Visualization Project Report - Group 2

Student Name
1. Vu Tuyet Vy - V202401789
2. Le Minh Huyen – V202401743

Introduction

The primary dataset used in this project is derived from the Carbon Majors database, which tracks greenhouse gas emissions from fossil fuels and cement produced by companies historically, dating from as early as 1854. The methodology for emissions calculation follows detailed protocols initially established by Rick Heede in his 2014 paper "Carbon Majors: Methods & Results Report". While the original accounting methods and calculation protocols remain foundational, recent data storage, processing methods, and output formats have evolved significantly since their initial release by the Climate Accountability Institute. Emission factors utilized primarily originate from the Intergovernmental Panel on Climate Change (IPCC) guidelines and other reputable sources like the IEA and EIA. Calculations include scope 3 category 11 emissions (combustion of marketed products) and direct scope 1 emissions (venting, flaring, fugitive methane, and own fuel usage emissions).

To address disparities arising from varied production measurement units, all commodities were standardized to tonnes of oil equivalent (TOE), allowing meaningful comparisons across different fuel types. Additionally, exploratory data analysis (EDA) revealed that emissions data were secondary, calculated from production data using commodity-specific coefficients. Therefore, to establish a common basis for comparing emission levels, we used the emission rate per TOE to compare across different fuel types (Table 1).

Table 1. TOE rate by commodities

TOE rate by commodities

Commodity	Emission rate per unit	Emission rate per TOE
Lignite Coal	1.34	4.46
Metallurgical Coal	2.97	3.96
Bituminous Coal	2.71	3.88
Sub-Bituminous Coal	2.02	3.67
Thermal Coal	2.37	3.65
Natural Gas	0.07	3.06
Anthracite Coal	2.92	2.98
Oil & NGL	0.40	2.93
Cement	0.64	

To enhance analysis and insights, additional datasets concerning country-specific participation and compliance with international climate agreements (Kyoto and Paris agreements) were incorporated. These datasets detailed each country's commitment levels, agreement participation timelines, and compliance status, enriching the context for interpreting emissions trends and evaluating international policy effectiveness.

Data analysis

Question 1: Emission Trends and Regional Characteristics

This analysis focuses on understanding global and regional emission trends, emphasizing the significant emitters and their fuel-specific emissions contributions. Analyzing total emissions, commodity types, country-specific data, and company-specific details provides comprehensive insights into how global emissions patterns evolve. Identifying regions and commodities responsible for significant emissions is essential to develop targeted and effective climate mitigation strategies.

Three analytical dimensions were addressed using appropriate visualizations: Overall Analysis, Country-specific Analysis and Company-level Analysis. The visualization employs multiple interactive charts to comprehensively explore global emission trends and distributions from 2000 to 2022. Each chart was specifically selected to effectively represent distinct aspects of the data, providing both granular insights and a comprehensive overview

Overall Analysis

For overall data analysis, the interactive emission map visualizes the geographical distribution of CO2 emissions by country. In this map, each circle's size indicates the magnitude of a country's total emissions, allowing an intuitive understanding of which nations are the most significant contributors. Simultaneously, pie charts within each circle encode the composition of emissions by fuel type (commodity), represented through distinct colors. This dual encoding of size and color in the map is highly effective because it allows users to simultaneously assess geographic concentration, the diversity of fuel usage, and each country's emission scale. For instance, large circles in China, the USA, and Russia immediately highlight these as major emission contributors, with distinct internal pie compositions signaling varied energy profiles: China predominantly emitting from coal, the USA largely from oil and natural gas, and Russia presenting a relatively diverse emission profile.

