SILESIAN UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE

SCIENTIFIC PAPERS OF SILESIAN UNIVERSITY OF TECHNOLOGY 2024
ORGANIZATION AND MANAGEMENT SERIES NO. 210

CREATOR AND VIEWER ACTIVITY ON YOUTUBE IN POLISH CONTENT ABOUT ELECTRIC VEHICLES

Marcin WYSKWARSKI

Institute of Economy and Informatics, Faculty of Organization and Management of Silesian University of Technology; marcin.wyskwarski@polsl.pl, ORCID: 0000-0003-2004-330X

Purpose: This study aims to examine the activity of content creators and viewers on YouTube in the context of Polish-language videos related to electric vehicles (EVs). The study aims to explore trends in video publication, user engagement through comments, and the relationships between video characteristics (such as duration and views) and user interaction.

Design/methodology/approach: The research uses a dataset of 1152 videos and 125,680 comments collected from YouTube using Python and R tools. The videos were identified through specific search phrases related to EVs. The methodology includes descriptive analysis of video publication trends and linear regression models to investigate the relationships between video duration, views, and comment count. Data were cleaned and pre-processed to remove non-Polish comments, duplicate entries, and outliers.

Findings: The analysis revealed a growing trend in user engagement with EV-related content on YouTube over the years. A small group of highly active users contributed the majority of videos, while a significant portion of user interaction consisted of replies, indicating a high level of viewer-to-viewer engagement. Regression models showed a statistically significant but weak correlation between video duration and comments, as well as between video views and comments, but found no significant relationship between video length and views.

Research limitations/implications: This study is limited to videos and comments in the Polish language, which restricts the generalizability of the findings to other linguistic or cultural contexts. Furthermore, data were collected up until May 2024, making the analysis of trends in that year incomplete. Future research could include a broader range of content and explore additional factors that influence user engagement, such as video content quality or audience demographics.

Practical implications: The findings suggest that content creators in the EV sector on YouTube may benefit from building more engaging discussions among viewers. Understanding that longer videos or videos with more views may attract slightly more comments can help creators tailor their content strategies. For platform managers and marketers, fostering user interaction through replies and discussions may be a key way to increase engagement with content.

Originality/value: This study is among the first to analyse the dynamics of creator and viewer activity on YouTube specifically within the Polish-language content about electric vehicles. It contributes to a better understanding of how creators and viewers engage with this topic, providing insights into content trends and user interaction that may guide future content development strategies.

Keywords: YouTube, Electric vehicles, User engagement, Viewer comments, Creator activity, Regression analysis.
Category of the paper: research paper.

1. Introduction

The European Union has implemented various initiatives aimed at fostering the adoption of electric vehicles (EVs), including financial incentives, infrastructure development, and strategic plans designed to encourage consumers to transition to EVs (Hawkins et al., 2013; Pasaoglu et al., 2012; Sierzchula et al., 2014). Research shows that widespread adoption of EVs is difficult to achieve without government subsidies (Rahmani, Loureiro, 2018). Incentives, particularly in the early stages of new technology introduction, play a crucial role in increasing consumer adoption (Davies et al., 2016), (Figenbaum et al., 2015), and (Trip et al., 2012).

Electric vehicles (EVs) have emerged as a critical technology in reducing greenhouse gas emissions. Alongside government incentives and infrastructure developments, social media platforms like YouTube have played an increasingly influential role in shaping public opinion and consumer behaviour towards EVs. This paper explores the activity of creators and viewers within Polish-language content on YouTube, providing insights into the patterns of engagement and interaction surrounding electric vehicles (EVs).

EVs are considered a key component in reducing greenhouse gas emissions and improving urban air quality, benefiting public health by reducing pollutants such as nitrogen oxides and particulate matter (Hannan et al., 2014; Ma et al., 2012; Van Mierlo et al., 2006). Despite these benefits, many consumers prioritize personal convenience and cost over environmental concerns when making purchasing decisions (Buenstorf, Cordes, 2008).

Given the rapid growth of the electric vehicle market, which is projected to make up over 30% of the light-duty vehicle market in the United States by 2030 (Wolinetz, Axsen, 2017) understanding public perception and sentiment towards EVs is increasingly important for both policymakers and manufacturers. One avenue for understanding these perceptions is through online platforms, particularly YouTube, where creators and viewers discuss and engage with content related to electric vehicles.

2. YouTube as a Platform for EV Discussion

YouTube has become a vital platform for disseminating information and shaping opinions about new technologies, including electric vehicles. The platform not only allows creators to upload videos about the latest EV models, charging infrastructure, and technological advancements but also provides a space for viewers to engage in discussions through comments. This interaction offers valuable insights into the public's perception of electric vehicles, as well as trends in viewership and content creation.

Research on user interactions and content creation on YouTube regarding EVs is still in its infancy, but the platform provides a rich source of data for understanding how people discuss and react to EV-related topics. Previous studies have shown that online platforms such as Twitter and discussion forums play a significant role in shaping consumer attitudes toward EVs. For example, (Suresha, Tiwari, 2021) used topic modelling to analyze tweets about EVs, finding that "Tesla" was one of the most frequently used hashtags. Similarly, sentiment analysis conducted by (Bhatnagar, Choubey, 2021) indicated that Tesla was viewed more positively compared to other EV manufacturers.

While studies like those (Ha et al., 2021), (Jiang, Everts, 2021), and (Asensio et al., 2008) have analyzed user sentiments and opinions about EVs using natural language processing (NLP) tools, YouTube remains an underexplored platform for understanding consumer engagement with EV-related content. This study aims to fill this gap by examining creator and viewer activity on YouTube, specifically focusing on Polish-language content about electric vehicles. The research explores trends in video publication, user engagement through comments, and the relationships between video characteristics (such as duration and view count) and user interaction. By analyzing the interactions between creators and viewers, this study contributes to a better understanding of how electric vehicles are perceived and discussed in Polish-speaking online communities.

