**ANL488 PROJECT PROPOSAL**

**SPATIO-TEMPORAL ANALYSIS OF PUBLIC SENTIMENT IN EUROPE AND THE USA ON LOW-CARBON ENERGY SOURCES**



**Submitted by**

**CALEB CHIA WE KEAT**

**SCHOOL OF BUSINESS Singapore University of Social Sciences**

**Presented to Singapore University of Social Sciences in partial fulfillment of the requirements for the**

**Degree of Bachelor of Science in Business Analytics**

**2022**

[Chapter 1: Introduction 1](#_Toc111420620)

[Chapter 2: Literature Review 3](#_Toc111420621)

[2.1 Twitter as a Source for Environmental Studies 3](#_Toc111420622)

[2.2 Sentiment Analysis 5](#_Toc111420623)

[2.3 Applications of Sentiment Analysis and Environmental Studies 7](#_Toc111420624)

[Chapter 3: Data Understanding and Preparation 9](#_Toc111420625)

[3.1 Data Collection 9](#_Toc111420626)

[3.2 Data Understanding 12](#_Toc111420627)

[3.3 Data Preparation 16](#_Toc111420628)

[Chapter 4: Proposed Modelling and Evaluation 19](#_Toc111420629)

[Chapter 5: Proposed Schedule 20](#_Toc111420630)

[References 21](#_Toc111420631)

# Chapter 1: Introduction

Secretary-General of the United Nations (UN), António Guterres, made a profound statement that “making peace with nature is the defining task of the twenty-first century” (United Nations, 2020). Climate change has been one of the most prominent issues in recent decades, sparking contentious debates across the globe as growing emphasis is placed on protecting and restoring our planet. Every aspect of human lives – food, water, ecosystems, living environments, and health – are directly and indirectly tied to and impacted by the earth’s climate. Despite this, human activities are the predominant contributors to the rapid degradation of the environment. In fact, over the last 150 years, human activities were responsible for almost the entire increase in greenhouse gases (United States Environmental Protection Agency, 2022). In the United States (US), the burning of fossil fuels was the largest contributing source of greenhouse gas emissions from human activities (United States Environmental Protection Agency, 2022).

One mitigating solution to burning fossil fuels for energy is to adopt low-carbon energy sources which enable energy generation with reduced amounts of carbon emissions. Low-carbon energy encompasses renewable, green, and clean energy sources. Renewable energy is energy that is produced from natural sources, such as wind and sunlight, which replenish themselves faster than they are consumed and thus, produces much lower emissions compared to burning fossil fuels (United Nations, n.d.). Some argue that renewable, green, and clean energy sources differ slightly by definition as clean energy, albeit producing zero emissions, is not necessarily renewable. For instance, some forms of bio-gas sources like manure, household waste, or organic matter which may not always be completely renewable but are regarded as clean energy sources. Hence, in this study, we look at low carbon-producing energy sources as a whole.

The successful implementation of greenhouse gas mitigations like low-carbon energy sources is highly dependent on the public’s willingness and active involvement in adapting energy consumption behaviours (Zhang et al., 2022). The public’s views and opinions on low-carbon energy development benefits and burdens hold valuable insights that influence their decision making to adopt low emission sources. This information will be vital to corporations, governments, policymakers, and international organisations like the UN, which can tap on the knowledge of public preferences and perceptions to design better adoption strategies and low-carbon emission policies (Zhang et al., 2022). As such, public support and acceptance for low-carbon energy are crucial determinants for the low-carbon energy transition (Kim et al., 2021).

In today’s digital age, a large portion of people’s social interactions are shifted online. Social media has become an important platform where people interact, shared ideas, raise awareness, and find a sense of belonging with likeminded communities. As such, social media has become an increasingly important tool to collect, analyse, and present data for research that involves understanding communications amongst online users (Zhang et al., 2022). Using social media to understand public perception curbs the selection and response biases that arise from methods like surveys and interviews (Kim et al., 2021). For instance, people who have more interest in the low-carbon energy transition are more likely to respond for interviews and surveys which may yield bias (Kim et al., 2021). Furthermore, social media offers massive data volumes which will require a lot of time and costs to collect the same amount of data through conventional means. Of the various social media platforms, Twitter is a popular platform where users interact and has been increasingly used by researchers to analyse public opinion (Reyes-Menendez et al., 2018).