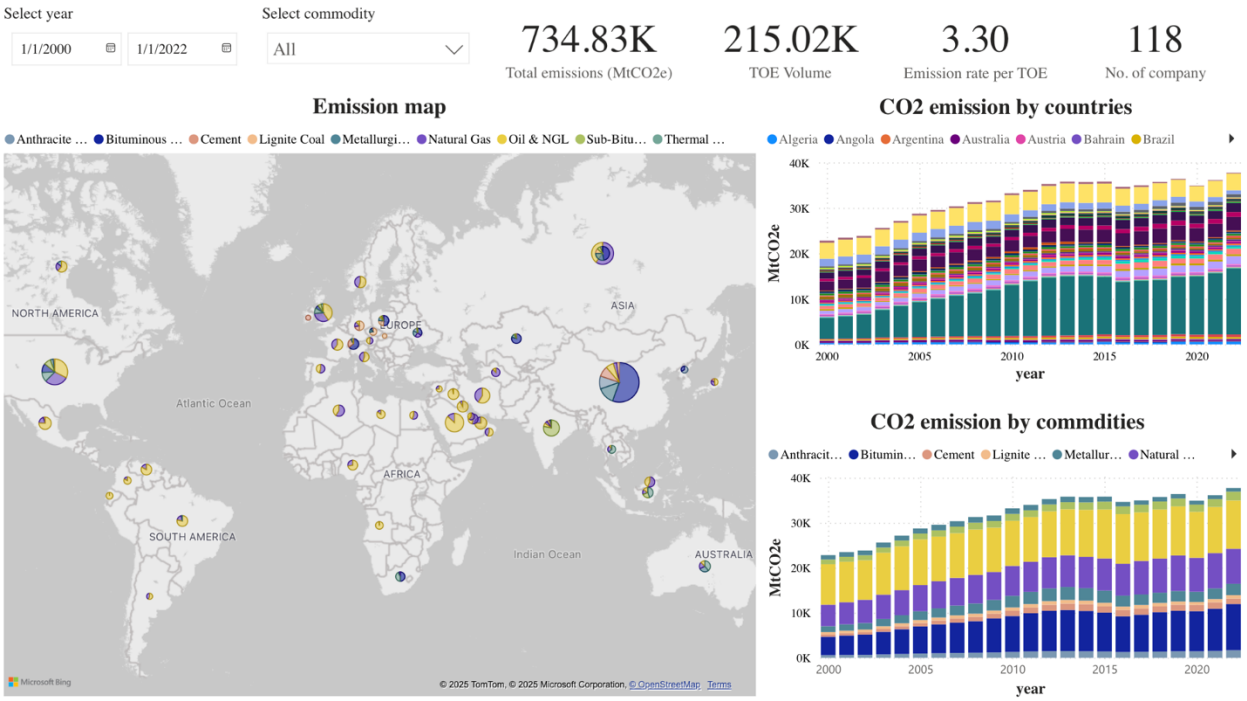

Figure 1. Overview Analysis

The two area charts complement the geographical insights from the map by illustrating trends and changes over time. The first area chart ("CO2 emission by countries") details how emissions from individual countries have evolved annually, using distinct colored segments for each nation. This visualization choice effectively illustrates both the absolute emissions by each country and their relative proportions to the global total. Observing this chart clearly highlights the growth or decline of each nation's emissions trajectory over time, enabling a nuanced analysis of global emission dynamics and identifying influential emitters across different historical periods. For example, China's rapidly growing contribution is prominently visible, reflecting its accelerated industrial growth, while emissions from countries such as the United States appear to stabilize or exhibit marginal declines, likely reflective of energy efficiency improvements or shifts towards alternative energies.

The second area chart ("CO2 emission by commodities") similarly provides a clear temporal view but focuses on the fuel sources themselves. It tracks the emissions associated with each commodity type - such as oil, coal (categorized into anthracite, bituminous, lignite, metallurgical, etc.), natural gas, and cement - by stacking their volumes over the years. This stacking enables users to readily discern overall emission growth and the specific fuel contributions within that growth. Notably, emissions from bituminous coal, oil, and natural gas prominently stand out:

- Bituminous coal emissions exhibit a robust growth trajectory, driven primarily by China's expanding industrial base and significant power-generation reliance on coal. Despite international calls to phase out coal, bituminous coal emissions remain notably high and

have continuously increased, underscoring the global dependency on this high-emission fuel source, particularly in developing economies.

- Oil & NGL (natural gas liquids) emissions illustrate a steady increase during the early 2000s, stabilizing somewhat in recent years. This trend suggests continued global dependence on oil for transportation, industrial use, and energy production, despite growing concerns and policy efforts targeting fossil fuel reduction.
- Natural gas emissions demonstrate a consistent upward trend over the analyzed period, reflecting natural gas's growing role as a transitional fuel towards cleaner energy sources. Its increased usage can partly be attributed to lower CO2 emissions per energy unit compared to coal, making it a favorable interim solution in energy transition strategies.

Country-specific Analysis

To effectively analyze country-specific emissions, the chosen visualizations include multiple interactive area and donut charts, strategically selected to depict emission trends, fuel composition, company types, and company-specific contributions clearly and comprehensively.

- Stacked area charts: Employed to illustrate the trends of emissions over time, facilitating easy visualization of the growth, stability, or decline of emissions segmented by commodity type and company ownership structures. Importantly, vertical dashed lines representing key dates of entry into international climate agreements (e.g., Kyoto and Paris Agreements) help evaluate the compliance and effectiveness of these commitments.
- Donut charts: Selected to offer a clear snapshot of proportional distributions at a glance, showing the relative contribution of each commodity type and the significance of emissions from different companies. These charts succinctly summarize extensive data into intuitive visuals, enabling quick comparisons between components.

This combination of visualizations delivers multi-dimensional insights: temporal trends, policy impact analyses, and granular breakdowns by fuel source and corporate contributions, offering a holistic understanding of emissions at a national scale.