3. Research Methodology

This section outlines the methods and tools used to gather, preprocess, and analyse data from YouTube related to electric vehicles (EVs) in Polish-language content. The methodology covers two main groups of data: videos and comments associated with the selected videos.

3.1. Data Collection

The dataset was collected in two stages, focusing on video data and user comments:

- YouTube video data: A total of 37,317 YouTube video links related to electric vehicles were gathered using Python's scrapetube package (Twersky, n.d.). The search was conducted using 74 phrases associated with electric vehicles, which included both car brand names and model names. These phrases were compiled based on information from an online platform dedicated to EVs (*No Title*, n.d.). Videos that did not contain the search phrase in the title or were not produced by Polish channels were excluded, resulting in a refined list of 1152 videos.
- YouTube comment data: On May 24, 2024, comments for each of these 1152 videos were collected using the Python package youtube-comment-downloader (*Youtube-Comment-Downloader*, n.d.). This tool allowed the retrieval of 128,851 comments without the need for the official YouTube API which allowed for the retrieval of 128,851 comments without the need for the official YouTube API.

3.2. Data Preprocessing

The data was preprocessed to ensure accuracy and relevance for analysis:

- Language filtering: Non-Polish comments were removed from the dataset to focus on Polish-language content.
- Duplicate removal: Repeated comments from the same user, such as advertisements or promotions, were treated as text strings and compared using the '==' operator to identify and remove duplicates.
- Comment filtering: Comments of zero length or those that contained fewer than one word after preprocessing were excluded.
- Content cleaning: Several steps were applied to clean the comment text:
  - Removal of URLs, hashtags, emojis, usernames, and non-letter characters to retain only textual content.
  - Verification of word count for each cleaned comment, and exclusion of comments with fewer than one word.

The final dataset consisted of 125,680 comments associated with 992 videos. Videos that did not have any comments or whose comments became empty after cleaning were excluded.

3.3. Data Transformation

Once the data was preprocessed, several transformations were applied to prepare the dataset for analysis:

- Video data:
  - Video duration, initially formatted as “HH:MM” was transformed into a numerical value representing the total duration in minutes to enable consistent analysis.

  - The publication dates were used to calculate the number of months and weeks since each video's release, which facilitated the calculation of average monthly and weekly views.
- Outlier removal: Extreme outliers in view counts and comment counts, such as videos with unusually high numbers of views or comments, were excluded. For instance, four videos with exceptionally high average monthly views (ranging from 231,894 to 1,682,899 views per month) were removed from specific analyses.
- Statistical transformations: Square root transformations were applied to certain variables (e.g., views, comments) to address skewness and ensure a more linear relationship in the regression models.

3.4. Data Analysis

The analysis consisted of two key approaches:

- Visual Descriptive Analysis: Descriptive visual analysis was used to explore trends in video publication and user engagement, utilizing box plots and bar charts to visually represent the distribution of comments, videos, and user activity over time.
- Regression Analysis: Three linear regression models were applied to examine the relationships between video characteristics and user engagement:
  - Model 1: Relationship between the number of comments and the duration of the video.
  - Model 2: Relationship between the number of comments and the number of views.
  - Model 3: Relationship between the number of views and the duration of the video.

Each model's statistical significance and R-squared values were assessed to determine the proportion of variance explained by the independent variables. Outliers were filtered to improve the models' accuracy.

3.5. Tools Used

- Python: Used for scraping video and comment data from YouTube, specifically with the scrapetube and youtube-comment-downloader packages.
- R programming language: Used for data manipulation, visualisation, and statistical analysis, employing packages like dplyr (Wickham et al., 2023), ggplot2 (Wickham, 2016) and tidyverse (Wickham et al., 2024).
- Microsoft Excel: Used during the data cleaning and preprocessing stages to manually inspect the dataset in a tabular format.

4. Results

Figure 1 presents the number of videos published each month from 2015 to May 24, 2024. The data reveals a steady increase in video publications over the years, with a noticeable peak in 2022, where 295 videos were published. The colours represent different months, showing the distribution of video publications throughout the year. The year 2024, although incomplete, shows a significant decline in the number of videos published compared to the peak years.