In this research, we prove that Twitter is a viable source for collecting real-time data pertaining to low-carbon energy sources which can serve as social indicators for the low-carbon energy transition. We will collect data related to low-carbon energy sources from users in Europe and the US. This data will be analysed to understand how public perceptions, areas of interests and concerns related to low-carbon energy change over time. Through spatiotemporal analysis, we can uncover meaningful insights on how low-carbon energy is perceived by online users and whether these sentiments changed over time. The findings can empower decision makers to better understand the developments and obstacles of the low-carbon energy transition faced by the general public.

# Chapter 2: Literature Review

As growing emphasis is being placed on protecting our environment, many studies have been conducted to continually assess efforts towards attaining sustainable development goals. Such studies often provide quantitative data and measures that corporations, governments, and international organisations like the UN can rely on to evaluate their progress regarding the challenges faced within the broad umbrella of sustainability – ranging from pollution reduction to better managing natural resources.

## 2.1 Twitter as a Source for Environmental Studies

In order to obtain information for decision-making on a topic, numerous studies tap on social media to assess the opinions of individuals. Among the social media platforms available, many researchers tend to lean towards Twitter as it most favours opinion sharing and is useful for the extraction of factors from public opinion in social media (Reyes-Menendez et al., 2018). Compared to Facebook, which has limitations with private and semi-private groups and profiles, Twitter facilitates open communication and allows users to share opinions in a manner that is the most accessible publicly – providing researchers more information compared to other platforms (Reyes-Menendez et al., 2018).

Information can easily be found on a given topic by searching for specific keywords, which returns all tweets containing these words. Among other relevant studies is the research conducted by Kim et al. (2021), which examined public sentiment towards solar energy in the United States (US) by extracting tweets that contained keywords specific to solar energy. The motivation of the study was to determine public support and acceptance for renewable energy, which are vital determinants for the low-carbon energy transition. Likewise, Ballestar et al. (2020) analysed Tweets containing the words “sustainability” or “sustainable” to investigate how these terms were used on social media and the content it entails. Their research also tried to find out the extent of which users’ perceptions on social media are aligned with that of scientists.

Twitter also allows users to group published tweets on topics through hashtags (represented by ‘#’) which facilitates relevant discussions surrounding a given topic. Tweet replies and retweets with a hashtag can also be visualised together, which allows users with common interests to communicate easily and allows researchers to easily identify discussions surrounding the topic of interest. One such study conducted by Reyes-Menendez et al. (2018) looked into the hashtag #WorldEnvironmentDay to identify economic, social, cultural, and environmental factors related to public health and environmental sustainability that most concern Twitter users.

Another study by Dahal et al. (2019) tapped on timestamps and geographic coordinates of tweets to analyse public opinion on climate change, which enabled the comparison of climate change discussions between different countries over time. This was possible as geotagged tweets contain geographical coordinates, like longitude and latitude, as well as timestamps. Similarly, research by Zhang et al. (2022) analysed people’s perceptions of greenhouse gas emissions and preferences for renewable energy resources in the US, Australia, and Europe to better understand their reactions and views towards mitigation policies. Using a sample of Twitter data, the study was able to discover themes of discussion, measure sentiments towards these resources, and compare sentiments with interests to identify policy-improvement opportunities.

The aforementioned studies are amongst many that have paved the way for environmental studies by using data from Twitter to research people’s preferences, perceptions, and emotions, thereby overcoming limitations of collecting data through conventional methods. This study complements previous studies by looking into online public perceptions surrounding renewable energy to identify social, economic, or political gaps as well as potential opportunities for improvement.

## 2.2 Sentiment Analysis

Sentiment analysis has been increasingly utilised to identify the polarity of sentiments, or emotions, expressed in texts (Zhang et al., 2022). It focuses on extracting the positive or negative opinions in texts as well as the intensity of these sentiments (Zhang et al., 2022). Although it is often used to evaluate consumers’ opinions towards products and services, it can also be used to understand communication in online environments and to conduct research on the impact of social media events (Reyes-Menendez et al., 2018). Essentially, sentiment analysis can be used to analyse user opinions expressed in topic-based comments on Twitter through machine learning models, as demonstrated in the study by Reyes-Menendez et al. (2018). In the study, the researchers trained a Support Vector Machine (SVM) algorithm using a sample of tweets related to sustainability and the environment to classify tweets’ sentiments. The analysis allowed Reyes-Menendez et al. (2018) to determine key factors that most concerned users by identifying specific topics with negative sentiment. The study concluded that factors with negative public opinion were ​​water and air pollution, massive industrialisation, climate change, deforestation, displacement of communities, and diminishing biodiversity. Reyes-Menendez et al. (2018) concluded in their research that positive factors are those where right actions were taken, whereas neutral and negative factors needed to be improved.

