DEPARTMENT OF INFORMATICS
OF THE TECHNICAL UNIVERSITY OF MUNICH

Bachelor's Thesis in Informatics

Using Twitter-Data to Analyze Consumer Acceptance of Alternative Fuel Vehicles

Johannes Fuest

DEPARTMENT OF INFORMATICS
OF THE TECHNISCHEN UNIVERSITÄT MÜNCHEN

Bachelor's Thesis in Informatics

Using Twitter-Data to Analyze Consumer Acceptance of Alternative Fuel Vehicles

Verwendung von Twitter-Daten zur Analyse der Verbraucherakzeptanz von Fahrzeugen mit alternativem Kraftstoff

Author:	Johannes Fuest
Supervisor:	Prof. Dr. Helmut Krcmar
Advisor:	Michael Weber
Submission Date:	15. September 2022

I confirm that this bachelor’s thesis in informatics is my own work and that I have documented all sources and materials used.

Munich, 15. September Johannes Fuest

Abstract

This bachelor's thesis conducts an analysis of the sentiment of tweets on alternative fuel vehicles in English and German over the past five years. The motivation behind this is to deepen our understanding of public opinion and consumer acceptance of the underlying technologies. The thesis describes the collection methods used to create databases of English and German tweets on alternative fuel vehicles, as well as their subsequent scoring for base sentiment by natural language processing. We then conduct a set of analyses to discover relationships and trends in the data. The results show that overall sentiment towards alternative fuel vehicles has slightly worsened over the past five years. Our analysis of relevant subtopics discussed in previous academic literature suggests this decline is driven by tweets referencing range and charging, batteries, and costs, with costs being the strongest driver of negativity. Tweets referring to sustainability do not exhibit this negative trend. We also find evidence that tweets with negative sentiment and more polarizing content on Twitter receive higher engagement than moderate tweets. Finally, we also discover a highly negative response of sentiment towards synthetic fuels in German tweets in response to the German government's decision to grant them exceptional status in its recent ban on combustion engines from 2035 onwards.

Keywords: Electric Vehicles, Synthetic Fuels, Twitter, Sentiment Analysis, Mobility, Sustainability

Table of Contents:

List of Figures

List of Tables

1 Introduction

1.1 Motivation

1.1.1 Environmental Crisis and Future Mobility

The global climate crisis requires urgent action across all spectrums of human activity in order to drastically reduce greenhouse gas emissions. To limit global warming to 1.5 degrees Celsius until 2100 and avoid a global catastrophe, the world must reach greenhouse gas neutrality by 2050 and have negative emissions from the onwards (IPCC, 2019). A key sector that will need to undergo major changes in this regard, is the transportation sector. In 2019, transportation alone was responsible for almost a third of US C02 emissions (EPA, 2019).

The modernization and change towards more sustainability of the sector has already begun. Future mobility will be geared towards environment-friendly, automated, and personalized travel on demand (Atasoy et al., 2015). The emergence of Tesla as the world's most valuable automotive producer is just one example of this trend. Both electric vehicles and synthetic fuels are promising technologies for reducing the carbon footprint of the mobility sector. Electric vehicles emit no CO_2 during operation, whilst synthetic fuels can be produces with electricity from renewable energy sources. The positive effect on global emissions these technologies have are likely to increase dramatically over time (Slowik and Lutsey, 2015).

The success of this shift in mobility will depend in large part on consumer willingness to adapt to the changes the associated new technologies bring. Consumer resistance to new technologies is generally one of the biggest barriers to their success (Egbue and Long, 2012). In the absence of favorable public opinion of these technologies, the adoption of electric vehicles and synthetic fuels is unlikely to succeed, which would mean losing the potential climate benefits associated with their use. A deeper understanding of the drivers of public opinion on these technologies would thus be a helpful contribution towards the fight against global climate change. This thesis will attempt to advance our understanding of public opinion on electric vehicles and synthetic fuels through an analysis of tweets posted on the matter.

1.1.1 Social Media and Public Opinion

Prior research indicates that social media is one of the richest sources of public opinion and thus ideal for measuring consumer acceptance (Kohl et al., 2018). Due to high user numbers, international spread, as well as ease of access, it offers multiple advantages as a source for gaging consumer sentiment. The first is that social media portals such as twitter are publicly accessible. Millions of publicly available tweets are generated daily, many of them containing user opinions on current issues, including the future of mobility. Secondly, due to its low costs, the user base of social media sites, such as Twitter, includes individuals across the entire soci-oeconomic and political spectrum (Duggan, Maeve and Brenner, 2012). This thesis will make use of these advantages and use more data collection and natural language processing methods to answer the research questions outlined in the next section.

1.2 Research Questions

Over the course of this thesis, we will answer three main questions, which are described below.

1.2.1 What is the current state of research on consumer acceptance of electric vehicles and synthetic fuels?

This question will be used to set a starting point for the analysis of consumer acceptance of electric vehicles and synthetic fuels. It will establish what prior research has been done on electric vehicle and synthetic fuel acceptance and describe which factors have already been identified as key drivers behind it. The resulting findings will be used further on to form a basis for comparison against the results from the analysis of the social media gathered during later stages of the project. This will allow for a better assessment of the contribution of this thesis to the literature on consumer acceptance of electric vehicles and synthetic fuels.

1.2.2 What Twitter data on consumer acceptance of battery electric vehicles and synthetic fuels can be collected and analyzed systematically?

This question will be used to develop a data collection and data mining approach for Twitter data on electric vehicles and synthetic fuels. Through web-scraping, natural language processing, and data mining, we will create a dataset of significant quantity of tweets on electric vehicles and synthetic fuels in both German and English. In creating this dataset, we will make choices regarding the tools used for gathering the data, the searching mechanisms used for locating the data, such as hashtags, as well as the techniques used for analyzing the data, such as sentiment analysis.

1.2.3 Which conclusions can be derived from the trends and key factors identified in the analysis of the collected data?

This question will be used as a guide throughout the analysis of the generated data. We will look to discover any trends and associations between the data we collect that might be relevant to any stakeholders in the alternative fuel vehicle landscape. There are many potential ways that trends or other insights from the analysis may be relevant. Firstly, any factors identified to have a significant impact on consumer sentiment towards electric vehicles or synthetic fuels could be used by policymakers to make advancements in sustainability. Knowing which issues are most likely to trigger negative sentiment or skepticism among consumers would allow a better evaluation of different policies. Secondly, producers of electric vehicles and synthetic fuels could benefit from a more nuanced understanding of consumer acceptance of electric vehicles and synthetic fuels by adjusting their strategies to address customer sentiment more effectively.

1.3 Outline

This chapter will give a brief outline of the rest of this thesis and explain each of the upcoming chapters.

In chapter two, we will provide some context and background on topics relevant to this thesis. We will briefly cover the definitions of electric vehicles and synthetic fuels, as well as providing some context on their current perception in society and politics. We will also give some more background information and context on the social network Twitter, which will serve as the data source for our analysis. Finally, chapter two will also cover the current state of academic literature on consumer acceptance of electric vehicles and synthetic fuels.

In chapter three, we will describe the methods of our analysis. This will cover the tools and techniques used for both data collection and data analysis. For data collection, we will describe Twitter's official API, as well as the python library tweepy we will use to manage the API calls. We will also briefly describe the python library pandas that we will use to manipulate and store the data gathered from Twitter. For data analysis we will offer a basic explanation of the base sentiment analysis models we will use. We will also explain the method of relative word frequency detection for mining events and give an overview of a range of subtopics within our databases that we will examine. Finally, we will also describe the relationships between various variables in our data and the trends that we will examine throughout the observed period.

In chapter four we will then present and describe the outcomes of the methods described in chapter three. The results will be broken up into several sections. First, we will cover the central, high-level results found in our main data-basis include. After this, we will cover the events found by our relative word frequency analysis, before analyzing notable relationships between sentiment and various variables found in the data. Finally, we will also describe the result of our subtopic analysis and compare this to our main findings.

In chapter five we will then discuss the results from the prior chapter and offer explanations and interpretations for our observations. We will discuss each section of the results chapter separately and address the significance of the results for our previously outlined research questions. Chapter 6 will then conclude this thesis by giving a summary of our main findings and recapitulating the steps taken to reach them.

2 Background

2.1 Electric Vehicles

Electric vehicles are vehicles whose main method of propulsion is through electric motors. The source of the electricity for an electric vehicle may come from various sources. Some electrical vehicles run on solar panels or similar collector systems, while most run on batteries charged with extravehicular electricity.

The first electric car was most likely invented by Robert Anderson some time between 1832 and 1839 (Roth, 2011). In the 19th and early 20th century, electric cars were somewhat popular, as they were originally more comfortable and easier to operate than internal combustion engine cars of the time (Encyclopedia Britannica, 2014). At their earliest peak, the global electric car fleet comprised around 30000 vehicles (Gerdes, 2012). Over the next 100 years however, the global auto industry was heavily dominated by internal combustion engine (ICE) cars, as lower costs through mass production and much quicker refueling times made them more popular than their electric counterparts (Gerdes, 2012).

Over the past 30 years however, internal combustion engine vehicles have come under increasing scrutiny, due to their significant contribution to the global climate crisis. Combustion engine cars emit greenhouse gasses, such as CO2, when burning the fuel required to run them. These emissions making them non climate friendly. Consequently, governments around the world have been increasingly leaning towards promoting electric vehicles over their more polluting counterparts, by using subsidies and other economic leavers (Jin et al., 2021). Coupled with advances in battery technologies, this has led to a new advent of electric cars around the world (Sandalow, 2009). The past 10 years alone have seen extreme growth in the sales numbers of electric vehicles. In all of 2012 130 000 electric cars were sold globally. In 2021, a mere 9 years later, this number was at 6.6 million, representing a more than 50-fold increase (Paoli and Gül, 2022).

In terms of costs, electric vehicles have also made significant progress over the past decades. The price per kWh of electric car batteries decreased from €605 per kWh in 2010, to €100 in 2019 (Hauri, 2019) and the total costs of EV ownership are now often lower than those of internal combustion engine vehicles due to lower maintenance and fueling costs (Shahan, 2020). The overall purchase price of an electric vehicle is expected to fall below that of a new ICE car when battery costs fall below US$100 per kWh, which is expected to happen in the mid-2020s (Hindustan Times, 2020).

Overall, electric vehicles are currently becoming cheaper and more popular at a high velocity, whilst also benefitting from broad political support, due to their zero emissions.

2.2 Synthetic Fuels

Synthetic fuels differ from conventional fuels in their production method. The term ‘synthetic’ refers to the production of these fuels, which is artificial and man-made, rather than extracted from nature. Synthetic fuels cover a wide spectrum of fuels with different origins and purposes, ranging from use in cars to rocket fuels, such as Hydrazine. There is no exact definition or clear categorization for these fuel types. When discussed in the context of mobility, synthetic fuels are often proposed as an alternative source of energy for cars that might make them more sustainable. Synthetic fuels enriched from sources relying on renewable energy could theoretically allow for internal combustion engine cars to achieve greater sustainability, paving a way for them to stay relevant in times of greater sensitivity to sustainability-related issues (Royal Society, 2019). Many automakers, including Porsche and Audi have recently made investments into synthetic fuels (Patrascu, 2020) (Audi MediaCenter, 2021).

2.3 Twitter

Twitter is a California-based microblogging site that was launched in 2006. The site allows its users to post media or short texts limited to 240 characters, as well as the options of replying to, liking or retweeting the content posted by others. Account creation is anonymous and requires only an email. Users can choose to follow others and thus create a personalized newsfeed consisting only of tweets posted by followed accounts.

According to its own Q4 2019 report, in 2019, Twitter had more than 200 million active users, posting more than 500 million tweets per day. The high number of users and simplicity of the site make it a powerful tool for gauging public opinion on selected topics. This is not without limitations however, as Twitter is an imperfect representation of the general public. A 2019 survey conducted by the Pew Foundation found that Twitter users are more likely to have a college degree than the general public and also have higher income than the average U.S. adult. The survey also suggested a higher rate of Twitter usage among Democrats than Republicans, showing an imperfect representation of the political public. In terms of age demographic, Twitter users are also far younger than the general public. The survey further found that 10% of users on Twitter create 80% of all tweets, further skewing the representation (Wojcik and Hughes, 2019).

A further limitation of Twitter as a reliable information source are bot-accounts. Twitter bots are accounts that are automatically programmed to tweet, retweet, and follow other accounts. The existence of Twitter's public API makes the presence of bots on the site possible. Whilst some bots can generate creative content and relevant product updates, many are malicious in nature and confound the twitter landscape by spreading misinformation, spam or slander (Miners, 2014) Bots amass significant influence and have been noted to sway elections, influence the stock market, public appeal, and attack governments (Urbina, 2013). It is therefore not unlikely that the landscape of tweets on electric vehicles over the past 5 years has been influenced heavily by bots.

Despite all of these flaws however, twitter remains ideal for mining public opinions (Mumtaz and Ahuja,) he aggregated sentiment expressed by twitter users often has predictive power on current events and correlates with traditional measures for assessing public opinion (Cody et al., 2016). This shows that Twitter remains a powerful source of mining public opinion, even with imperfect representation of the general public and the presence of bots.

2.4 Consumer Acceptance of Electric Vehicles and Synthetic Fuels

While the technology behind electric vehicles is surprisingly old, the advancements that are now making it available and viable for mass-production are new and consumer acceptance will play a large role in scope and speed of the uptake of the resulting new wave of electric vehicles.