Analyzing China's emission profile through these visualizations provides valuable insights into its dominant coal consumption and limited effectiveness of international climate commitments. The stacked area chart reveals a substantial growth trend, sharply accelerating since the late 1990s, coinciding with China's rapid industrialization and economic expansion. The dominance of coal-particularly bituminous coal-is visually explicit, accounting for a significant majority of total emissions. The donut charts reinforce these insights, showing coal as the overwhelmingly dominant fuel source. Moreover, the company's emissions chart highlights a concentrated structure where state-owned entities, notably China Coal, CNPC, and Sinopec, dominate emissions, clearly reflecting China's centralized governance and energy-security-focused policies. Critically, marked dates of China's participation in the Kyoto (2002) and Paris Agreements (2016) do not correspond

to noticeable declines or even stabilization in emissions, suggesting limited practical compliance or impact. Despite international commitments, the emission trajectory continues upward, indicating internal policy imperatives emphasizing economic growth and energy security over emission reductions.

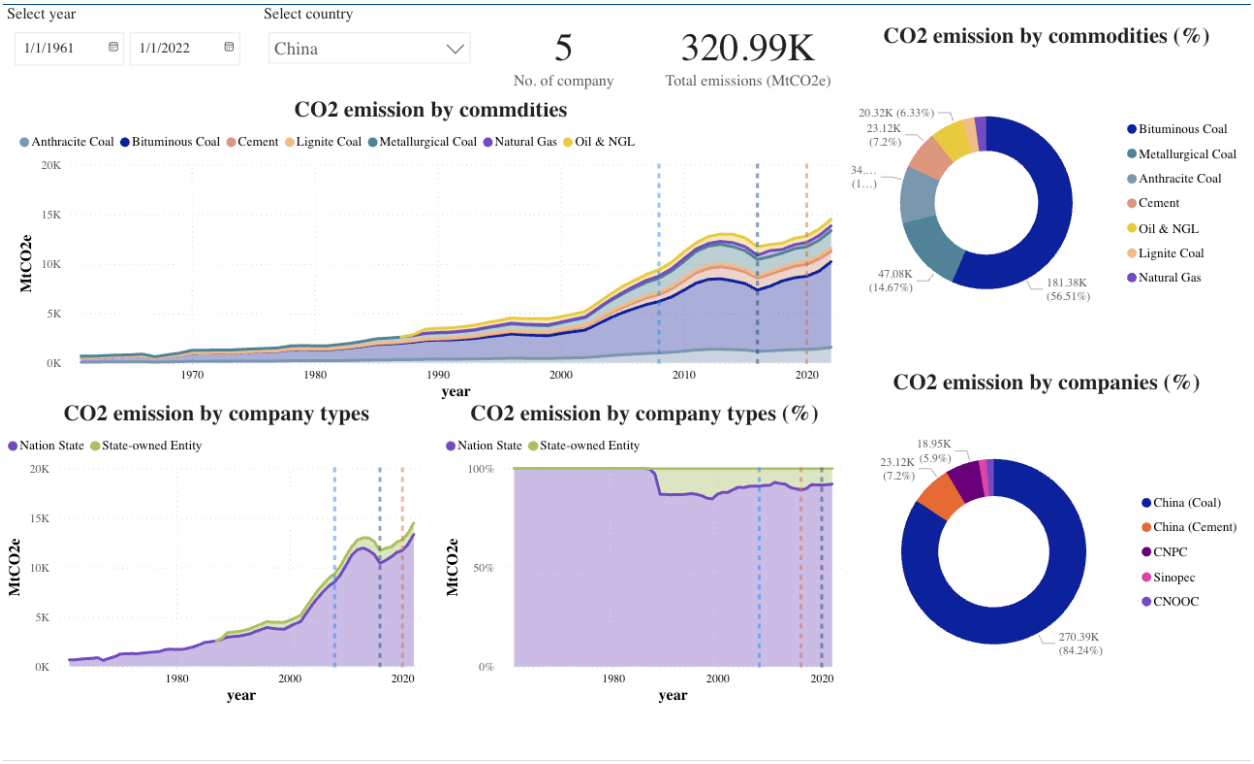

Figure 2. Country-specific Analysis - China

In contrast, the visualization for the United States illustrates a diversified energy mix predominantly reliant on oil, natural gas, and coal, with substantial emissions attributable to investor-owned companies. The stacked area chart highlights a fluctuating yet relatively stable trend over recent decades, reflecting efforts in technology improvements, regulatory pressures, and the growing use of renewable energies. The donut charts vividly illustrate the U.S.'s diversified emission profile, indicating that emissions primarily stem from oil and natural gas liquids, followed closely by coal and natural gas. The presence of multiple companies - ExxonMobil, Chevron, ConocoPhillips, among others - depicts a competitive market-driven structure. Analyzing the compliance aspect, the chart indicates U.S. participation in the Paris Agreement (2016). Still, similar to China, the visual evidence does not indicate substantial reductions directly linked to the commitment. The U.S.’s relatively consistent emissions trajectory suggests that voluntary agreements without binding enforcement mechanisms have limited effectiveness, especially within a free-market, investor-driven corporate structure.