Sentiment analysis on social media data can help identify socio-economic, cultural, and political factors for improvement and empower important decision-making. The study by Kim et al. (2021) utilised a “Robustly optimised Bidirectional Encoder Representations from Transformers pre-training Approach (RoBERTa)” sentiment classification model to perform the classification task. Their research found that public sentiment was more positive in the Northeast US region compared to the South and also in states with more democratic voters in the 2020 presidential election. States with a more mature solar market and consumer-friendly net metering policies also had more positive public sentiment towards solar energy (Kim et al., 2021). Zhang et al. (2022) used a lexicon-based sentiment analysis method, which has a list of all sentimental words and maps the words to specific values, for their study. The distribution of sentiments towards the various renewable energy sources were plotted to analyse the overall sentiments across Australia, Europe, and the US. Thereafter, Zhang et al. (2022) looked into the differences in people’s preferences for energy resources in these countries by using their sentiments to gauge satisfaction. The study concluded that people’s origin was a determinant for their energy source preference which will enable policy-makers to consider civilians’ perceptions when establishing sustainable solutions. This study further advocates the use of social media to evaluate sentiment in real-time and demonstrates the practical application of social media analysis.

Furthermore, sentiment analysis can aid in comparing and contrasting the nature of topic discussions. Dahal et al. (2019) and Ballestar et al. (2020) applied “Valence Aware Dictionary and sEntiment Reasoner (VADER)”, a rule-based model that can manage a variety of social media content and return their sentiment polarity, to determine the overall attitudes and feelings of tweets. The sentiment analysis by Ballestar et al. (2020) found that 84.64% of tweets about sustainability were either positive or neutral and 15.35% of tweets were negative. Their research also identified topics linked to sustainability for positive and negative sentiments. Ballestar et al. (2020) found that the word ‘new’ was highly linked to positive sentiments, and negative sentiments were highly associated with ‘climate crisis’. This was taken a step further by Dahal et al. (2019) who compared the change in sentiment of tweets pertaining to climate change between different countries over time. The sentiment analysis showed that the overall sentiment was negative, predominantly because of demeaning terms used in heated discussions and negative responses towards current events. Having added temporal and geospatial factors, Dahal et al. (2019) was also able to uncover that there were spikes in sentiments, which had a direct correspondence with significant real-world events. They found that climate change discussions occur mainly as part of a negative reaction to current events, more so when users are discussing political or extreme weather events.

## 2.3 Applications of Sentiment Analysis and Environmental Studies

The study by Reyes-Menendez et al. (2018) illustrated that the analysing sentiments expressed through tweets can help identify cultural, economic, social, and environmental factors that users were concerned about. The information can then be utilised by companies or non-profit organisations with decision making and environment-related initiatives. Moreover, these findings can serve international organisations like the UN as well as government bodies in strategizing and decision-making towards attaining sustainability goals. For instance, US states that wish to garner public support for solar energy can consider supporting the growth of solar businesses and implementing consumer-friendly net metering policies, as found by Kim et al. (2021). Likewise in the research by Zhang et al. (2022), which will enable policy-makers to determine crucial development directions based on people’s attitudes and preferences towards renewable energy sources.

**2.4 Limitations in Previous Studies**

Firstly, there were limitations in using Twitter as a platform for data collection. As acknowledged by Kim et al. (2021), data from tweets may not be fully representative of all US citizens due to the demographics of the platform. Twitter users are younger and usually more politically liberal, hence tend to favour renewable energy development and possibly inflating the estimated sentiment score against the true sentiment score. As the sentiment classification is highly dependent on the social media platform’s demographics, this limitation can be addressed by using multiple platforms for data (Kim et al., 2021). Zhang et al. (2022) also highlighted that low internet diffusion rate in some regions causes a notable selection bias for areas with less education, lower income, or more old people. Data collection from social media also has a bias as people who are not vocal or whose views are doubtful may not be inclined to contribute to tweeting activity (Zhang et al., 2022).

Dahal et al. (2019) and Ballestar et al. (2020) faced inefficiencies as only tweets in English were examined. For Dahal et al. (2019), sentiment analysis was developed with English in mind and classified non-English tweets as neutral. Whereas Ballestar et al. (2020) deliberately limited the sample to only English tweets. Dahal et al. (2019) suggests a way to overcome this is by translating the tweets or adopting a supervised sentiment analysis technique that can handle multiple languages if there is enough data in foreign languages.

