Report LAB 2.2 - AI1703

Activity 1:

Dashboards are tools that help aggregate, display, and analyze data in a visually intuitive and comprehensible manner. They are used to present key information succinctly and in an easy-to-follow format, enabling users to quickly grasp the situation and make effective decisions. By utilizing charts, graphs, and other visual elements, dashboards transform complex data into straightforward and easy-to-understand information, enhancing analytical and decision-making capabilities.

Advantages of Interactive Dashboards:

- Interactive dashboards allow users to interact directly with the data, enabling customization, filtering, and detailed analysis according to specific needs.
- Dashboards help save time in searching and analyzing information by quickly and visually displaying important data.
- Dashboards support rapid and accurate decision-making by providing detailed and timely information.
- Dashboards can integrate with various data sources and automatically update information, ensuring users always have the most recent data.

Activity 2:

The "Greenhouse Gas Giants" dataset comprises three CSV files (Emission High, Medium, and Low granularity) sourced from the project website Carbon Majors and made available on Kaggle through the following link: Greenhouse Gas Giants on Kaggle.

Our focus is on the Medium Granularity dataset, which contains 12,551 rows and 7 features. These features are:

1. **year**: The year of the recorded data.
2. **parent_entity**: The name of the parent entity responsible for the emissions.
3. **parent_type**: The type of the parent entity (e.g., state-owned, investor-owned, nation state).
4. **commodity**: The type of commodity produced by the entity.
5. **production_value**: The value of commodity production.
6. **production_unit**: The unit of measurement for the production value.
7. **total_emissions_MtCO2e**: The total emissions measured in million tonnes of CO2 equivalent (MtCO2e).

	year	parent_entity	parent_type	commodity	production_value	production_unit	total_emissions_MtCO2e
0	1962	Abu Dhabi National Oil Company	State-owned Entity	Oil & NGL	0.91250	Million bbl/yr	0.363885
1	1962	Abu Dhabi National Oil Company	State-owned Entity	Natural Gas	1.84325	Bcf/yr	0.134355
2	1963	Abu Dhabi National Oil Company	State-owned Entity	Oil & NGL	1.82500	Million bbl/yr	0.727770
3	1963	Abu Dhabi National Oil Company	State-owned Entity	Natural Gas	4.42380	Bcf/yr	0.322453
4	1964	Abu Dhabi National Oil Company	State-owned Entity	Oil & NGL	7.30000	Million bbl/yr	2.911079

Figure 1: The Dataset of "Greenhouse Gas Giants"

The dataset is comprehensive and does not contain any missing values, providing a solid foundation for analysis. So, we can skip that processing to step-by-step guide for performing Exploratory Data Analysis (EDA) by Dashboard

```
data.info()
✓ 0.0s
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12551 entries, 0 to 12550
Data columns (total 7 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   year                    12551 non-null  int64
 1   parent_entity           12551 non-null  object
 2   parent_type             12551 non-null  object
 3   commodity               12551 non-null  object
 4   production_value        12551 non-null  float64
 5   production_unit         12551 non-null  object
 6   total_emissions_MtCO2e  12551 non-null  float64
dtypes: float64(2), int64(1), object(4)
memory usage: 686.5+ KB
```

Figure 2 : The information of dataset

Firstly, we import the necessary libraries to show the dashboard about the datasets

```
import dash
from dash import dcc, html
from dash.dependencies import Input, Output
import plotly.express as px
import pandas as pd
```

Figure 3: Necessary libraries of the dataset

We divided our main dataset into two smaller subsets to facilitate easier visualization in a dashboard, as the parent entities often overlap across the years. By doing so, it becomes more manageable to visualize data over time.

