Report on data mining and statistical analysis

Introduction

The first activity of the ALMONDO project involved collecting data on climate lobbying and the climate-related public debate, with two main objectives in mind. The first objective was to obtain a better understanding of the lobbying process and the formation of opinions on climate issues. The second aim was to extract knowledge that can inform the opinion dynamics model to be developed in subsequent project activities.

To achieve these objectives, we focused on two types of data, representing two different actor types involved in the climate debate. The first type is climate lobbying data, which describes the lobbying activities of companies around the globe. This includes dedicated meetings with policymakers or other organizations and activities on social media platforms, specifically Twitter

(formerly known as X). The second type of data focuses on individuals discussing climate issues in public social media forums, with a particular emphasis on Reddit.

Our data collection process drew from four primary sources: the EU Transparency Register, Lobby Map, EU public consultations, and LobbyFacts. Additionally, we incorporated data from X (former Twitter) and Reddit to capture user activity and interactions related to climate lobbying. By analyzing these diverse data sources, we aim to provide a comprehensive overview of the climate lobbying landscape and the dynamics of public opinion on climate issues.

This report details the data collection process and outlines the main characteristics of the resulting datasets. Through this analysis, we aim to shed light on the intricate processes of climate lobbying and public opinion formation, providing valuable insights to support the development of an informed opinion dynamics model in future project phases.

Climate lobbying data sources

The data for this report is drawn from a combination of four primary sources: the EU Transparency Register, Lobby Map, EU public consultations, and LobbyFacts. These sources provide public data on the activities of various companies engaged in climate lobbying. To capture a complete view of these activities and the actors involved, we supplemented this information with data from X and Reddit, focusing on user activity and interactions related to climate lobbying. By integrating these diverse datasets, we aim to offer a comprehensive overview of the lobbying landscape and its influence on public debate around climate issues.

Data sources

The first source we refer to is the EU Transparency Register. This database collects information on interest representatives who influence EU policies and decision-making processes. This register aims to provide transparency on lobbying and advocacy activities, indicating which organizations, interests and resources are used for these activities.

The register is managed by secretaries of the European Parliament, the Council and the European Commission. The data is updated periodically and represents the updated situation of the

company, it is not possible to have information on the past activities of the organization. The information contained in the register is of a general nature about the company and more specific about its lobbying activity. Among the first we find contact details, Head of relations with the EU, purpose. Among the latter, however, there is the list of policies of interest to the organization, on which the lobbying activity is carried out, number of people involved in this activity, sectors of interest and budget used for them.

The second source we consider is the LobbyMap. Lobby Map collects and publishes data through the approach to monitor, evaluate and score companies and industry groups based on their involvement in climate policy. The objective is to provide insight into the company's commitment to climate-related issues. It analyzes almost 750 of the largest companies and industrial associations in the world following the principles of objectivity, transparency and ease of understanding. Through an in-depth analysis of available data, it assigns scores and ranks companies based on the degree and direction of climate lobbying carried out.

The information contained in Lobby Map is mainly textual in nature. The policies most and least supported by organizations are reviewed, specifically providing an overview of positions taken on regulations related to climate and energy transition. Additionally, corporate participation in industry associations is highlighted. Finally, the dataset provides summary indicators on the use and direction (support / non-support of policies) of companies' climate lobbying activity scaling it to numbers and letters.

Our third source of lobbying information concerns the EU public consultations. The website provides a database of initiatives that are subject to public consultation. By taking part, citizens can contribute to the EU legislative process by providing feedback and consulting related documents. By entering the name of an organization on the site it is possible, among other things, to obtain information about it thanks to a link with the transparency register, read any reports present or measures taken by the European Commission against an organization. It is possible to keep track of these measures over time, if any.

The fourth source used is the LobbyFacts - a project in collaboration with Corporate Europe Observatory and LobbyControl, that provides essential data on lobbying in European institutions. It offers information on the lobbying activity of various entities, including academic institutions,

associations, companies, non-governmental organizations, religious and public organizations, and commercial associations. It allows you to monitor lobbying costs over different years. The present time frame is from 2015 to 2024 although there is information for a very low number of organizations since 2011. For each organization, general information is provided, such as goal, address, website and more specific information on lobbying activity such as meetings attended, annual budget used.

Finally, we have used X (formerly Twitter) to extract data about the social media activities of companies involved in climate lobbying. X is a social media platform that enables companies and individuals to share content publicly with friends and followers through posts. An API is available to download posts for different users upon payment of a monthly fee.

Data on lobbyists

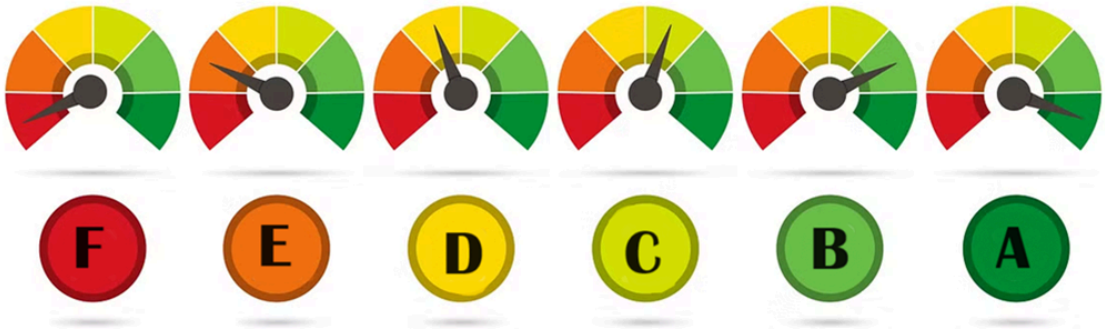
To create a dataset containing all the organizations that have carried out lobbying activities in Europe, data from different sources were merged. The organizations listed in Lobby Map operating in Europe were taken into consideration, focusing on companies involved in climate lobbying activities. Information regarding lobbying spending was subsequently integrated by conducting a search by name on the EU Transparency Register.

Using the information from various data sources, a new dataset was created containing a total of 201 observations, including 146 companies and 55 industrial associations. The dataset includes information such as the name of the organization, the type (company or industrial association), the country where the organization's headquarters are registered, the countries in which it operates, sector of activity, and X ID (used to collect posts through the X API). Furthermore, the dataset contains engagement indicators on climate lobbying, pointing the direction of lobbying activity (in support or against EU policies), lobbying budget, and policies on which the organization has lobbied to identify the companies that invested more in climate matters. All information is updated as of March 2024. Furthermore, to identify the type of lobbying activity carried out by the various organizations, the textual data provided by LobbyMap were analyzed; a qualitative analysis of the data was carried out using the MAXQDA software, which allows identifying codes through keywords and classifying the textual information. Firstly, the type of positive or negative lobbying was identified (in support or against community policies),

subsequently the policies linked to the climate most affected by the lobbying activity were identified. Finally, dummies were added for entities carrying out lobbying linked to circular economy issues and for those carrying out negative climate lobbying (against EU policies), also indicating the type. It turns out that many companies carrying out negative lobbying oppose stringent EU policies, for example promoting the use of fossil gasses alongside renewable energy.

Within the dataset, there is an influence score, which is a measure linked to conditional lobbying performance assigned by LobbyMap. Each company is assigned one of the following scores: A+, A, A-, B+, B, B-, C+, C, C-, D+, D, D-, E+, E, E-. The letter A indicates that the company carries out lobbying activities mainly in support of community climate policies, while the letters D, E, and F indicate that the company opposes them.

Figure 1: Influence scores present in LobbyMap.



Source: own elaboration based on creativemarket.com.

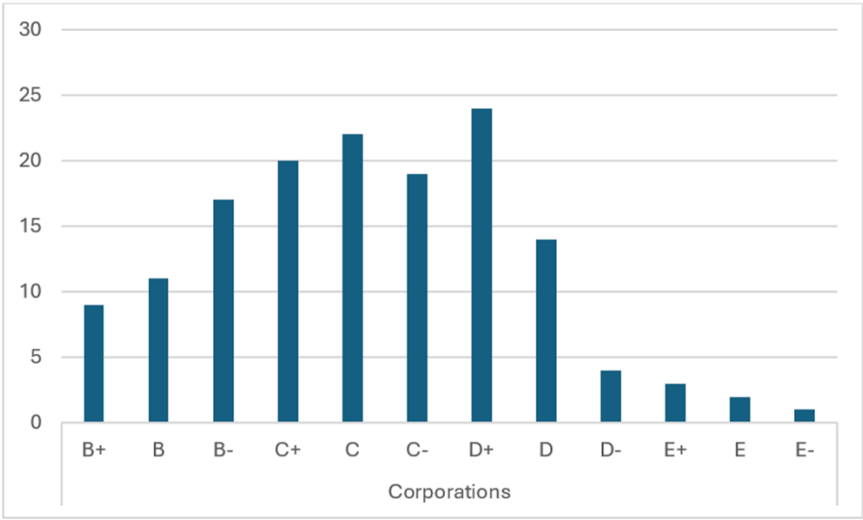
Descriptive Statistics

To provide more information about the composition of the dataset, some descriptive statistics are provided. First we note that there are no corporations with influence score A, although climate lobbying activity is in favor of community policies for the most virtuous. It is observed that many companies take central and not extreme positions (see Figure 2).

While considering the industrial associations, the situation changes, some of them are very supportive of community policies on climate matters, but many are opposed to them (Figure 3).