Both studies by Reyes-Menendez et al. (2018) and Ballestar et al. (2020) acknowledged inefficiencies in the time horizon of the data. The data should be validated such that there is no seasonal bias by analysing tweets in a different time period to ensure conclusions remain stable (Ballestar et al., 2020). Further, Ballestar et al. (2020) indicated that there could be more depth in their analysis by understanding how discussions evolve over time and through a geospatial view of the discussions, however, this limitation was addressed through the work of Dahal et al. (2019).

There were also weaknesses in the techniques used for sentiment analysis as faced by Kim et al. (2021) and Reyes-Menendez et al. (2018). Kim et al. (2021) highlighted that the model was insufficient in capturing fine-grained emotions such as frustration, fear, sarcasm, etc. Likewise, Reyes-Menendez et al. (2018) acknowledged that ironies and sarcasm posed a challenge as the algorithm tended to classify such tweets as neutral. Both studies suggested that such emotions be better captured through machine learning which can be trained to mitigate this weakness.

# Chapter 3: Data Understanding and Preparation

## 3.1 Data Collection

For data pertaining to low-carbon energy sources, we collected tweets that contained ‘renewable energy’, ‘clean energy’, and ‘green energy’, along with the hashtags ‘#renewableenergy’, ‘#cleanenergy’, and ‘#greenenergy’. The data collected consisted of tweets from June 1, 2021, to June 1, 2022. This will allow us to better understand how perceptions change over a one year period as well as to identify any external factors that may have influenced public sentiment. Through a preliminary scrape, it was found that a lot of tweets did not have location data and the data comprised tweets with various languages. As such, the search query for the scrape was filtered to specifically retrieve only tweets that were in English and from Europe and the US. By adding the specific location in the filter, the scraped tweets were all geotagged and contained location data – which is required for this study. Only tweets in English were scraped due to the researchers’ linguistic limitations and to facilitate the analysis method which may not be able to handle multiple languages accurately.

The data was collected with snscrape, a scraping tool for social networking services, which runs on Python. An alternative tool was to scrape data using the Twitter Application Programming Interface (API), however, Twitter’s API limits requests every 15 minutes, which limits the number of tweets one can extract and limits extracting data from archives. Thus, snscrape was the preferred choice. The scraped data was then stored into a DataFrame to facilitate visualisation, cleaning, and analysis. Figure 1 shows the code to collect the relevant tweets and Table 1 shows how each field in the dataset can be interpreted. Figure 2 shows the output of the scraped data and Table 2 shows a sample of tweets in the dataset. There was a total of 6 scrapes, which generated 6 datasets. The changes made in the query for each scrape are shown in Table 3.



Figure 1: Code to scrape renewable energy tweets from the US

|  |  |
| --- | --- |
| **Field Name** | **Interpretation** |
| date\_time | Date and time that the tweet was posted |
| content | Tweet content |
| hashtags | The hashtags in the tweet (if any) |
| language | Tweet language |
| user\_location | User’s location |
| coord | Geographical coordinates where the tweet was posted |
| place | Country and city where the tweet was posted |

