PPOL 5203 FINAL PROJECT
Public Perspectives on Electric Vehicle Policy: Insights from YouTube Data

Ann Lian, Yuhan Ma, Junling Shao

This manuscript was compiled on December 13, 2023

In this study, we analyze public opinion on electric vehicle (EV) policy using data-driven methods applied to YouTube content from 2019 to 2023. Employing BERT topic modeling on video titles identifies prevalent EV policy aspects. We then conduct sentiment analysis on comments to gauge public attitudes. Additionally, TF-IDF uncovers significant positive and negative words in comments, revealing nuanced public expressions related to EV policy. Our research tracks evolving sentiments over time, aiming to comprehensively outline public views, highlighting consensus and contentious areas. This study provides insights for policymakers, industry stakeholders, and researchers interested in the evolving landscape of EV policy and public opinion.

Sentiment | Text Analysis | Public Policy | Eletric Vehicles

1. Introduction

The emergence of Electric Vehicles (EVs) signifies a significant turning point in the automotive sector, spurred by a heightened consciousness regarding environmental issues and the imperative for sustainable transportation alternatives. This initiative is propelled by the pressing necessity to comprehend and tackle the ramifications of EV policies, notably amidst their escalating global acceptance. Our focus is driven by a keen interest in exploring how these policies are perceived by the public, particularly their expressions on social media platforms, which serve as vital indicators of public sentiment and opinion.

Significance of Electric Vehicles. Electric vehicles (EVs) are increasingly seen as a crucial element in the global strategy to reduce greenhouse gas emissions and combat climate change. By utilizing electricity instead of traditional gasoline or diesel fuels, EVs can significantly decrease the amount of CO2 and other harmful pollutants released into the atmosphere. This transition is vital in urban areas, where vehicle emissions contribute substantially to air pollution and health issues. In addition to their environmental benefits, EVs are also transforming the automotive industry with their innovative technology. Electric motors offer instant torque, resulting in quicker acceleration and a smoother driving experience compared to internal combustion engines. This performance aspect is increasingly appealing to consumers who no longer have to compromise on driving enjoyment for environmental consciousness.

The advancements in battery technology play a critical role in the EV revolution. As batteries become more efficient, cost-effective, and capable of longer ranges, the practicality of EVs skyrockets. These developments address one of the main concerns potential EV owners have had in the past – range anxiety. With the increased range and growing network of charging stations, EVs are becoming a viable option for long-distance travel, further enhancing their appeal. Economically, the shift towards EVs opens up new opportunities and challenges. On one hand, it paves the way for new jobs and industries related to electric vehicle manufacturing, battery production, and charging infrastructure. On the other hand, it poses a significant challenge to traditional automotive manufacturers and oil industries, which must adapt to stay relevant in this changing landscape. Government policies and incentives play a pivotal role in accelerating the adoption of EVs. Subsidies, tax rebates, and investments in charging infrastructure are essential in making EVs more accessible and affordable to the general public. Additionally, some governments are setting ambitious targets for phasing out fossil-fuel-powered vehicles, further pushing the automotive industry towards electrification.

In summary, the significance of electric vehicles extends beyond their environmental benefits. They represent a major shift in automotive technology, driving patterns, and energy consumption. This transition not only promises to reduce our carbon footprint but also leads us towards a more sustainable and technologically advanced future in transportation.

Overview of Major EV Policies. The global push towards electric vehicle (EV) adoption is a multi-faceted endeavor, with countries around the world implementing a variety of policies and incentives to encourage the shift away from fossil fuel-powered vehicles. These efforts are pivotal in the global strategy to reduce greenhouse gas emissions and combat climate change.
In the United States, the approach to fostering EV adoption is comprehensive, involving both federal and state-level initiatives. The federal government offers significant incentives, such as tax credits up to 7,500 dollars for new EV buyers, alongside funding for EV infrastructure development to expand the network of charging stations. Additionally, the U.S. Department of Energy supports research and development in EV technology, focusing on improving battery efficiency and reducing costs. State policies further enrich this landscape. California's Zero Emission Vehicle (ZEV) program, which mandates a certain percentage of vehicle sales to be zero-emission vehicles, is a notable example. Other states offer additional incentives like tax credits, rebates, and reduced vehicle registration fees. Local governments complement these efforts with initiatives such as EV-only parking spaces, EV car-sharing programs, and municipal charging stations. In Europe, countries have adopted varied strategies. Norway and the Netherlands, for example, offer tax exemptions, free parking, and toll exemptions for EV owners. The European Union collectively aims for carbon neutrality by 2050, relying heavily on electrifying transportation. Asia's approach is equally diverse. China, aiming to dominate the global EV industry, combines subsidies and tax exemptions for consumers with quotas for electric vehicle production. India's focus on two-wheelers and three-wheelers through the FAME scheme reflects its unique transportation needs.

These global efforts highlight the importance of a multi-layered approach in the transition to electric vehicles. While financial incentives and subsidies are crucial, the success of these policies also depends on the development of adequate charging infrastructure and ongoing advancements in EV technology. As the cost of batteries decreases and infrastructure expands, these comprehensive policy frameworks across different nations and regions are set to accelerate the shift towards a more sustainable and technologically advanced future in transportation.

Research Questions. This project seeks to delve into the nuances of public opinion regarding EV policies as reflected on social media platforms. The central research questions are:

- What EV policies are most commonly discussed on social media, and what criticisms emerge during their implementation?
- What are the dominant sentiments expressed on these platforms concerning Electric Vehicles Policy?
- How can insights from sentiment analysis on social media inform the development of future EV policies and strategies?

By addressing these questions, this project aims to provide a comprehensive understanding of public opinion towards EV policies, thereby informing and shaping future strategies in this crucial area of sustainable transportation.