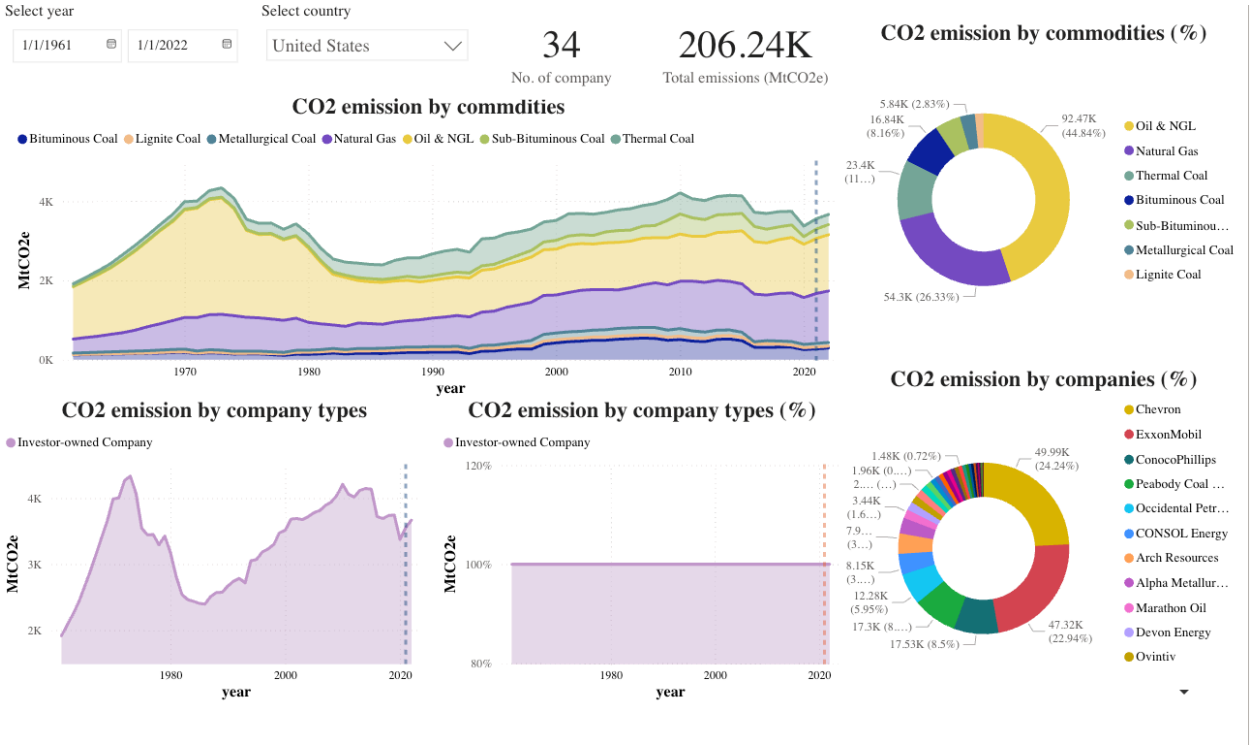

Figure 3. Country-specific Analysis - US

The U.K.'s emission analysis starkly contrasts the two previous examples, demonstrating significant emissions reduction and strong compliance with international agreements. The stacked area charts show a clear and sustained downward trend from the early 2000s onwards, coinciding with strong regulatory frameworks, proactive policy measures, and commitments under international climate accords. Visual markers for Kyoto (2002) and Paris (2016) agreements correlate closely with visible and sustained reductions in emissions, reflecting genuine policy effectiveness. The fuel composition highlights a substantial reduction in coal use, transitioning predominantly to oil and natural gas liquids. The donut chart further underscores a more concentrated corporate structure dominated by companies such as BP and Shell, entities which themselves have committed to substantial emissions reductions aligning with governmental policies. These data-driven insights underscore the effectiveness of stringent regulatory frameworks and corporate responsibility initiatives, confirming the U.K.’s successful alignment with international climate targets.

Examining other European nations, such as Switzerland and Spain, further validates the observed relationship between stringent regulatory frameworks, supranational agreements (notably through the European Union), and effective emission reductions. Both countries exhibit notable downward emission trajectories and diversified fuel mixes, transitioning toward lower-emission commodities and renewable sources. The effectiveness of emissions control in these countries is strongly correlated with EU-wide directives such as the European Green Deal, rigorous carbon pricing

mechanisms, and legally binding national commitments to renewable energy transitions. This supranational regulatory environment compels high compliance and accountability among member states, fostering coordinated actions and substantive emissions reductions. Moreover, visualizations demonstrate that in nations like Switzerland and Spain, investor-owned corporations dominate the emission landscape. These companies, under stringent regulatory scrutiny, are transitioning faster towards sustainable practices compared to state-controlled entities seen elsewhere.