Table 1: Data field interpretations

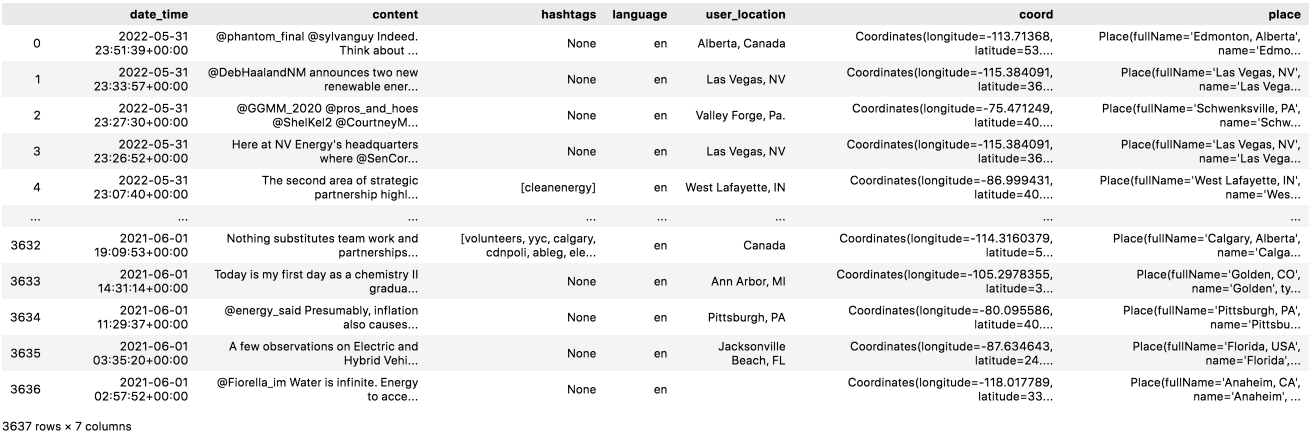


Figure 2: DataFrame containing scraped tweets

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **date\_time** | **content** | **language** | **user\_location** | **coord** | **place** |
| 2022-05-31 23:51:39+00:00 | @phantom\_final @sylvanguy Indeed. Think about it, years from now, new EVs should be made with nearly 100% recycled batteries. Combined with energy generated from renewable sources, and it is clear how personal transportation's impact on the the environment will be greatly reduced. | en | Alberta, Canada | Coordinates(longitude=-113.71368, latitude=53.395531) | Place(fullName='Edmonton, Alberta', name='Edmonton', type='city', country='Canada', countryCode='CA') |
| 2022-05-31 23:33:57+00:00 | @DebHaalandNM announces two new renewable energy initiatives at NV Energy on Tuesday afternoon, including new renewable energy offices that will streamline permitting for projects. Story coming. https://t.co/5bg8X49Mq2 | en | Las Vegas, NV | Coordinates(longitude=-115.384091, latitude=36.129459) | Place(fullName='Las Vegas, NV', name='Las Vegas', type='city', country='United States', countryCode='US') |

Table 2: Sample of scraped tweets

|  |  |
| --- | --- |
| **Keywords searched** | **Location** |
| (renewable energy) OR (#renewableenergy) | USA |
| (clean energy) OR (#cleanenergy) | USA |
| (green energy) OR (#greenenergy) | USA |
| (renewable energy) OR (#renewableenergy) | Europe |
| (clean energy) OR (#cleanenergy) | Europe |
| (green energy) OR (#greenenergy) | Europe |

Table 3: Changes in each query

## 3.2 Data Understanding

Upon looking through the collected tweets, it was found that the field ‘user\_location’ is unreliable for our analysis in the later stage. This field contains the location that users have on their profile but they can enter anything they want. Some users entered locations like ‘Planet Earth’ or ‘I am lost’ as their profile location. It was verified, however, that the fields ‘coord’ and ‘place’ are accurately geotagged according to the location that was queried and did not contain any missing geotags. Hence, we will use the data from ‘coord’ and ‘place’ for the geospatial portion of analysis in the later stage.

For the next part of this subsection, we will be analysing tweet volume. The number of tweets in the 6 datasets contain geotagged tweets from June 1, 2021, to June 1, 2022 and are shown in Table 4.

|  |  |
| --- | --- |
| **Dataset** | **Number of tweets** |
| Renewable Energy tweets from the US | 3,637 |
| Clean Energy tweets from the US | 5,317 |
| Green Energy tweets from the US | 2,138 |
| Renewable Energy tweets from Europe | 3,689 |
| Clean Energy tweets from Europe | 1,091 |
| Green Energy tweets from Europe | 2,344 |

Table 4: Number of tweets in each dataset

We will now look at the distribution of tweet data volume over the duration of one year for each dataset, starting with the 3 datasets containing US tweet data. Figure 3 shows the volume of renewable energy tweets in the US. Figure 4 shows the volume of clean energy tweets in the US. Figure 5 shows the volume of green energy tweets in the US.