2. Data

Data sources. The data for this project was sourced using the YouTube API, primarily to access video IDs and associated comments. The YouTube API serves as a crucial gateway to extract extensive information from the platform, allowing us to gather user-generated content efficiently. Specifically, the $commentThread()$ function within the YouTube API facilitated the retrieval of comment threads, including replies, providing a comprehensive dataset for analysis. This method was chosen due to YouTube's position as one of the largest video-sharing platforms globally, hosting a diverse array of users and a multitude of comments. Leveraging the API offered us an opportunity to tap into this reservoir of public opinions and sentiments, essential for understanding various policies and garnering insights from user-generated discussions. However, while the YouTube API provided valuable data, limitations were acknowledged. These limitations include potential rate limiting, which imposed restrictions on request frequency and volume, as well as the possibility of incomplete datasets due to filters or privacy settings. Additionally, the ever-evolving policies and terms of the YouTube API may impact data accessibility and necessitate continual adaptation to ensure data accuracy and completeness. Despite these limitations, the YouTube API remains a critical resource for obtaining valuable insights into public sentiments and opinions expressed within the platform's vast and dynamic ecosystem.

Data Acquisition. For our data collection process, we relied on the $youtube-data-api$ library, which provides a comprehensive set of functions designed to streamline access to the YouTube API. Our initial step involved obtaining videos from each year, followed by a manual filtration process. And then get the comments of each video.

Get Video IDs. Using the $yt.search()$ function, we set specific parameters to target videos relevant to our research theme of EV policy. This involved employing keywords $('EV, policy')$ within the query parameter $('q =')$ to filter videos pertinent to our focus. Additionally, we utilized the $'publishedAfter'$ and $'publishedBefore'$ parameters to select videos within designated time frames, segregating them year by year. Furthermore, setting $'regionCode = `US'$' ensured that the collected videos primarily used English, aiding subsequent text analysis.
During our data collection process, we encountered varying degrees of relevance among the videos obtained, despite our efforts to target EV policy-related content using specific keywords. To address this, we meticulously examined the dataset and identified a distinct threshold that demarcated relevant videos from those unrelated to our research theme. This manual assessment allowed us to establish a clear boundary or "divider" within the dataset, enabling us to discern and subsequently eliminate videos that did not align with our research focus on EV policy. By systematically reviewing the dataframe, we efficiently segregated the videos, retaining only those directly relevant to our study while discarding the unrelated ones. This

meticulous filtration process ensured the integrity and relevance of the dataset, refining it to comprise solely content pertinent to our research objectives.

Get Comments of Videos. Subsequent to obtaining the video IDs, our next step involved retrieving the comments associated with each video. Leveraging the `yt.get_video_comments()` function, we implemented a restriction by setting the parameter `max_result` to 500. This parameter served to limit the retrieval to a maximum of 500 comments per video. This decision was informed by our observation that the majority of videos typically contained fewer than 500 comments. However, a few videos exhibited an extensive number of comments, surpassing 10,000 in some cases. To prevent potential bias in our subsequent analysis, we opted to cap the number of retrieved comments per video. Without this limitation, a disproportionately large portion of the total comments could stem from a handful of videos, potentially skewing our analytical outcomes. Thus, by imposing a cap on the retrieved comments per video, we aimed to ensure a more balanced representation across the dataset, fostering an unbiased and comprehensive analysis of the collected comments. To prevent interruptions caused by videos with closed comments, we employed a try-except approach in our data retrieval process. This method allowed us to identify and skip videos that had closed comments, ensuring a seamless data collection process without errors or interruptions.

Data wrangling. Although the raw data sets have already been provided in tidy, tabular format, some columns still contain nonsensical observations due to inaccurate entries and/or other reasons. For that reason, we conduct a thorough wrangling process on the key columns that are important in our analysis. The wrangling process includes the following steps:

Cleaning Comments without `video_id`. Within the comments dataset gathered, a significant observation emerged—comments lacked `video_id` but featured `parent_comment_id`. However, given our requirement to merge videos and comments based on `video_id`, comments devoid of this crucial identifier were deemed incompatible. Therefore, as a primary step in our data refinement process, we systematically removed comments that lacked the essential `video_id`, ensuring a cohesive and functional dataset for subsequent analyses.

Refinement by Comment Length. Furthermore, within the comments dataset, a substantial portion comprised brief entries consisting of emojis or single words, potentially offering limited insights and introducing unnecessary noise into our analyses. Hence, to maintain data quality and ensure the relevance of information, comments containing fewer than 100 characters were excluded. This action aimed to retain more substantial comments, providing a richer source of information for our subsequent analysis.

Recoding Date-Time Variables for Sentiment Trend Analysis. Acknowledging the importance of analyzing sentiment trends over distinct time periods, we performed a recoding of the date-time variables into year and quarter divisions. This transformation aimed to enable a more granular analysis, facilitating a nuanced exploration of sentiment shifts over various temporal contexts. This structured approach to data wrangling ensured a refined and organized dataset, primed for comprehensive sentiment and text analyses across different periods.

Merging Videos and Comments Data. Recognizing the significance of information encapsulated within the videos dataset, notably `video_publish_date` and `video_title`, critical for sentiment and text analyses, we proceeded to merge these two datasets. This merging operation allowed for a more comprehensive dataset, amalgamating essential video details with comment-related information, facilitating more robust and contextually enriched analyses.

Table 1 presents a chronological summary from 2019 to 2023 of EV policy content on YouTube, showcasing annual video uploads, user engagement through comments, and the year-on-year variations in these metrics.

