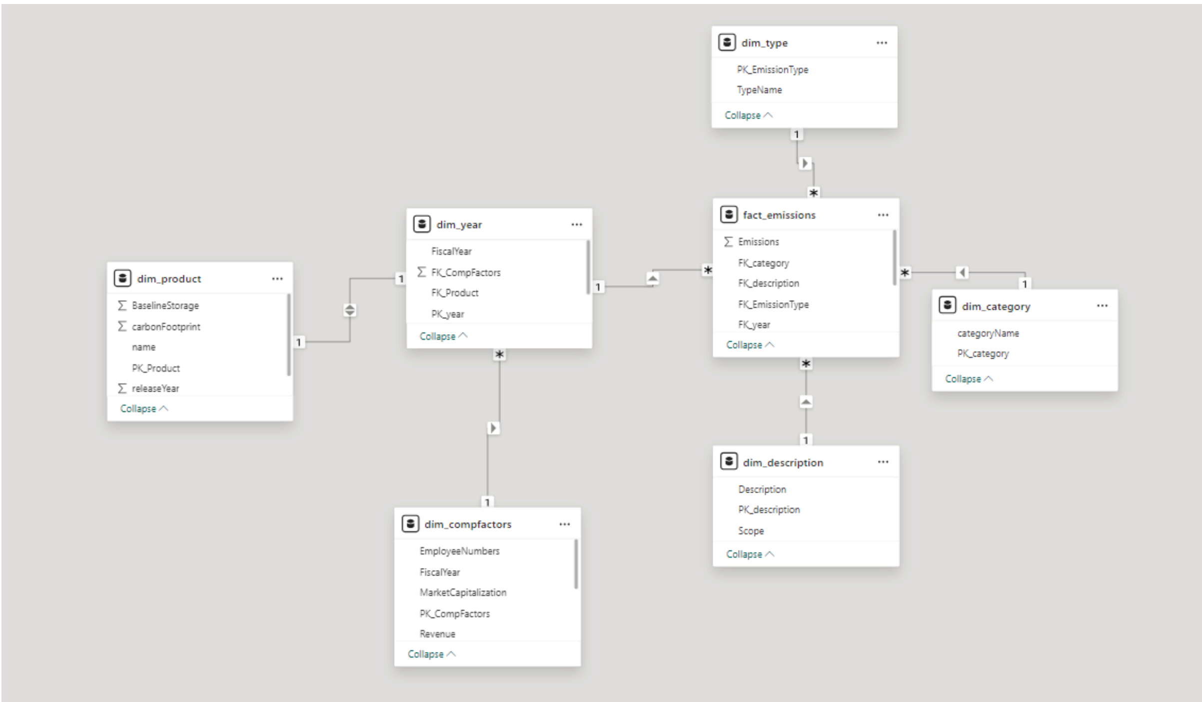


figure 32:Snowflake Schema

Chapter 6: Deployment

1-Introduction:

This chapter will specifically present the final step of the CRISP-DM methodology: **Deployment**. In this stage, we will implement the dashboards and deploy them using Power BI Service, allowing us to analyze and visualize the data through various interactive reporting features

2-Creation Of Dashboards :