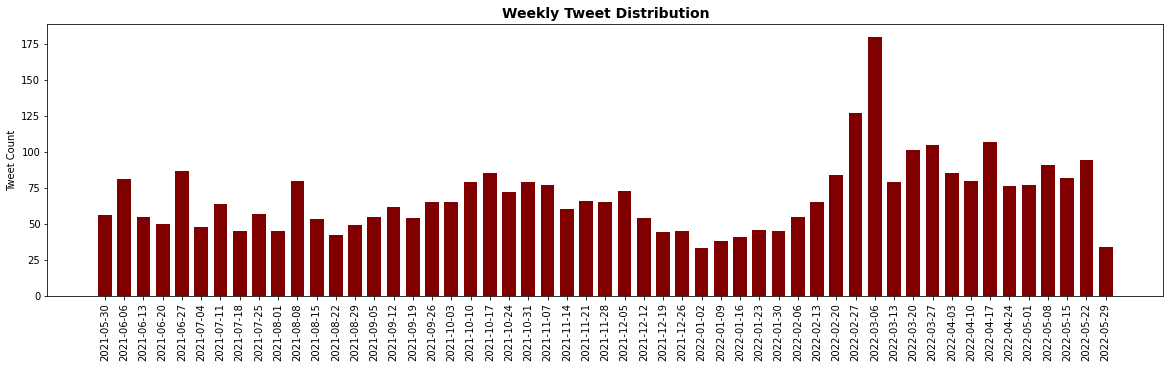


Figure 3: Volume of renewable energy tweets in the US

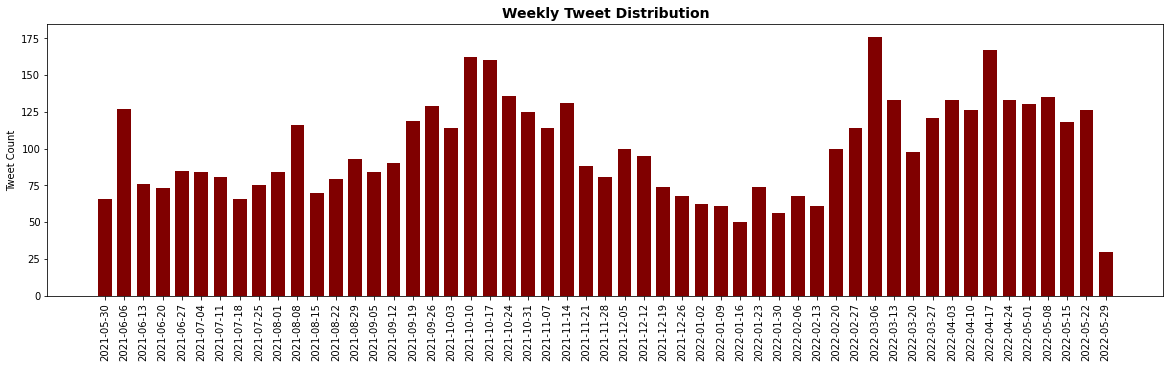


Figure 4: Volume of clean energy tweets in the US

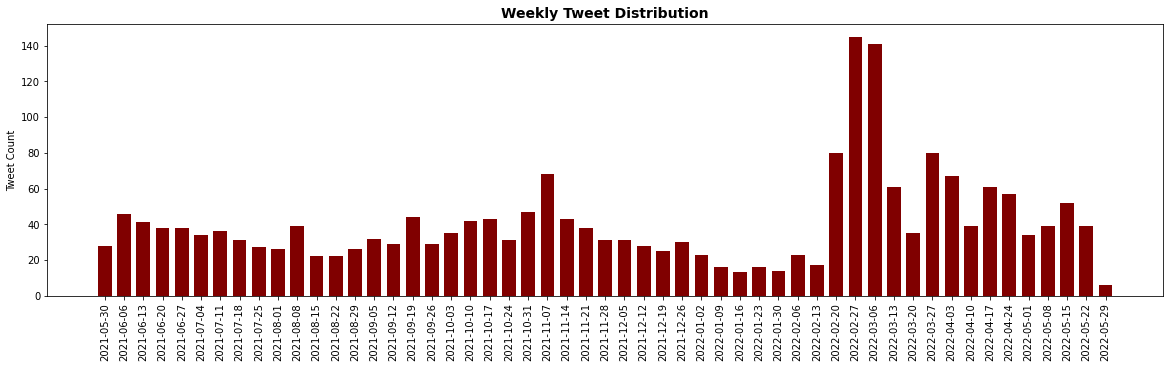


Figure 5: Volume of green energy tweets in the US

Of these 3 datasets, clean energy tweets had the highest volume – 5,317 tweets as shown in Table 4. Green energy tweets had the lowest volume (2,138 tweets) compared to renewable energy and clean energy. From the distribution of tweet volume in the 3 datasets, a significant spike in volume can be seen around early March 2022. This is the period when the Russia-Ukraine war intensified, likely sparking discussions around the impending energy crisis and spikes in energy prices. As the collected tweets were related to low-carbon energy sources, we can infer that the discussions stemming from the Russia-Ukraine war also created significantly more discussions around renewable energy, clean energy, and green energy. Russia is a major player in global energy markets as the world’s largest gas exporter (International Energy Agency, 2022). Hence, the spike in volume suggests that users could be raising awareness about the need for diversifying energy sources or discussing about alternative energy sources to consider as a result of the war.

We will now look at the 3 datasets containing tweet data from Europe. Figure 6 shows the volume of renewable energy tweets in Europe. Figure 7 shows the volume of clean energy tweets in Europe. Figure 8 shows the volume of green energy tweets in Europe.