Traditional approaches for explaining technology acceptance center around the technology acceptance model (TAM) (Davis, 1980, which explains technology uptake using the technology's perceived ease of use, as well as its perceived usefulness. While several studies find this basic explanation to still hold for new technologies to this day (Müller, 2019), the exact quantification of perceived ease of use and perceived usability is complicated and comprised of individually noteworthy components, such as costs, range, charging times, environmental concerns and other issues.
How exactly these factors impact consumer acceptance of electric vehicles is the subject of much debate in the academic literature. Some surveys find costs and infrastructure to be of the highest relevance (Linzenich et al., 2019) (Liao, Molin and van Wee, 2017) (Wicki, Brückmann and Bernauer, 2022), whereas others deem personal identity and environmental attitudes more crucial (Rezvani, Jansson and Bodin, 2015) (Wang, Cao and Zhang, 2021). There have also been attempts to identify distinct customer segments for electric vehicles (Higueras-Castillo et al., 2020) (Ferguson et al., 2018) with mixed results. A recent study examining 94 research papers on BEV acceptance between 2010 and 2019 found that many of the examined studies are based on unavailable data and are thus non-replicable. In cases where data is available, the studies on facilitators and obstacles are ambiguous and often contradictory (Wicki et al., 2022). Overall, there is a vague consensus that acceptance is driven by psychological and economic factors, but this appears difficult to quantify and formalize.

For synthetic fuels, the literature is far more sporadic. Given the much greater role electric vehicles currently play in the debate around sustainability of transportation, this is not surprising. A Forsa survey from 2021 revealed that among German consumers, 82% would be happy to use e-fuels (Forsa, 2021), while a FuelsEurope Survey of citizens of ten European countries showed that 69% agree that alternative fuels could be an affordable and efficient solution to reduce emissions (FuelsEurope, 2019). It thus appears that there is a general willingness among consumers to consider e-fuels under the right circumstances. Given the lack of concrete synthetic fuel options currently available to consumers, it is unclear how this sentiment might translate to acute purchasing decisions in the future.

Overall, there is currently a significant gap in the academic literature explaining consumer acceptance and consumer sentiment towards electric vehicles and synthetic fuels. There is no firm consensus on the defining underlying factors that drive acceptance of electric vehicles.

3 Methods

This chapter outlines the methods used to conduct our data collection and analysis. It details the process and tools used during data collection to create four main databases on electric cars and synthetic fuels in German and English respectively. It then describes the analyses carried out on those databases, spanning relative word frequency, sentiment analysis and subtopics examined.

3.1 Data Collection

To start our analysis, we will be collecting tweets on electric vehicles and synthetic fuels. The architecture for our data collection is to access the official Twitter API through the open-source python library 'Tweepy' and then save the pulled data into csv files using the python pandas library. This method and its building blocks are described in the upcoming sections in more detail. The end result of this data collection method will be four csv files, containing all tweets on electric vehicles and synthetic fuels in German and English that were posted over the past five years. For every tweet, the tables will contain the publication timestamp, positive, negative, and neutral sentiment probabilities, number of likes, comments, and retweets, author follower and following counts, as well as the total number of posted tweets by the author.

3.1.1 Twitter API

The Twitter API is Twitter's official, publicly accessible application programming interface. It offers a host of features and access points, including for recent tweets, full-archive tweets, and user data. Its features are however limited depending on access level. Given the academic nature of this project, it uses the academic research access tier, which is currently the least restrictive available tier, offering its users the ability to retrieve up to ten million archived tweets per month with no limit on how far back in time the requests go.

For this thesis, we seek both tweet data, such as the tweets text, its author id, likes, comments and retweets generated, as well as the user data of the tweets' authors, such as follower counts, age of account and total tweets posted. For these two types of data, the twitter API offers two different endpoints. The results of requests to these endpoints must then be combined to get the complete dataset required for the project.

The first endpoint is the full-archive search endpoint, which supplies data on the full archive of tweets since March 2006. For every request a user makes, the endpoint can return data on 100 tweets. It offers pagination token functionality for searches resulting in larger volumes of tweets, which can then be served across multiple requests. The total number of tweets that can be obtained from this endpoint per month per user is ten million. The data returned by this endpoint covers the author's id, the tweet text, the exact timestamp of its publication, as well as the numbers of likes, comments, and retweet.

The users lookup endpoint meanwhile, offers the API client the ability to look up data on public profiles of twitter users from either their user handle or user id. This endpoint returns data on users in increments of 100 users per requests, including timestamp of the account creation, verification status, follower and following count as well as number of total tweets posted by the account. The number of requests for this endpoint is currently limited to 75 per 15-minute window, regardless of the client's access level.

3.1.2 Tweepy and Pandas

Tweepy is an open-source python library that drastically simplifies interactions for a client with the Twitter API. It covers all of twitter's RESTful API methods, including the users lookup and full-archive search endpoints. Using its client class, the library allows programmers to easily generate their API requests and format them as a single string including search terms and filters for language. It also handles all authentication aspects of the API and allows for easy modification of the observed timespan of tweets searched. The library also has a custom response class which allows for easy parsing of the Twitter API's responses.
For the data storage and conversion of the responses, we use pandas, a python library for handling and manipulating datasets. Pandas' dataframe class allows for easy manipulation of the data, data table merging, as well as permanent storage of the data as .csv files.

With these libraries, we build four search queries directed at the full-archive search endpoint to give us tweets on electric vehicles and synthetic fuels over the past five years in German and English respectively. The table below details the search terms included in the queries. Note that the Twitter API works with exact word matching, which necessitates specifying search terms in singular and plural form. The keyword search is not case-sensitive. Given that many hashtags use English terms even when the rest of the tweet is German, English search terms are also included in the German tweet searches. The requests are further specified to exclude retweets, as these offer little additional value for analysis, but still take up request capacity in the Twitter API's monthly cap of ten million tweets. We also exclude replies for capacity reasons. The cap of ten million tweets also necessitates a restriction to the past five years only.

Database	**Search Terms**
Electric Vehicles (English)	electric vehicles, electric vehicle, electric cars, electric car, e-cars, e-car, ecars, ecar
Electric Vehicles (German)	electric vehicles, electric vehicle, electric cars, electric car, e-cars, e-car, ecars, ecar, e-autos, e-auto, eautos, eauto , efahrzeuge, efahrzeug, e-fahrzeuge, e-fahrzeug, elektroautos, elektroauto
Synthetic Fuels (English)	synthetic fuels, synthetic fuel, e-fuels, e-fuel, efuels, efuel, synfuels, synfuel, artificial fuels, artifical fuel
Synthetic Fuels (German)	synthetic fuels, synthetic fuel, e-fuels, e-fuel, efuels, efuel, synfuels, synfuel, artificial fuels, artifical fuel, e-treibstoffe, e-treibstoff, etreibstoffe, etreibstoff, ekraftstoffe, ekraftstoff, e-kraftstoffe, e-kraftstoff, synthetische kraftstoffe, synthetischer kraftstoff, synthetischem kraftstoff

Table 1: Search Terms Used for Database Creation

The resulting requests are then sent to the full-archive search endpoint. For each request, the Twitter API sends back the desired data on a hundred tweets, as well as a pagination token. The one hundred tweets and their accompanying data are passed into a pandas dataframe, while the pagination token is used to send out the next request. In order for the resulting .csv file to be readable later on, commas, newlines and other punctuation are removed from the tweet texts. This process continues until we had collected every tweet in the desired category over the past 5 years. Once the dataframes are complete, we use pandas' .to_csv function to permanently save the collected tweets as .csv files.

The next part of the process is to collect all user ids in the previously created databases. This is done by launching queries to the users lookup endpoint of the Twitter API. The users lookup endpoint takes requests of up to 100 user IDs and returns the user account creation date, their amounts of followers and followed accounts, as well as their total tweet count and whether or not they were verified. We generate the requests to this endpoint in several steps. First, we create a unique list of user IDs from the data earlier pulled from the full archive search endpoint. We then divide this list into chunks containing 100 user IDs. For each such chunk, we then send a request to the users lookup endpoint and iteratively add the results parsed from the reply to a pandas dataframe. Once all user data has been gathered, we again save the data to a .csv file as our user database.

In the final data collection step, we use pandas' merge function to add the user details to our previously created databases of tweets, which are then ready for analysis.

3.2 Data Analysis

In order to extract new insights from the assembled Twitter data, we conduct two main analyses of all tweets. First, for every word in the databases, we calculate its relative appearance frequency on a monthly basis. The goal of this is to identify trending topics or events that might explain observed changes in sentiment over time. Secondly, all tweets are analyzed using sentiment analysis models, which return percentage estimates on whether the model rates a tweet's base sentiment to be positive, neutral, or negative. We then use the generated sentiment data to see if we can find any strong relationships between tweet sentiment and other factors in the data. For the databases on electric vehicles, we further inspect subsets of the databases that only include tweets discussing the subtopics of batteries, range or charging, costs, and sustainability respectively. This section outlines the details of these analyses.

3.2.1 Relative Word Frequency

The first part of the analysis aims to understand which words appeared unusually often each month, in order to extract significant events that could have impacted sentiment towards the examined technologies on Twitter during that month.

To find the most relevant words per month we take several steps. We first count the total number of appearances of each word across all tweets in our database. We ensure that each word receives exactly one frequency by removing all punctuation and transforming all tweets to lower case. For every word found in any tweet of the database, we thus receive the exact number of times it appears in the database. We then divide this number by the total number of tweets in the database, yielding each word's overall frequency.

$$Overall\ Word\ Frequency = \frac{Total\ Appearances\ of\ Word\ in\ Database}{Total\ Number\ of\ Tweets\ in\ Database}$$

After this, we aggregate the data by month, with each month's row in the resulting dataset containing a concatenation of all tweet texts for that month. For each month we then repeat the aforementioned process, again counting how often each word appears per tweet and dividing this by the number of tweets in that month. For every word appearing anywhere in our database, this yields a word frequency in every month of the five-year period covered.

$$Monthly\ Word\ Frequency = \frac{Total\ Appearances\ of\ Word\ in\ Month}{Total\ Tweets\ in\ Month}$$

We then divide these monthly appearance counts per tweet for every word in every month by the word's average appearance per tweet across the whole database. The resulting number describes how many more times on average that word appears during the month it refers to, compared to how often it appeared on average across all months in our observations. We call this number the relative word frequency.

$$Relative\ Word\ Frequency = \frac{Monthly\ Word\ Frequency}{Overall\ Word\ Frequency}$$

After this, we take the twenty words with the highest positive relative frequency for each month under the condition that it appears at least a minimum amount of time. The minimum threshold is set to avoid outliers, such as misspelled words, which only appeared a handful of times in total. The threshold for this is adjusted for each database according to its overall size. For each of these words we then check whether they give clues to notable events or topics on their respective database in the month they appear frequently. This yields a timeline of the most-discussed topics for each database per month and will allow a better explanation of what is driving the observed sentiment changes over time.

3.2.2 Sentiment Analysis

Sentiment Analysis describes the computational treatment of opinions, sentiments, and subjectivity of a text (Medhat, Hassan and Korashy, 2014). Twitter and other microblogging sites are particularly suitable for this type of analysis due to their user's habit of regularly expressing their personal opinions on these sites (Pak and Paroubek, 2010). Sentiment analysis models are a class of natural language processing models that take text as input and output a classification or probability score that mathematically describes the sentiment of the text input. This output can then be used to monitor changes in sentiment over time and quantify effects of other variables on sentiment.

Twitter text poses some unique challenges for such models, as the user-generated, conversational, and idiosyncratic text typically seen in social media content is very different to other text forms, such as news articles or books (Hussein, 2018). For this reason, this thesis relies on sentiment analysis models designed specifically for Twitter data. The first was developed by the University of Cardiff and called twitter-roberta-base-sentiment (Barbieri et al. 2020). It was pre-trained on over 100 million tweets. This model takes text in English as input and outputs probabilities of the text's underlying sentiment being positive, negative, or neutral respectively. For example, the text "Covid cases are increasing fast!" would output probabilities of 72.36% for negative, 22.87% for neutral and 4.77% for positive sentiment. For German Tweets we use a multilingual version of the same model, which was published a year later (Barbieri et al., 2021). Both sentiment analysis models are based on Google's BERT NLP approach (Devlin et al., 2018) and are available as part of the python transformers library provided by the hugging-face website (Wolf et al., 2019). By using this library, we can comfortably download both models and apply them to every row in our databases and add the three generated scores as new columns, labeled positive, neutral and negative respectively.

Given the computational intensity of these models and the high number of rows in our data, this process is too time-consuming to be performed on a personal laptop within the timeframe of this thesis. Because of this, the python code of this part of the project is deployed on a cloud server with more computational power. Furthermore, the process is parallelized by breaking up the data into sections that the models could work on simultaneously using python's Thread library. Due to ease of use and extensive documentation of the service, we choose an AWS EC2 Ubuntu instance to perform the required tasks in the cloud. The version chosen is Amazon's c6a.xlarge instance which offers 32vCPUs and enough storage to handle the data.

3.2.3 Examined Relationships

With the sentiment analysis complete, we then explore several associations between the variables in our tables and map these out on charts using matplotlib.

The first examined relationship is between engagement percentile and sentiment. For this we rank tweets based on their percentile of likes, comments, and retweets and round the percentile of the rank to three decimal places. We then plot this against the average sentiment for each percentile. The same process is repeated for neutrality. In a similar manner, we also examine the relationship between follower percentile and sentiment, as well as user activity percentile and sentiment. User activity is measured by the total number of tweets a user has posted on their account.