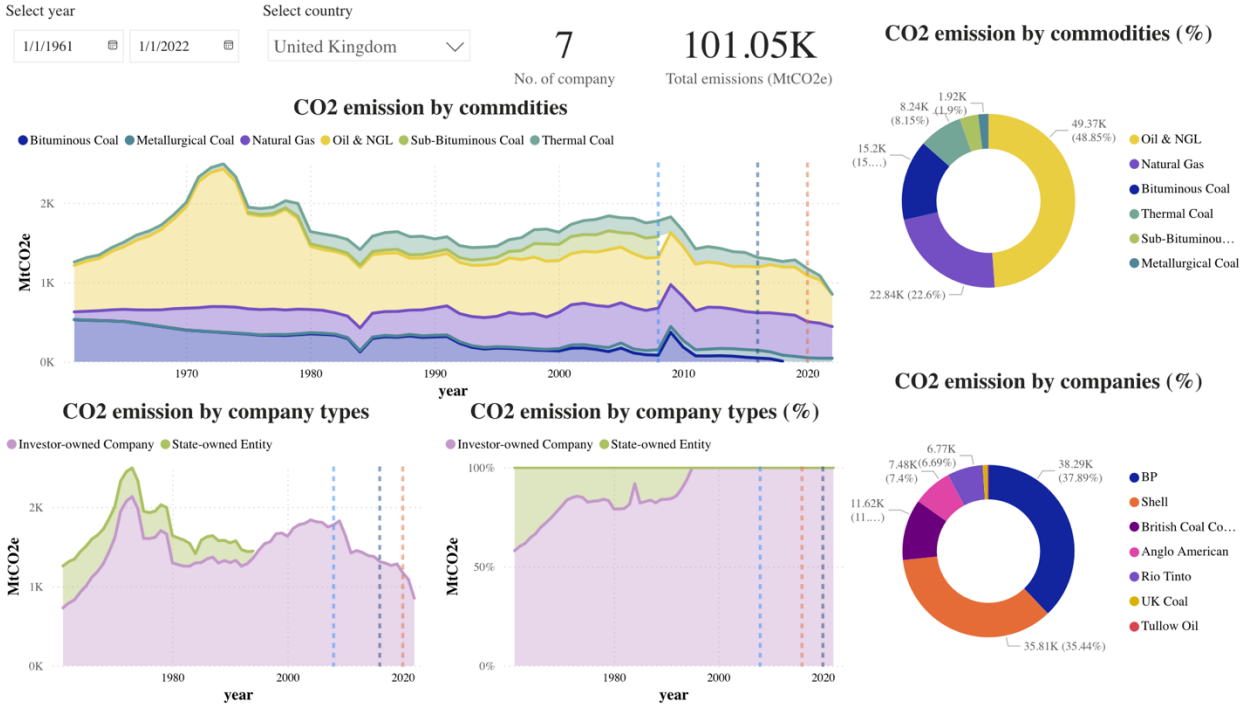

Figure 4. Country-specific Analysis - UK

Company-level Analysis

To explore emissions at the entity level, the selected visualizations include tabular summaries and bar charts, chosen for their clarity in ranking and comparing emissions among various companies globally. These visualizations explicitly communicate each company's total emissions, their contribution percentage to global emissions, and their respective home countries. Using tabular data enables straightforward comparisons and ranking, essential for identifying top emitters and analyzing the organizational nature of emissions (state-owned vs. investor-owned entities).

The entity-focused visualization highlights several crucial insights into global CO2 emissions:

- China Coal emerges distinctly as the leading emitter globally, alone responsible for approximately 26% of total emissions. This significant proportion underscores China's continued reliance on coal, specifically bituminous coal, which is the primary contributor to its emissions. Such prominence in emissions from a single state-owned entity

emphasizes both the scale of China's industrialization and the centralization of its energy policy.

- Notably, the top five emitting entities-China Coal, Saudi Aramco, Gazprom, National Iranian Oil Company, and Coal India-are all state-owned enterprises. This trend strongly suggests the significant influence state control has on energy production policies and highlights the intersection between national interests and energy security. Such dominance by state-owned enterprises also illustrates the limited market-driven flexibility these entities possess, potentially hindering rapid transitions toward renewable energy sources.
- The visualization distinctly reveals the dominance of entities involved in oil and gas production. Companies such as Saudi Aramco, Gazprom, National Iranian Oil Company, and ExxonMobil feature prominently in the global rankings. This dominance underscores global reliance on oil and natural gas, reflecting their critical roles in transportation, industrial energy, and global economic structures. It also highlights potential vulnerabilities in international energy transition efforts, considering the deeply entrenched interests and infrastructure associated with these fossil fuels.

Overall, the entity-level analysis, supported by clear visualization choices such as ranking tables and comparative bar charts, highlights critical strategic challenges. It clearly identifies the central role of state-owned companies and oil and gas corporations, thereby providing policymakers, investors, and environmental strategists with actionable insights for targeting regulatory reforms, corporate accountability, and transitioning efforts toward sustainable energy systems.