Year	2019	2020	2021	2022	2023
Video	154	106	168	162	175
Differences	0	-48	62	-6	13
Comments	4329	4139	4116	4167	8266
Differences	0	-190	-23	51	4099

Table 1. Number of Videos and Comments

3. Analysis Methods

Bert Topic model. Electric vehicle (EV) policy involves a range of strategies designed to promote the adoption and production of EVs. In our research, we utilized the YouTube API to gather data, but a key question arises: what insights can we derive from the limited text information available? To address this, we employed the Bert topic model, a method that uses BERT embeddings to form dense, meaningful clusters, complete with scores and counts. This technique not only facilitates the identification of easily interpretable topics but also retains crucial words within the topic descriptions. Consequently, we applied the Bert topic model to video titles as a means to extract information relevant to EV policies.

Video titles are concise summaries of their content. Using the Bert topic model on these titles allows for the efficient extraction of key topics and themes. This is particularly useful for quickly identifying the primary focus areas of a large number of videos. Further,By analyzing video titles with the Bert topic model, it's possible to identify broader trends and patterns in the content

related to EV policies. This can provide insights into public interest, prevailing sentiments, and emerging topics in the field of electric vehicles.

Upon successful installation package, we configure BERT topic to utilize the "paraphrase-MiniLM-L3-v2" embedding model,capturing semantic meanings and allowing us to derive high-quality embeddings for the textual data.To ensure a focused analysis, we set the minimum topic size to 3, allowing the model to discard topics with less than three occurrences, thereby enhancing the significance and distinctiveness of the identified topics. With the BERT topic model instantiated, we proceed to fit the model to our dataset comprised of video titles in different years.Further, we extract the frequency distribution of the topics with "gettopicinfo()", which provides us with an overview of the topic representation across the dataset from 2019 to 2023. Finally, we use buit-in visualization model ".visualizebarchart" to visualize parts of our topic cluster.

Sentiment Analysis. The sentiment analysis process involves several sequential steps applied across multiple DataFrames representing different years of comments. Initially, the $SentimentIntensityAnalyzer()$model is instantiated as 'sid' to facilitate sentiment analysis. This model is then used to analyze the comments of each year, generating polarity scores for each comment's text. These polarity scores are stored in a new 'sentiment' column within their respective DataFrames. Subsequently, the sentiment-analyzed DataFrames are vertically concatenated using 'pd.concat()' to create a unified DataFrame encompassing sentiments across multiple years. Within the $'combined_df'$, the $'compound_values$ from the $'sentiment'$ column (stored as dictionaries) are extracted and added into a new column labeled $'compoundvalues'$. This extraction process allows for the capture of compound sentiment scores associated with each comment. To further categorize these compound values, a function is defined, setting thresholds (-0.2 and 0.2) to label sentiments as "negative," "positive," or "neutral." This function is then applied to the 'compound values' column, resulting in the creation of a new column named 'sentiment label' in the 'comments' DataFrame. This 'sentiment label' column assigns specific sentiment labels to each comment based on their respective compound scores. By systematically conducting sentiment analysis, extracting compound sentiment values, and categorizing sentiments based on defined thresholds, this process enables the creation of a comprehensive DataFrame facilitating the exploration and interpretation of sentiment trends across different years of comments.

TF-IDF. TF-IDF (Term Frequency-Inverse Document Frequency) is a numerical statistic that reflects how important a word is to a document in a collection or corpus. The TF-IDF analysis identified words that are significant in distinguishing between positive and negative sentiments within the comments on EV policy. This information can be valuable for understanding public opinion, identifying areas for policy improvement, and tailoring communication strategies. We utilized TF-IDF for analyzing text data. Initially, the text is meticulously preprocessed—cleaned of URLs, special characters, and transformed to lower case, ensuring a standardized dataset. This preprocessing is crucial for the effective application of the TF-IDF Vectorizer, a tool that identifies and scores the most relevant words in the text. Each year's data is treated as a distinct dataset, allowing for year-specific insights. The TF-IDF scores are meticulously calculated to pinpoint the most significant words, a critical step for understanding textual trends. Additionally, sentiment analysis adds another layer of depth. By calculating sentiment polarity for each comment, the project assesses not just the prevalence of words, but also the emotional tone they convey. The final step is visualizing this rich dataset. Word clouds and bar plots are employed to vividly display the most impactful words, combining their TF-IDF and sentiment scores. These visualizations not only highlight the key words but also reveal the underlying sentiment trends, providing a comprehensive picture of the textual landscape over the four-year period. This project stands out for its thorough approach to text analysis, offering nuanced insights into the evolving themes and sentiments in the data.

4. Results

Bert Topic model.

The Conclusion of Top Four Categories of EV policy. From the model results, we could draw some very useful conclusions. The figure 1 is a results sample:

	Topic	Count	Name	Representation	Representative_Docs
0	-1	46	-1_ev_policy_delhi_car	['ev', 'policy', 'delhi', 'car', 'electric', 'vehicle', 'laos', 'ಎಸ್', 'trying', 'revolt']	['Powerful Vested Interests Trying t
1	0	28	0_pakistan_vehicles_approved_cars	['pakistan', 'vehicles', 'approved', 'cars', 'electric', 'car', 'policy', 'ever', 'policy2020', 'jolta']	['Electric Car Policy2020 Pakistan |
2	1	27	1_delhi_government_vehicle_electric	['delhi', 'government', 'vehicle', 'electric', 'of', 'policy', 'boost', 'economy', 'to', 'mobility']	['Mobility Plug by OMI, E2- Future
3	2	26	2_auto_zone_singh_news	['auto', 'zone', 'singh', 'news', '2020', 'update', 'ev', 'launch', 'sonet', 'kia']	['NEW DELHI EV POLICY LATEST U
4	3	20	3_episode_hub_transportation_live	['episode', 'hub', 'transportation', 'live', 'technology', 'video', 'the', 'on', 'week', '2040']	['EV Hub Live Episode #1: Electric
5	4	17	4_charging_stations_station_van	['charging', 'stations', 'station', 'van', 'amazon', 'swapping', 'charger', 'review', 'ev', 'kerala']	['Electric Vehicle News India #5, De
6	5	11	5_sri_ktr_ministers_summit	['sri', 'ktr', 'ministers', 'summit', 'launching', 'telangana', 'at', 'ajay', 'kumar', 'of']	['Ministers KTR & Sri Ajay Kun
7	6	11	6_खर_सरक_दन_पर	['खर', 'सरक', 'दन', 'पर', 'भव', 'delhi', 'हन', 'आर', 'लन', 'ki']	['Delhi सरकार NEW VEHICLE खरीदने
8	7	11	7_vis_isuw_programs_policies	['vis', 'isuw', 'programs', 'policies', 'international', 'scenarios', 'and', 'scenario', 'india', '2020']	['ISUW 2020 | Shantanu Jaiswal, BI
9	8	9	8_india_model_bikes_nahak	['india', 'model', 'bikes', 'nahak', 'mercedes', 'jaguar', 'fast', 'ep', 'entering', '11']	['Electric Vehicle News #8 Jaguar i-

Fig. 1. results sample

We conducted a manual review of the outcomes generated by the topic model spanning the years 2019 to 2023. As depicted in Figure 2, a consistent trend emerged, showcasing prevalent topics such as tax, subsidy, charging stations, and battery swapping among others. These recurring clusters led us to identify four significant EV policy mechanisms commonly implemented in practical scenarios.

1. **Infrastructure:** Government policies focused on infrastructure play a crucial role in promoting the adoption of electric vehicles (EVs). This involves supporting the installation of public charging stations, which are essential for EV users who need to recharge their vehicles while away from home.

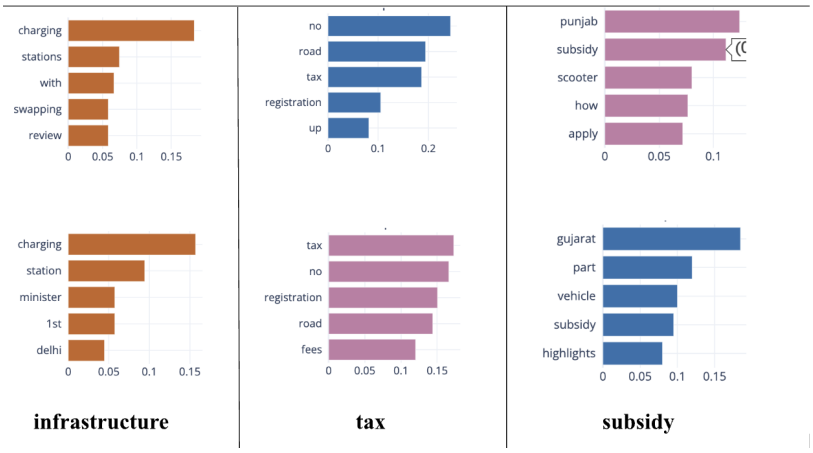

Fig. 2. Three Categories Policy

2. **Subsidies:** Subsidies serve as a direct financial incentive to lower the entry barrier for consumers interested in purchasing EVs. By reducing the upfront cost, subsidies make EVs more accessible to a broader segment of the population. These financial incentives can also extend to businesses involved in the design and production of EVs, fostering innovation and growth within the EV industry. By subsidizing research and development or manufacturing costs, governments can stimulate technological advancements and increase the competitiveness of EVs in the market.

3. **Tax Benefits:** Tax incentives are a strategic tool used by governments to make electric vehicles more financially appealing. These benefits might include reductions or exemptions from vehicle registration taxes, import duties, and value-added tax (VAT) on EV purchases. Such measures lower the overall cost of ownership for EVs, making them a more attractive option compared to traditional vehicles. In addition, for businesses, tax benefits could include favorable depreciation schedules for EVs or tax credits for installing charging infrastructure, thereby encouraging corporate investment in environmentally friendly transportation solutions.

4. **Regulation and Mandates:** This aspect of EV policy is unique as it targets specific goals rather than focusing directly on EV manufacturers or consumers. Such policies may encompass mandates that require a certain percentage of new vehicle sales to be electric by a designated date, enforcing a shift towards cleaner transportation. Additionally, emissions regulations can be implemented that necessitate the use of electric vehicles to meet environmental standards. In the realm of social media, as observed on YouTube, the topic related to President Biden's mandates on this matter generates considerable discussion, indicating in 3.

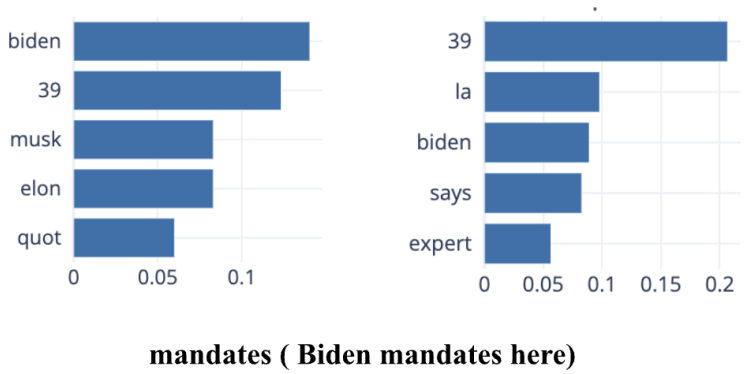

Fig. 3. Mandates

Our analysis reveals that in the early or developing phases of electric vehicle (EV) adoption, nations primarily focus on infrastructure development, subsidies, and tax incentives. By checking representative documents could find that these discussions centered around countries like India, China, and Pakistan. These strategies lay the groundwork for a supportive ecosystem for both consumers and manufacturers, alleviating initial expenses, enhancing the availability of charging facilities, and providing financial incentives to bolster the appeal and feasibility of EVs. In contrast, countries with more established EV markets, such as the United States, are inclined to adopt a more regulatory-driven approach. That's why we can observe the dominant topic about "Biden" in our model. Policies like the Biden Mandates and stringent emissions standards are examples

of such strategies. These regulatory measures are designed to expedite the shift towards electric mobility and stimulate further innovation in the EV sector, demonstrating a different stage in the evolution of EV policy implementation.