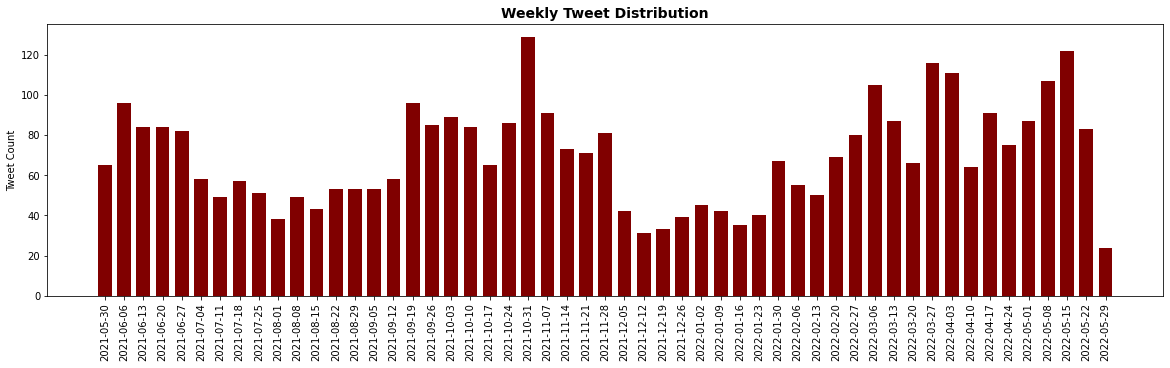


Figure 6: Volume of renewable energy tweets in Europe

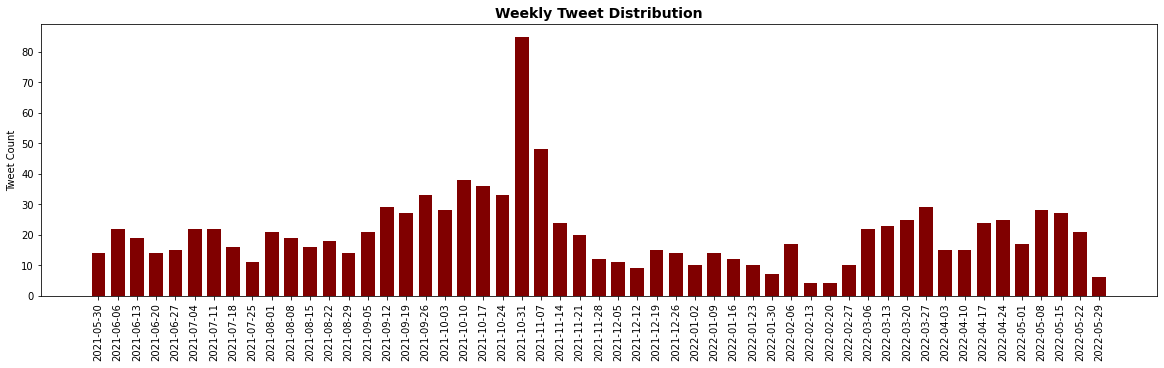


Figure 7: Volume of clean energy tweets in Europe

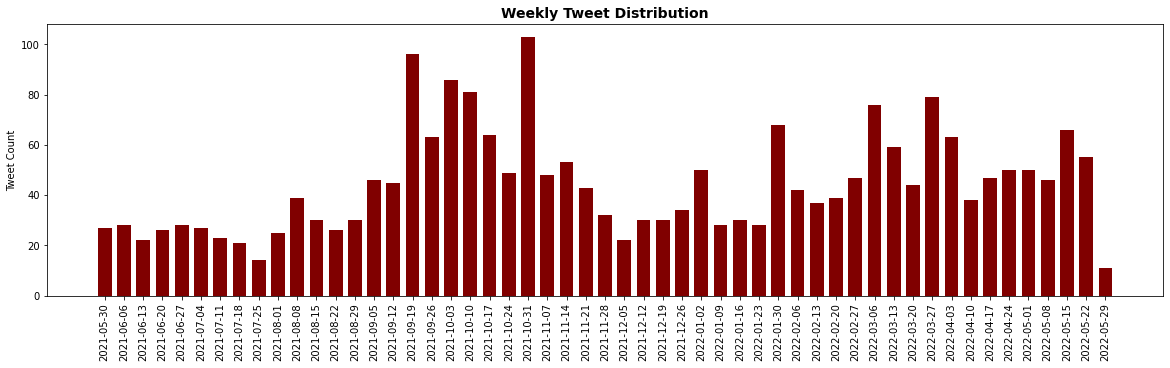


Figure 8: Volume of green energy tweets in Europe

In terms of overall volume, geotagged low-carbon energy tweets were less in Europe compared to the US. In Europe, renewable energy tweets had the highest volume, with 3,689 tweets over the time period. Clean energy tweets were the least in Europe with 1,091 tweets. Similar to tweet data from the US, there is a significant spike in tweet volume, but at the end of October 2021 instead. This is in line with the UN Climate Change Conference in Glasgow (COP26) which took place from October 31, 2021, to November 12, 2021. This highlights that significantly more discussions were sparked in Europe surrounding COP26, possibly regarding the outcomes and expectations of the parties. Renewable energy tweets also saw increased volumes after February 2022, however, upon further inspection, the volume is likely inflated by an account that started posting weather updates and raising awareness for solar energy approximately thrice a day.

## 3.3 Data Preparation

The first part of data preparation was to combine the tweet data by their region as we wish to analyse public sentiment towards low-carbon energy sources as a whole rather than as 3 separate sources. This means that the 6 datasets will be combined to become 2 datasets (one for tweets in Europe and the other for tweets in the US). The combination was done by concatenating the 3 DataFrames for each region using pandas. There were several duplicate tweets as some tweets contain more than one of the searched keywords and was scraped accordingly. For instance, a tweet may contain both ‘renewable energy’ and ‘clean energy’ and thus will be present in 2 datasets. These duplicates were removed from the dataset. Table 5 shows the initial number of tweets, number of duplicates, and the number of tweets after removing duplicates for each dataset.

|  |  |  |
| --- | --- | --- |
|  | **Tweets from the US** | **Tweets from Europe** |
| Before removing duplicates | 11,092 | 7,124 |
| Number of duplicates | 500 | 395 |
| After removing duplicates | **10,592** | **6,729** |