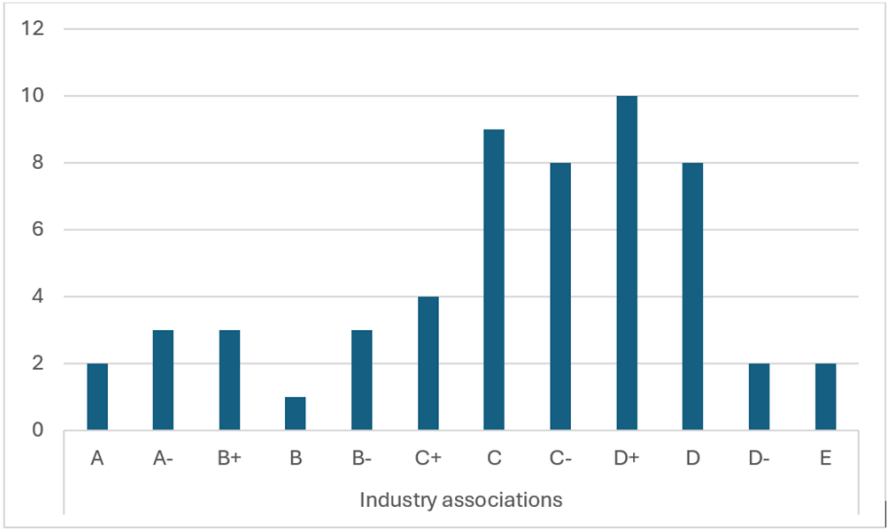
In both cases the score received by multiple organizations is D+ (the organization carries out mainly negative climate lobbying). In fact, within the dataset it was observed that many organizations support the majority of climate policies, however they do not show support for policies that impose stringent constraints on their sector.

Figure 2: Number of corporations for each influence score



Source: own elaboration based on Lobby Map.

Figure 3: Number of industry associations for each influence score



Source: own elaboration based on Lobby Map.

The tables presented below illustrate how lobbying budgets vary based on influence scores, with notable differences in investments depending on the type of organization. Industry associations, on average, allocate larger budgets for lobbying, totaling 1,322,339 compared to the average of 993,305.5 for corporations.

Table 1 displays the lobbying budgets of corporations. It is evident that, on average, corporations with an influence score of C invest more in climate lobbying. Similarly, Table 2, which details the budgets of industry associations, shows that those with an influence score of C also allocate higher average budgets for lobbying. However, significant differences can be observed between the two types of organizations. For corporations, the lobbying budgets for those supporting and opposing community policies are relatively similar. In contrast, for industry associations, there is a marked difference: those engaged primarily in negative climate lobbying, particularly in opposition to community policies, have a substantially higher budget, reflected in an influence score of D. It is important to note that there are some missing values due to the absence of certain companies in the EU transparency register.

Table 1: Lobbying budget of corporations based on their influence score.

	Corporations				
	Obs	*Mean*	*Std. dev.*	*Min*	*Max*
B	*32*	*685155.3*	*596726.9*	*24999*	*2249999*
C	*56*	*1375892*	*1422737*	*99999*	*7999999*
D	*36*	*672082.4*	*802625.4*	*10000*	*3499999*

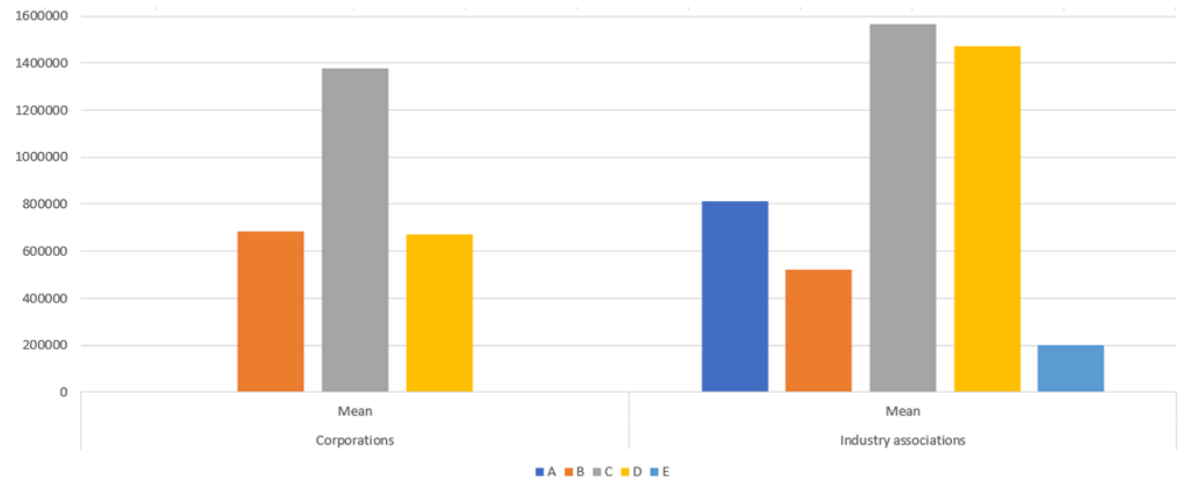
Source: own elaboration based on EU Transparency Register and LobbyMap.

Table 2: Lobbying budget of industry associations based on their influence score.

	Industry associations				
	Obs	*Mean*	*Std. dev.*	*Min*	*Max*
A	*4*	*812499*	*606045.9*	*49999*	*1499999*
B	*5*	*519999*	*216794.8*	*199999*	*799999*
C	*18*	*1563888*	*2416278*	*24999*	*10000000*
D	*19*	*1471052*	*1653558*	*99999*	*4999999*
E	*1*	*199999*	*.*	*199999*	*199999*

Source: own elaboration based on EU Transparency Register and LobbyMap.

Figure 4: Lobbying budget of corporation and industry associations based on their influence score



Source: own elaboration based on EU Transparency Register and LobbyMap.

Expanding the data on lobbyists

To broaden the scope of organizations considered, the LobbyFacts dataset was utilized. This dataset exclusively includes organizations based in Europe and provides extensive information on meetings and lobbying budgets, encompassing both numerical and textual data. The dataset spans the period from 2015 to 2024. The number of registered organizations varies annually: from 2015 to 2021, it includes approximately 1,000 entities per year, increasing to around 12,000 organizations in 2022 and 2023. For 2024, the dataset only covers the months from January to March, with approximately 5,000 organizations represented.

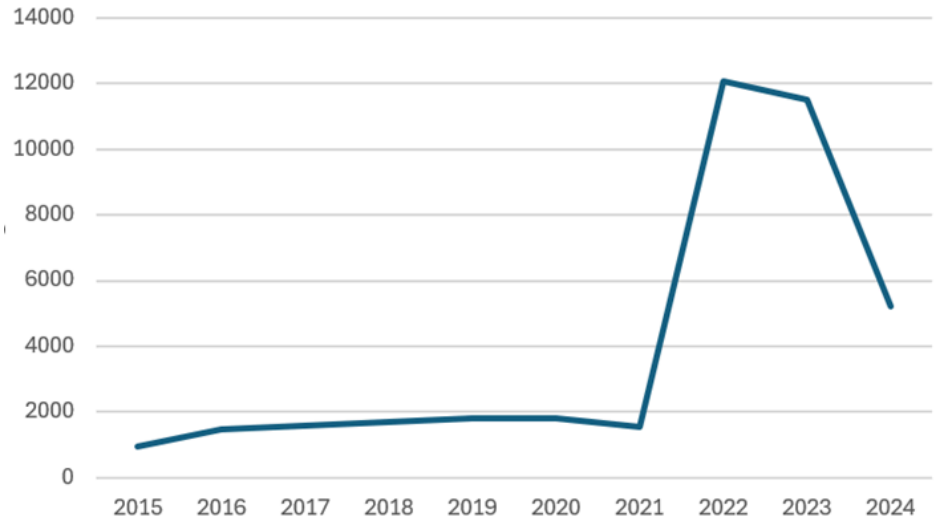
The relevant information for the analysis was selected and processed. Among the selected data are the average annual lobbying budget of the various organizations and the number of meetings with the European Commission in which each organization participated. By analyzing the textual information in the dataset, a dummy variable was created that takes a value of 1 if the organization participated in at least one meeting related to climate issues during the year. This variable serves as a proxy for whether the organization engages in climate lobbying activities. The dataset also includes information such as the type of organization and its name. Additionally, a dummy variable identifies companies that are present in both the LobbyFacts and LobbyMap

datasets. The available data enable the identification of organizations that participate in the same meetings, facilitating the potential for future network analysis.

Descriptive Statistics of LobbyFacts

In this paragraph, we will present some descriptive statistics of the dataset to facilitate understanding of the information it contains. Figure (5) illustrates that in the years 2022 and 2023, many organizations were registered in the dataset. Considering only the first months of 2024, the number of entities present exceeds that of the years prior to 2022.