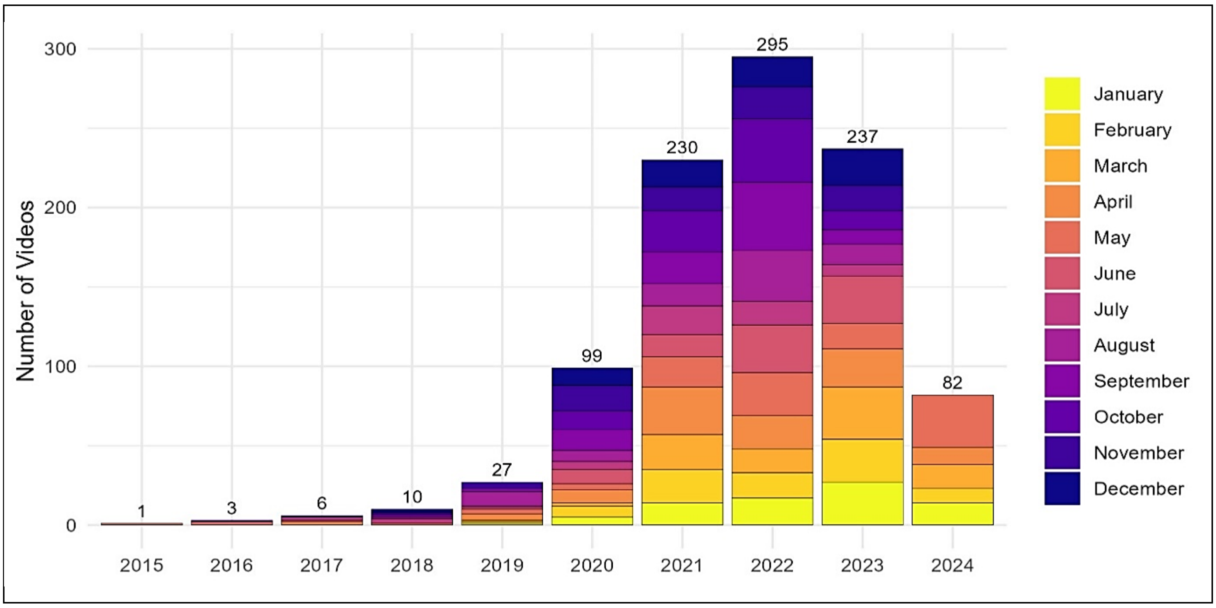

Figure 1. Number of Videos Published per Month in Each Year (data as of May 24, 2024).
Sources: original research.

Figure 2 illustrates the distribution of YouTube videos published by users each year. The box plot displays the median, interquartile range, and outliers for the number of videos published annually. A noticeable increase in publication activity is evident starting around 2019, with a significant rise in 2021 and 2022. The plot also emphasizes the variability in video publication, with 2021 and 2022 showing a wider distribution, suggesting more diverse activity among users during these years.

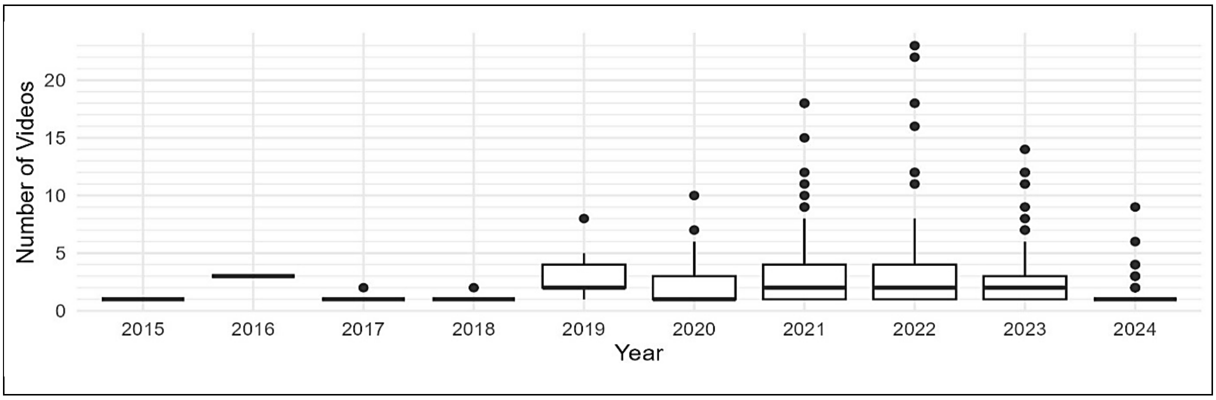

Figure 2. Distribution of YouTube Videos Published by Users per Year (data as of May 24, 2024).
Sources: original research.

Figure 3 shows the number of videos published by users each year, grouped by quantile. Each quantile group represents a different segment of users, categorized based on the number of videos they published in a given year. The chart displays how many videos were published by users in each quantile, with the groups arranged from those who published the fewest videos (quantile 1) to those who published the most (quantile 4). In the chart for the year 2021, it is clear that the largest segment of the bar (representing the 4th quantile) is significantly bigger than the others. This suggests that more than 50% of the videos were published by the top 25% of the most active users (those who published the most videos). Quantile 4, representing these highly active users, dominates the number of videos published that year.

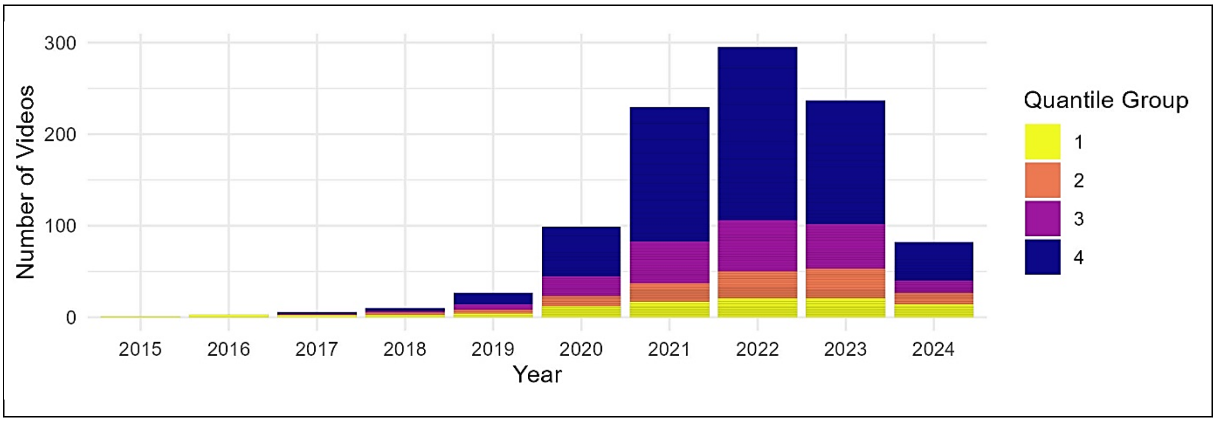

Figure 3. Quantile Distribution of YouTube Videos by Users per Year (data as of May 24, 2024).
Sources: original research.