For each of these relationships we further calculate the correlation coefficient, as well as the p-value, in order to interpret the statistical significance of the observed relationship. The correlation coefficient describes the magnitude of the relationship between the two variables, whereas the p-value describes the likelihood of observing the encountered data in a sample under the assumption that underlying population correlation is zero. A very low p-value therefore informs us that encountering the observed data under the assumption that the two variables have no underlying correlated relationship is low.

3.2.4 Subtopic Analysis

For the large electric vehicle databases, we take a step further in our analysis and filter the databases on four categories based on content of the tweets. The categories are batteries, costs, range and sustainability. We perform this filtering through a simple search of the search terms in the table below, using python's substring method.

Topic	**Search Terms**
Battery	'battery', 'batterie'
Range/Charging	'sustainabl', 'cO2', 'sustainability', 'global warming', 'climate change', 'klima', 'nachhaltig', 'erderwärmung', 'umwelt'
Costs	'range', 'ranges', 'distance', 'reichweite', 'fahrtweite', 'load', 'charging', 'charge', 'infrastructure', 'lade', 'infrastruktur', 'säule', 'wirkungsgrad', 'effizienz', 'effic', 'energy', 'energie'
Sustainability	'range', 'ranges', 'distance', 'reichweite', 'fahrtweite', 'load', 'charging', 'charge', 'infrastructure', 'lade', 'infrastruktur', 'säule', 'wirkungsgrad', 'effizienz', 'effic', 'energy', 'energie'

Table 2: Search Terms Used for Subtopic Filtering

For each of the thus generated databases, we again observe sentiment over time, tweet volumes over time and the relationships between various tweet characteristics and sentiment as outlined in section 3.2.3.

Analyzing these subtopics will lend greater context to the picture painted by our analysis of the EV databases as a whole. It may be the case, for example, that we see a significant improvement in sentiment over time in the entire database. Being able to see the development of sentiment of tweets for each subtopic in the database over time will allow us to see how sentient on the subtopics are most contributing to this change. It may be the case that the significant improvement is driven by only one or multiple subsets and it may be also be the case that despite an overall improvement in sentiment, one subtopic is moving in the opposite direction. It may also be the case that overall sentiment is very stable, but sentiments on subtopics are exhibiting significant trends that are simply cancelling each other out. Having the sentiment on subtopics at hand therefore greatly enriches the original analysis.

4 Results

4.1 High-Level Observations

4.1.1 Tweet Volumes

Database Description	Database Name	Number of Tweets
EV English	EV_EN_FINAL	3.580.873
EV battery-related English	EV_EN_BATTERY	214.680
EV cost-related English	EV_EN_COSTS	208.558
EV sustainability-related English	EV_EN_SUSTAINABILITY	52.553
EV range/charging-related English	EV_EN_RANGE	457.633
EV German	EV_DE_FINAL	412.415
EV battery-related German	EV_DE_BATTERY	2.993
EV cost-related German	EV_DE_COSTS	25.466
EV sustainability-related German	EV_DE_SUSTAINABILITY	10.631
EV range/charging-related German	EV_DE_RANGE	34.598
Synthetic Fuels English	SF_EN_FINAL	28.012
Synthetic Fuels German	SF_DE_FINAL	13.193

Table 3: Total Tweet Volumes of Generated Databases

The tweet volumes collected in total for our four databases are detailed in table 3. Unsurprisingly, English databases include more tweets than their German counterparts, although this effect is less noticeable for synthetic fuels, suggesting that these are relatively more discussed in German tweets.

The development of these volumes over time is charted in Fig. 1. Two things are most noticeable here. Firstly, for English tweets, conversation levels on Twitter around electric vehicles declined sharply in early 2020 but have been increasing ever since. For German Tweets, this. Has been different, with conversation levels strongly fluctuating around an average of 6500 tweets per month. For both databases, the lowest tweet volumes were recorded in early 2020, which is likely due to the outbreak of the coronavirus pandemic dominating the conversation on Twitter during those months. Secondly, there has been a relatively steady increase in tweet volumes on synthetic fuels for both German and English tweets. For German Tweets especially, there has been an explosion in tweet volume on synthetic fuels over the past two months.

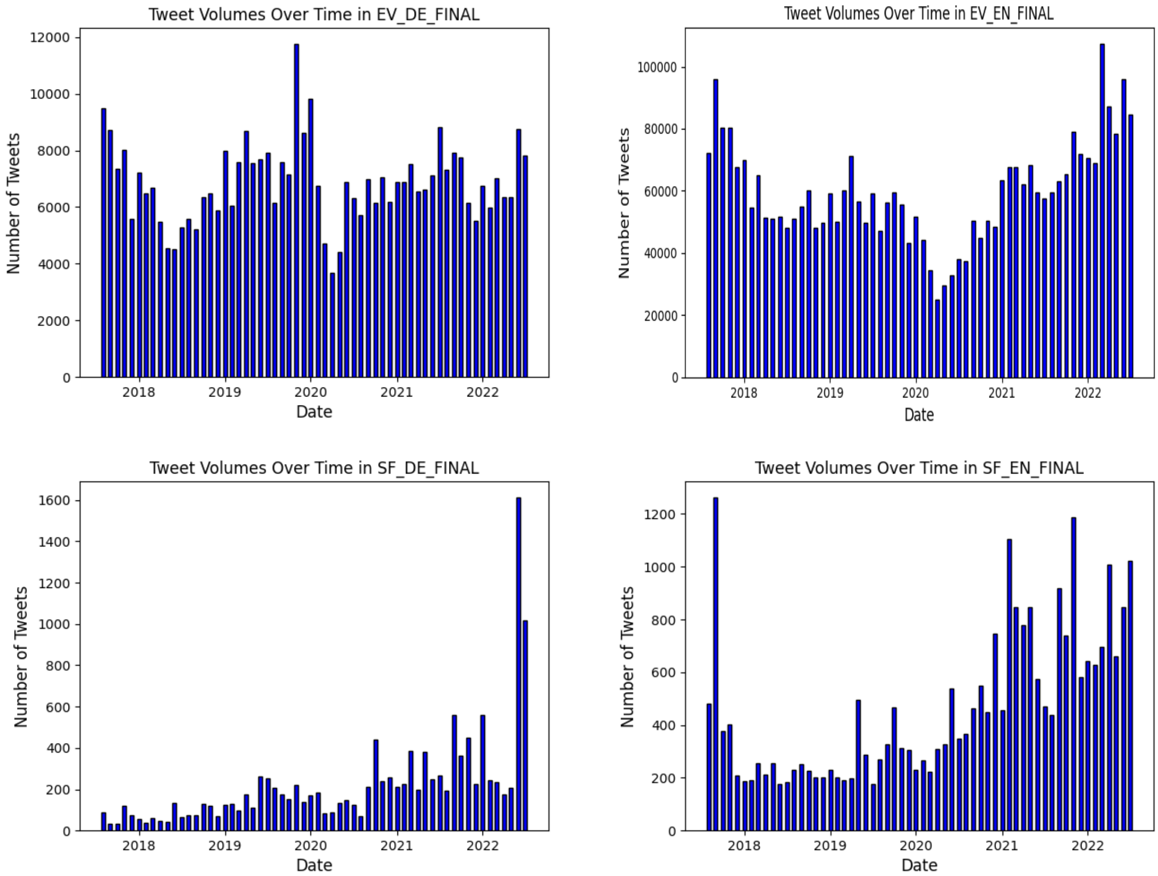


Fig 1: Tweet Volumes Over Time in Main Databases
Source: Own Analysis

4.1.2 Overall Tweet Sentiment Distribution

The tweet sentiment distribution for our databases is shown in Fig. 2. Two things stand out here. First, tweets in English generally appear to be far more positive than German tweets. Secondly, among German tweets, those on electric vehicles are generally more positive than those on synthetic fuels, which is not the case for English Tweets. Across all four databases, there are far more tweets that at the extremes of sentiment positivity than in the centre, showing the polarizing nature of the topic.

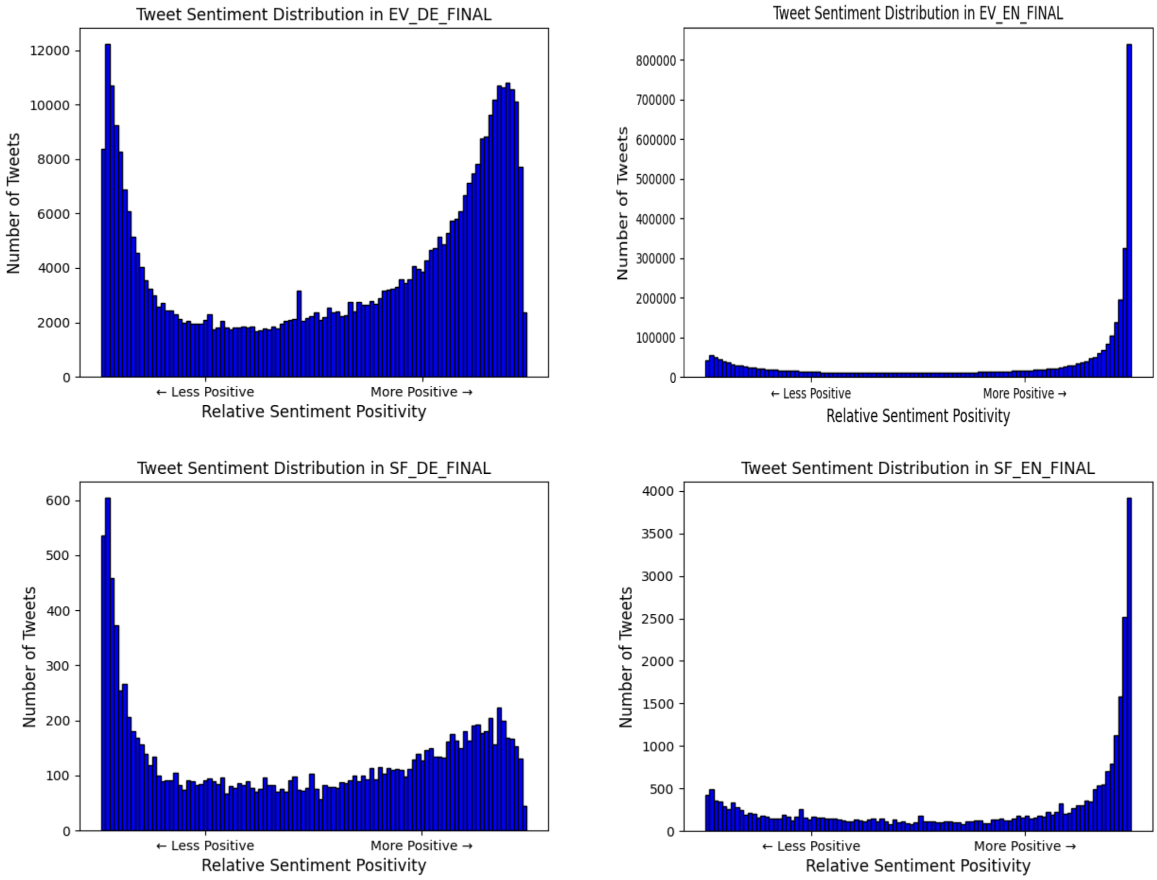


Fig. 2: Tweet Sentiment Distributions in Main Databases
Source: Own Analysis

4.1.3 Sentiment Distribution Over Time

The development of the average sentiment of our tweets over time is shown in Fig. 3. Several things stand out here. Firstly, Sentiment across all four databases has either remained relatively steady or become more negative over the past five years. Secondly, sentiment of German tweets has been consistently more negative than that of English tweets. The most significant recent development has been a sharp increase in negative sentiment of German tweets on synthetic fuels. Whilst the average probability that a German tweet on synthetic fuels is negative predicted by our model was 15 % in Germany, this number jumped up to 37% and 38% in June and July respectively, representing the biggest observed spike in sentiment.

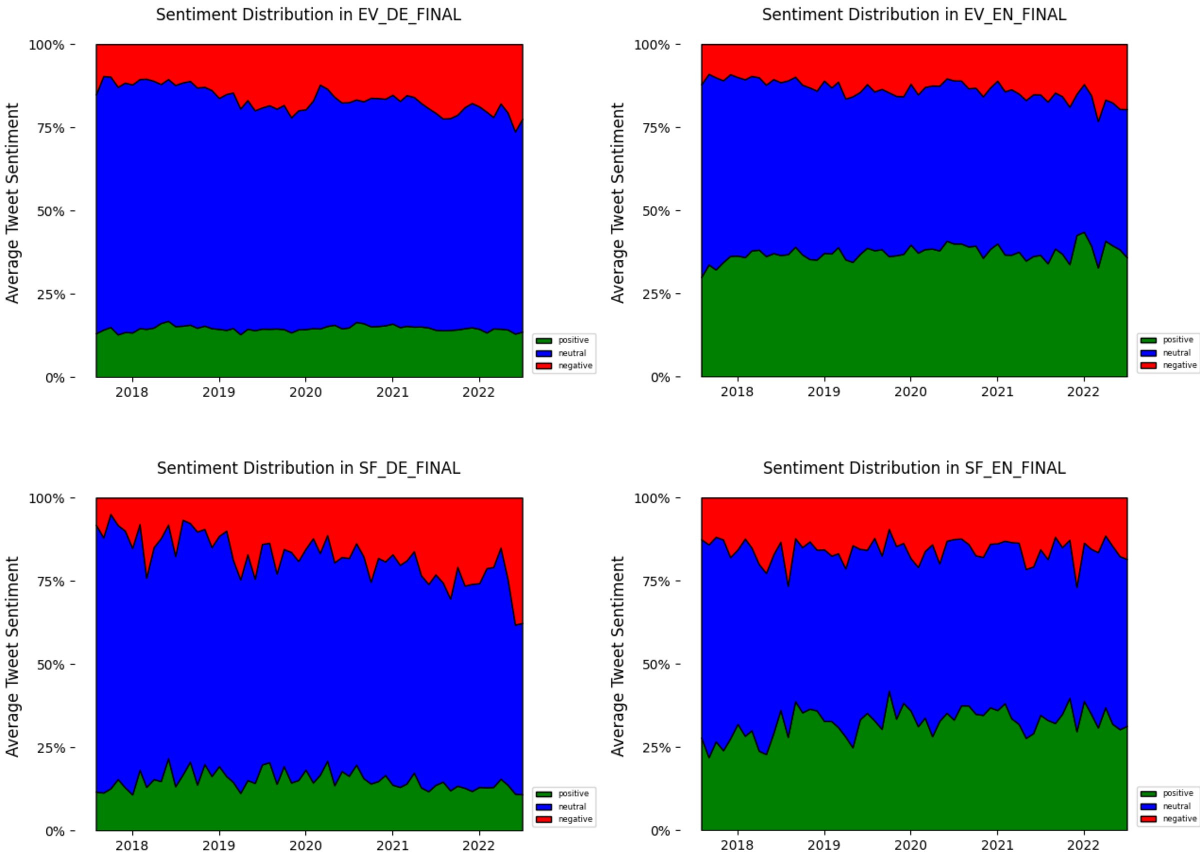


Fig. 3: Sentiment Distribution Over Time in Main Databases
Source: Own Analysis

4.2 Relative Word Frequency

Tables 4, 5, 6, and 7 show the results generated by our relative word frequency analysis for our main databases. For each month, the tables contain the most significant event reflected in the trending words, as well as the key words out of the top 20 with high relative frequency used to identify the event. For the synthetic fuel databases, only the last months are shown, as no significant events appeared before then.