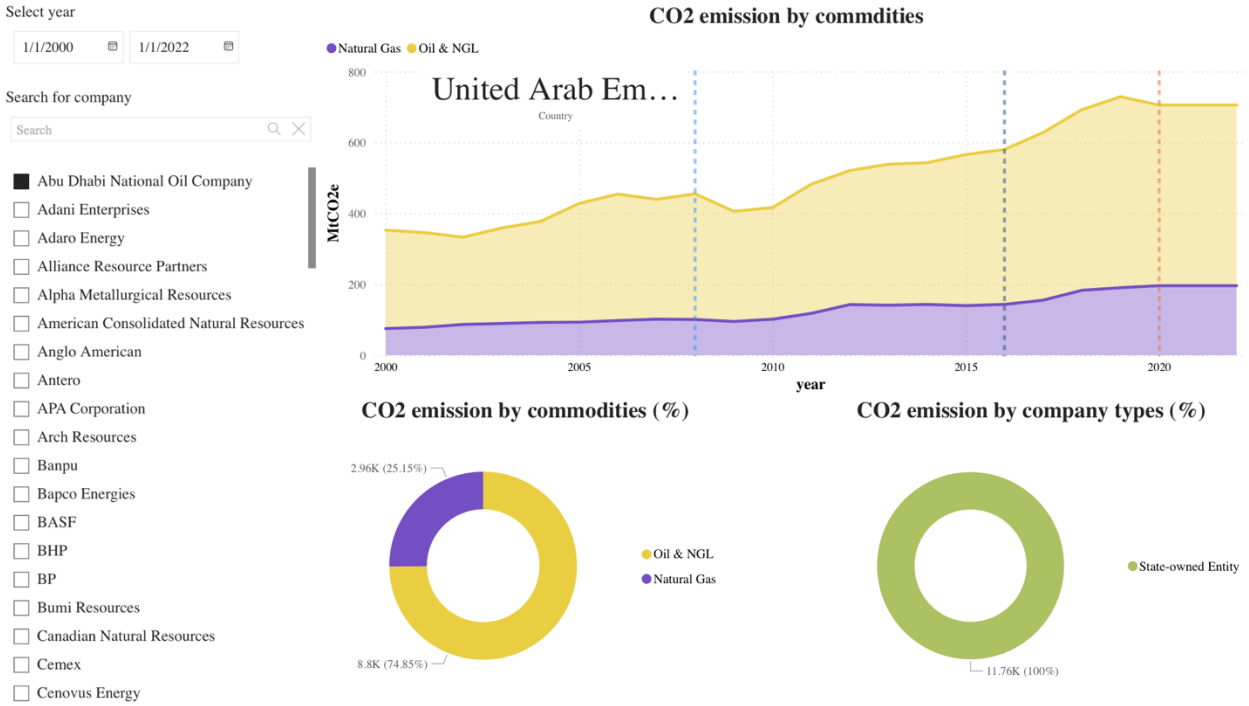

Figure 5. Company-level Analysis example (Abu Dhabi National Oil Company)

To explore emissions at the entity level, the selected visualizations include tabular summaries and bar charts, chosen for their clarity in ranking and comparing emissions among various companies globally. These visualizations explicitly communicate each company's total emissions, their contribution percentage to global emissions, and their respective home countries. Using tabular data enables straightforward comparisons and ranking, essential for identifying top emitters and analyzing the organizational nature of emissions (state-owned vs. investor-owned entities).

Conclusion for Question 1

The analysis provides a comprehensive overview of global CO_2 emission trends from major entities and clearly highlights how emissions have evolved, along with key insights into commodities and the influence of policy milestones. Overall, total emissions from major entities have steadily increased from 2000 to 2022. Bituminous coal, oil (Oil & NGL), and natural gas consistently emerge as dominant sources contributing significantly to global emissions.

Examining emissions against policy milestones such as the Kyoto Protocol and Paris Agreement reveals varied degrees of effectiveness:

- China shows little practical responsiveness to its international commitments. Despite its entry into global agreements, China's emissions, driven by state-controlled enterprises, have continued to rise sharply, particularly from coal consumption. This indicates domestic policy priorities largely oriented towards energy security and economic expansion, rather than emission reduction commitments.
- The United States demonstrates a marginally stable emissions trajectory despite participation in the Paris Agreement. The decentralized, market-driven nature of its energy sector suggests voluntary international commitments lack enforceable power to significantly reduce emissions, especially in the absence of stringent domestic regulations or mandates.
- In contrast, the United Kingdom and other European nations demonstrate notable alignment with international climate commitments, evidenced by substantial emission reductions following their adherence to agreements. This effective compliance likely results from robust regulatory environments, market adaptations, and proactive corporate responsibility initiatives, underscored by strong EU-wide directives like the European Green Deal and rigorous carbon-pricing frameworks.