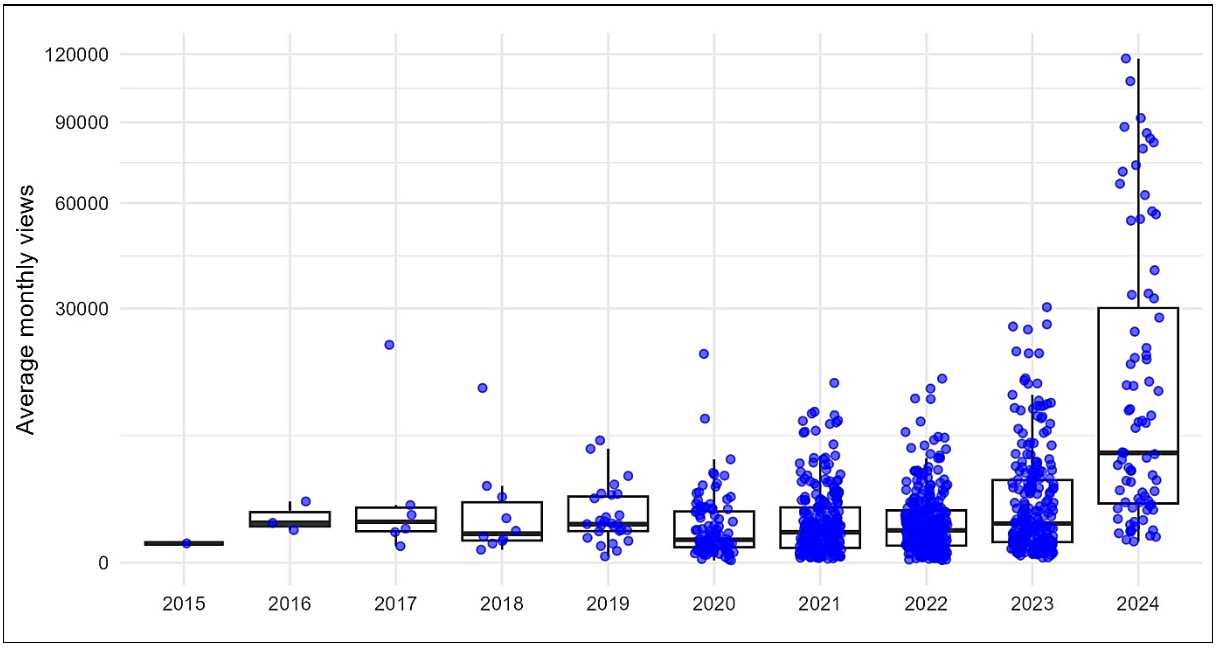

Figure 4. Distribution of average monthly views of videos by year (data as of May 24, 2024).
Sources: original research.

Figure 4 shows the distribution of average monthly views of videos over the years. Each point represents a video and indicates its average monthly views, which were calculated by dividing the total number of views by the number of months from its publication date up to May 24, 2024. This metric reflects the average audience interest per month, providing a more precise assessment of viewing trends over time. The increasing spread and higher median values in recent years suggest growing interest and viewership for the content, particularly in 2024, where a significant rise in average monthly views is evident. This trend highlights increased engagement with these videos in recent years. To ensure a more accurate representation of the overall data, four extreme outliers were excluded from this chart. These excluded videos had exceptionally high average monthly views, with one video from 2023 reaching 231,894 views per month, and three videos from 2024 reaching 346,415.2, 556,472.8, and 1,682,899.1 views per month, respectively. Additionally, the data shows a growing trend in video viewership, with an increase in engagement from 2021 to 2023. This highlights that recent content attracts more attention from viewers.

Figure 5 illustrates the total number of comments made on videos over the years, broken down by month. The vertical bars represent each year from 2015 to 2024, with colours corresponding to the different months of the year. The height of each bar reflects the cumulative number of comments made in that year, revealing an increasing trend in user engagement over time. This trend is especially notable in 2021, 2022, and 2023, which show a significant increase in the number of comments compared to previous years. In 2024, while there is a visible decline, the number of comments still indicates considerable activity despite being a partial year, as data collection stops on May 24, 2024. This chart highlights the increase in user interaction with the video content, showing a peak in activity starting in 2021.