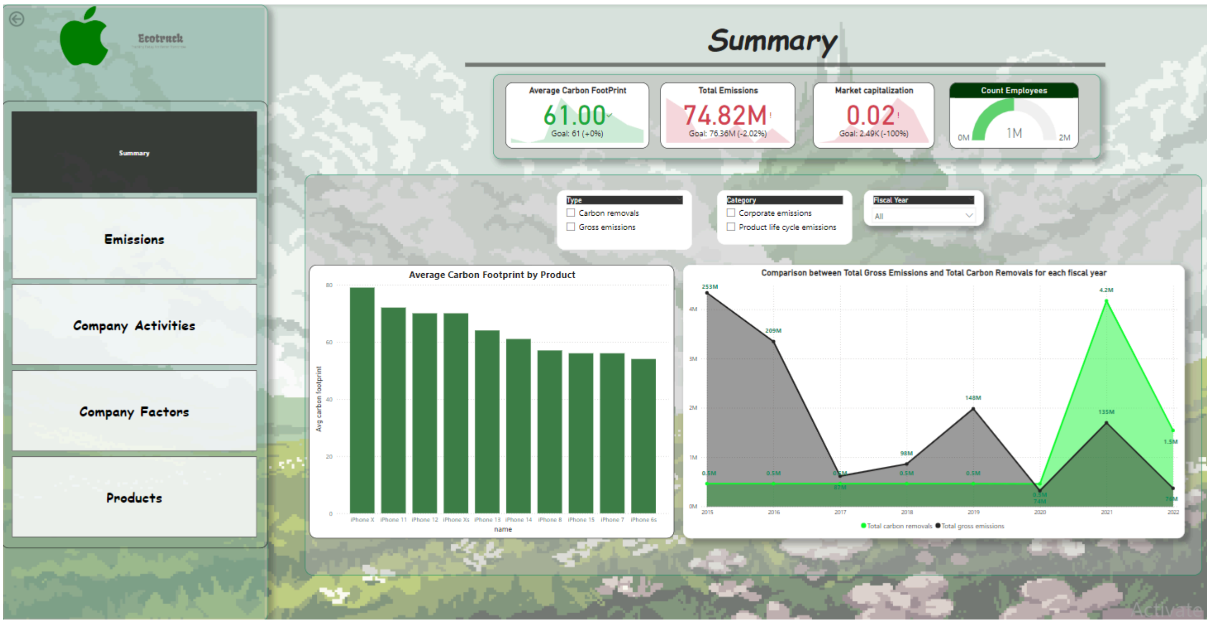
- **Overview of Dashboard Design:**

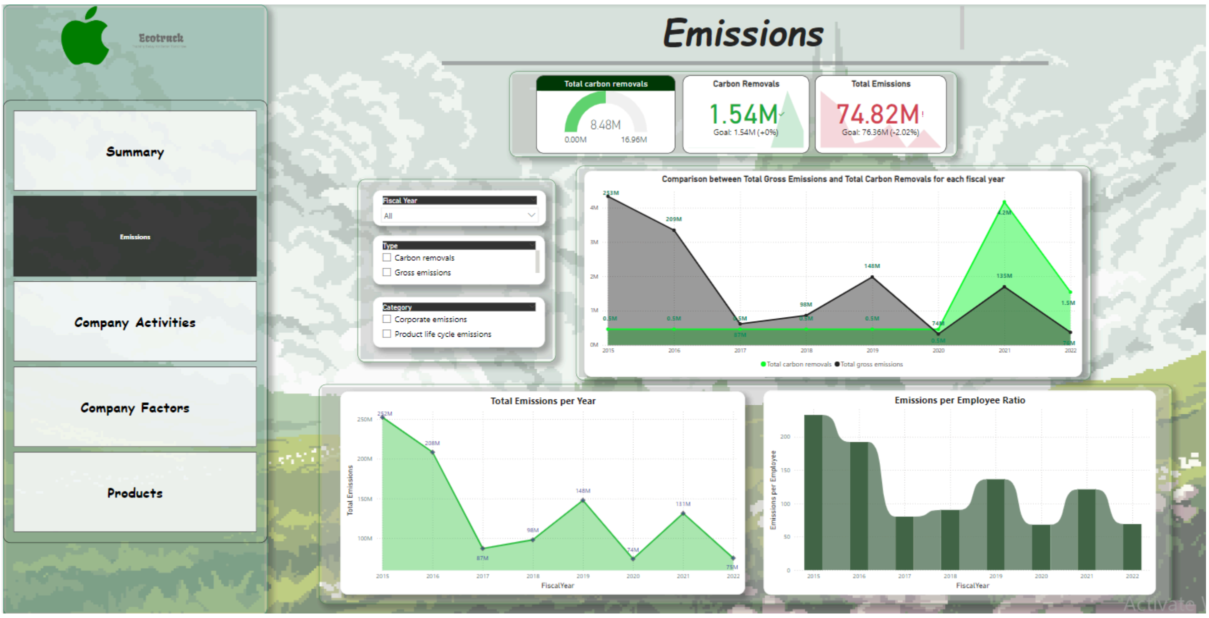
  We have finally created our dashboards using Microsoft Power BI Desktop before deploying them with Power BI Service:

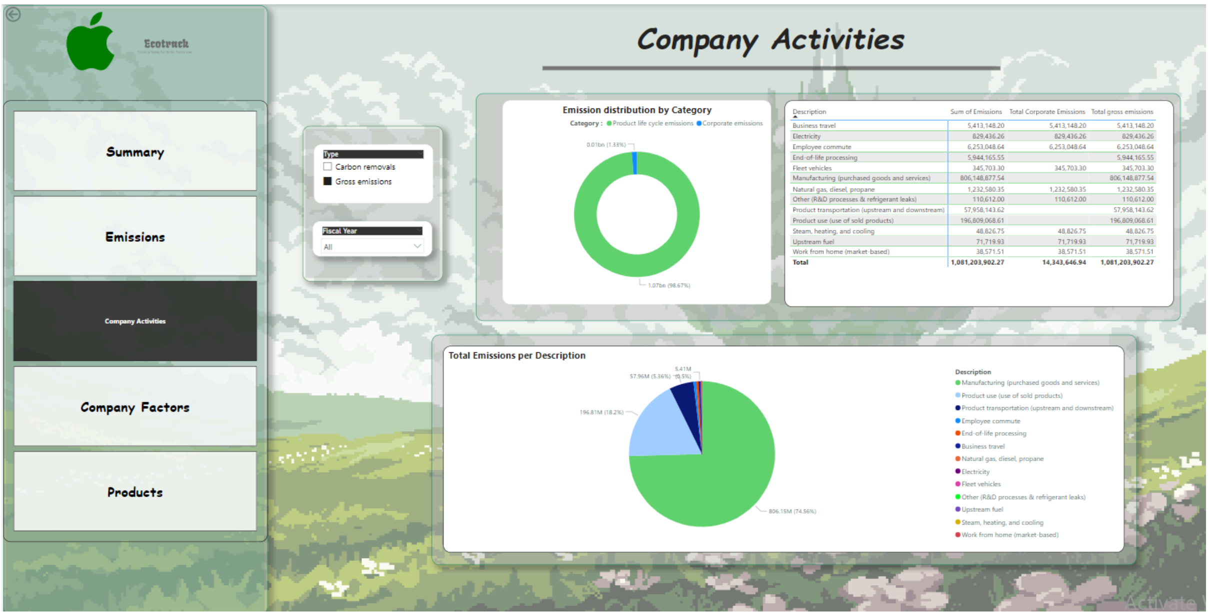
  1- Sustainability Team PowerBi Dashboard :

  The Sustainability Team Dashboard provides comprehensive insights into Apple's environmental impact, focusing on carbon footprint and greenhouse gas emissions. This dashboard includes visualizations that display the total carbon footprint, trends over time, and breakdowns by product and emission type. It highlights high-emission products, tracks emission reduction initiatives.