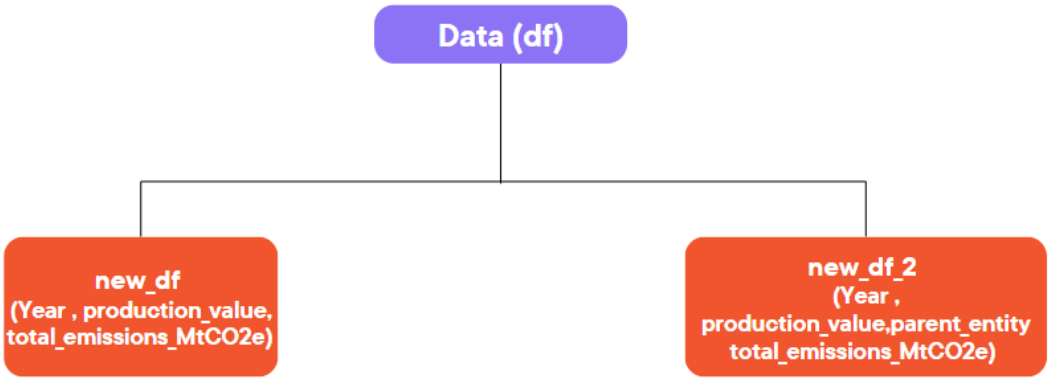


Figure 4: Divided datasets with 2 small datasets

Secondly, exploring the dataset by creating line charts that show the total production value and total CO2 emissions over the years with **new_df** dataset

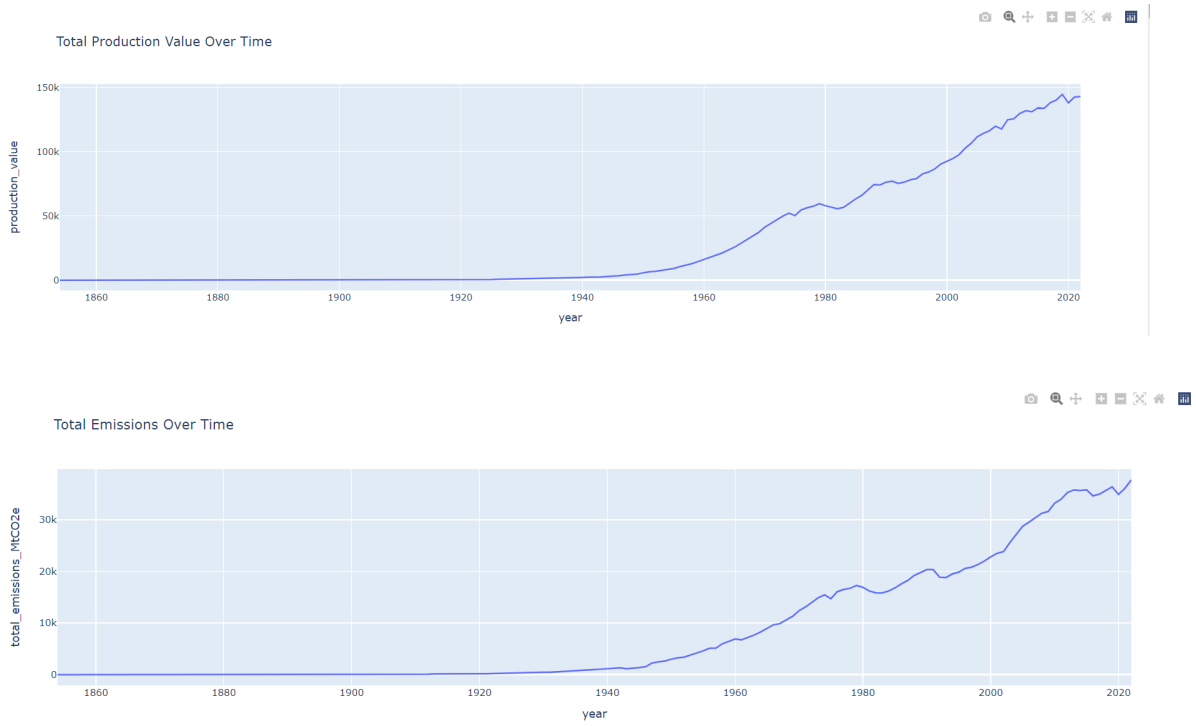


Figure 5 : Line plots showing Total Production and Total Emissions over time

The line charts for Total Emissions and Total Production Value reveal a consistent upward trend over the years. This indicates that both the production activities and the associated CO2 emissions have been steadily increasing. Such a pattern suggests a direct correlation

between industrial growth and environmental impact, emphasizing the need for sustainable practices to mitigate emissions while accommodating production demands.