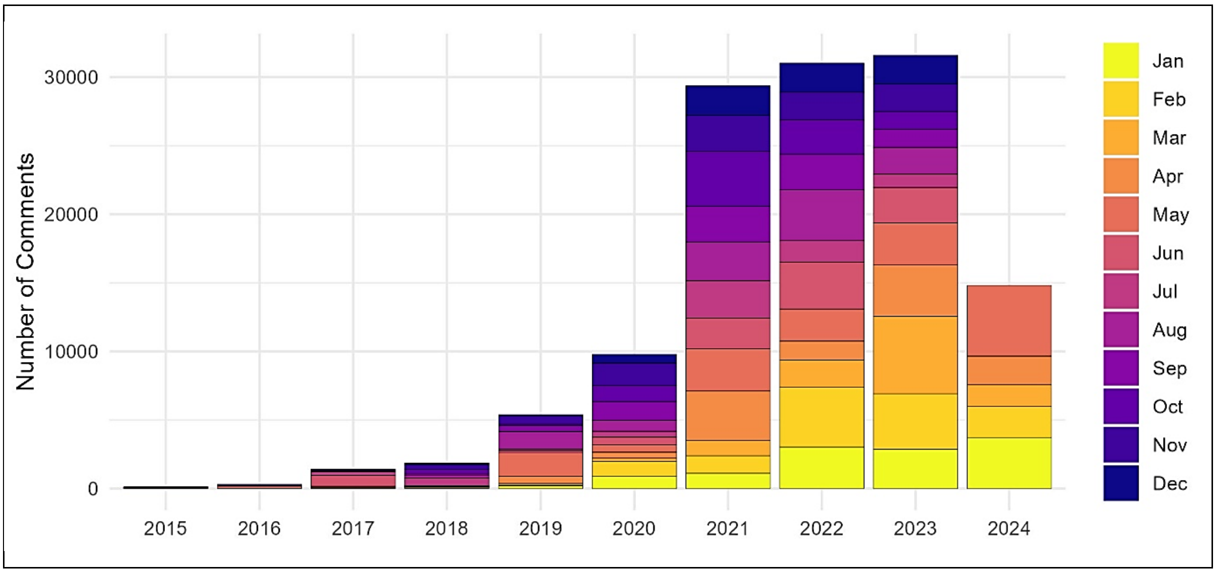

Figure 5. The number of comments.

Sources: original research.

Figure 6 shows the proportion of replies to original comments by year and overall. The high percentage of replies suggests that users are not only posting original comments but are also actively engaging in discussions with one another. This indicates a higher level of interaction and community engagement, as users respond to others' comments and debate among themselves, rather than just reacting to the video content itself. A reply rate of 48% for all comments highlights a strong tendency towards dialogue among users, reflecting a vibrant and interactive community.

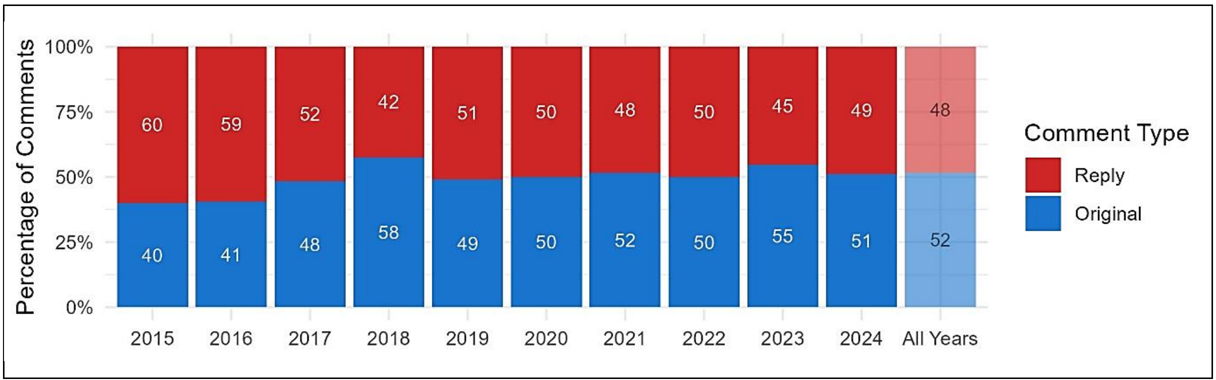

Figure 6. The proportion of Replies to Original Comments (data as of May 24, 2024).

Sources: original research.

Figure 7 depicts the distribution of original comments by the number of replies they have received, grouped into categories. The x-axis represents the number of replies, categorized into groups (1, 2-5, 6-10, 11-20, and 21+), while the y-axis indicates the count of original comments within each group. The chart reveals that a significant portion of original comments (11,240) receive only one reply, followed by 8748 comments that receive between two to five replies. The number of comments significantly decreases as the number of replies increases, with only 1553 comments receiving between six to ten replies, 553 comments receiving between eleven to twenty replies, and just 173 comments receiving more than twenty-one replies. This distribution highlights the pattern of engagement with original comments, showing that while most comments receive some level of interaction, only a few become the centre of more extensive conversations.

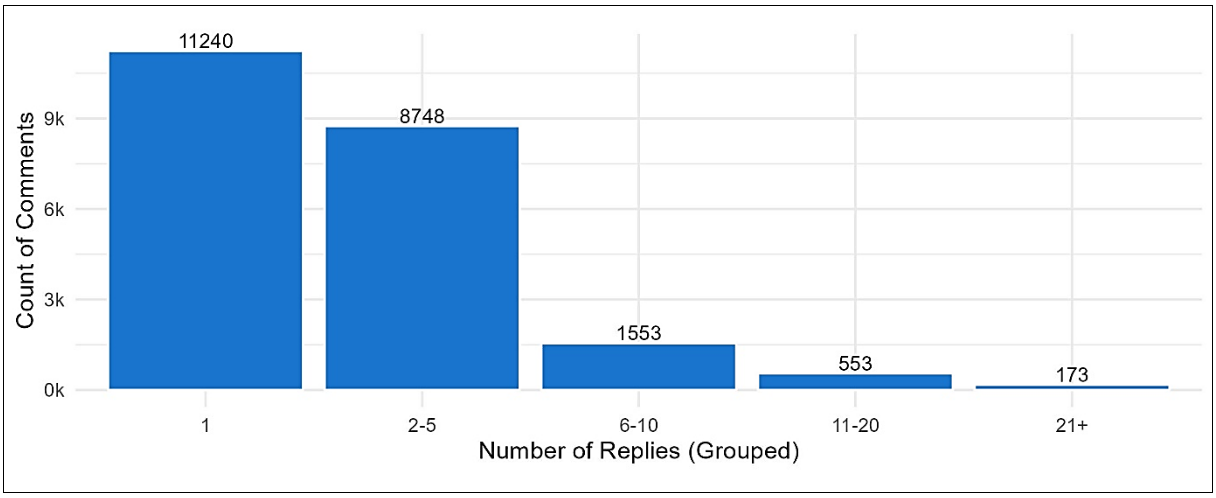

Figure 7. Distribution of Original Comments by Number of Replies (data as of May 24, 2024).
Sources: original research.