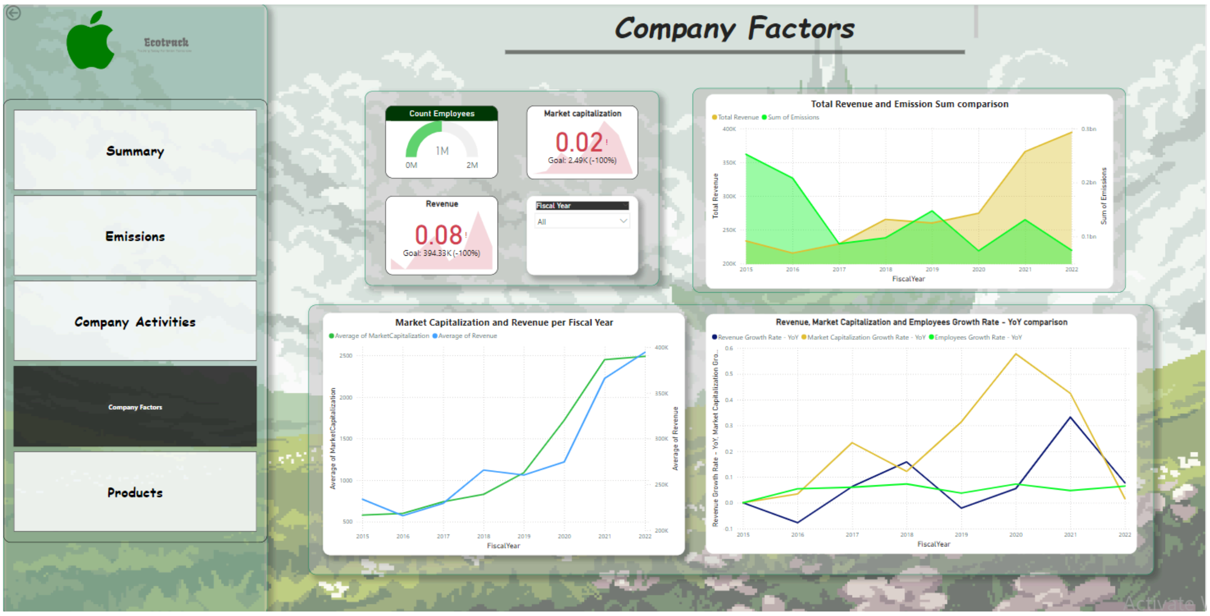
  Additionally, it includes compliance indicators and the impact of sustainability efforts, offering a detailed view to help the team make informed decisions and drive sustainability initiatives effectively.

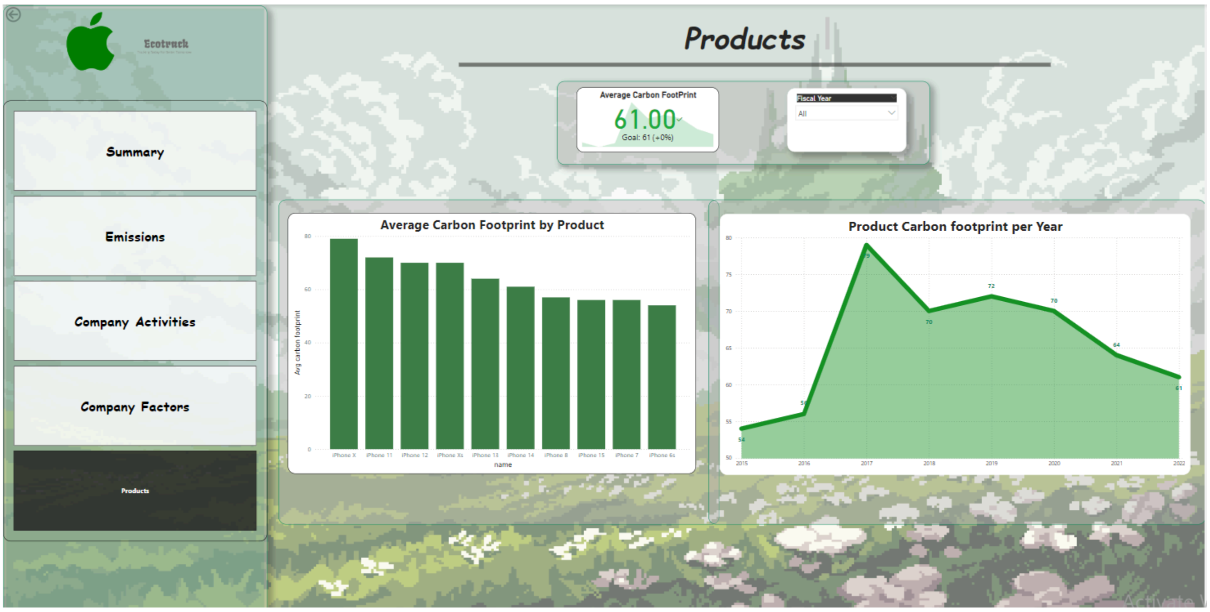

Ecotrack
Summary
Emissions
Company Activities
Company Factors
Products
Summary
Average Carbon FootPrint
61.00
Goal: 61 (+0%)
Total Emissions
74.82M
Goal: 76.36M (-2.02%)
Market capitalization
0.02
Goal: 2.49K (-100%)
Count Employees
1M
Type
Carbon removals
Gross emissions
Category
Corporate emissions
Product life cycle emissions
Fiscal Year
All
Average Carbon Footprint by Product
Comparison between Total Gross Emissions and Total Carbon Removals for each fiscal year