We can see more clearly the total output value of parent entities through the following chart over the years.

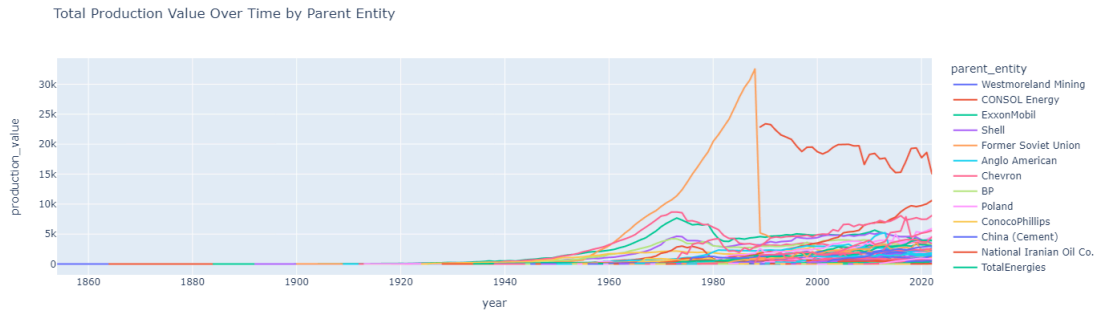


Figure 6: Total Production Value by Parent Entity over time

We can see that the top 5 largest production values belong to Former Soviet Union and Gazprom. The largest was 27.192 bcf/yr of Former Soviet Union in 1988. Beside that we can see the top 5 largest total emission CO2 belong to China (Coal) with the largest was 8646 million tones in 2022.

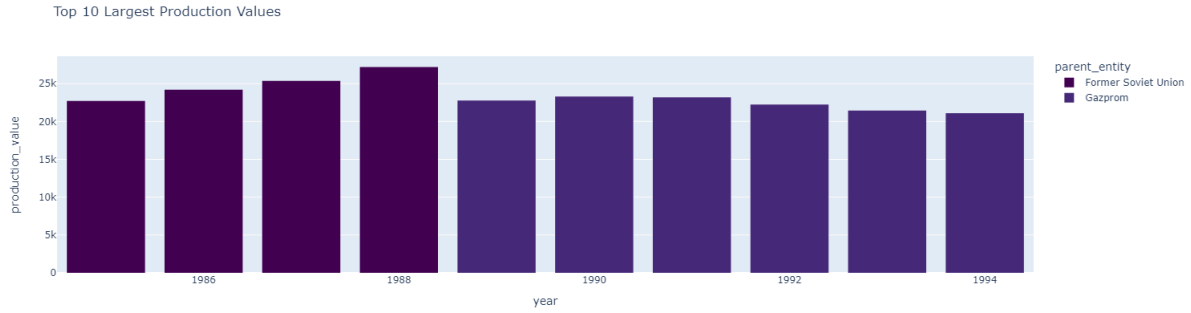


Figure 7: Bar chart of top 10 largest production values

We can see throughout that the total production value of Russian and Soviet corporations is always very high and accounts for a very large proportion of the total production values of the entire world.