Figure 8 illustrates the distribution of the number of comments per video by year, using a boxplot to display the spread of data. The y-axis represents the number of comments and employs a square root scale to better visualize the distribution, especially given the presence of outliers. Each boxplot shows the interquartile range of comments, highlighting the median and spread of data within each year. The use of a square root scale on the y-axis helps to compress the range of data, allowing for clearer observation of the central tendency and spread of comments, particularly in years with high variability. This chart indicates that while most videos have a relatively low number of comments, a few videos each year receive a significantly higher number, demonstrating substantial engagement on select content.

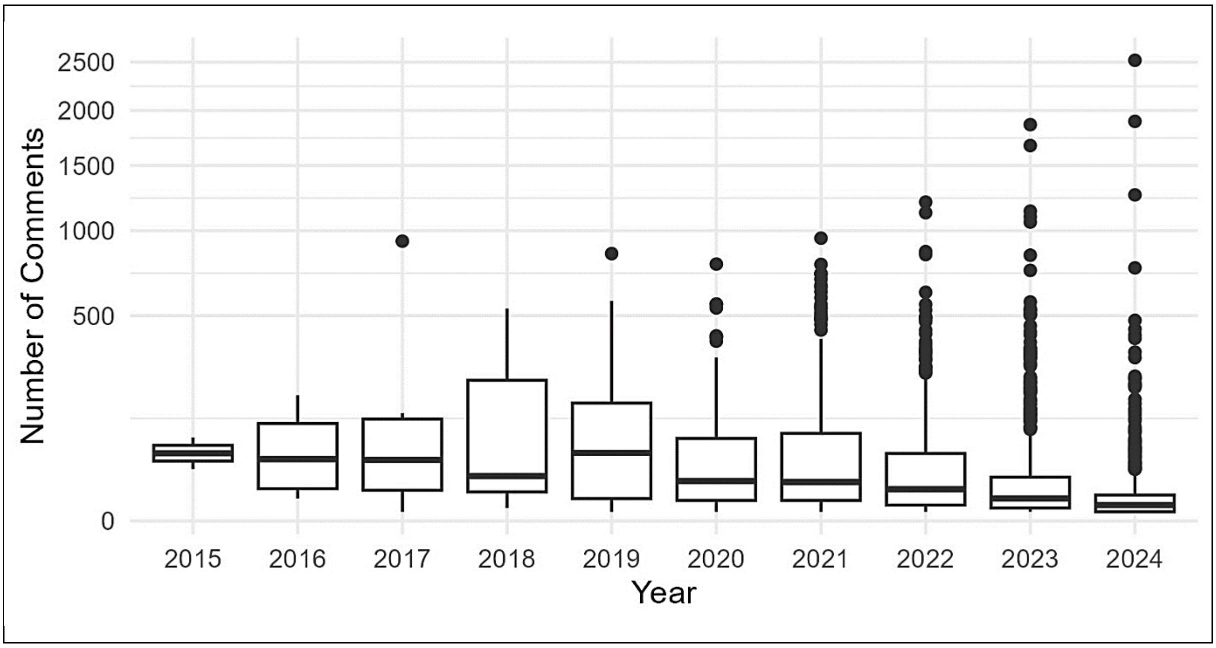

Figure 8. Comments per Video by Year (Data as of May 24, 2024).
Sources: original research.

Figure 9 presents three scatter plots illustrating the relationships between video duration, views, and comments. To better visualise the data, a square root scale has been applied to both the x and y axes. The top left plot shows the relationship between video duration and the number of comments, indicating that longer videos tend to receive more comments, which might suggest that longer content generates higher viewer engagement. The top right plot depicts the relationship between the number of views and the number of comments, showing that videos with a higher number of views generally have more comments. The bottom left plot illustrates the relationship between video duration and the number of views, where no strong correlation is evident, suggesting a weak or negligible relationship between the length of a video and its view count. The points on the scatter plots are colour-coded according to the year of video publication, allowing for an analysis of trends and changes over time. These colours highlight potential differences in user behaviour across different periods.

Table 1 provides a summary of the regression model results based on a dataset containing 990 videos and 125,680 comments. The table includes information on whether each model is statistically significant and the proportion of variance explained by each model.

- Model No. 1: This model investigates the relationship between the number of comments and the duration of the video. The model is statistically significant, with a p-value of 2.91e-08. It explains approximately 3.07% of the variance in the number of comments, as indicated by an R^2 value of 0.0307. This suggests that while the model is significant, video duration only accounts for a small portion of the variability in comment counts.
- Model No. 2: This model examines the relationship between the number of comments and the number of views. The model is highly statistically significant, with a p-value of 1.18e-55. It explains about 22.13% of the variance in the number of comments, as indicated by an R^2 value of 0.2213. This indicates a moderate level of explanatory power for the number of comments based on the number of views.
- Model No. 3: This model explores the relationship between the number of views and the duration of the video. The model is not statistically significant, as reflected by a p-value of 0.365. It explains only 0.08% of the variance in the number of views, as indicated by an R^2 value of 0.0008, demonstrating that video duration has almost no explanatory power for predicting view counts.