Ecotrack
Summary
Emissions
Company Activities
Company Factors
Products
Emissions
Total carbon removals
8.48M
Carbon Removals
1.54M
Goal: 1.54M (+0%)
Total Emissions
74.82M
Goal: 76.36M (-2.02%)
Fiscal Year
All
Type
Carbon removals
Gross emissions
Category
Corporate emissions
Product life cycle emissions
Comparison between Total Gross Emissions and Total Carbon Removals for each fiscal year
Total Emissions per Year
Emissions per Employee Ratio



Description	Sum of Emissions	Total Corporate Emissions	Total gross emissions
Business travel	5,413,148.20	5,413,148.20	5,413,148.20
Electricity	829,436.26	829,436.26	829,436.26
Employee commute	6,253,048.64	6,253,048.64	6,253,048.64
End-of-life processing	5,944,165.55		5,944,165.55
Fleet vehicles	345,703.30	345,703.30	345,703.30
Manufacturing (purchased goods and services)	806,148,877.54		806,148,877.54
Natural gas, diesel, propane	1,232,580.35	1,232,580.35	1,232,580.35
Other (R&D processes & refrigerant leaks)	110,612.00	110,612.00	110,612.00
Product transportation (upstream and downstream)	57,958,143.62		57,958,143.62
Product use (use of sold products)	196,809,068.61		196,809,068.61
Steam, heating, and cooling	48,826.75	48,826.75	48,826.75
Upstream fuel	71,719.93	71,719.93	71,719.93
Work from home (market-based)	38,571.51	38,571.51	38,571.51
Total	**1,081,203,902.27**	**14,343,646.94**	**1,081,203,902.27**





- **<u>DAX measures:</u>**

DAX stands for Data Analysis Expressions. It is a collection of functions, operators, and constants that can be used in formulas or expressions to calculate and return values in Power BI. DAX is designed to work with relational data, making it an essential tool for creating sophisticated calculations, data models, and visualizations in Power BI.DAX is crucial for several reasons like Dynamic Calculations,Enhanced Data Models,Efficiency, Interactivity..

Here is a detailed explanation of the DAX measures used in our visualizations:

- `Total Revenue = SUMX(dim_compfactors,dim_compfactors[Revenue])`:This measure calculates the total revenue by summing up the Revenue column in the dim_compfactors table.

- `Total Product Life Cycle Emissions =CALCULATE(SUM(fact_emissions[Emissions]),FILTER( fact_emissions, RELATED(dim_category[categoryName]) = "Product life cycle emissions" ))`: This measure calculates the total emissions related to the product life cycle by filtering the fact_emissionstable where the dim_category is "Product life cycle emissions".

- `Total gross emissions =CALCULATE( SUM(fact_emissions[Emissions]), FILTER( fact_emissions,RELATED(dim_type[TypeName]) = "Gross Emissions”))`:This measure calculates the total gross emissions by filtering the fact_emissions table where the dim_type is "Gross Emissions".

- `Total Emissions =CALCULATE( SUM(fact_emissions[Emissions])):` This measure calculates the total corporate emissions by filtering the fact_emissions table where the Category is "Corporate Emissions".