Month	Significant Keywords in top 20	Most Significant Event Mined
Aug 17	Dongfeng, RenaultNissan, Joint	Renault-Nissan and Dongfeng joint venture
Sep 17	Dyson, Vacuum, James, Inventor, Radically	Inventor James Dyson announces plan to build radically different EV
Oct 17	Shell	Shell opens EV charging points
Nov 17	Roadster	Tesla reveals second generation roadster
Dec 17	Mexico, Fords	Ford shifts EV production from Michigan to Mexico

Jan 18	CES	Consumer Electronics Show Awards go to EVs
Feb 18	Porsche	Porsche announces 6bn dollar investment in EV range until 2022
Mar 18	Geneva	2018 geneva motor show dominated by EVs
Apr 18	Sweden, Road, Highway, Charges	Sweden unveils world's first EV charging road
May 18	AAA, Study, Americans	new AAA survey shows that 20 percent of Americans will likely purchase an EV as next car
Jun 18	Pikes, Peak	VW breaks pikes peak hill climb record with custom-built EV
Jul 18	Germany, Federal	German government announces subsidies for municipal electromobilitiy
Aug 18	Kalashnikov	Kalashnikov unveils electric car
Sep 18	Mercedes, Mercedesbenz, Unveils	Mercedes launches first all-electric car
Oct 18	Chooses, Singapore, Dyson	dyson to build electric vehicle in singapore
Nov 18	Trump, Threatens, Subsidies	trump threatens to end EV subsidies
Dec 18	Trump, Subsidies	Trump says carmakers who go all-electric will fail
Jan 19	Cadillac	Cadillac shows ist first EV
Feb 19	AAA, Cold, Weather	AAA study shows cold weather reduces EV range
Mar 19	Geneva	Geneva motor show unveils new Evs
Apr 19	Auspol, Campaign	EV debate in Australian election campaigns
May 19	Dyson	Dyson reveals details of its EV plans
Jun 19	Rover, Land, Jaguar, Develop	BMW and Jaguar Land Rover to work together on Evs in UK
Jul 19	Noise, EU, Council	New EU regulation dictates all new electric vehicles to be fitted with acoustic vehicle alerting system (AVAS)
Aug 19	Taycan, Porsche	Porsche Taycan impresses in endurance tests
Sep 19	Frankfurt	Frankfurt Motor Show 2019
Oct 19	Schumer, Proposes,	Chuck Schumer NY Times opinion piece on EV policy
Nov 19	Cybertruck	Tesla unveils Tesla Cybertruck
Dec 19	Rivian, Investment	Rivian receives 1.3bn investment
Jan 20	Sony, CES, Concept	Sony reveals EV prototype at CES 2018
Feb 20	Super, Bowl	Super Bowl commercials include many EVs
Mar 20	Coronavirus	Coronavirus pandemic breaks out
Apr 20	Coronavirus, Covid, Pandemic	Coronavirus pandemic dominates news
May 20	Vonderleyen, Greenrecovery	European ‚Green Deal‘ will be motor for recovery in EU says von der Leyen

Jun 20	Lyft, Ecar	Lyft announces electric cars only from 2030 onwards
Jul 20	Fisker	US Startup Fisker announces electric SUV to go on sale in 2021
Aug 20	Chinese, Maker	Chinese EV maker Xpeng raises $1.5 billion in U.S. IPO
Sep 20	Worldevday	World EV day introduced
Oct 20	Hummer, Announces	Hummer supertruck EV revealed by GM
Nov 20	Borisjohnson, Ban,	Boris Johnson announces petrol and diesel vehicle ban from 2030 onwards
Dec 20	Apple, Secret, Expected	Speculation about an Apple EV to launch in the future
Jan 21	Baidu, Geely,	Chinese giant Baidu and Automaker Geely invest heavily in their joint EV venture
Feb 21	Lucid	Lucid goes public and raises $4bn
Mar 21	Tesla, Bitcoin	Tesla to accept bitcoin as payment
Apr 21	Voltswagen	April fools Voltswagen name change
May 21	Lightning	Ford lightning EV revealed
Jun 21	Infrastructure, Taxes	US invests 15bn in EV infrastructure
Jul 21	Stellantis, Bn	Stellantis announces 40bn dollar EV investment plan
Aug 21	Recall, Fires, Chevrolet	Chevrolet Bolt Evs recalled due to battery fire risk
Sep 21	Worldevday	World EV day number 2
Oct 21	Hertz	Hertz orders 100000 Tesla Model 3 for its rental fleet
Nov 21	Scott, Morrison	Scott Morrison campaigns pro EV
Dec 21	AABB	AABB launches world's fastest EV charger
Jan 22	CES	Consumer electronics show 2022
Feb 22	Mokkae, Vauxhall,	Opel/Astra/Vauxhall reveals Mokka-E car
Mar 22	Gasprices, Ukraine, Russian, Putin	Gas prices rise due to Ukraine war
Apr 22	Earthday	Drive Electric Earth Day
May 22	Georgia	Georgia to fund EV charging network using federal cash
Jun 22	Stabenow, Debbie, Senator, Democrat	Senator Debbie Stabenow confirms demo-crats EV subsidy plans cancelled
Jul 22	Buttigieg, Pete	Pete Buttegieg pleads for electric vehicles subsidies

Table 4: Events and Significant Words Found in EV_EN_FINAL

Month	Keywords	Event Mined
Aug 17	Dieselgipfel, Dieselgate, Fordert, SPD, Elektroautoquote	Martin Schulz fordert E-Auto Quote nach Dieselskandal
Sep 17	Dyson	Dyson produziert bald E-Autos
Oct 17	Elektroautoquote, Eukommission	EU erwägt Elektroauto-Quote für Neuwagen
Nov 17	Rohstoffknappheit, Warnt, Industrie	Industrie warnt vor Rohstoffknappheit durch E-Autos
Dec 17	Uniti, Schwedische, Startup	Schwedisches Start-Up Uniti stellt erstes E-Auto vor
Jan 18	CES	Consumer Electronics Show 2018
Feb 18	Telekom	Telekom plant Zapfsäulen für E-Autos
Mar 18	Ipace, Jaguar	Jaguar i-Pace E-Auto bestellbar
Apr 18	Byton	Europaprämiere für Bytons Elektro-SUV
May 18	Sixtchef, Erich, Sixt, Fehler, Hält	Sixt-Chef Erich Sixt hält E-Autos für politische Fehler
Jun 18	Pikes Peak	VW bricht mit E-Auto Pikes Peak Rekord
Jul 18	Thüringen, CATL, chinesische	Chinesischer Hersteller CATL öffnet Produktionsstandort in Arnstadt, Thüringen
Aug 18	Börse, Droht	Tesla-Chef Elon Musk erwägt nach eigenen Angaben, den Elektroautopionier von der Börse zu nehmen
Sep 18	Etron, Audi	Audi enthüllt E-Tron Elektro-SUV
Oct 18	Singapur, Dyson, Fabrik	Dyson wird seine E-Autos in Singapur herstellen
Nov 18	Emden, Hannover	Emden und Hannover VW Fabriken werden zu. E-Auto Werken
Dec 18	Telekom, Umstellung	Telekom startet Aufbau von Ladenetz für Elektroautos
Jan 19	Tempolimit	Experten raten Bundesregierung zu Tempolimit und Pflichtquote für E-Autos
Feb 19	Verlängerung,	Bundesregierung verlängert Kaufprämien für E-Autos
Mar 19	Genfer, Autosalon	E-Autos dominieren Genfer Autosalon 2019
Apr 19	Ifoinstituts, Ifoinstitut, Ifostudie, Ifo	Kontroverse Studie des Ifo-Instituts: E-Autos verursachen mehr CO2 als Verbrenner
May 19	Shell, Schnellladesäulen, Tankstellen	Shell bietet Aufladungsstationen für E-Autos an Tankstellen an
Jun 19	Fridaysforfuture	Puls-Umfrage legt nahe, dass E-Autos durch Greta Thunberg und Fridays For Future beliebter werden
Jul 19	Kobaltabbau, Kinderarbeit, Kongo	E-Autos wegen Kinderarbeit beim Kobaltabbau in Afrika unter Kritik

Aug 19	Regenwaldkiller, Rettetregenwald	E-Autos wegen Ursprungs der gebrauchten Rohstoffe im Regenwald unter Kritik
Sep 19	Weltpremiere, Taycan	Weltprämiere des Porsche Taycans
Oct 19	Kongo, Kobalt	Kobalt aus dem Kongo: Hier sterbe Menschen für Akkus
Nov 19	Autogipfel, Kanzleramt	Autogipfel: Kanzleramt kündigt Erhöhung von Kaufprämien für E-Autos an
Dec 19	Erdogan	Erdogan stellt türkischen E-Auto Plan vor
Jan 20	Sony, CES, Las, Vegas	Sony überrascht mit eigenem E-Auto bei CES 2020 in Las Vegas
Feb 20	Ausnahmen, Grünes, Tempolimit	Grüne erwägen Ausnahme für E-Autos bei Tempolimit
Mar 20	Coronavirus	Coronavirus
Apr 20	Höllmüller, Narkus, Chef, Efahrer	E-Fahrer.com Chef Höllmüller fährt statt BMW nun Tesla.
May 20	Abwrackprämie, Kaufprämie, Kaufprämien	Erhöhte Kaufprämien für E-Autos treten in Kraft
Jun 20	Konjunkturpaket, Kaufprämie, Verdoppelt	Konjunkturpaket der großen Koalition verdoppelt Kaufprämie auf E-Autos
Jul 20	Grünheide, Gigafactory, Brandenburg	Elon Musk Twittert Bild der Gigafactory in Berlin Grünheide
Aug 20	Forscher, Verursachen, Angenommen	Niederländische Studie: Aktuelle E-Autos für weniger CO2 verantwortlich als Verbrenner
Sep 20	Scheuer, Bundestag	Bundesverkehrsminister Scheuer kritisiert Elektroauto-Lieferprobleme
Oct 20	Dacia	Dacia Spring vorgestellt
Nov 20	Autogipfel, Umweltsünder, Autobranche	Autogipfel resultiert in Umfangreichem Unterstützungspaket der Bundesregierung für E-Mobilität
Dec 20	Apple, Experten, Könnte	Gerüchte um Apple E-Fahrzeug im Jahr 2024
Jan 21	Wirtschaftsministerium, Gesetzentwurf	Wirtschaftsministerium zieht Gesetzentwurf zur Spitzenglättung zurück
Feb 21	Feuergefahr, Tiefgarage	Erste Stadt verbietet wegen Brandgefahr E-Autos aus Tiefgaragen
Mar 21	Kia	Kia stellt EV6 Elektroauto vor
Apr 21	Huawei, Xiamo, Shanghai	Chinesische Autohersteller bei Auto Shanghai 2021 legen Fokus auf Extras und Accessoires bei E-Autos
May 21	Baerbock, Stromkosten	Grüne für Doppelmoral bei SUVs unter Kritik
Jun 21	Elektrogate, Rechenfehler, wissenschaftler	Rechenfehler bei CO2-Emissionen von E-Autos, Wissenschaftler diskutieren.
Jul 21	Unlogisch, Laschet	Laschet kritisiert Tempolimit als unlogisch
Aug 21	Triell, Laschet, Baerbock	"Triell" der Kanzlerkandidaten auf RTL von E-Autos dominiert

Sep 21	Iaamobility, IAA	IAA Mobility 2023 in München von E-Autos dominiert
Oct 21	Wolfsburg, Boss	Baurat des VW Werks in Wolfsburg fordert Bau eines E-Autos ab 2024
Nov 21	Koalitionsvertrag,	Koalitionsvertrag enthält 15 Millionen Elektroautos bis 2030, sowie 1 Million Ladesäulen als feste Ziele
Dec 21	Toyota	Toyota erhöht E-Auto Absatzziele und stellt 15 neue Modelle vor
Jan 22	Sony, CES, Vision	Sony stellt auf CES 2022 erneut E-Auto vor
Feb 22	Delorean	DeLorean feiert Wiedergebrut als E-Auto
Mar 22	Ukrainekrieg, Spritpreise, Spritpreis, Benzinpreis, Ukraine, Krieg, Russland	Spritpreise werden durch Ukraine-Krieg in die Höhe getrieben
Apr 22	Homeofficepflicht, Verkehrsaufkommen	Ende der Homeofficepflicht könnte Verkehrsaufkommen erhöhen
May 22	Wissing, Volker, Verkehrsminister, Abwrackprämie	E-Autos: Bundesverkehrsminister Volker Wissing plant Abwrackprämie
Jun 22	Lindner, Kaufprämien, abschaffen, Verbrennerverbot, Verbrenneraus	Lindner will Kaufprämien für E-Autos abschaffen und EU beschließt Verbrennerbervot ab 2035
Jul 22	Gaskrise, Energiekrise, Stromversorgung	Drohende Energiekriese

Table 5: Events and Significant Words Found in EV_DE_FINAL

Month	**Keywords**	**Event Mined**
Jul 22	Lindner, Oliver, Diess, Porschebacked	Porschegate scandal: Porsche CEO Oliver Blume and Finance minister Christian Lindner’s private messages leaked

Table 6: Events and Significant Words Found in SF_EN_FINAL

Month	**Keywords**	**Event Mined**
Jun 22	Euumweltminister, Verbrenneraus, Verbrennerverbot	EU-Kommision beschließt Verbrenneraus ab 2035
Jul 22	Porschegate, Lindnerrücktritt, Lindnergate	Porschegate-Skandal: Christian Lindner und Prosche-Chef Oliver Blume in zu engem Kontakt bei Lobbyarbeiten?