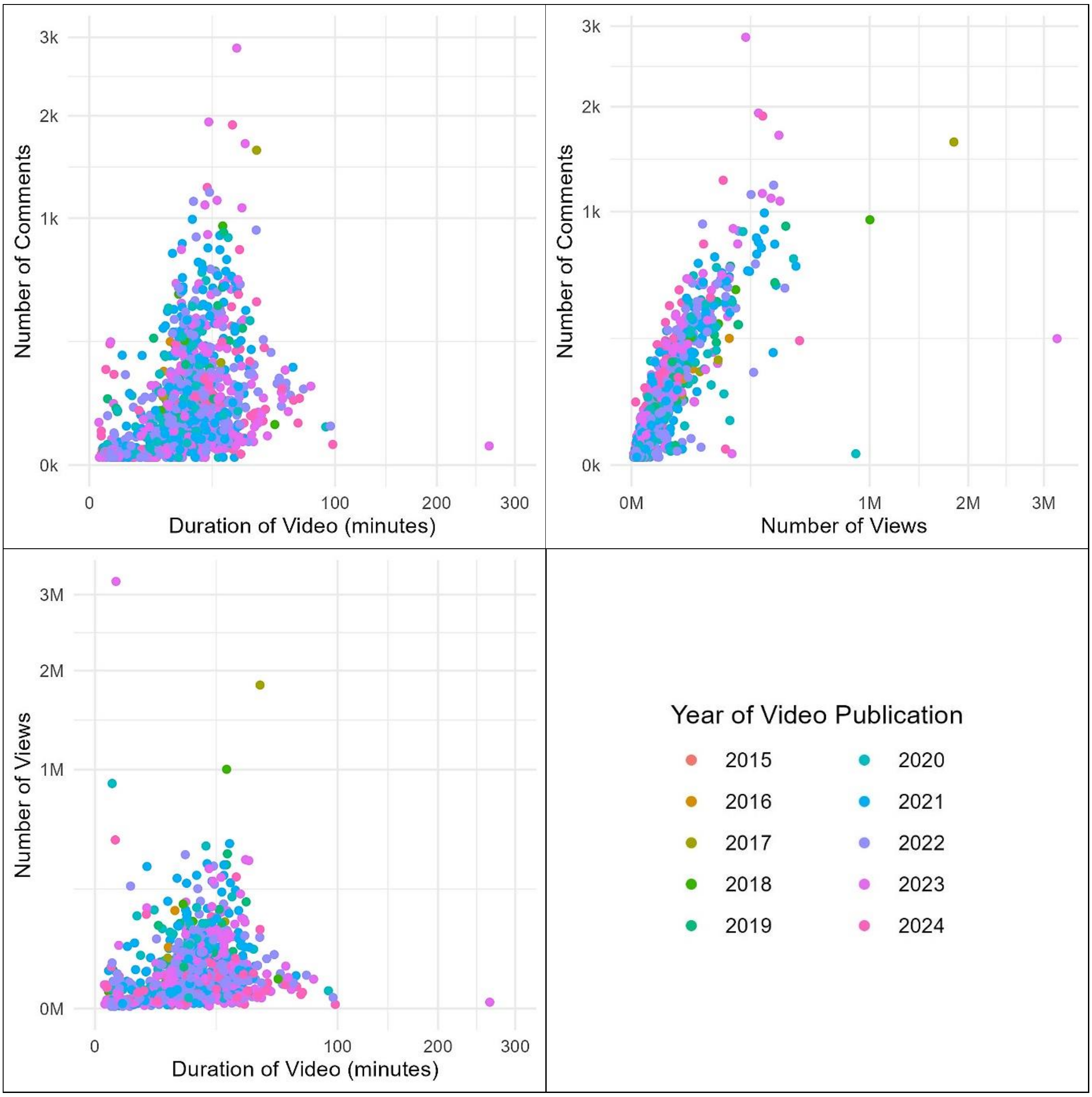

Figure 9. Relationships Between Video Duration, Views, and Comments (Data as of May 24, 2024).

Table 1.
Summary of Regression Model Results

	Model		
	No. 1: Number of Comments ~ Duration of Video	**No. 2: Number of Comments ~ Number of Views**	**No. 3: Number of Views ~ Duration of Video**
Intercept	78,51	94,46	37384,43
Slope	2,5622	0,0008	257,9255
Residual SE	222,94	199,82	138476,36
R squared	0,0307	0,2213	0,0008
p-value	2,91e-08	1,18e-55	3,65e-01
Comments	Low R squared, moderate positive relationship between video duration and comments, statistically significant	Moderate R squared, strong positive relationship between views and comments, highly significant	Very low R squared, no statistically significant relationship between video duration and views

5. Conclusion

This study investigated the activity of content creators and viewers on YouTube, focusing on Polish-language videos related to electric vehicles (EVs). The analysis spanned multiple aspects of user engagement, including video publication trends, comment activity, and the relationship between video characteristics and user interaction.

Several key findings emerged from the analysis:

- Increasing Engagement Over Time: The data revealed a clear upward trend in both the number of videos published and the level of user interaction (measured through comments and views) over the years. Particularly, the years 2021, 2022, and 2023 showed a significant increase in engagement, as reflected by the growing number of comments and views. However, early data from 2024 suggests a decline in activity.
- Dominance of Active Users: The distribution of videos published across quantile groups showed that a small portion of highly active creators (quantile 4) were responsible for the majority of the content. This trend was most evident in recent years, where the top 25% of users contributed more than 50% of the published videos, indicating that a concentrated group of creators is driving most of the content production.
- High User Interaction: The comment analysis demonstrated that a significant portion of user engagement comes from responses to other comments rather than original comments. This trend, particularly evident in recent years, highlights a high level of interaction and discussion among viewers, pointing to the creation of a more engaged community around EV content. Replies accounted for 48% of the total comments, indicating a vibrant environment where users not only react to the videos but also actively engage in discussions.
- Relationships Between Video Characteristics and Engagement: Regression models were used to assess the relationships between video duration, views, and comments. The analysis found statistically significant but relatively weak relationships between video length and comment count, as well as between video views and comments. This suggests that while longer videos and those with higher view counts tend to attract more comments, these factors only explain a small portion of the variability in user engagement. The relationship between video length and views, however, was found to be statistically insignificant, indicating that video duration does not have a significant impact on the number of views.