- `Total carbon removals = ABS(CALCULATE( SUM(fact_emissions[Emissions]), FILTER( fact_emissions, RELATED(dim_type[TypeName]) = "Carbon Removals" ))):` This measure calculates the total carbon removals by filtering the fact_emissions table where the Type is "Carbon Removals".

```
Revenue Growth Rate - YoY =VAR CurrentYear =
MAX('dim_compfactors'[FiscalYear])

VAR CurrentYearRevenue = CALCULATE(
SUM('dim_compfactors'[Revenue]),'dim_compfactors'[FiscalYear] = CurrentYear
)

VAR PreviousYearRevenue = CALCULATE(
SUM('dim_compfactors'[Revenue]),'dim_compfactors'[FiscalYear] = CurrentYear
- 1  )RETURN  DIVIDE(CurrentYearRevenue - PreviousYearRevenue,
PreviousYearRevenue, 0)
```

This measure calculates the year-over-year revenue growth rate by comparing the current year's revenue to the previous year's revenue.

```
Market Capitalization Growth Rate - YoY =

VAR CurrentYear = MAX('dim_compfactors'[FiscalYear])

VAR CurrentYearMarketCap =

    CALCULATE(

        SUM('dim_compfactors'[MarketCapitalization]),
'dim_compfactors'[FiscalYear] = CurrentYear

    )

VAR PreviousYearMarketCap =

    CALCULATE(

        SUM('dim_compfactors'[MarketCapitalization]),
```

```
        'dim_compfactors'[FiscalYear] = CurrentYear - 1
    )
RETURN
    DIVIDE(CurrentYearMarketCap - PreviousYearMarketCap,
PreviousYearMarketCap, 0)
```

This measure calculates the year-over-year market capitalization growth rate by comparing the current year's market capitalization to the previous year's.

```
   -  Employees Growth Rate - YoY =
VAR CurrentYear = MAX('dim_compfactors'[FiscalYear])
VAR CurrentYearEmployees =
    CALCULATE(
        SUM('dim_compfactors'[EmployeeNumbers]),
        'dim_compfactors'[FiscalYear] = CurrentYear
    )
VAR PreviousYearEmployees =
    CALCULATE(
        SUM('dim_compfactors'[EmployeeNumbers]),
        'dim_compfactors'[FiscalYear] = CurrentYear - 1
    )
RETURN
    DIVIDE(CurrentYearEmployees - PreviousYearEmployees,
PreviousYearEmployees, 0)
```

:This measure calculates the year-over-year employee growth rate by comparing the current year's number of employees to the previous year's.

-`Count Employees = sum(dim_compfactors[EmployeeNumbers])`:This measure counts the total number of employees by summing up the Employees column in the dim_compfactors table.

-`Avg carbon footprint = AVERAGE(dim_product[carbonFootprint])` :This measure calculates the average carbon footprint by averaging the Carbon Footprint column in the carbon_footprint_by_product table.

3-Publishing dashboards on Power BI Service :

At this stage, we have already created dashboards using Power BI Desktop. Using a professional Microsoft account, we will present the various reports created on the web portal with Power BI Service. The Figure below shows the publication of dashboards on the Power BI Service.

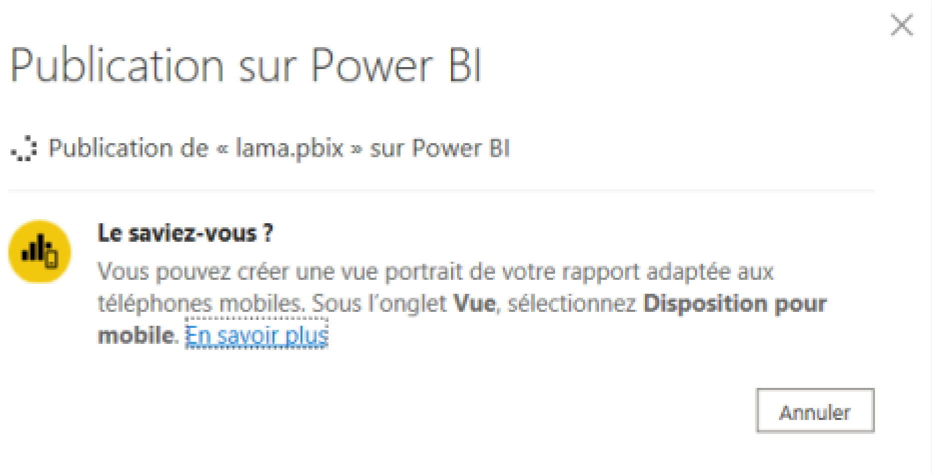


figure 37:Share of the report via Power BI Service