Conclusion of Top Four Concerns on EV Policy. We also detected some negative attitudes towards the EV policy. By checking the representative documents, four concerns about EV policy are concluded in figure 4.

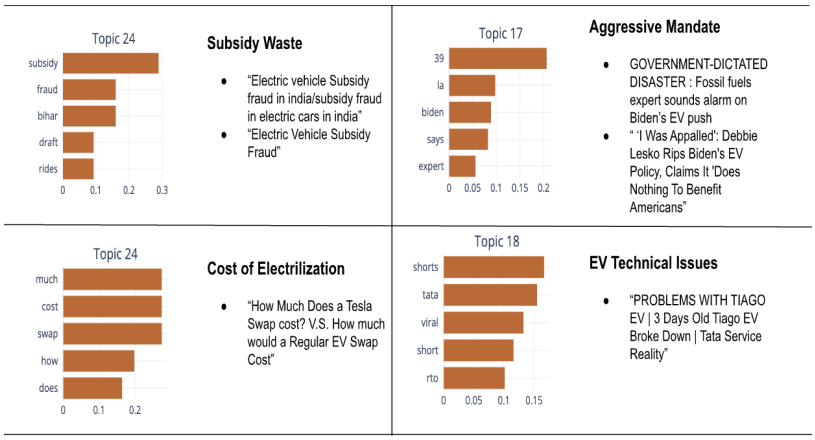

Fig. 4. Top Four Concerns

1. **Subsidy Waste:** Concerns have emerged over the misuse of government subsidies by some EV companies and beneficiaries, leading to a wastage of resources. Particularly, the term "subsidy fraud" underscores instances of deception within subsidy programs. Such fraudulent activities not only undermine the integrity of these programs but also result in the inefficient allocation and expenditure of vital funds, which are intended to facilitate and encourage the adoption of electric vehicles. This topic might spurred by the subsidy fraud in India's electric cars company.

2. **Cost of Electrilization:** The financial cost of transitioning to electric vehicles has become a growing public concern, underscored by the topic "swapmuchcost." This reflects the escalating apprehension about the costs associated with the electrification of vehicles, notably the expense of converting from traditional fuel systems to electric ones.

3. **Aggressive Mandate :** In recent years, a prominent example of a mandate in the EV sector is the "Biden Mandate." During 2022 and 2023, discussions around EV policies frequently featured mentions of "Biden," indicating a Biden related EV political atmosphere. Key objectives of this mandate include an ambitious target of 67 percent of new vehicles being electric by 2032. Analysis of public opinion reveals a generally negative sentiment towards the Biden administration's approach. Critics argue that these targets are excessively ambitious and question their feasibility, leading to skepticism about the actual benefits these policies might bring to Americans.

4. **EV Technical Issues:** Moreover, certain technical challenges specific to electric vehicles (EVs) themselves have become increasingly prominent, highlighted by discussions around topics such as "shortstara," which alludes to Tara, an Indian EV manufacturer. Given the electric vehicle industry's relatively recent emergence and its continuously developing technical capabilities, EVs are encountering notable issues in terms of safety and functionality, especially when compared to traditional, combustion-engine vehicle.

Sentiment. The sentiment analysis conducted aimed to discern the prevailing sentiments within comments linked to the videos, culminating in two pivotal graphs depicting trends and distributions. The figure 5 outlines the trend of sentiment scores, primarily focusing on compound values. Predominantly, data points align above 0, indicating an overarching positive sentiment prevailing within the comments. However, a discernible trendline showcases a gradual decline in sentiment scores over time, suggesting a noteworthy shift towards less positive expressions or a potential evolution in the nature of discussions associated with the videos.

The figure 6 illustrates the distribution of sentiments across quarters, categorized into Positive (¿0.2), Neutral (between 0.2 and -0.2), and Negative (¡-0.2) labels based on the compound values derived. Within this breakdown, a stable trend emerges for Positive Sentiment, signifying consistent positivity within comments over time. Conversely, Negative Sentiment exhibits an increasing trend, indicating a rise in negative expressions within the discussions related to the videos. Conversely, the trend in Neutral Sentiment showcases a decline, suggesting a decrease in the neutrality of sentiments expressed over the assessed periods.

These insights gleaned from the sentiment analysis offer valuable perspectives on the evolving nature of sentiments within the video-associated comments. The assessment of trends and sentiment distributions provides a comprehensive view, shedding light on shifts in sentiments and potential alterations in the tone or direction of discussions across different time spans.