The findings confirm that while video duration and the number of views are weak predictors of comment activity, they do show that viewer engagement has increased over time. The large proportion of replies to original comments points to a dynamic community, indicating that discussions about electric vehicles are growing in depth and interaction.

References

1. Asensio, O.I., Alvarez, K., Drorc, A., Hollauere, E.W.C. (2008). Evaluating popular sentiment of electric vehicle owners in the United States with real-time data from mobile platforms. *Annali Dell'Istituto Superiore Di Sanita*, *44*(2), 135-144.
2. Bhatnagar, S., Choubey, N. (2021). Making sense of tweets using sentiment analysis on closely related topics. *Social Network Analysis and Mining*, *11*(1), 1-11. https://doi.org/10.1007/S13278-021-00752-0/TABLES/2
3. Buenstorf, G., Cordes, C. (2008). Can sustainable consumption be learned? A model of cultural evolution. *Ecological Economics*, *67*(4), 646-657. https://doi.org/10.1016/J.ECOLECON.2008.01.028
4. Davies, H., Santos, G., Faye, I., Kroon, R., Weken, H. (2016). Establishing the Transferability of Best Practice in EV Policy across EU Borders. *Transportation Research Procedia*, *14*, 2574-2583. https://doi.org/10.1016/J.TRPRO.2016.05.350
5. Figenbaum, E., Assum, T., Kolbenstvedt, M. (2015). Electromobility in Norway: Experiences and Opportunities. *Research in Transportation Economics*, *50*, 29-38. https://doi.org/10.1016/J.RETREC.2015.06.004
6. Ha, S., Marchetto, D.J., Dharur, S., Asensio, O.I. (2021). Topic classification of electric vehicle consumer experiences with transformer-based deep learning. *Patterns*, *2*(2), 100195. https://doi.org/10.1016/j.patter.2020.100195
7. Hannan, M.A., Azidin, F.A., Mohamed, A. (2014). Hybrid electric vehicles and their challenges: A review. *Renewable and Sustainable Energy Reviews*, *29*, 135-150. https://doi.org/10.1016/J.RSER.2013.08.097
8. Hawkins, T.R., Singh, B., Majeau-Bettez, G., Strømman, A.H. (2013). Comparative Environmental Life Cycle Assessment of Conventional and Electric Vehicles. *Journal of Industrial Ecology*, *17*(1), 53-64. https://doi.org/10.1111/J.1530-9290.2012.00532.X
9. Jiang, X., Everts, J. (2021). *Making sense of electrical vehicle discussions using sentiment analysis on closely related news and user comments*. https://arxiv.org/abs/2112.12327v4
10. Ma, H., Balthasar, F., Tait, N., Riera-Palou, X., Harrison, A. (2012). A new comparison between the life cycle greenhouse gas emissions of battery electric vehicles and internal combustion vehicles. *Energy Policy*, *44*, 160-173. https://doi.org/10.1016/J.ENPOL.2012.01.034
11. *No Title* (n.d.).
12. Pasaoglu, G., Honselaar, M., Thiel, C. (2012). Potential vehicle fleet CO2 reductions and cost implications for various vehicle technology deployment scenarios in Europe. *Energy Policy*, *40*(1), 404-421. https://doi.org/10.1016/J.ENPOL.2011.10.025

13. Rahmani, D., Loureiro, M.L. (2018). Why is the market for hybrid electric vehicles (HEVs) moving slowly? *PLOS ONE*, *13(3)*, e0193777. https://doi.org/10.1371/JOURNAL.PONE.0193777
14. Sierzchula, W., Bakker, S., Maat, K., Van Wee, B. (2014). The influence of financial incentives and other socio-economic factors on electric vehicle adoption. *Energy Policy*, *68*, 183-194. https://doi.org/10.1016/J.ENPOL.2014.01.043
15. Suresha, H.P., Tiwari, K.K. (2021). Topic Modeling and Sentiment Analysis of Electric Vehicles of Twitter Data. *Asian Journal of Research in Computer Science*, *12*(2), 13-29. https://doi.org/10.9734/AJRCOS/2021/V12I230278
16. Trip, J.J., Lima, J., Bakker, S. (2012). *Electric mobility policies in the North Sea Region countries. Report for E-Mobility NSR*. https://www.yumpu.com/en/document/view/7530898/electric-mobility-policies-in-the-north-sea-region-e-mobility-nsr
17. Twersky, C. (n.d.). *scrapetube*. Retrieved from: https://pypi.org/project/scrapetube/, June 13, 2023.
18. Van Mierlo, J., Maggetto, G., Lataire, P. (2006). Which energy source for road transport in the future? A comparison of battery, hybrid and fuel cell vehicles. *Energy Conversion and Management*, *47*(17), 2748-2760. https://doi.org/10.1016/J.ENCONMAN.2006.02.004
19. Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag, https://ggplot2.tidyverse.org
20. Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org
21. Wickham, H., Vaughan, D., Girlich, M. (2024). *tidyr: Tidy Messy Data*. https://tidyr.tidyverse.org
22. Wolinetz, M., Axsen, J. (2017). How policy can build the plug-in electric vehicle market: Insights from the REspondent-based Preference And Constraints (REPAC) model. *Technological Forecasting and Social Change*, *117*, 238-250. https://doi.org/10.1016/j.techfore.2016.11.022
23. *youtube-comment-downloader* (n.d.). Retrieved from: https://pypi.org/project/youtube-comment-downloader/, June 13, 2023