Table 7: Events and Significant Words Found in SF_DE_FINAL

4.3 Observed Sentiment Relationships

4.3.1 Sentiment and Tweet Engagement

The relative sentiment positivity of a tweet is the chance our models assign it of being positive, divided by the sum of the probability of it being negative plus the probability of it being positive.

$$Relative\ Sentiment\ Positivity = \frac{\%positive\ sentiment}{(\%\ positive\ sentiment + \%negative\ sentiment)}$$

It describes the share of positive sentiment among the extreme sentiment probabilities assigned by our base sentiment model. Fig. 4 shows the relationship between positivity and tweet engagement. Tweet engagement is measured by looking at the rank of a tweet's likes, comments and retweets compared to all other tweets in our databases.

Markers are sized according to the frequency of tweets in the percentile they represent. The largest dots in all four charts represent tweets that receive no engagement at all, meaning zero likes, zero comments and no retweets. As can be seen in the chart, they make up around 15-25% of all tweets in our databases.

The correlation between the engagement percentile and relative sentiment positivity is negative for all three databases and statistically significant at the 5% for all databases except SF_EN_DB. Tweets with negative base sentiment thus seem to generate higher engagement among Twitter users. This effect also seems to be stronger among German tweets than among English tweets.

For the relationship between tweet neutrality and engagement, the data is similar. Neutrality is negatively correlated to engagement to a statistically significant degree across all databases. The only difference to the relationship between positivity and engagement is that this effect appears to be equally strong for German and English tweets. The data thus shows that less neutral tweets on electric vehicles and synthetic fuels generated significantly more engagement on Twitter over the past five years.

Overall, less neutral and more negative content generally seem to have lead to higher engagement in tweets English and German tweets on electric vehicles and synthetic fuels over the past five years.

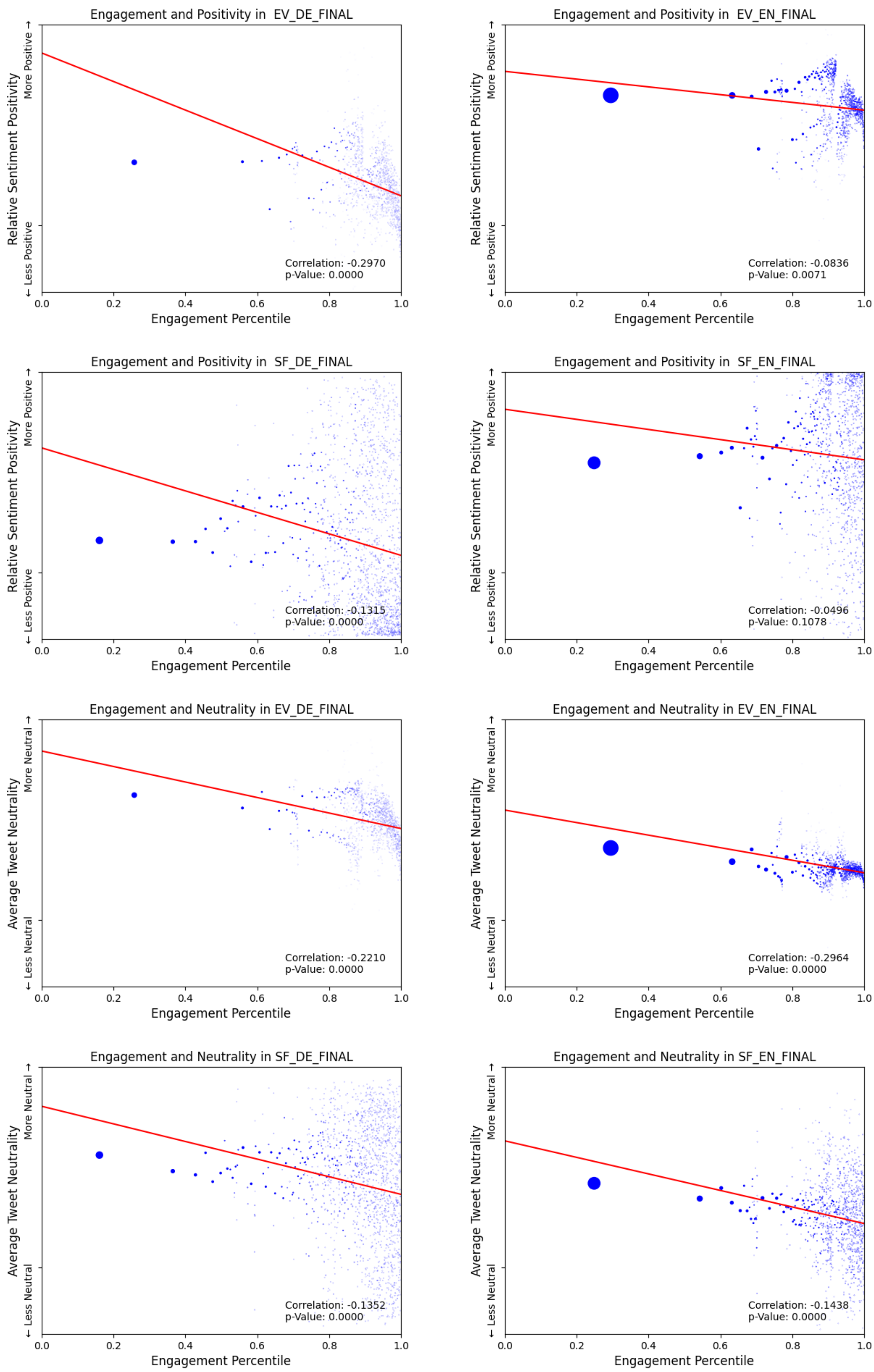
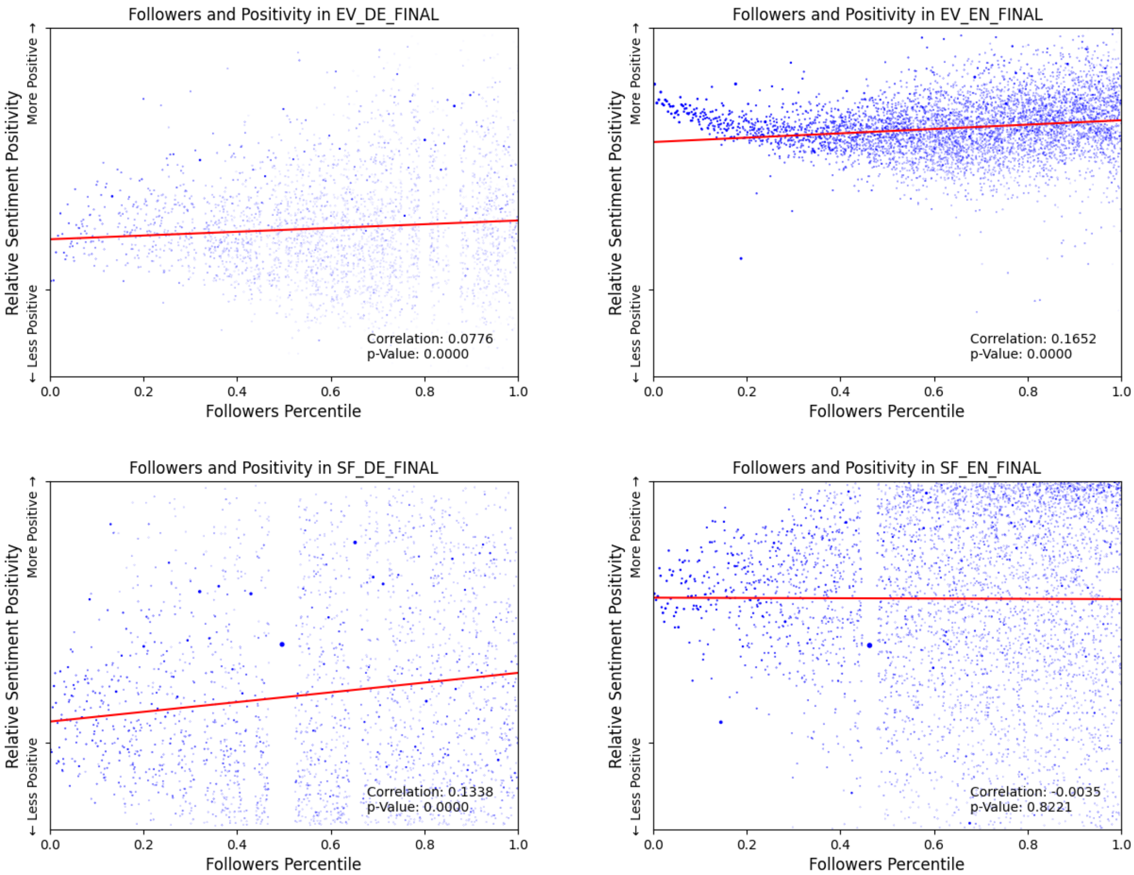


Fig 4: Sentiment and Engagement in Main Databases
Source: Own Analysis

4.3.2 Sentiment and User Characteristics

Fig. 5 below shows the relationship between follower numbers and sentiment. Once again there is consensus between all databases other than SF_EN_FINAL, where no statistically significant relationship can be observed. Across all other databases there appears to be a slight positive correlation between follower numbers and relative sentiment positivity, suggesting more influential users on Twitter have slightly more favorable sentiment towards alternative fuel vehicles. The slight positive, yet also statistically significant correlation between follower numbers and neutrality of tweets suggests that more influential users tend to post more neutral tweets when referring to alternative fuel vehicles.



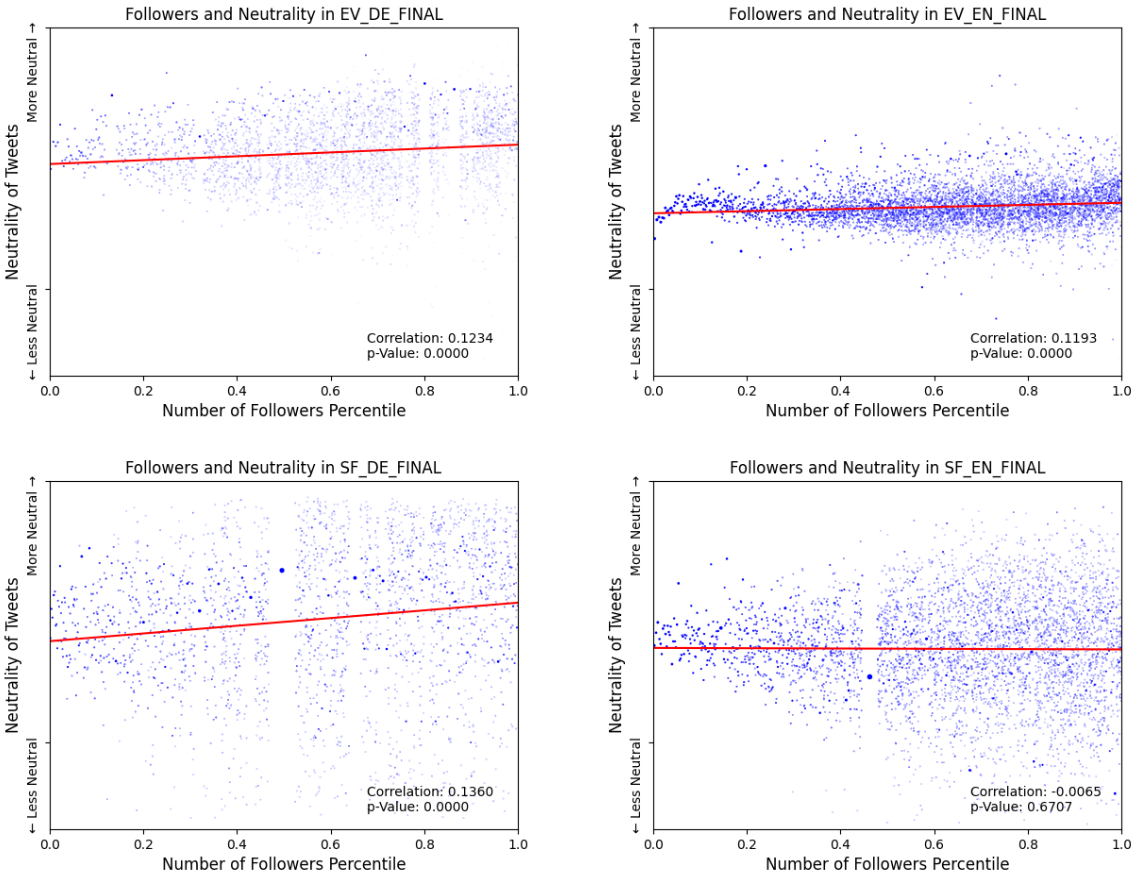


Fig. 5: Followers and Sentiment in Main Databases

Source: Own Analysis

In Fig. 6 we see the relationship between sentiment and users’ total number of tweets posted. Across almost all databases we see two statistically significant relationships. Firstly, higher levels of user activity, using users’ tweet counts as the defining statistic, are negatively correlated to relative sentiment positivity of tweets. This correlation is significant at the 5% level across all databases. This shows that users who post frequently on Twitter generally display a less favorable sentiment towards alternative fuel vehicles when doing so. Secondly, there appears to be a slightly positive correlation between user activity and neutrality of tweets. This correlation is statistically significant at the 5% level across all databases apart from SF_DE_FINAL. There the relationship is only statistically significant at the 6% level. This shows that users who post more frequently on Twitter tend to post more neutral content.