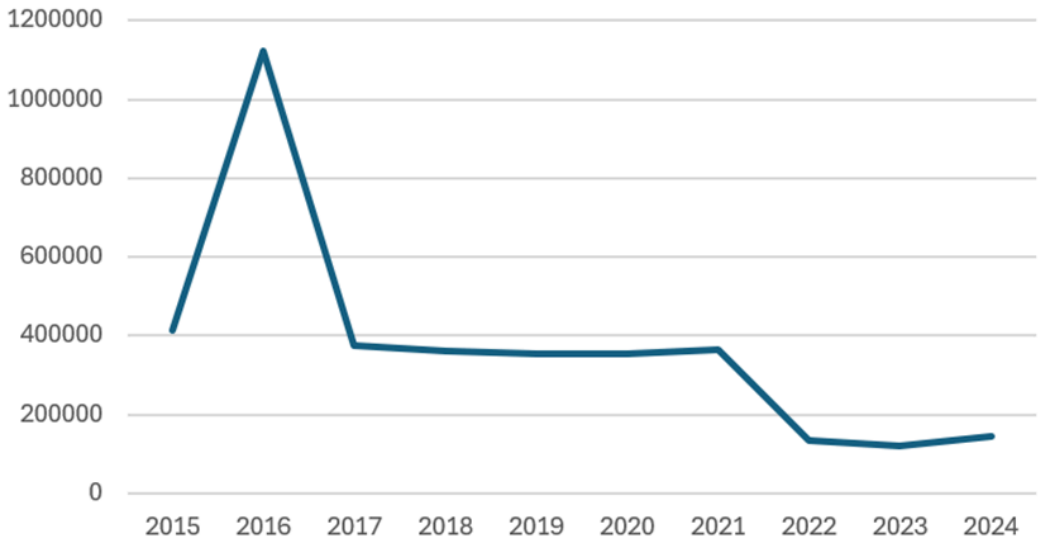
Figure 5: Number of organizations in the different years.



Source: own elaboration based on LobbyFacts

Observing the average budget for lobbying in different years (Figure 6), it is notable that it reaches a peak in 2016, remaining relatively stable from 2017 to 2021 and continuing similarly from 2022 to 2024.

Figure 6: Lobbying budget (mean)



Source: own elaboration based on LobbyFacts

Table (3) contains the organizations present within the dataset, each identified with a number to make the graphs in figures (7) and (8) more readable.

Table 3: Types of organizations

ID	Organization
2	Companies & groups
4	I - Professional consultancies/law firms/self-employed consultants
5	Professional consultancies
6	III - Non-governmental organizations
7	Non-governmental organizations, platforms and networks and similar

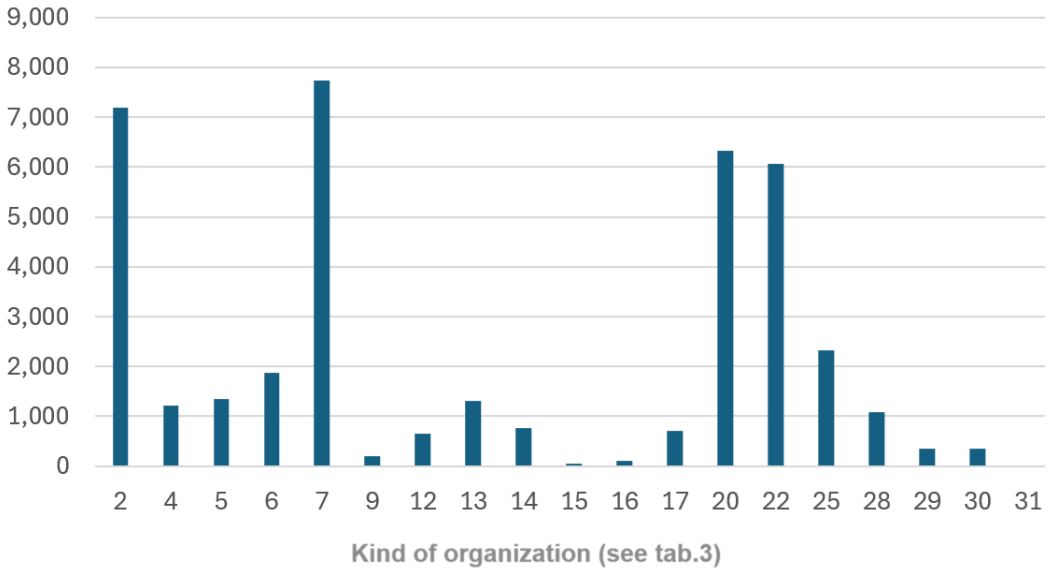
9	Law firms
12	IV - Think tanks, research and academic institutions
13	Think tanks and research institutions
14	Academic institutions
15	V - Organisations representing churches and religious communities
16	Organizations representing churches and religious communities
17	VI - Organisations representing local, regional and municipal authorities, other public or mixed entities, etc.
20	II - In-house lobbyists and trade/business/professional associations
22	Trade and business associations
25	Trade unions and professional associations
28	Other organizations, public or mixed entities
29	Associations and networks of public authorities
30	Self-employed individuals
31	Entities, offices or networks established by third countries

Source: own elaboration based on LobbyFacts

From Figure (7), it is evident that the most prevalent organizations in the dataset are Companies & groups, Non-governmental organizations, platforms, and networks, and similar entities. In-house lobbyists, trade/business/professional associations, and trade and business associations also exhibit significant presence, with over 6000 registered entities

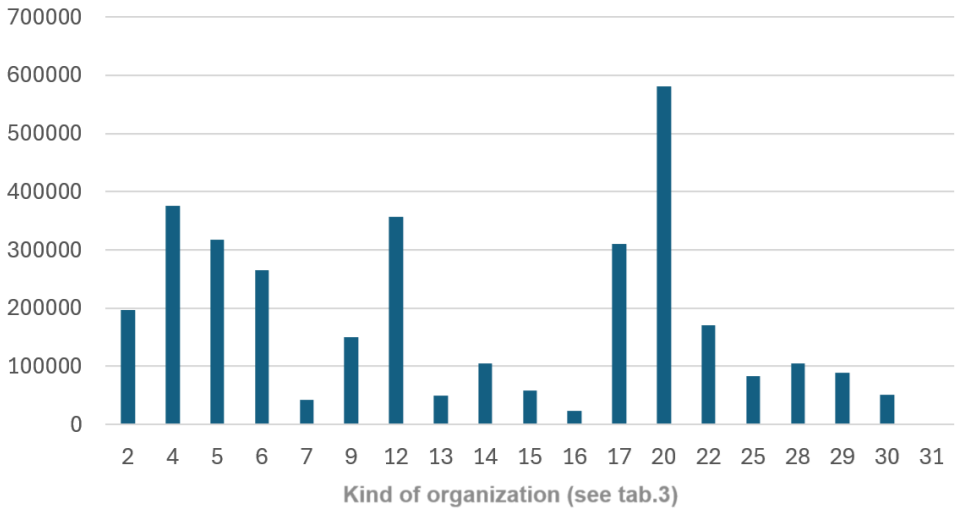
Figure 8 illustrates that the organizations making the highest investments in lobbying are in-house lobbyists and trade/business/professional associations, with an average investment of nearly 600,000 euros. Additionally, other organizations with substantial budgets include professional consultancies/law firms/self-employed consultants, professional consultancies, research and academic institutions, and organizations representing local, regional, and municipal authorities, along with other public or mixed entities, exceeding the budget threshold of 300,000 euros.

Figure 7: Number of organizations in the different categories.



Source: own elaboration based on LobbyFacts

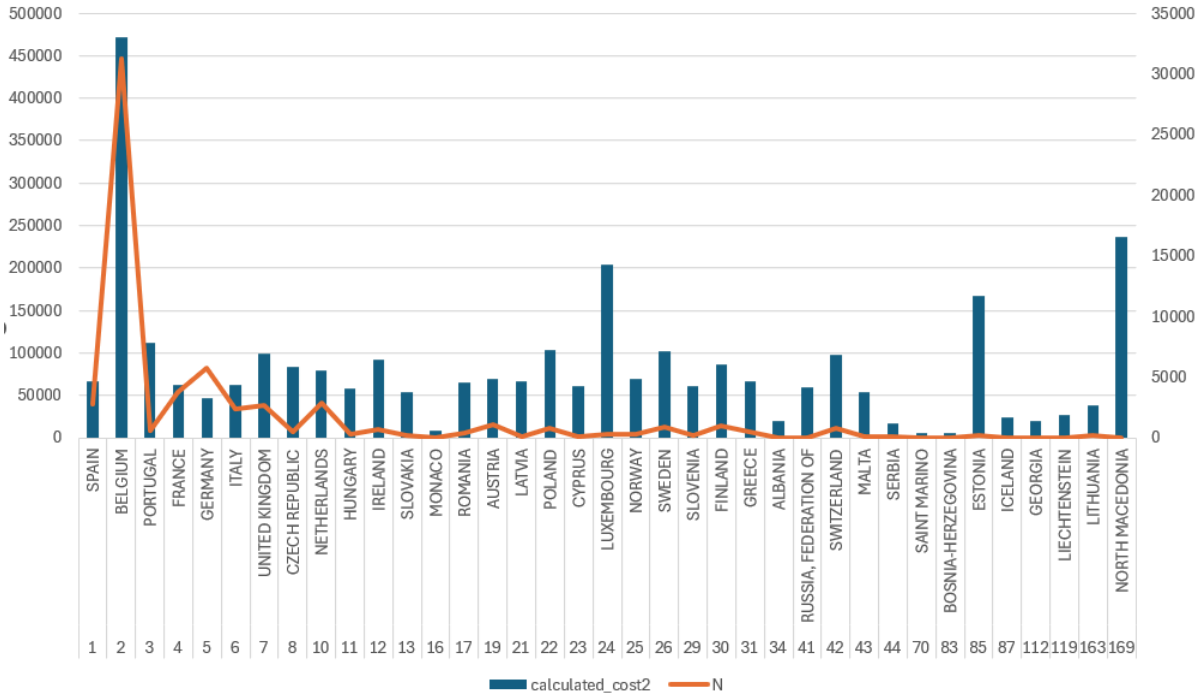
Figure 8: Lobbying budget (mean) for the different categories



Source: own elaboration based on LobbyFacts

Figure (9) depicts the cost per lobbying (calculated_cost2) and the number of organizations (N) active in lobbying and present in each country.