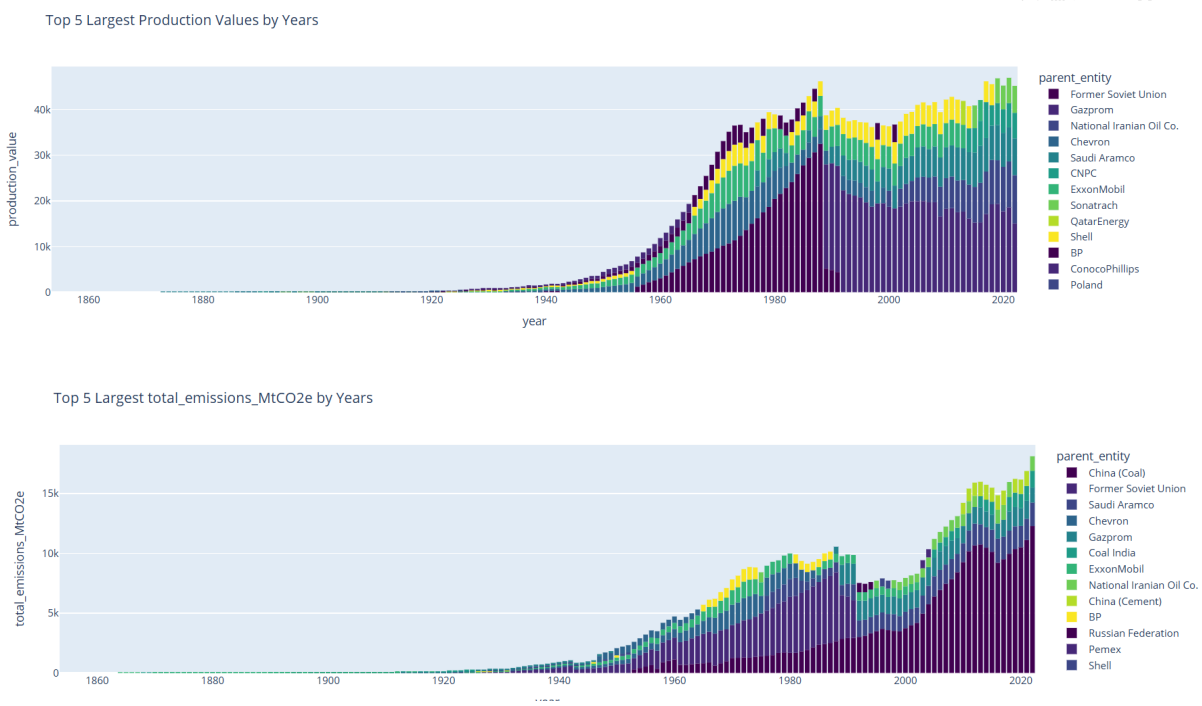


Figure 8: Top 5 Largest Total Production Values and Total Emissions by Parent Entity

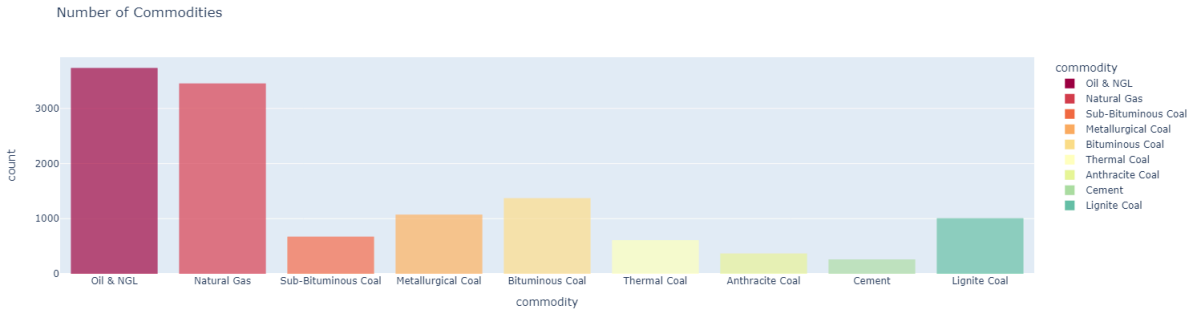


Figure 9: Bar Chart of Number of Commodities

In this dataset, we can see that commodities focus heavily on Oil & NGL and Natural Gas, in which commodities, especially related to coat, also account for a large proportion.