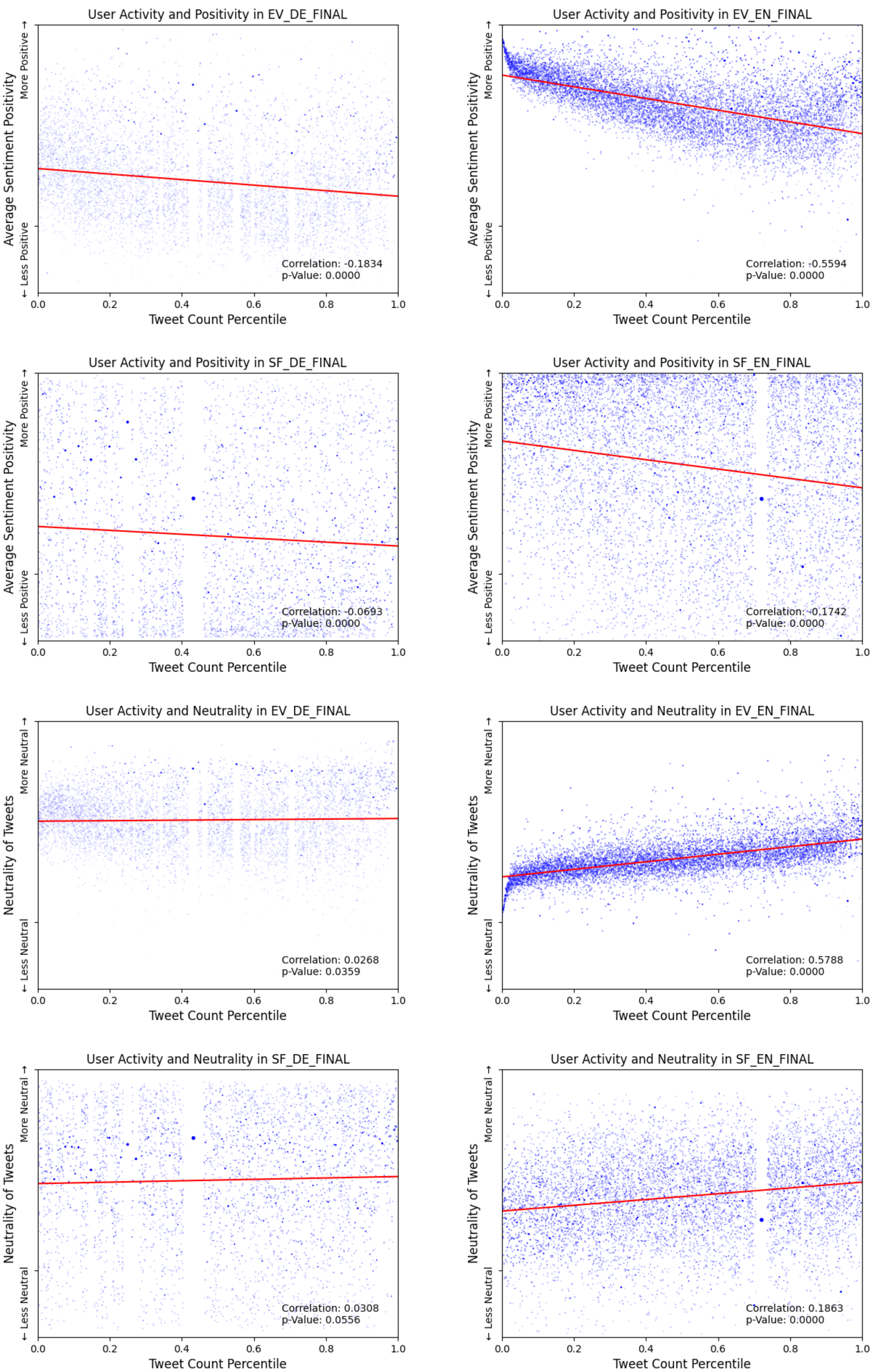


Fig. 6 Sentiment and User Activity Levels in Main Databases
Source: Own Analysis

4.4 Subtopics

4.4.1 Batteries

Fig. 7 shows the overall distribution of sentiment of tweets on electric vehicles to do with batteries over time for both German and English tweets. German tweets appear to be slightly more varied in their sentiment on batteries, whereas English tweets are extremely positive overall.

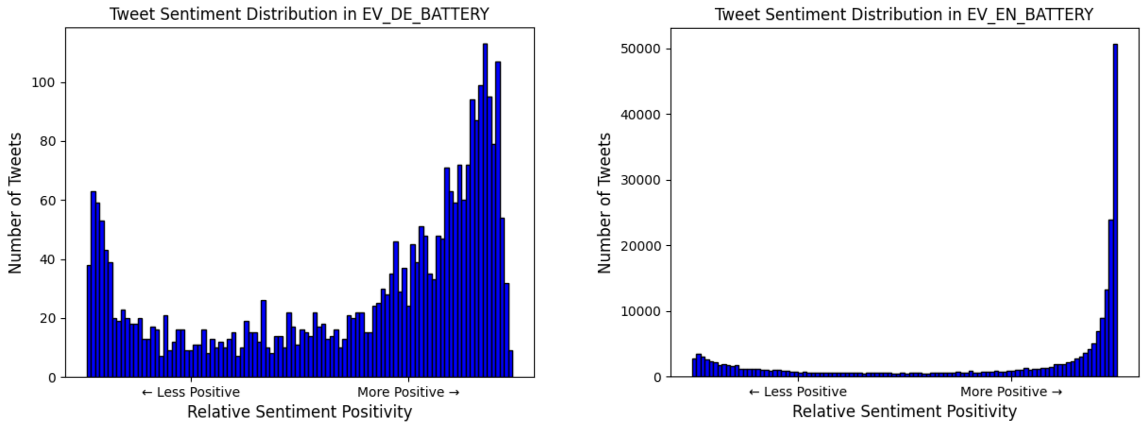


Fig. 7: Sentiment Distribution of Battery Related EV Tweets
Source: Own Analysis

Fig. 8 shows how this sentiment developed over time. Two things stand out here. First, German tweets on batteries are far more neutral overall. Secondly, German sentiment on electric vehicle tweets on battery-related issues has remained relatively stable, while there has been a downward trend in English tweet sentiment on the matter over the past five years. For German tweets, this is not in line with overall base sentiment in the EV database, which had worsened slightly over the past five years. For English tweets this observed trend is in line with the overall database.

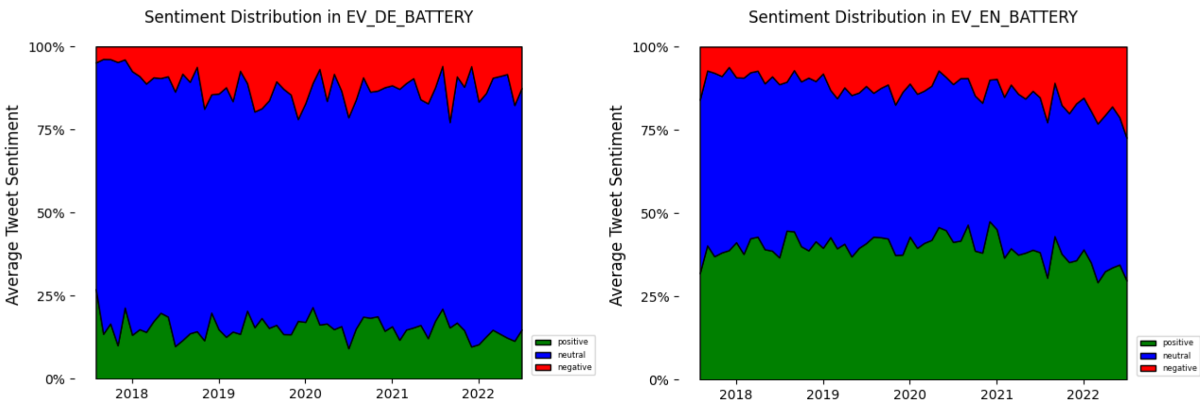


Fig. 8: Sentiment of Battery Related EV Tweets Over Time
Source: Own Analysis

4.4.2 Range and Charging

Fig. 9 shows overall distribution of sentiment of tweets on electric vehicles to do with range or charging over time for both German and English tweets. Once again German tweets appear to be generally more varied in their sentiment on batteries, whereas English tweets are extremely positive overall.

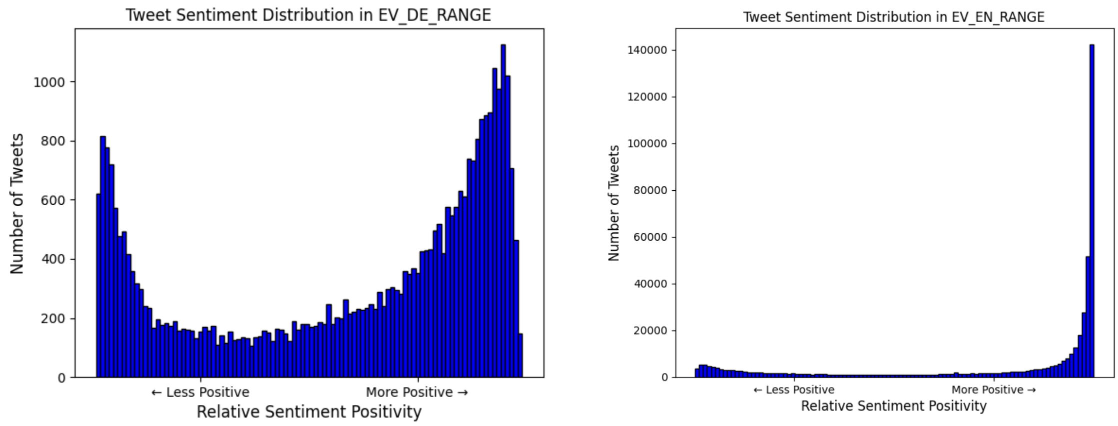


Fig. 9: Sentiment Distribution of Range or Charging Related EV Tweets
Source: Own Analysis

Fig. 10 shows how this sentiment developed over time. Both for German and English tweets, sentiment on range and charging related issues has become slightly more negative over the past five years. This development is in line with the overall change in sentiment across the two full EV databases over the past five years.

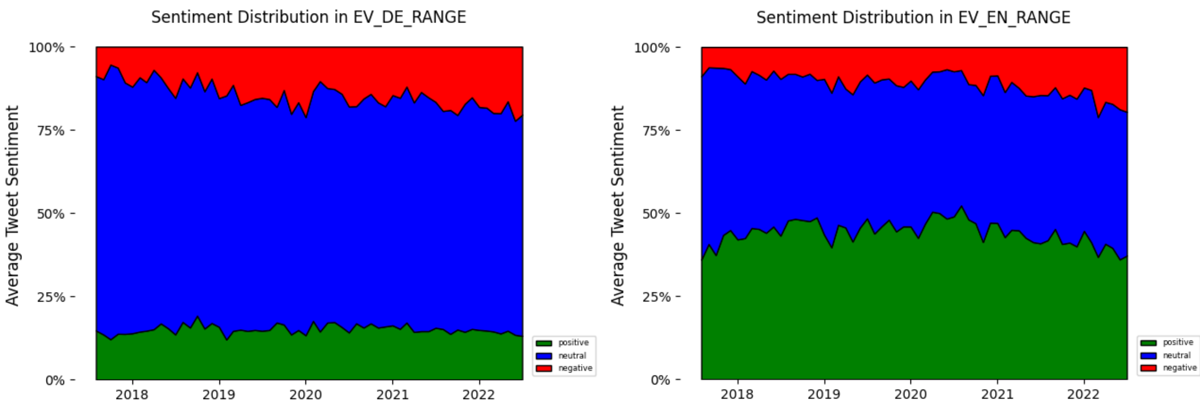


Fig. 10: Sentiment of Range or Charging Related EV Tweets Over Time
Source: Own Analysis

4.4.3 Costs

Fig. 11 shows the overall distribution of sentiment of tweets on electric vehicles to do with costs and prices over time for both German and English tweets. German tweets appear to be far more negative in their sentiment on costs, whereas English tweets are once again extremely positive. Note that despite their relative positivity compared to German Tweets, English tweets on costs are generally far more negative when compared to tweets on other subtopics reviewed this far.