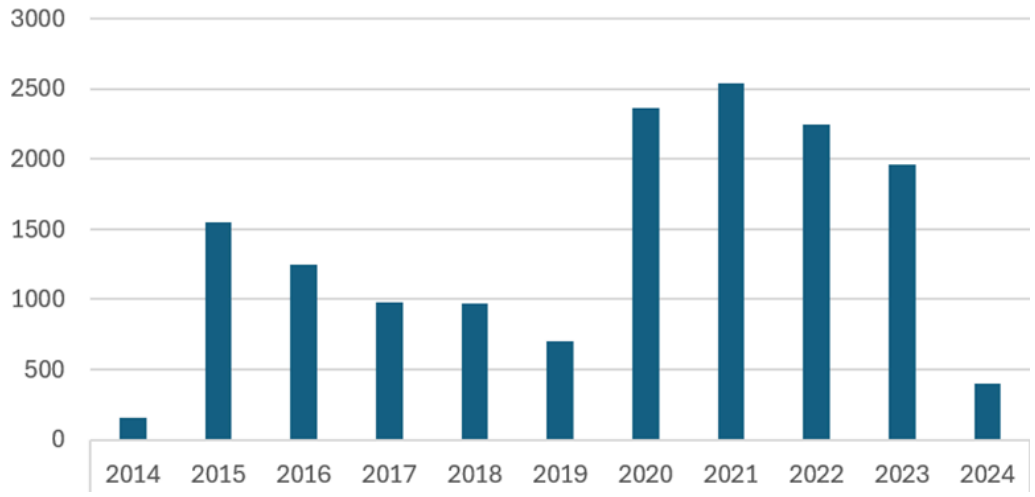
Figure 9: Numbers of meeting about climate



Source: own elaboration based on LobbyFacts

A subset of meetings within the dataset pertains to climate-related issues, indicating a discernible trend of increased frequency since 2020. These meetings encompass topics such as energy, emissions, sustainability, and green initiatives, identified through textual analysis. Initially, the number of such meetings was relatively low, gradually rising to over 1,500 in 2015. Notably, this escalation coincides with the signing of the Paris Agreement. Subsequently, the frequency of climate-related meetings surged, surpassing 20,500 in 2021, a milestone year marked by the signing of the European Climate Law (refer to Figure 10).

Figure 10: Numbers of meeting about climate



Source: own elaboration based on LobbyFacts

Social media data

The dataset presented so far contains data from official lobbying activities. To complete the information present there, we have started to download the X posts of the 201 organizations included. This dataset will serve as a foundational resource for analyzing the communication patterns and engagement strategies of lobbyists on social media.

We started from the X usernames of the 201 organizations, and downloaded all their posts from January 1, 2023, to December 31, 2023. This timeframe selection aimed to balance the acquisition of a substantial dataset while mitigating X's rate limits. In the pursuit of original content, only posts originating directly from the accounts were included, excluding reposts. Notably, there were no language restrictions imposed during the collection process, acknowledging the multilingual nature of corporate communication on the platform.

The data collection procedure is still ongoing, due to limitations imposed by the X API. At the moment of publication of the report we have downloaded data for 40 lobbyists, for a total of 10,562 posts. We will continue collecting data until all 201 accounts will be processed.

Data Retrieval Details

For each post, we collected the following attributes:

- *Post ID:* Unique identifier for the post
- *Author ID:* Unique identifier for the author
- *Created At:* Timestamp of when the post was published
- *Edit History Post IDs:* List of post IDs representing the edit history
- *Attachments Info:* Information about media attachments
- *Entities Info:* Details on entities (e.g., users mentioned, hashtags, cashtags)
- *Annotations:* Annotations added to the post

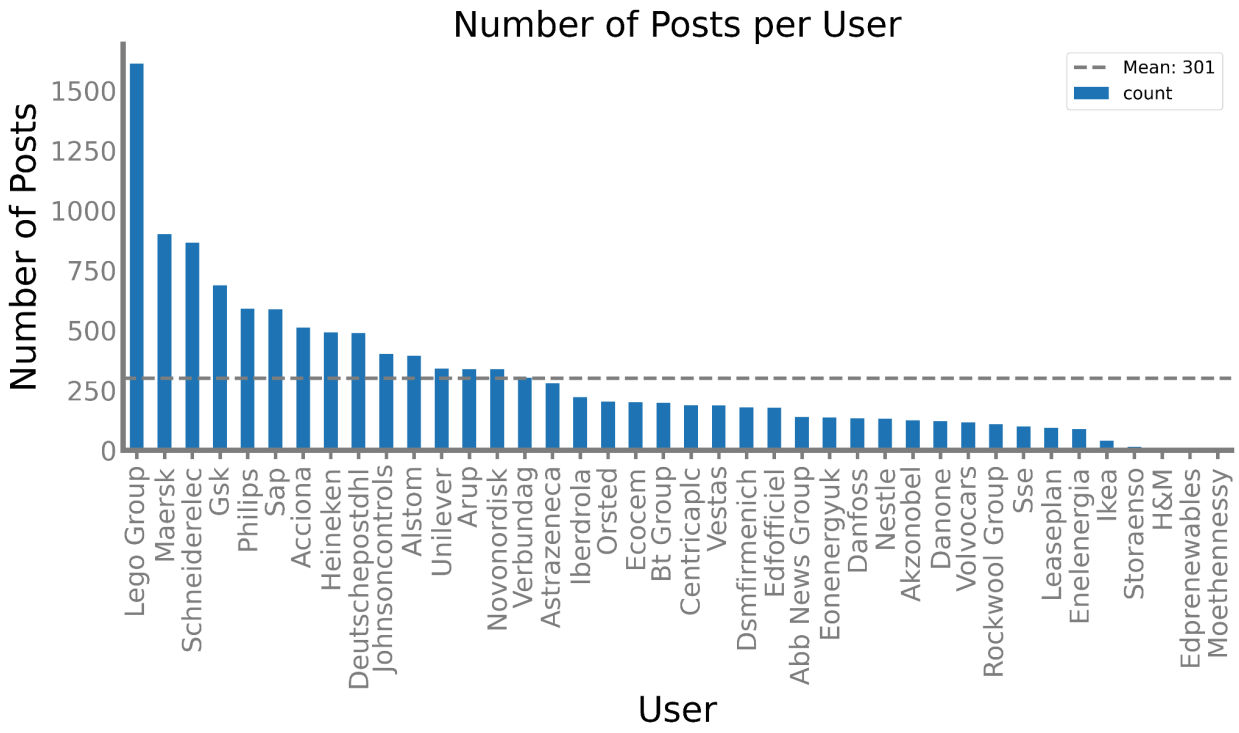
- *Hashtags:* List of hashtags included in the post
- *URLs:* URLs included in the post
- *Language:* Language of the post
- *Public Metrics:* Engagement metrics including:
    - *Bookmark Count*
    - *Impression Count*
    - *Like Count*
    - *Quote Count*
    - *Reply Count*
    - *Repost Count*
- *Text:* The full text of the Post

Dataset characteristics

In the rest of this section we will present the main characteristics of the X data collected.

We first study the activity volume of lobbyists in terms of the number of posts. Figure 11 shows the total number of posts per lobbyist, for the year 2023. Notably, the most active accounts are LEGO Group with 1,613 posts, Maersk with 902 posts, and Schneider Electric with 866 posts. Conversely, the least active accounts are H&M with 7 posts, Moët Hennessy with 1 post, and EDP Renewables with 1 post.

Figure 11: Numbers of Posts per Lobbyist



Figures 12 and 13, instead, look at the weekly and monthly distribution of the number of posts, divided by the various accounts. In general, we see an apparent increase in activity in the second part of the year. We observe local peaks of activity in the months March, June and October.

Figure 12: Weekly distribution of posts per account (count)