About the parent types, parent entities belonging to Investor-owned companies account for the largest proportion followed by state-owned entities and finally nation states account for the least proportion.

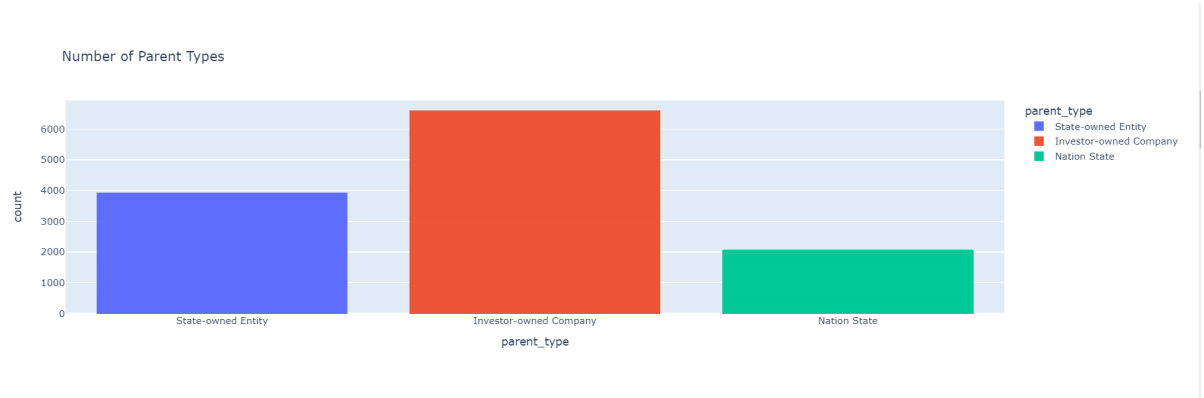
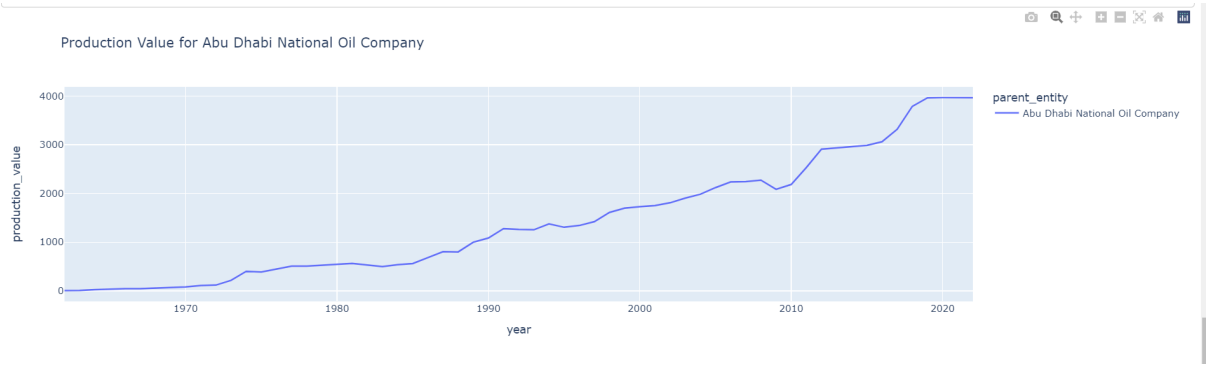
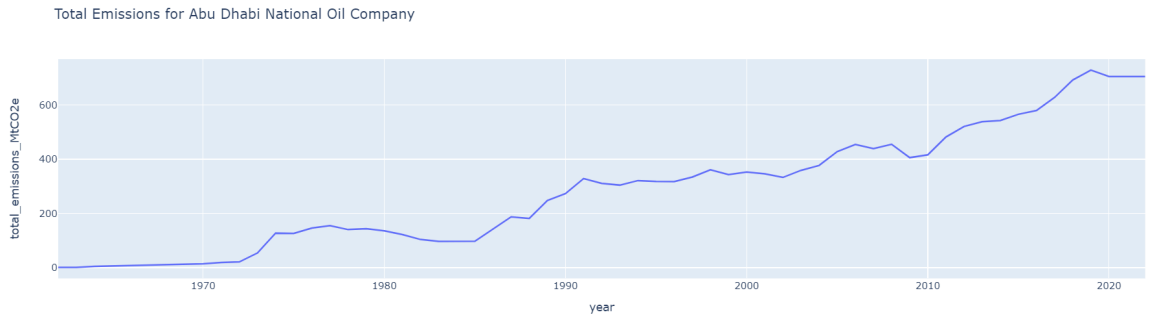