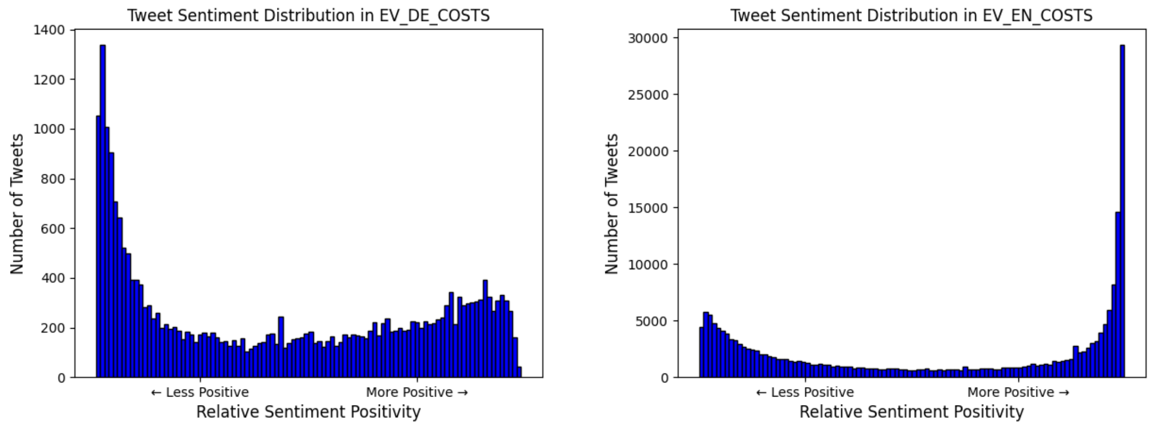


Fig. 11: Sentiment Distribution of Cost Related EV Tweets
Source: Own Analysis

Fig. 12 shows how this sentiment developed over time. Both for German and English tweets, sentiment on cost related issues has become more negative over the past five years. This development is in line with, but more extreme than the overall change in sentiment across the two full EV databases over the past five years. The levels of sentiment have also been consistently more negative than for the overall EV databases.

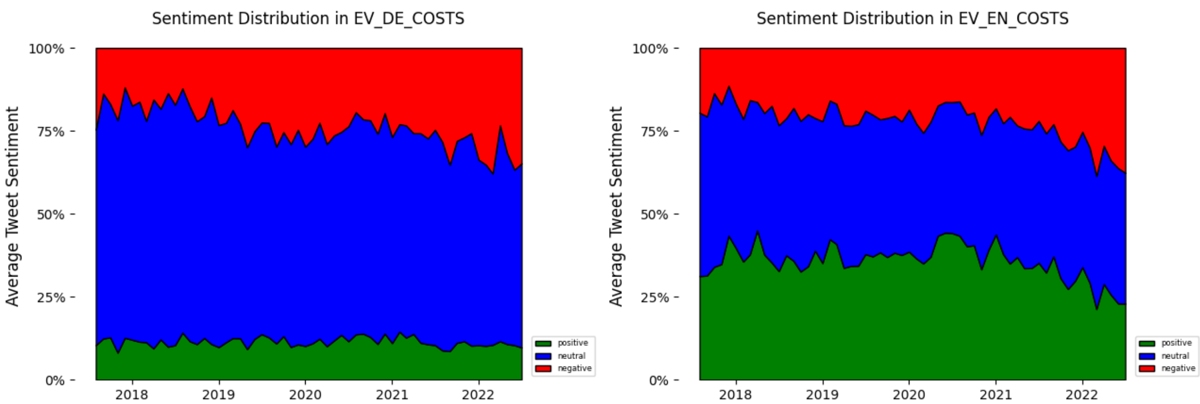


Fig. 12: Sentiment Distribution of Cost Related EV Tweets Over Time
Source: Own Analysis

4.4.4 Sustainability

Fig. 13 shows the overall distribution of sentiment of tweets on electric vehicles to do with sustainability over time for both German and English tweets. German tweets appear to be far more highly polar their sentiment on sustainability, with a near even split between negative and positive sentiment, whereas English tweets are once again extremely positive. Compared to the other subtopics examined, this subtopic results in the most polarizing distribution seen so far for German tweets.

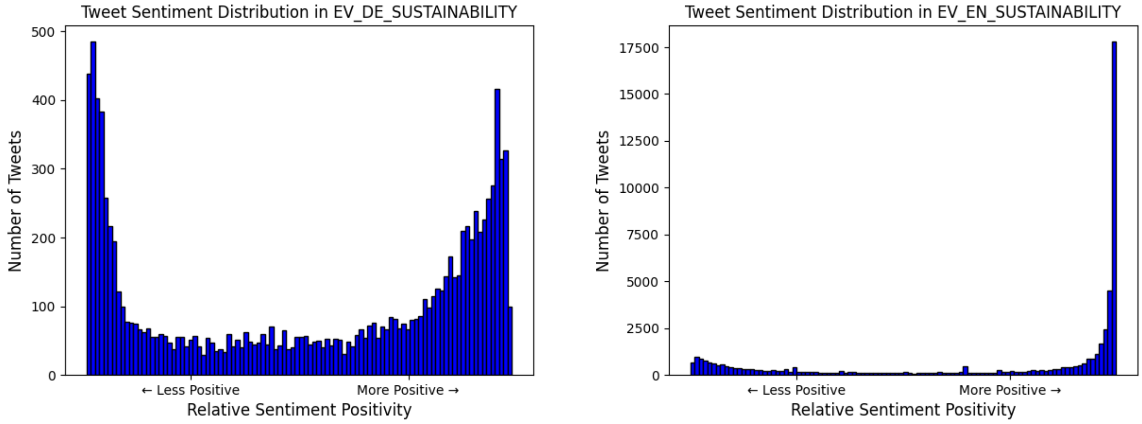


Fig. 13: Tweet Sentiment Distribution of Sustainability Related EV Tweets
Source: Own Analysis

Fig. 14 shows how this sentiment developed over time. Both for German and English tweets, sentiment on sustainability related issues has remained relatively stable over the past five years, with only minor overall declines in sentiment over this period. This trend is in line with the overall databases, although the observed change is much weaker.

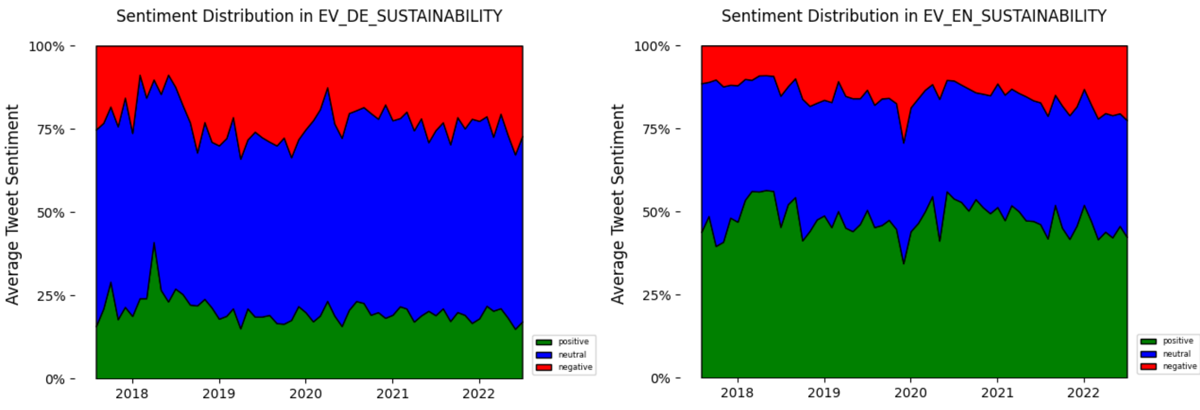


Fig. 14 Sentiment of Sustainability Related EV Tweets Over Time
Source: Own Analysis

Overall, it appears that the sentiment on subtopics within the EV conversation generally moved in the same direction as the overall sentiment, with costs exhibiting the most extreme, and sustainability exhibiting the least extreme downward trend in sentiment.

5 Discussion

5.1 Overall Sentiment

5.1.1 Electric vehicles

There are two main discussion points regarding the general sentiment of tweets on electric vehicles yielded by our analysis.

First, there is a large difference between the base level of sentiment of EV tweets in German and that of EV tweets in English. As can be seen in Fig. 3, the average share of positive base sentiment of German EV tweets never reached 20%, staying consistently near the 15% mark. English EV tweets were far more positive, never posting a share below 25% positive base sentiment. There are many possible explanations for this. It could be that German-speaking Twitter users, the majority of whom will be Germans, have more ambivalent feelings towards electric vehicles, due to the potential threat they pose to Germany's existing automobile industry, which is heavily dominated by combustion engine vehicles. Another reason might simply be that Germans are generally more negative when posting on Twitter. It could also be that the multilingual model generally outputs more negative sentiment than the English only model. Further analysis is required to determine the exact cause of this difference.

Secondly, sentiment of tweets on electric vehicles has generally worsened over the past 5 years, with peaks in negative sentiment for EV tweets coming in 2022 for both German and English tweets. Our subtopic analysis shows that this trend is mostly driven by tweets related to batteries, costs, range and charging, with costs being the most significant driver of sentiment downturns. There are once again many possible reasons for this. It could be that increasing costs and prices due to inflation or other worsening economic circumstances have led to this downturn in sentiment. It is also possible that consumers have been expressing negative sentiment as a response to reductions in subsidies for electric vehicles. Further analysis is needed to identify the root causes of this trend.

While there were plenty of events found by our relative word frequency analysis, it appears that none of them were significant enough to leave a strong mark on overall sentiment in the database, or that events cancelled each other out.

5.1.2 Synthetic Fuels

As can be seen in Fig. 3 there is a significant difference between the sentiment of German and English on synthetic fuels. Once again, English tweets are far more positive, with the share of positive sentiment not dropping between 25% since 2018, whereas the German share of positive sentiment never once rose to that number. A second difference between tweets on synthetic fuels in English and German is the trend of sentiment in the two databases. English tweet sentiment on synthetic fuels has remained relatively stable, with negative sentiment around the 15% mark and positive sentiment of around 30%. The fluctuations around these means have been stronger than in the EV database, but this can be explained by the smaller size of the database. For German tweets however, there has been a strong trend towards more negative sentiment. In August 2017, Tweets were given an average negative sentiment score of 8%. This number more than quadrupled to 37% and 38% in June and July of 2022 respectively. This is the strongest shift in overall sentiment in any of the main databases over the past five years. The bulk of this downward shift occurred in the past two months, with the share of negative sentiment jumping from 25% in May to its peak of 38% in June, representing a one-month rise of over 50%.

The most likely explanation for this downward trend are two key events that occurred in the German political sphere over the past two months.

Firstly, the European Union's transportation ministers recently agreed on a ban of all combustion engine cars from 2035 onwards. This ban did include an exception for those cars who run exclusively on E-Fuels that can be produced in a carbon neutral way. For many German Twitter users, this represented an unwelcome compromise, as evidenced by the presence of the words 'Kompromiss' and 'Hintertür' in the top 20 words of our relative word frequency analysis for June 2022". The highly negative sentiment this event generated might be indicative of anger towards the use of synthetic fuels as a backdoor attempt of keeping combustions engines alive, rather than be a sign of bad sentiment towards the technology per se.

The second significant event that worsened this effect was the "Porschegate" political scandal in July 2022. This scandal involved the revelation that German finance minister Christian Lindner, a self-professed Porsche Fan, and Porsche CEO Oliver Blume have been in very close, even hourly contact since Mr. Lindner's party's coalition negotiations to form the current government and that Mr. Blume heavily influenced the inclusion of the synthetic fuel exception in the recent combustion engine ban (n-tv.de, 2022). Our relative word frequency analysis reflects this, with the over half of the top 20 words for July relating to this incident ("Porschegate", "Lindnerrücktritt", "Lindnergate", "Volkswagenchef", "Lobbyist", "Lindnerruecktritt", "Fdp-chef", "Korruption", "stündlich", "Kontakt", and "entschuldigt").

The significance of these two related events is further reflected in the explosion of German tweet volumes on synthetic fuels over the past two months, as can be seen in Fig. 1. While April and May had seen a mere 200 tweets on the topic respectively, this number exploded to 1600 in June and 1000 in July. It thus seems that the general sentiment towards E-fuels has been significantly impaired by the porschegate affair and the EU's decision to include them as an exception to the combustion engine ban from 2035 onwards.

5.2 Subtopics

This thesis investigated four key subtopics within the two databases for electric vehicle tweets in English and German respectively. In order to identify whether these subtopics displayed any significant trends or whether noticeable differences could be observed between English and German tweets on the respective subtopic.

5.2.1 Batteries

For batteries, the development of sentiment was different for German and English tweets. As has been the case throughout the majority of our observations, English tweets on the whole were more positive. This can be seen in fig. 8. There are many possible explanations for this, but none which can be further supported using evidence from this thesis. While English tweets on the whole were more positive, they exhibited a downward trajectory in sentiment, which was absent from German battery related EV tweets. Once again, our analysis offers no way of explaining this downturn in sentiment, as no battery-specific events were found in the relative word frequency analysis. Further research is needed here. What it does mean is that batteries are unlikely to have contributed to the negative development in sentiment of German EV tweets as a whole but may have played a part in the changes in sentiment of English EV tweets.