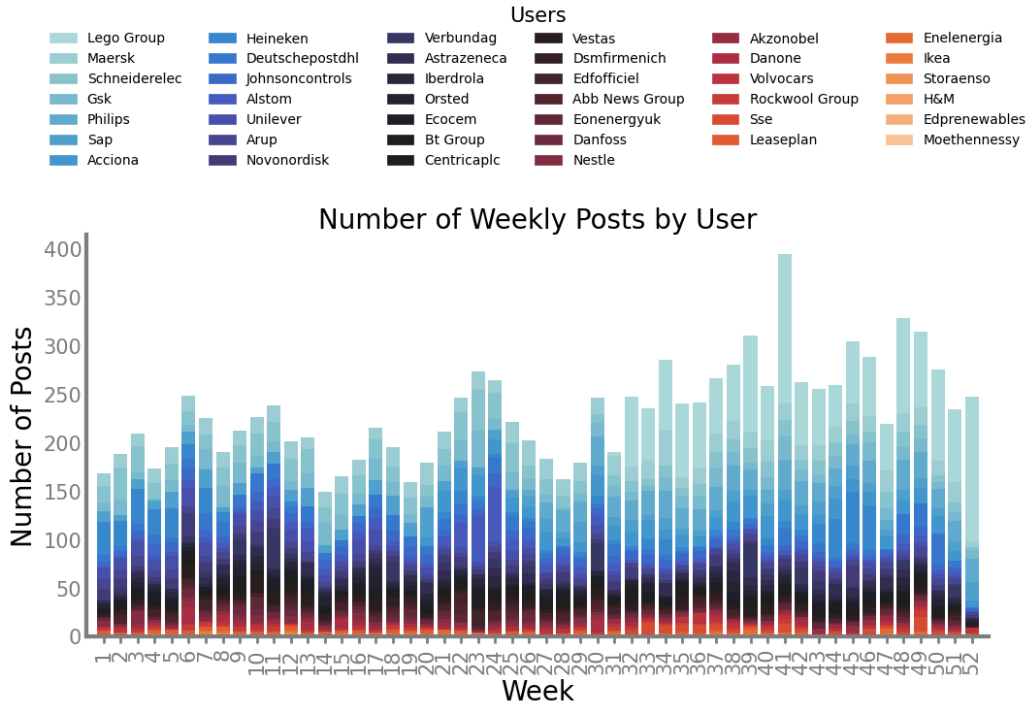
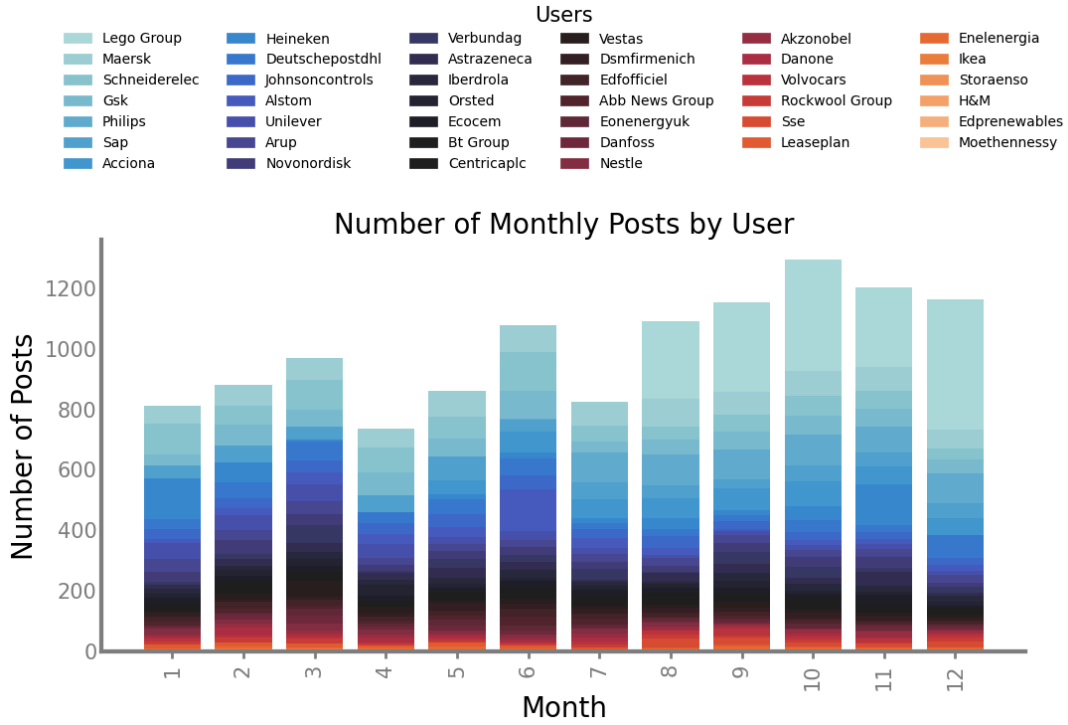
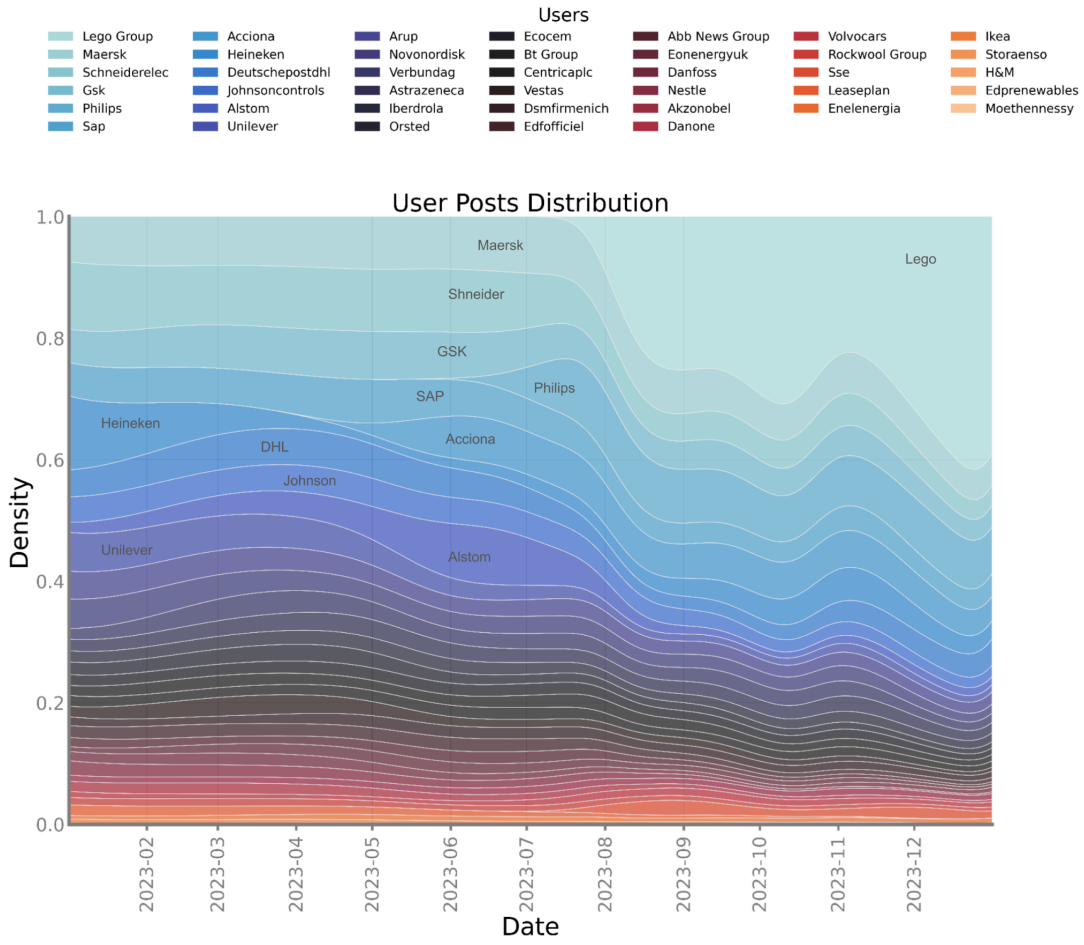


Figure 13: Monthly distribution of posts per account (count).



For a better understanding of the patterns in time, Figure 14 displays the distribution of posts per lobbyist, in time. We note that most companies tend to maintain a stable amount of posts in time, however, there are exceptions. For a few companies, the data appears to be missing before a certain month, such as Lego, Philips, Acciona. This is due to limitations of X data: only the last 3200 posts are available for each account. Clearly these companies exceed this level so data for the first months of 2023 are not available any more. Other non uniform patterns, not due to data limitations, exist. For instance Alstom appears to have an increased activity during the summer, while Heineken seems to have a larger activity during the first three months of the year.

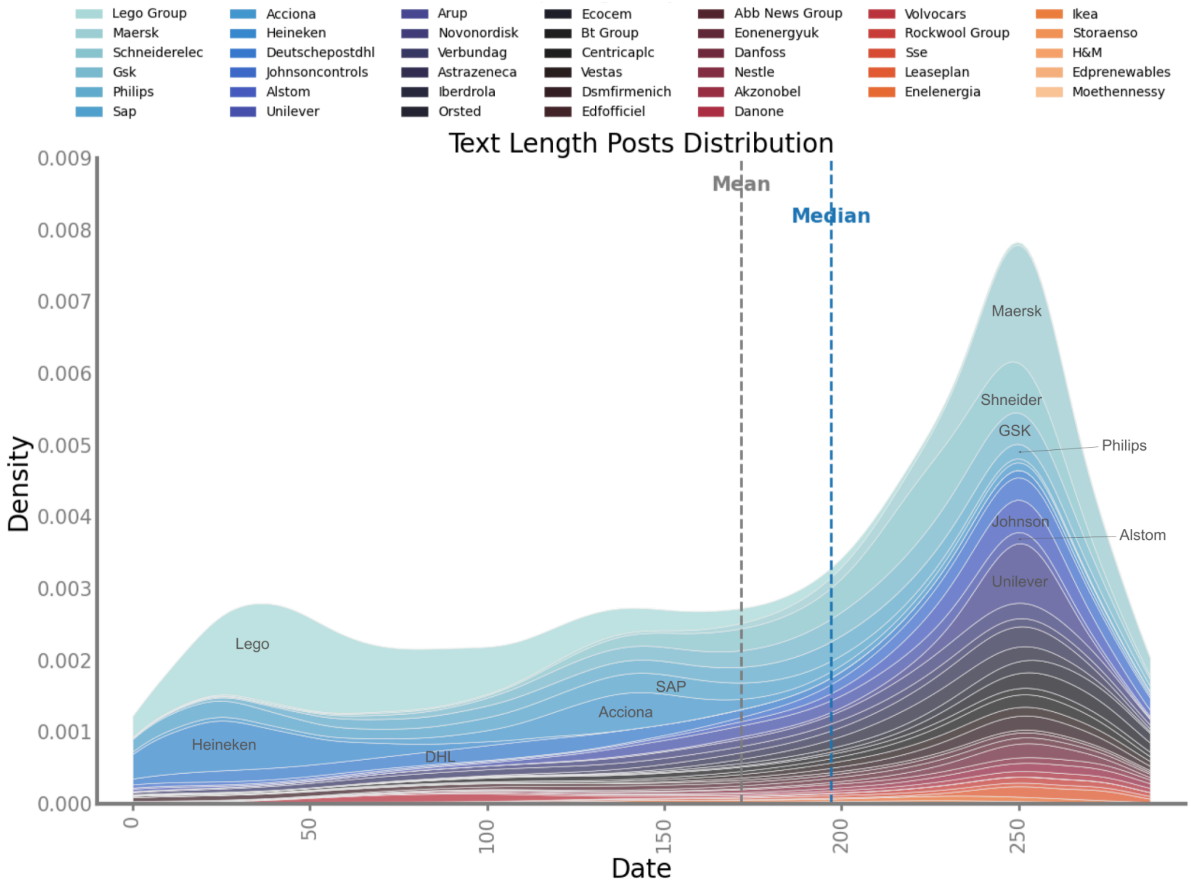
Figure 14: Numbers of Posts per Lobbyist