Figure 10: Bar Chart of Number of Parent types

To add details to the parent entity section, we will use a chart to show the total production value and total CO2 emission of a parent entity. As in this case of Abu Dhabi National Oil Company, we can see that both of the above charts have increased over the years and the emission per value has also fluctuate greatly.





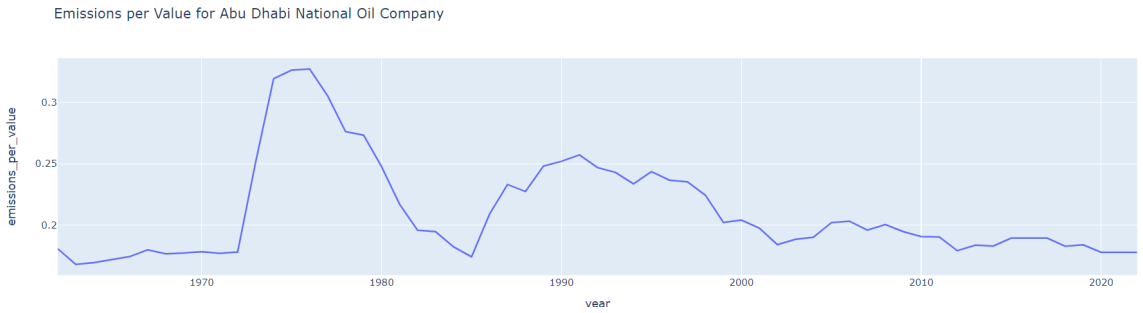


Figure 11: Line chart of Parent entity

Figure 12: Total emissions and Total Production Values of Parent Entities over time with Scatter Plot

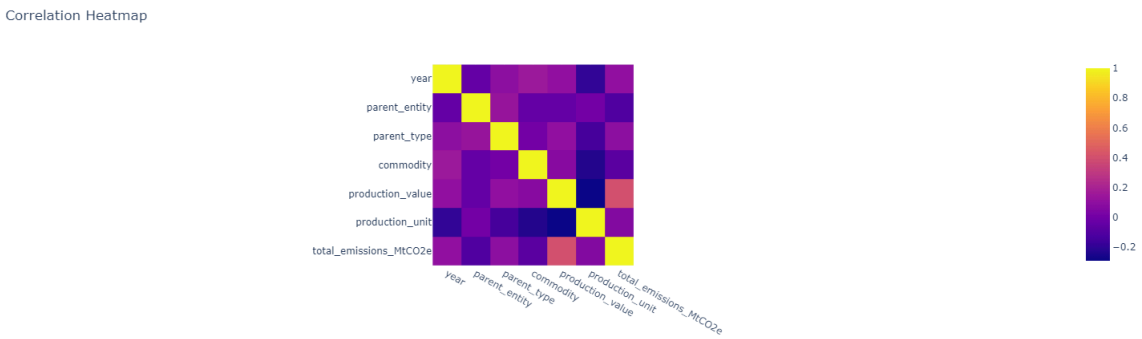


Figure: Correlation of Heatmap of the Dataset

We use correlation heatmaps to clearly visualize the correlation between pairs of variables, making it easier to recognize distinctive patterns or relationships between them.

To conclude, we completed basic EDA about the dataset with many useful features such as total production values and total emission CO2 of parent entities over time.