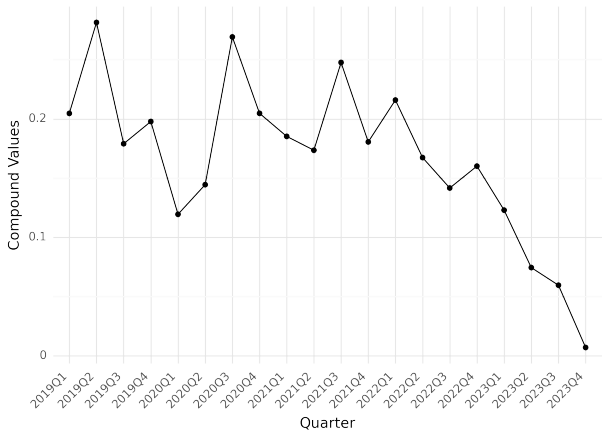

Fig. 5. The Trend of Sentiment Scores

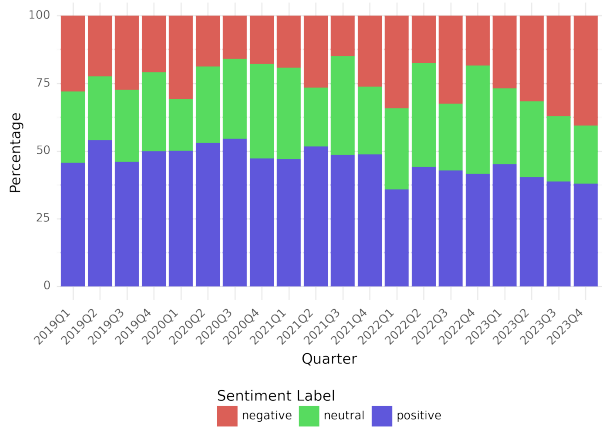

Fig. 6. The Distribution of Sentiment

TF-IDF. Figure 7 displays two horizontal bar graphs representing the ”Top Words in Strongly Positive Comments” and ”Top Words in Strongly Negative Comments” from the year 2019, using a TF-IDF model combined with sentiment analysis. The left graph with green bars shows words that are most strongly associated with positive comments. Words like ”best,” ”good,” ”great,” and ”video” have the highest TF-IDF scores, suggesting they are frequently used in positive contexts and are important in distinguishing positive comments from other comments. Other words like ”thank,” ”please,” ”love,” and ”happy” are also significant, which aligns with the language typically found in positive feedback.

The right graph with red bars shows the words most strongly associated with negative comments. Here, we see words like ”hai,” ”bad,” ”rental,” and ”never,” which are generally associated with negative sentiments. There are also words like ”car,” ”Pakistan,” ”China,” and ”battery,” indicating that specific subjects or topics might be associated with negative comments. The presence of certain words like ”never” and ”bad” clearly indicates a negative sentiment.

Figure 8 displays two horizontal bar graphs showcasing the ”Top Words in Strongly Positive Comments” and ”Top Words in Strongly Negative Comments” for the year 2021.

In the positive comments graph (green bars), words such as ”good,” ”best,” ”great,” and ”video” indicate a general sense of approval and appreciation, similar to the findings in 2019. The specific mention of ”EV,” ”electric,” ”battery,” and ”India” suggests that the discourse remains focused on electric vehicles, possibly with a significant portion of the conversation relating to the Indian market or EV developments in India. The inclusion of words like ”please,” ”thank,” and ”bro” reflect a conversational tone in the positive feedback, which may imply a community-like environment among proponents of EVs.

On the negative comments side (red bars), the word ”stupid” tops the list, followed by ”bad,” ”bike,” and ”price,” pointing to criticisms likely regarding the practicality, cost, and perhaps comparison to traditional bikes or motorbikes. Notable terms like ”expensive,” ”investment,” and ”quality” suggest concerns about the affordability and perceived value of EVs. Words such as ”government,” ”Toyota,” and ”sell” could be referencing specific policies, companies, and market dynamics that are under scrutiny or debate.

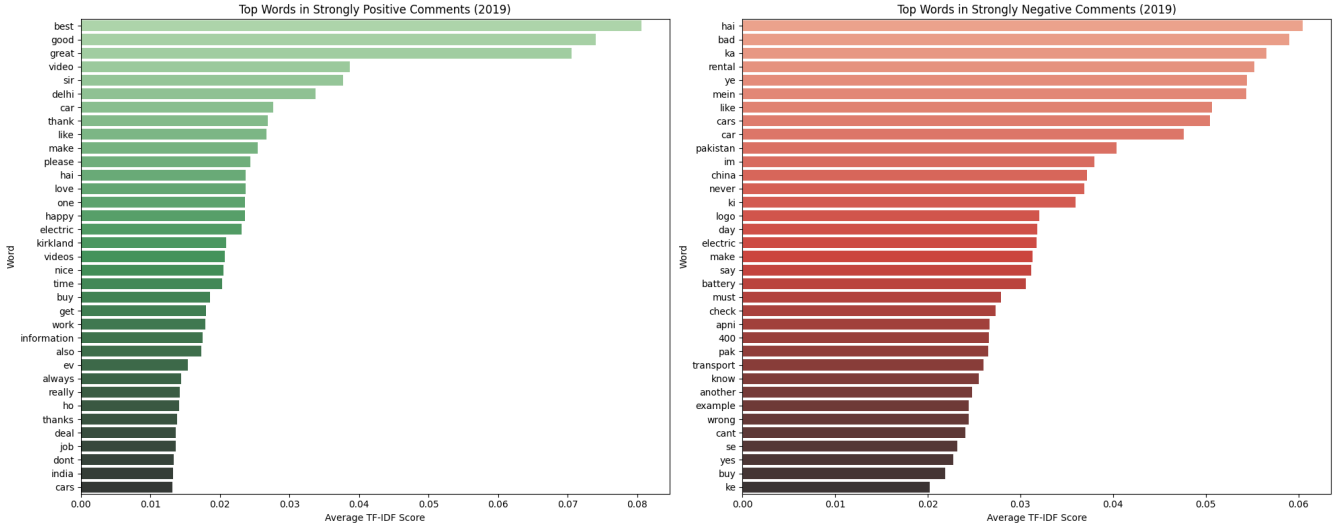

Fig. 7. Top Words in Strongly Positive and Negative Comments (2019)