5.2.2 Charging and Range

For EV tweets related to charging and range, the results are very similar to those for the overall databases, with English tweets once again more positive than their German counterparts, as can be seen in Fig. 9. The sentiment trend visible in Fig. 10 is also very similar to that of the overall trend in Fig. 3 for both German and English tweets. Charging and Range related tweets thus do not seem to differ much from other tweets on electric vehicles regarding sentiment. The slight downward trend in their sentiment makes it plausible to assume that charging and range related issues contributed to the overall downturn in sentiment of electric vehicle tweets over the past five years.

5.2.3 Costs

For EV tweets related to costs, English tweets were once again generally more positive, whilst both German and English tweets exhibited a downwards trend in line with the sentiment development of the overall EV databases. This downward trend was stronger than among other subtopics however and suggests that the general downturn in sentiment of EV tweets over the past five years was at least partly driven by more negative EV tweets relating to costs of electric vehicles. The reasons for this are unclear and require further examination.

5.2.4 Sustainability

For EV tweets related to sustainability the findings very closely resemble those of charging and range. Once again English tweets were more positive and both English and German EV tweets related to sustainability exhibited minorly worsening sentiment over the past five years, which is in line with the overall sentiment changes in the database.

5.3 Social Media Characteristics

As can be seen in Fig. 4, engagement was negatively correlated with neutrality and positivity of tweets across all our databases. These results suggest that negative tweets on alternative fuel vehicles generate higher engagement than positive tweets. This is line with analyses of Twitter posts on other issues and appears to be a general phenomenon of modern social media (Lee and Xu, 2018), rather than explicitly related to alternative fuel vehicles.

Fig. 5 meanwhile shows that high follower numbers are associated with more positive sentiment on alternative fuel vehicles. There might be several explanations for this. It may be the case that accounts with high follower numbers are engaged in monetization activity, which would include advertisements, which would naturally exhibit positive sentiment about the product they are advertising. Another explanation might be that accounts with higher follower counts are less likely to post negative content on alternative fuel vehicles in order to avoid a backlash from otherwise inclined Twitter users. Further analysis would be required to support either of these theories.

In Fig. 6 we can see that users who post more, post more negative content on average. This means that much of the negative sentiment we see in the databases is likely to be driven by a smaller subset of Twitter users. The direction of causality, should this relationship be causal at all, is unclear here. It may be the case that users who post often and engage with lots of Twitter content on alternative fuel vehicles become more negative in their tweets over time, but it is equally plausible that users with more negative sentiments towards alternative fuel vehicles feel the need to post more frequently on the issue. Further analysis would be required to determine the details of this observed relationship. In terms of interpretability however, this relationship suggests that more active users post more negative content, which skews the overall impression of Twitter users' sentiment towards the negative.

5.4 Limitations

Despite the Conclusions drawn in earlier sections of this chapter, there are some limitations to the interpretability and scope of this analysis.

Firstly, as previously touched upon, there is no evidence that Twitter, or social media in general, give an accurate representation of the consumer EV landscape at large. Even if the userbase itself did perfectly represent the consumer landscape, there still may be a significant difference between users who post tweets on EVs and those who don't. Some initial evidence that tweet frequencies and user characteristics are related was discussed in section 5.3

Secondly, there may be differences between base sentiment of a user's posts on Twitter and that users' actual sentiment towards and acceptance of alternative fuel vehicles. There is currently no academic literature on how base sentiment on Twitter relates to overall consumer sentiment towards a technology.

Thirdly, comparisons between the results for tweets of different languages must be viewed against the backdrop of the fact that we used two different models for our analysis. It may simply be the case that the model used to analyze German tweets generally outputs more negative sentiments than that used for the analysis of English Tweets. Further investigations and comparisons are required here.

5.5 Further Research

This thesis has opens up several avenues for further research.

First, the time period of the analysis could be expanded to include tweets further in the past. The reason for the choice of five years as the observation period for this thesis were due to API request limits and computing resource constraints. Given more time and computing resources, the analysis could easily be expanded.

Secondly, the sentiment of tweets connected to individual events could be used to identify their significance and impact on overall sentiment. This thesis identified over 120 events that occurred in the twitter sphere around alternative fuel vehicles over the past 5 years. While this would have gone beyond the scope of this thesis, several steps could be taken to expand this part of the analysis. First, average sentiment scores for tweets relating to each event could be calculated and tracked. Secondly, events could be categorized to find relationships between certain event types and their impact on overall tweet sentiment. It is plausible for example, that launches of EV models of some brands have an overall positive impact on the perception of alternative fuel vehicles. It may also be the case that political decisions that support alternative fuel vehicles with taxpayer money are frowned upon, for example. A systematic categorization of events and subsequent analysis of their associated base sentiment impact could give clues to how future events might be received by the public.

Thirdly, there is scope for much deeper natural language processing analysis beyond base sentiment. Emotion detection could be user to see the main emotions associated with events or subtopics addressed in this analysis. The distribution of emotions of all tweets could also be tracked over time. An analysis like this might for example reveal that initial excitement around new EV launches has dropped over the past years and given way to more calculated considerations on costs. Furthermore, intent analysis could be used to refine the analysis and isolate differences between promotions, opinions, complaints, feedback, or other types of intents behind tweets. Both of these techniques could uncover aspects of consumer sentiment towards alternative fuel vehicles that base sentiment could not.

6 Conclusion

This bachelor's thesis collected data on four million tweets on electric vehicles and synthetic fuels using Twitter's official API and the tweepy and pandas python libraries. We analyzed the tweet texts using sentiment analysis models, which assigned every tweet a probability of its base sentiment being positive negative or neutral. With this data we then examined the relationships between sentiment and engagement, as well as between sentiment and user characteristics. We also calculated the top 20 words for every month that appeared unusually often in that month and used the results to map out significant real-world EV-related events that were spoken about on Twitter for that month.

Our analysis yielded several noteworthy results. Firstly, we found that sentiment of German Tweets on synthetic fuels plummeted massively in June and July of 2022. Our relative word frequency analysis indicated that this may have been due to the EU's decision to exclude synthetic fuel powered combustion engines from its recent bad and the subsequent porschegate scandal in German politics. We also found that despite the presence of many events, no Event impacted sentiment on electric vehicles as significantly as the aforementioned.

We also found that sentiment of electric vehicle tweets has slightly worsened over the past five years. This trend was also visible when filtering tweets according to prominent subtopics within the electric vehicle debate, namely batteries, range and charging, costs, and sustainability, with cost related tweets seeing the most significant downturn in sentiment. This, coupled with the absence of major events that impacted sentiment suggests that this downturn in sentiment is systematic and associated with the core essence of electric vehicles, rather than the result of a particular event or circumstance. We further found evidence that users who post more frequently tend to post more negative tweets and that negative and less neutral tweets generate higher engagement than more positive or neutral tweets.

While the results of our analysis go a long way towards answering our research questions there is still scope for more research and a deeper analysis of twitter sentiment towards alternative fuel vehicles. First, the scope of our analysis could be extended to a longer time period. Secondly, the events uncovered by our relative word frequency analysis could be examined in more detail. Thirdly, more natural language processing techniques could be applied to the tweet text gathered in order to extract more nuanced insights than base sentiment can offer.

7 Literature

Atasoy et al. (2015): “The concept and impact analysis of a flexible mobility on demand system”. In: Transportation Research Part C: Emerging Technologies 56 (2015), pp. 373–392.

Audi MediaCenter (2021): "Audi advances e-fuels technology: new "e-benzin" fuel being tested". Audi MediaCenter. Retrieved 2021-03-30.

Barbieri et al. (2020): Tweeteval: Unified benchmark and comparative evaluation for tweet classification. arXiv preprint arXiv:2010.12421.

Barbieri et al. (2021): Xlm-t: A multilingual language model toolkit for twitter. arXiv preprint arXiv:2104.12250.

Cody et al. (2016): Public opinion polling with Twitter.

Davis (1980): A technology acceptance model for empirically testing new end-user information systems: Theory and results (Doctoral dissertation, Massachusetts Institute of Technology).

Devlin et al. (2018): Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Duggan, Maeve, and Brenner (2013): The demographics of social media users, 2012. Vol. 14. Washington, DC: Pew Research Center's Internet & American Life Project, 2013.

Egbue and Long (2012): “Barriers to widespread adoption of electric vehicles: An analysis of consumer attitudes and perceptions”, Energy Policy, Volume 48, 2012.

Encyclopedia Britannica (2014): *"Electric automobile". Encyclopædia Britannica (online). Archived from the original on 20 February 2014. Retrieved 2 May 2014.*

EPA (2022): https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions (Online resource, last accessed on 10.04.2022)

Ferguson et al. (2018): How open are Canadian households to electric vehicles? A national latent class choice analysis with willingness-to-pay and metropolitan characterization, Transportation Research Part D: Transport and Environment, Volume 58, 2018"

Forsa (2021): Bekanntheit von und Einstellungen zu E-Fuels, available at https://www.uniti.de/fileadmin/artikel/forsa%20E-Fuels%20Ergebnisbericht.pdf, last accessed on 09. Sep. 2022

FuelsEurope (2019): Consumer Survey on Future Mobility Options, available at https://www.fuelseurope.eu/wp-content/uploads/DEF_EU_ConsumerSurvey_digital.pdf, last accessed on 09. Sep. 2022

Gerdes (2012): "The Global Electric Vehicle Movement: Best Practices From 16 Cities". *Forbes.* Archived from the original on 29 July 2017. Retrieved 20 October 2014.

Hauri (2019): "Wir arbeiten mit Hochdruck an der Brennstoffzelle" [We are working hard on the fuel cell]. Neue Zürcher Zeitung (in German).

Higueras-Castillo et al. (2020): Potential Early Adopters of Hybrid and Electric Vehicles in Spain—Towards a Customer Profile. Sustainability. 2020; 12(11):4345. https://doi.org/10.3390/su12114345

Hussein (2018): A survey on sentiment analysis challenges. Journal of King Saud University-Engineering Sciences, 30(4), 330-338.

Hindustan Times (2020): "Why are electric cars expensive? The cost of making and buying an EV explained". Hindustan Times. 23 October 2020. Retrieved 30 October 2020.

IPCC (2019): "Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems"

Jin et al. (2021): Driving a green future: a retrospective review of china's electric vehicle development and outlook for the future.

Kohl et al. (2018): Anticipating acceptance of emerging technologies using twitter: the case of self-driving cars. Journal of Business Economics, 88(5), 617-642.

Lee and Xu (2018): The more attacks, the more retweets: Trump's and Clinton's agenda setting on Twitter. Public Relations Review, 44(2), 201-213.

Liao, Molin and van Wee (2017): Consumer preferences for electric vehicles: a literature review, Transport Reviews, 37:3, 252275, DOI: 10.1080/01441647.2016.1230794

Linzenich et al. (2019): What fuels the adoption of alternative fuels? Examining preferences of German car drivers for fuel innovations, Applied Energy, Volume 249, 2019

Medhat, Hassan, and Korashy (2014): Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal, 5(4), 1093-1113.

Miners (2014): "Bot or Not? Researchers make an app to sniff out bots on Twitter". PC World. Retrieved May 15, 2014.

Mumtaz and Ahuja (2016): A lexical approach for opinion mining in twitter. IJ Education and Management Engineering, 4, 20-29.

Müller (2019): Comparing technology acceptance for autonomous vehicles, battery electric vehicles, and car sharing—A study across Europe, China, and North America. Sustainability, 11(16), 4333.

N-tv.de (2022): „Worum es beim „Porschegate"-Skandal geht", online source, https://www.n-tv.de/mediathek/videos/politik/Worum-es-beim-Porschegate-Skandal-geht-article23485079.html, last accessed on 09. Sep. 2022

Pak and Paroubek (2010): Twitter as a corpus for sentiment analysis and opinion mining. In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10).

Paoli and Gül (2022): Electric cars fend off supply challenges to more than double global sales.

Patrascu (2020): "Future Porsche Cars to Run on eFuels, Motorsport Machines Included". autoevolution. Retrieved 2021-03-30.

Rezvani, Jansson and Bodin (2015): Advances in consumer electric vehicle adoption research: A review and research agenda, Transportation Research Part D: Transport and Environment, Volume 34, 2015, Pages 122-136"

Roth (March 2011): *Das erste vierrädrige Elektroauto der Welt* [*The first four-wheeled electric car in the world*] (in German). pp. 2–3.

Royal Society (2019): Sustainable Synthetic Carbon Based Fuels for Transport. Pp. 9-13 London.

Sandalow (2009): Plug-In Electric Vehicles: What Role for Washington? (1st. ed.). The Brookings Institution. pp. 1–6

Slowik and Lutsey (2016): Evolution of incentives to sustain the transition to a global electric vehicle fleet. 10.13140/RG.2.2.36368.81920.

Shahan (2020): " Electric Cars Have Mostly Lower Total Cost in Europe". CleanTechnica.com

Urbina (2013): "I Flirt and Tweet. Follow Me at No. Socialbot". The New York Times. Retrieved May 15, 2014.

Wang, Cao and Zhang (2021): The influences of incentive policy perceptions and consumer social attributes on battery electric vehicle purchase intentions, Energy Policy, Volume 151, 2021"

Wicki, Brückmann and Bernauer (2022): How to accelerate the uptake of electric cars? Insights from a choice experiment,Journal of Cleaner Production, Volume 355, 2022

Wicki et al. (2022): What do we really know about the acceptance of battery electric vehicles? – Turns out, not much, Transport Reviews, DOI: 10.1080/01441647.2021.2023693

Wojcik and Hughes (2019): "Sizing Up Twitter Users". Pew Research Center.

Wolf et al. (2019): Huggingface's transformers: State-of-the-art natural language processing. arXiv preprint arXiv:1910.03771