Question 2: Characteristics of Emission Clusters

Cluster analysis was conducted to uncover structural emission patterns related to countries, commodities, and companies. Utilizing clustering methods allows us to identify specific emission behaviors and underlying relationships, thereby facilitating targeted climate policy

recommendations. Grouping emissions data into distinct clusters helps elucidate patterns that may not be immediately apparent through traditional statistical analyses.

K-means clustering, an unsupervised machine learning algorithm, was utilized in this study to systematically identify and interpret distinct patterns of emissions based on a combination of quantitative and categorical characteristics. The algorithm partitions the dataset into k distinct, non-overlapping clusters by iteratively assigning data points to clusters whose centroids (cluster centers) are nearest to the data points, based on Euclidean distance. After each assignment, the centroids are recalculated as the mean of the points within each cluster, and the process repeats until convergence-defined as minimal or no change in cluster assignments.

For this analysis, the features selected included one quantitative variable, total emissions measured in megatonnes of CO2 equivalent (total_emissions_MtCO2e), and several categorical variables: parent type, commodity type, and country. The inclusion of these categorical variables required preprocessing steps to convert them into numerical representations suitable for clustering. Specifically, categorical variables were processed using OneHotEncoder, converting categories into a binary matrix indicating the presence or absence of each category. The quantitative variable was scaled using StandardScaler to ensure uniform contribution to clustering outcomes, minimizing the distortion due to differences in data scales.

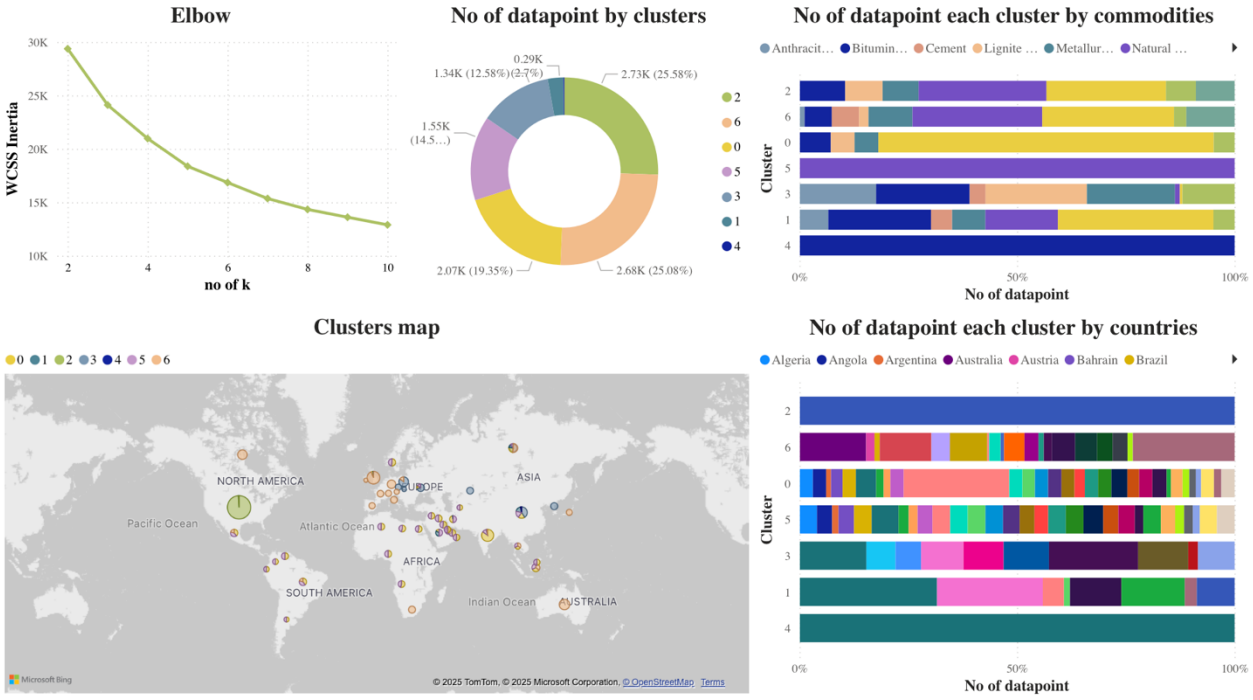

Figure 6. Elbow method and Clustering Analysis Result Visualizations

The optimal number of clusters was determined using the elbow method, an analytical approach that examines the within-cluster sum of squares (WCSS)-a measure of compactness of data points within each cluster. By evaluating WCSS across a range of clusters (from 2 to 10), the optimal

cluster number (k=7) was identified as the point where further increases in clusters yielded minimal reductions in WCSS, thus achieving a balance between cluster compactness and interpretability.