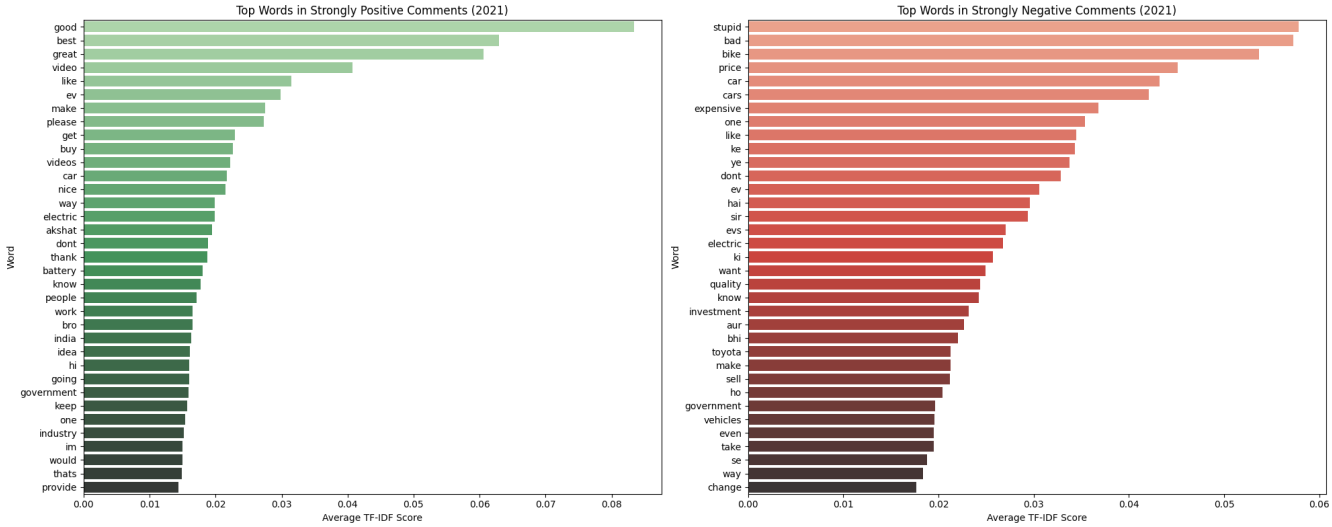

Fig. 8. Top Words in Strongly Positive and Negative Comments (2021)

The mention of ”change” at the bottom of the negative list hints at discussions about the need for, resistance to, or consequences of changes in the EV industry or policies. ”EV,” ”electric,” and ”EVs” appearing in both positive and negative comments indicate that electric vehicles are still a central topic of discussion, but with divided sentiments.

Compared to 2019, the 2021 data suggests a continued and perhaps more nuanced public engagement with the topic of EVs, where specific aspects such as pricing and quality are becoming prominent points of contention. This information can be particularly useful for policymakers and industry stakeholders who aim to understand and address public concerns, improve the reception of EVs, and strategically plan for interventions that could enhance the adoption of electric vehicles. Upon collecting the most frequent words from strongly positive and negative comments spanning from 2019 to 2023, we manually categorized each term based on its relevance to policymaking, with the aim of deciphering public sentiment on EV policy. This thorough analysis yields critical insights into the collective viewpoint, serving as a key resource for policymakers, automotive firms, and other interested parties who are actively shaping the future of electric vehicles and associated regulations. It equips these stakeholders with a clearer understanding of the facets of EV policy that are positively embraced by the public, as well as those areas eliciting concerns that must be addressed to enhance public support and policy efficacy.

Table 2 shows the list of top words from positive comments about EVs from 2019 to 2023. It reveals several key themes and sentiments in public discussions:

Battery	Carmax	China	Deal	Delhi
Electric	Government	India	Industry	Information
Job	Kirkland	Make	Oil	Save
Tesla	Time	Toyota	Work	

Table 2. Relevant Top Words in Positive Comments (2019 - 2023)

1. **Electric and Battery:** These words being at the top suggests that there is a positive sentiment towards the core technology of EVs. The term "electric" reflects the general subject of discussion, while "battery" indicates a specific focus on the technology that powers these vehicles.

2. **Geographical References:** The presence of "China", "India", and "Delhi" shows that there's a significant amount of positive discussion around EVs in these regions. China and India are two of the largest automotive markets globally, and positive comments might reflect successful policies, growing market penetration, or consumer enthusiasm in these areas.

3. **Brand Names:** "Tesla" and "Toyota" indicate positive sentiment towards these manufacturers. Tesla is known for pioneering the EV market, and Toyota has made significant advances in hybrid technology. Positive comments could reflect approval of their products, innovation, or contribution to the industry.

4. **Economic Factors:** Words like "oil", "save", "job", "industry", and "deal" suggest that economic considerations are a major part of the positive discourse around EVs. There could be a recognition that EVs can save money on fuel (oil), create jobs, impact the larger automotive industry positively, and offer good deals, potentially through incentives or lower operational costs.

5. **Policy and Government:** The word "government" implies that state policies and regulations regarding EVs are seen in a positive light, which might be due to subsidies, tax breaks, or supportive infrastructure policies that encourage EV adoption.

6. **Functionality and Reliability:** The word "work" may indicate that commenters positively view the functionality and reliability of EVs. This suggests that the practical day-to-day use of EVs meets or exceeds expectations.

7. **Environmental Impact:** The term "save" also has environmental connotations, possibly indicating that commenters are acknowledging the benefits of EVs in terms of reducing emissions and protecting the environment.

8. **Information and Awareness:** The term "information" could imply that there is a positive reception to the information being shared about EVs, whether through marketing, educational content, or news media. This suggests effective communication from stakeholders in the EV space.