A second aspect relevant for measuring the activity of lobbyists is the post length. While shorter posts are faster to produce, appealing to a younger audience, longer posts are more costly and

may appeal more to a mature audience. Figure 15 shows the length distribution of all posts (in terms of number of characters), for each lobbyist. From top to bottom we can observe the density distribution from the most active to the less active accounts. We generally observe a bimodal distribution, with most companies tending to have most posts between 200 and 300 characters (300 is the maximum allowed), with an overall median around 195. An obvious exception is Lego Group, who tends to have very short posts, and generates the second mode around 45 characters. Deepening our analysis we noticed that shorter posts tend to be replies to customers/other users while longer posts tend to be communications about companies' activities and principles. Lego is by far the most active user in our dataset to engage with customers/users. A similar pattern can be observed for Heineken, again posting short messages with links.

Figure 15: Text length distribution of posts per account



A third important aspect that can be an indicator of the influence of each lobbyist is the language used in their posts. Figure 16 shows the number of posts for each account in each language. We

note that English is by far the most used language, however the total number of languages used are 22. Most companies who use a language different from English also use English, to improve their international reach. However, a few lobbyists use almost exclusively the national language of their country or origin: Verbundag and Deutsche Post DHL in German, EDF Officiel in French, Acciona in Spanish, EnelEnergia in Italian. Figure 17 shows more closely the distribution of languages different from English. Among the 40 lobbyists studied, only 17 use a language different from English. Philips, Heineken and GSK use the most languages, probably also because they are multinationals with offices around the globe.

Figure 16: Languages used by each user

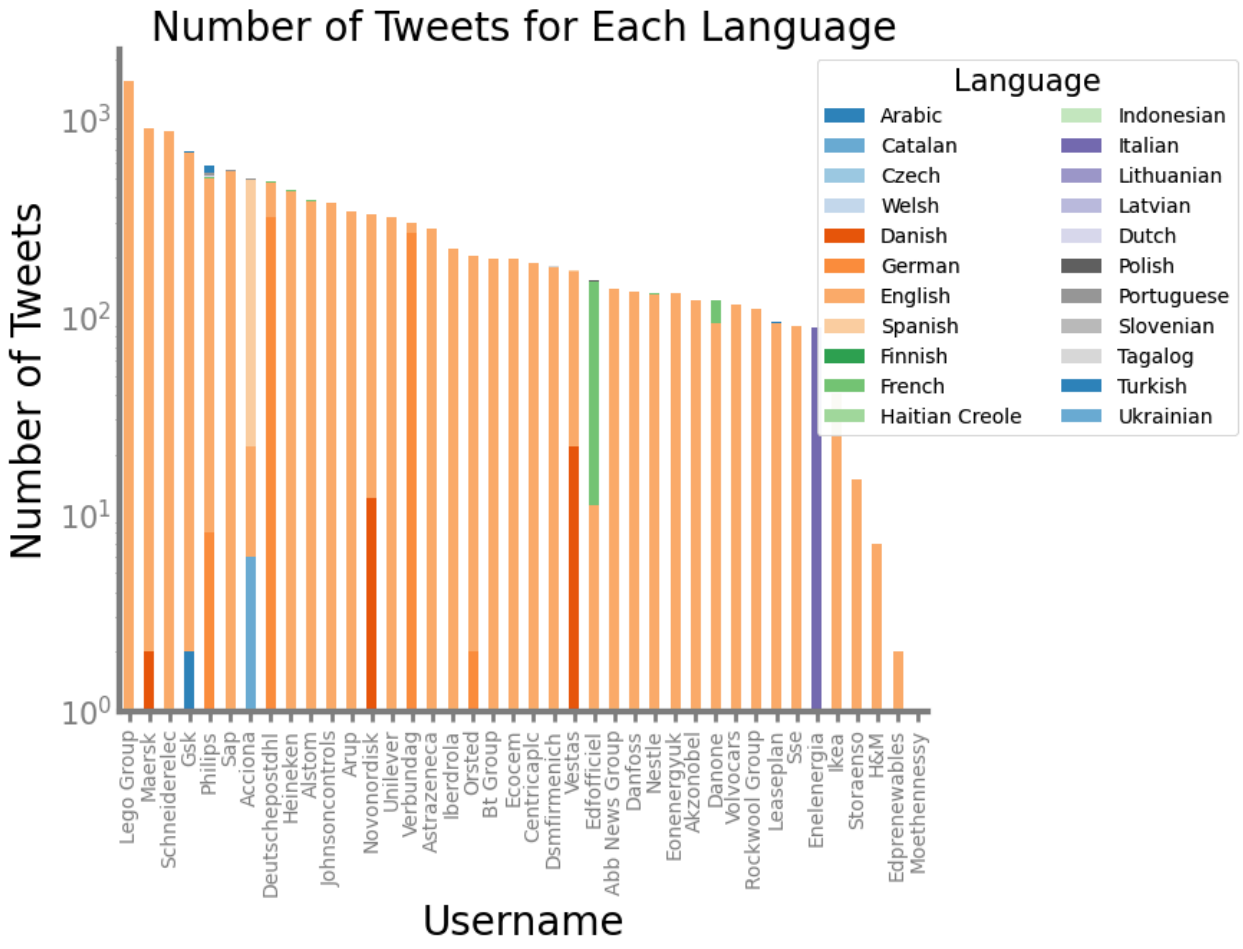
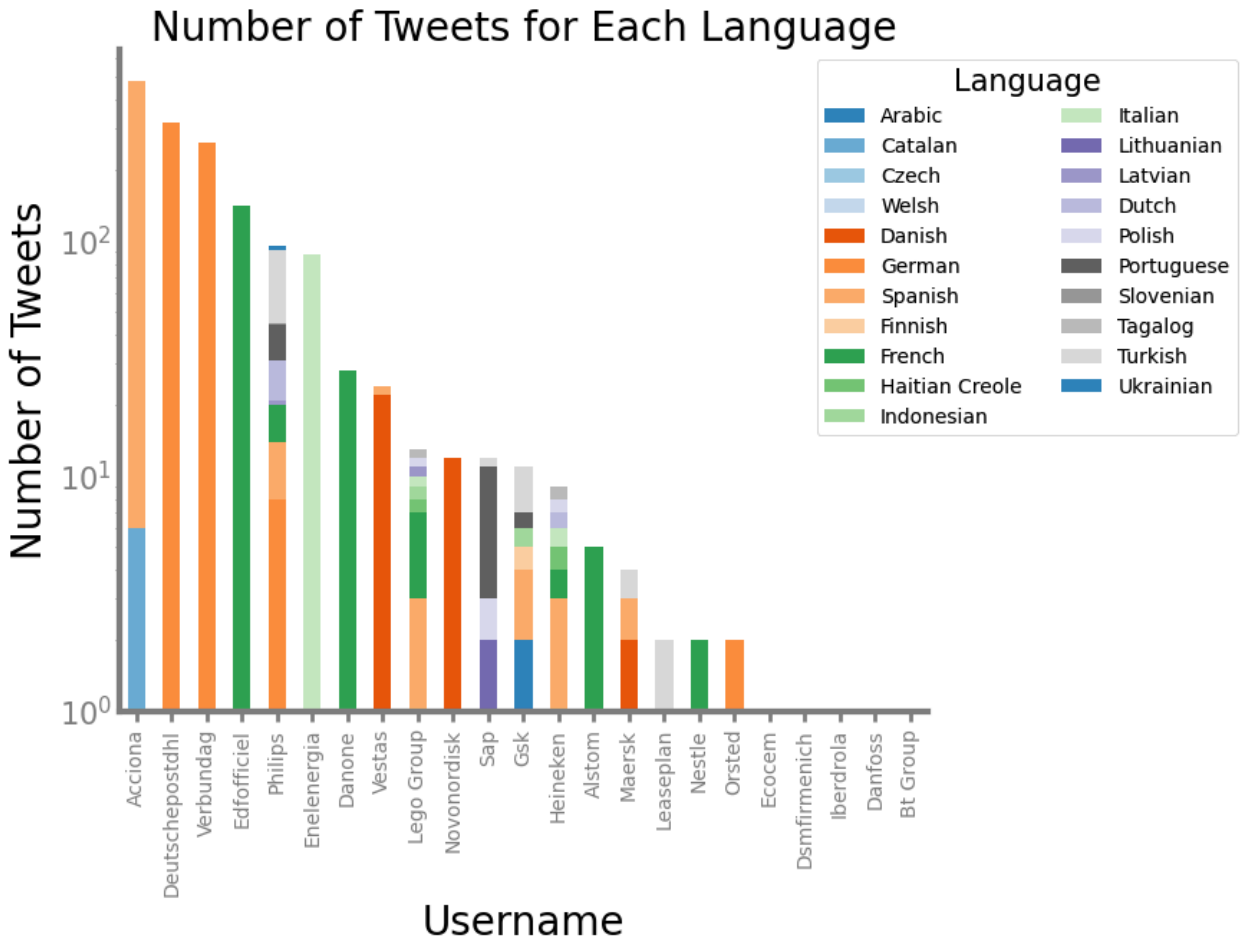
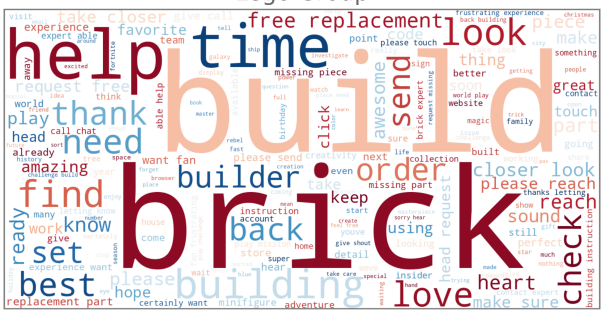