Distinct cluster characteristics emerged clearly from the analysis:

- Cluster 4: Exclusively comprised China's bituminous coal data, highlighting China's distinct coal-centric emission profile. This cluster underscores the specific challenge China faces in reducing coal dependency, critical for meeting global climate targets. This aligns with earlier findings in the first question, which identified China Coal as the leading global emitter, responsible for approximately 26% of total emissions.
- Cluster 5: Exclusively contained natural gas-related emissions, reflecting commodity-specific emission profiles. This specialization suggests a targeted potential for reducing emissions by shifting to lower-carbon fuels within this commodity group.
- Cluster 2: Dominated by US emissions, clearly highlighting national-specific emission attributes influenced by the US's free-market structure and heavy reliance on oil and natural gas.

Other clusters exhibited more mixed commodity and geographic characteristics, reflecting emission complexities driven by diverse factors such as market structures, geographic constraints, policy environments, and technological adoption rates. These findings illustrate the multifaceted nature of global emission dynamics and highlight the need for customized policy responses tailored to cluster-specific attributes. For instance, the dominance of bituminous coal in Cluster 4 (China) suggests that policies focused on promoting renewable energy adoption and phasing out coal-fired power plants, coupled with carbon capture technologies, would be particularly relevant for this group, differing significantly from strategies aimed at reducing natural gas emissions prevalent in Cluster 5. Cluster 2 (US-dominated) might require a greater emphasis on transportation electrification and promoting energy efficiency across various sectors, reflecting its different emission drivers.

Conclusion

The comprehensive analysis presented in this study provided detailed insights into global greenhouse gas emission trends, emphasizing key regional, national, and corporate drivers. By integrating data from the Carbon Majors database, standardized using tonnes of oil equivalent (TOE), the analysis facilitated meaningful comparisons across diverse fuel types and geographic contexts. Supplementary datasets detailing country-specific commitments and compliance with major international climate agreements (Kyoto Protocol and Paris Agreement) enriched the contextual understanding, allowing evaluation of policy effectiveness and compliance patterns.

Despite international efforts, total emissions from major emitting entities have consistently risen between 2000 and 2022, primarily driven by the persistent reliance on fossil fuels - particularly bituminous coal in China, oil globally, and increasingly natural gas. Significantly, the impact of global policy milestones such as the Kyoto Protocol and Paris Agreement varied substantially among different nations. China's substantial emissions growth, fueled predominantly by state-controlled enterprises heavily reliant on coal, notably demonstrated minimal practical responsiveness to these international commitments. Similarly, the United States displayed relatively stable emissions, reflecting the limited effectiveness of voluntary agreements within a market-driven economy. In sharp contrast, European nations, particularly the UK, demonstrated notable alignment and compliance, achieving substantial emission reductions attributable to stringent regulatory frameworks, corporate responsibility initiatives, and supranational directives like the European Green Deal.

Further deepening the analysis, K-Means clustering - an unsupervised machine learning algorithm - was applied to identify structural emission patterns that traditional statistical analyses might overlook. This algorithm partitions data points into distinct clusters by iteratively minimizing the within-cluster sum of squares (WCSS), based on Euclidean distance to cluster centroids. The clustering process utilized both quantitative and categorical variables, specifically:

- Quantitative: total emissions (total_emissions_MtCO2e), scaled using StandardScaler to ensure each variable equally influenced clustering results.
- Categorical: parent type, commodity type, and country, transformed via OneHotEncoder into numerical representations suitable for clustering analysis.

Employing the elbow method to determine the optimal number of clusters (k=7), distinct emission clusters emerged clearly:

- Cluster 4 was exclusively composed of China's bituminous coal emissions, underscoring China's unique coal-centric emission profile and highlighting the challenge of coal dependency for meeting global climate targets. This result aligns with earlier findings that identified China Coal as the world's leading single emitter, emphasizing the urgent need for targeted renewable energy policies and carbon capture technology adoption.
- Cluster 5 exclusively represented natural gas emissions, reflecting commodity-specific emission behaviors. This grouping provides clear opportunities to implement targeted strategies aimed at optimizing operational efficiency and transitioning to lower-carbon or renewable alternatives within natural gas-dependent industries.
- Cluster 2, predominantly featuring emissions from the United States, revealed distinct national-specific emission characteristics shaped by a free-market structure, energy-intensive lifestyles, and heavy reliance on oil and natural gas. Policy implications here

involve significant opportunities in transportation electrification, renewable energy expansion, and aggressive energy efficiency standards.

Other identified clusters were more heterogeneous, incorporating diverse commodities and geographies, indicative of complex emission dynamics influenced by varying market structures, regulatory environments, geographic constraints, and technology adoption rates. This complexity illustrates the necessity of tailored climate policy responses that effectively address the unique emission drivers within each cluster.

Overall, this integrative analytical approach combining detailed exploratory data analysis, geographical visualization, compliance assessment, and advanced machine learning clustering provided nuanced insights essential for developing informed, effective, and context-specific global climate policies. Future research should continue exploring more granular compliance patterns at both corporate and national scales, leveraging sophisticated analytical techniques to further refine strategies aimed at mitigating global emissions and achieving international climate objectives.