9. **Specific Locations and Retailers:** "Kirkland" and "Carmax" may refer to locations or companies that have a positive association with EVs in the minds of commenters. Kirkland could be related to a regional policy, a business, or an event, while Carmax might be mentioned positively in the context of purchasing or dealing with EVs.

Battery	Biden	Bike	Carbon	Carmax
China	Countries	Demand	Electric	Expensive
Government	Hybrid	Investment	Model	Oil
Pakistan	Petrol	Planet	Poor	Power
President	Price	Quality	Rental	Ride
Roads	Scooters	Sell	Solution	Tax
Time	Toyota	Transport	US	

Table 3. Relevant Top Words in Negative Comments (2019 - 2023)

Table 3 shows the list of top words from negative comments about EVs from 2019 to 2023. It reveals concerns and criticisms in the public discourse:

1. **Cost-Related Concerns:** Words like "expensive", "tax", "price", and "investment" suggest that a significant portion of the negative comments may be related to the financial aspect of EVs. This could reflect perceptions that EVs are not affordable for the average consumer or that the associated costs, such as taxes, are prohibitive.

2. **Energy and Environmental Issues:** "Oil", "petrol", "carbon", and "planet" indicate environmental considerations are being discussed, possibly pointing to doubts about the actual environmental benefits of EVs or contrasting them with traditional fossil fuels.

3. **Technical and Performance Issues:** "Battery", "poor", "quality", "power", and "model" likely relate to concerns about the performance and reliability of EVs. There might be skepticism about battery life, vehicle power, or the overall quality of certain models.

4. **Infrastructure and Practicality:** The presence of "roads", "transport", "bike", "scooters", and "ride" implies issues with the infrastructure and practicality of using EVs for everyday transportation.

5. **Brand and Model Specific Feedback:** Negative mentions of "Toyota" and "Carmax" might reflect specific grievances with these brands or the services provided by companies like Carmax.

6. **Political and Policy Criticism:** "Biden", "president", "government", and "countries" suggest that government policies and leaders are being criticized in the context of EV development and implementation.

7. **Global and Regional Challenges:** "China", "Pakistan", and "us" indicate geopolitical discussions, possibly reflecting concerns about the global distribution of EV technology and market dynamics.

8. **Market Dynamics:** "Sell", "demand", and "rental" point towards the market aspects of EVs, including sales performance and rental services, which might not meet the expectations of some consumers.

In summary, the negative comments point towards concerns about the cost and affordability of EVs, skepticism about their environmental impact and technical performance, challenges with infrastructure, and criticism of government policies and market dynamics. These points suggest areas where consumers feel the EV industry and policymakers need to focus their efforts to address the gaps between expectations and reality.

5. Discussion

Re-introduce Our Main Results. Utilizing the YouTube API, we have successfully extracted video titles and comment texts to conduct a comprehensive analysis of public sentiment towards electric vehicle (EV) policy. By applying the BERT topic model, we identified four key categories of EV policy, along with four primary areas of concern related to these policies. Furthermore, our sentiment analysis of video comments has provided valuable insights into the evolving trends of public opinion regarding EV policy. Additionally, by calculating TF (Term Frequency) scores, we have been able to pinpoint the most prominent words in both negative and positive comments, offering a nuanced understanding of public perceptions and attitudes.

Our Contribution.
Practical Meaning: our research provides valuable insights into public sentiment regarding EV policies, which can guide policymakers and industry stakeholders. Understanding public concerns and interests can help in the formulation of policies that are more aligned with societal needs and preferences, leading to higher acceptance and successful implementation. In addition, by identifying the key topics and sentiments surrounding EV policies, our study contributes to shaping a more informed public discourse. It can encourage more nuanced discussions among the general public, media, and advocacy groups, promoting a deeper understanding of the issues at play in the EV sector.
Methodology: By combining topic modeling and sentiment analysis, our research bridges the gap between qualitative insights and quantitative data. This comprehensive approach allows for a more holistic understanding of public opinion, enriching the field of policy analysis with both depth and scale. Secondly, we might provide an option when we need know the general information and public attitudes towards specific industry. Our approach to analyzing EV policies diverges from conventional methods that rely on article extraction or sentiment surveys. Instead focusing on video content, we specifically selected YouTube video title and comments as our primary data source. While the videos themselves may not yield extensive sentiment data, the user comments beneath them offer a wealth of insights. Videos, in comparison to articles and surveys, are more engaging and appealing to a broader audience, making them a potent medium for information dissemination. Consequently, the comment sections of these videos serve as a rich repository of public opinion and perspectives, providing us with ample data for our analysis.

Next Step.
In our future study, we plan to extend our research by incorporating data from additional social media platforms and online forums, aiming to encompass a wider spectrum of public opinions. This expansion is expected to enhance the robustness of our findings, although it will face the more challenges of managing and analyzing a much larger and possibly more heterogeneous dataset, which could be complicated to extract clear patterns and trends. Furthermore, we are considering the integration of more advanced techniques, such as video transcription and natural language processing (NLP), to analyze verbal content more thoroughly. This approach allows us to gain a more comprehensive understanding of the discussed topics and sentiments then merely focues on the video titles and comments, especially in connecting specific EV policies to corresponding sentiment changes. Additionally, we aim to refine our filtering methods to address the current data imbalance skewed towards Indian content. By broadening our dataset to include videos and comments from diverse geographical locations, we hope to perform a more targeted analysis of public sentiment towards EV policies across different countries. The main challenge here will be ensuring that our filtering methods can effectively normalize the data across various regions, considering the vast cultural and linguistic diversity, which might impact the interpretation and comparability of sentiments and opinions.