Figure 17: Languages different from English used by each user

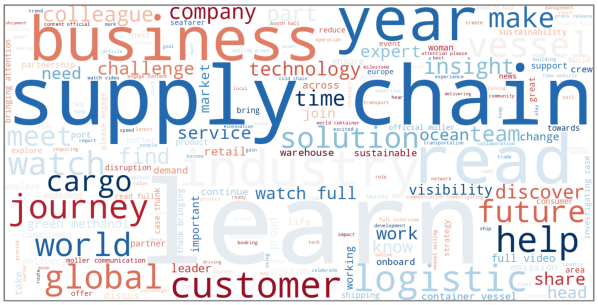


A final analysis of the social media presence of lobbyists concerns the topics covered by their posts. Below we include a series of word clouds, one for each lobbyist, ordered from the most active user overall to the less active one. The word clouds were obtained by processing the text of the posts, after translating all posts into English. We note that companies discussing **climate change**, **renewable energies** or **sustainability** are several in our dataset as we can simply see from word clouds. However, many companies included in climate lobby registers do not appear to discuss climate-related topics; several accounts mainly discuss topics related to the core business of the company, or are more inclined to relate with customers on social media.

Lego Group



Maersk



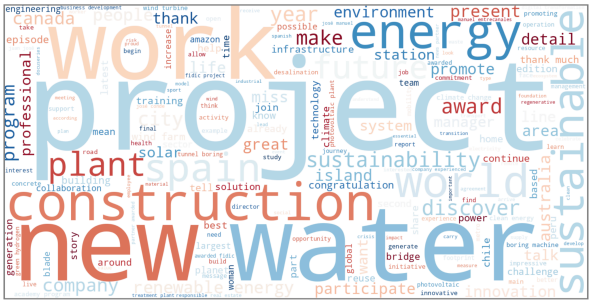
Schneiderelec

Gsk

Philips

Sap

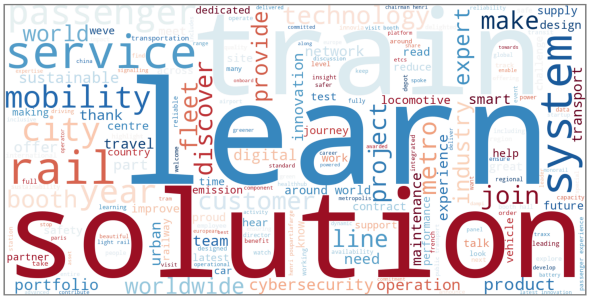
Acciona



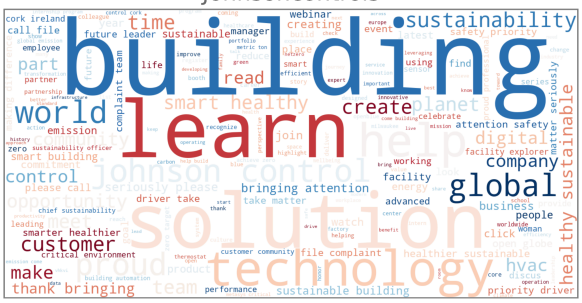
Deutschepostdhl

Heineken

Alstom



Johnsoncontrols

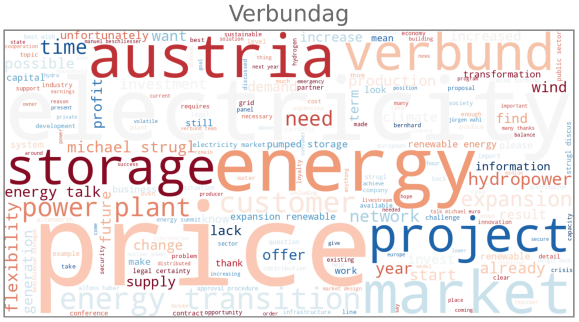


Arup

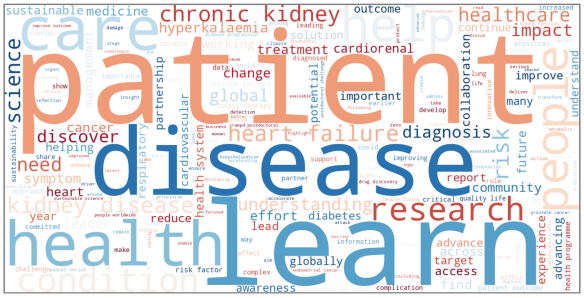
Novonordisk

Unilever

Verbundag



Astrazeneca

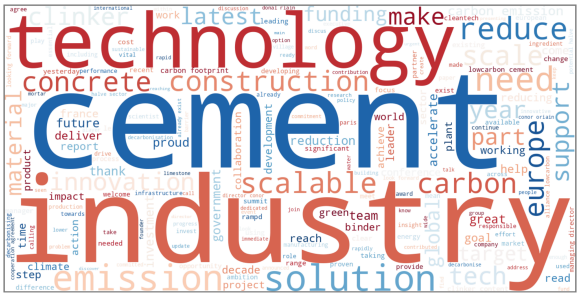


Iberdrola

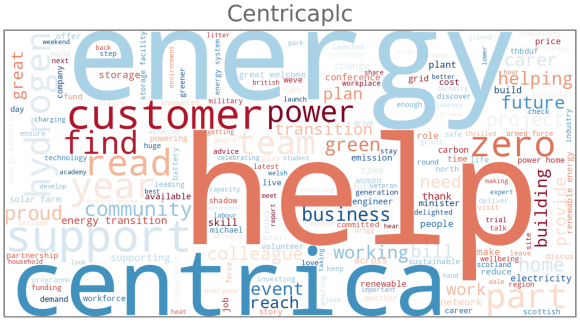
Orsted

Bt Group

Ecocem

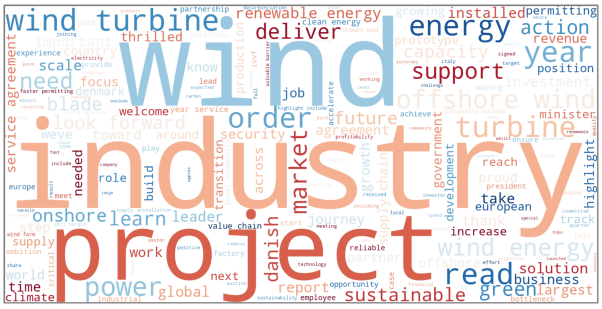


Centricaplc



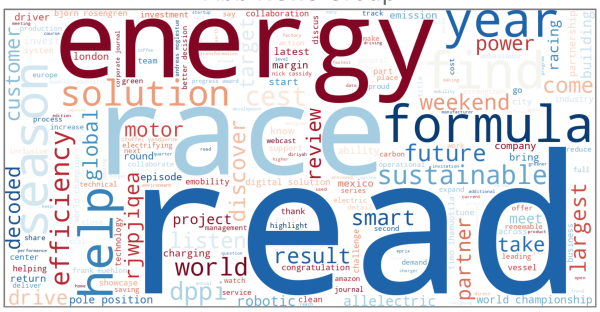
Dsmfirmenich

Vestas



Edfofficiel

Abb News Group



Danfoss

Eonenergyuk

Nestle

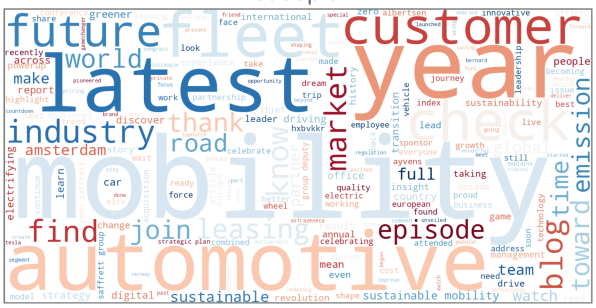
Akzonobel

Danone

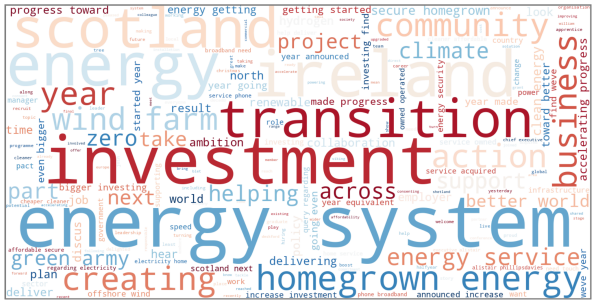
Volvocars

Rockwool Group

Leaseplan

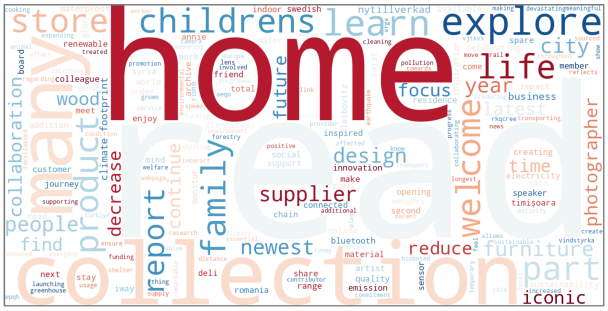


Sse



Enelenergia

Ikea



Storaenso

H&M

Edprenewables

Moethennessy

Reddit data

The Reddit dataset comprises posts and comments about Climate Change published by users during 2022. Here, users are individuals from the general population, unlike previously described for the X platform, where users were lobbyists. Individual users were annotated with labels reflecting their stance toward climate change (either as "skeptics" or "supporters"), and four user interaction networks, each covering discussion taking place during a quarter of 2022, were built. We selected Reddit due to the nature of interactions this platform allows to collect, which are primarily user-driven rather than influenced by complex algorithms.

To ensure a balanced analysis, we included subreddits on neutral topics that could capture interactions of both "supporters" and "skeptics." At the same time, we selected subreddits (i.e., thematic forums) that were slightly more oriented toward one of the two categories to more closely analyze the dynamics and relationships between the two groups, which were used for training models that can classify user stances. To further focus the data collection on the specific theme of interest, most subreddit contents were filtered using the keyword "climate." In particular, such filtering was applied to generalistic subreddits (e.g., r/askscience). The final dataset is then composed of 40,872 posts and 661,024 comments (in English) from the following subreddits:

> r/climate, r/ClimateChaos, r/climatedisalarm, r/conspiracy, r/ConspiracyII, r/conspiracytheories, r/TopConspiracy, r/Conservative, r/collapse, r/TrueAskReddit, r/changemyview, r/energy, r/environmentalscience, r/askscience, r/AskScienceDiscussion, r/environment, r/CollapseSupport, r/EverythingScience, r/sustainability, r/EvolveSustain, r/facepalm, r/WhitePeopleTwitter, r/Futurology, r/AskConservatives

The selection of subreddits was primarily guided by their presence and popularity on the archival site TheEye (https://the-eye.eu/redarcs/), which catalogs posts and comments from the most active subreddits.

User stance annotation

In order to annotate users, we leverage a BERT model fine-tuned to classify their expressed stance using textual features of their posts/comments.

We trained the model leveraging contents from explicitly aligned subreddits (i.e., for the "supporter" category: r/climatechange, r/ClimateActionPlan, r/ClimateOffensive; for the "skeptics" one: r/climateskeptics, r/skeptic): only posts from such selected subreddits have been used, covering the period from 2020 to 2022 while posts from 2019 were being employed for the model testing phase. The model resulted in an accuracy of 0.954 on the training set and 0.89 on validation. The final annotated dataset used to generate the quarterly interaction networks is composed of all subreddits not covered in the training set for the year 2022.

After obtaining the prediction probability values from BERT for user-generated contents of 2022, we rescaled them in the interval [-1, 1], thus better representing the nuances between extreme content generated by "supporters" (-1) and "skeptics" (1). Final users' scores were computed for each quarter as the average of their content values.

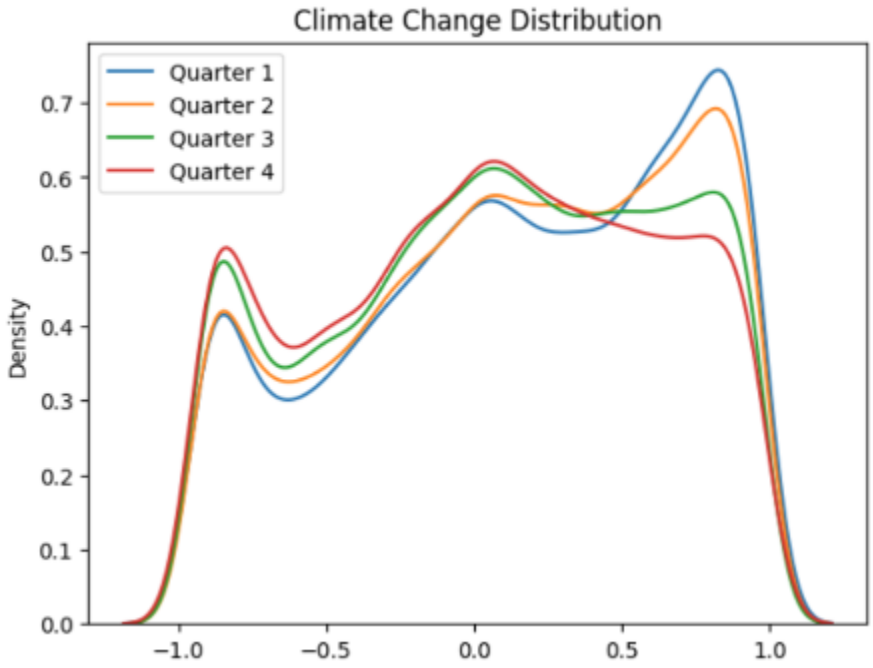
Social interaction network

Besides user opinions on climate change, we employed Reddit data to extract 4 social networks, one for each quarter or 2022. The users represented vertices in the network, and comments on posts represented edges between users.

Characteristics of the network

Figure 21 shows the distribution of opinions of users for each quarter. We observe that the opinions display a three-modal distribution, with the presence of two extremist groups (supporters and skeptics) and a group of moderates. In time, it appears that the fraction of extreme skeptics decrease, while moderates and extreme supporters increase.

Figure 21. Distribution of user opinions on climate change (-1=’supporter’, +1=’skeptic’)



As summarized in the following table, the number of nodes varies significantly between quarters, peaking in the third quarter and decreasing in the fourth quarter. The number of connections between nodes follows a similar trend to the number of nodes, with a steady increase up to the third quarter and a decrease in the fourth quarter. Such a trend could indicate increased interaction or activity during the middle months of the year, with a decline towards the end of the year. The table also includes the overall number of skeptics and supporters, classified by applying a threshold at 0 on the continuous opinions shown in figure 21. We note that the number of skeptics is always greater than supporters in every quarter.

Quarter	**Edges**	**Nodes**	**Supporters**	**Skeptics**
1	48939	26186	10227	15959
2	69212	32519	13003	19516

3	77527	34294	14927	19367
4	47217	23349	10607	12742

Conclusions and Discussion

The data collected so far represent a valid source of information on lobbying activity in Europe, they are very useful for monitoring the climate lobbying expenses of different organizations and understanding whether there are substantial differences between them. Furthermore, the textual data collected allow us to identify the policies on which lobbyists have mainly carried out their activity and, based on the direction of lobbying, understand how they could be influenced. Finally, the data can provide us with insights into lobbying intensity over time which is useful for highlighting the response of lobbyists to certain EU policies.

The report utilized a variety of data sources to provide a comprehensive view of climate lobbying activities and public opinion on climate issues. These sources include EU Transparency Register, Lobby Map, EU public consultations, LobbyFacts and social media platforms such as Twitter (X) and Reddit. The integration of data from different sources provides a nuanced understanding of lobbying activities. The EU Transparency Register and Lobby Map offer structured data on lobbying activities, while social media platforms offer insights into public discourse and opinion dynamics. In particular, the data from the EU Transparency Register and Lobby Map includes details such as the number of meetings with the European Commission, lobbying budgets, policies targeted, and types of organizations involved (e.g., corporations, non-governmental organizations). The Lobby Map data categorizes companies based on their support or opposition to community policies, with influence scores assigned to each company. These scores help identify companies that are more active or influential in climate lobbying, with notable differences in lobbying budgets based on their influence scores.

Analysis of Reddit data reveals the distribution of user opinions on climate change, showcasing a three-modal distribution with groups of supporters, skeptics, and moderates. Over time, extreme skeptic opinions have decreased, while moderates and extreme supporters have increased. Moreover, the Reddit data was used to create social interaction networks for each quarter of 2022, indicating varying levels of user activity and engagement. The number of nodes (users) and edges (interactions) peaked in the third quarter, suggesting increased engagement during the middle of the year.

All in all, the combination of structured lobbying data and unstructured social media data provides a thorough overview of climate lobbying and public opinion. This dual approach highlights both the formal lobbying efforts by organizations and the informal public discourse on climate issues. The data indicates that lobbying activities and budgets are substantial, with significant investments made by both corporations and industry associations. The analysis of lobbying intensity and the direction of lobbying efforts (support or opposition) offers insights into how lobbying may influence EU climate policies. Moreover, social media analysis reveals shifting trends in public opinion on climate issues, with a growing number of moderate and supportive voices over time. This shift could impact future lobbying strategies and policy responses.

The findings suggest that monitoring and understanding lobbying activities are crucial for policymakers to assess the influence of various stakeholders on climate policy. The data also highlights the need for transparency and regulation to manage lobbying efforts and their impact on policy-making. Overall, the report underscores the importance of leveraging diverse data sources to gain a holistic understanding of climate lobbying and public opinion, providing valuable insights for both researchers and policymakers.