Table 5: Number of tweets after removing duplicates

The second part of data preparation involves cleaning the content in each tweet. Tweets often contain information that have no meanings or do not add value to the analysis and, therefore, should be removed. Such things include the mentioning of other users (‘@user’), quoted retweets (‘RT: …’), URLs, unwanted characters, or whitespace, which may affect the sentiment analysis scoring later on. Emojis and emoticons were not removed from the content as some sentiment analysis algorithms, such as unsupervised lexical-based methods, are able to handle and provide a score for emojis and emoticons. Hashtags are also kept as users often use words with hashtags as part of their sentences. A function was written using Regular Expression on Python to check if a tweet contains certain patterns in the text and to replace or remove that pattern accordingly. Figure 9 shows the function definition on Python and Table 6 shows the specific Regular Expression pattern to be searched for and the outcome of finding said patterns – whether it is replaced with another string or removed. The function is then run on all tweets in both datasets and stored as in a new column for cleaned tweets. Table 7 displays a few tweet contents before and after they are cleaned.

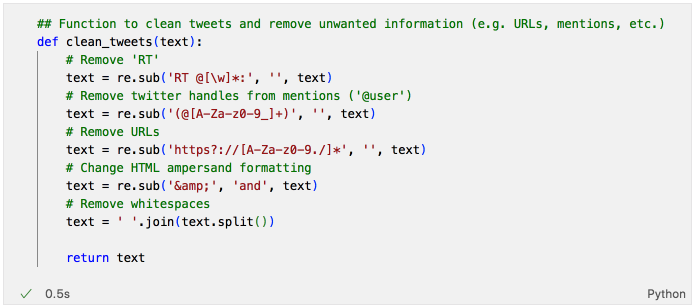


Figure 9: Function to clean tweet content

|  |  |  |
| --- | --- | --- |
| **Purpose of Regular Expression** | **Text pattern to search** | **Outcome** |
| Search for quoted retweets | RT @[\w]\*: | Remove |
| Search for mentioned users | (@[A-Za-z0-9\_]+) | Remove |
| Search for URLs | https?://[A-Za-z0-9./]\* | Remove |
| Search for HTML ampersand formatting | &amp; | Replace with ‘and’ |
| Search for whitespaces | ‘ ‘ | Remove |

Table 6: Regular Expression search pattern and outcome of search

|  |  |
| --- | --- |
| **Tweet content before cleaning** | **Tweet content after cleaning** |
| ‘TONIGHT: There are only a few hours left to pass a comprehensive #climate bill that prioritizes equity, jobs, &amp; justice! @DonHarmonIL, @RepChrisWelch - nearly 50 legislators won’t vote for an energy bill without a #FossilFreeFuture. #CleanEnergy #twill https://t.co/Z0D6yooaTw’ | ‘TONIGHT: There are only a few hours left to pass a comprehensive #climate bill that prioritizes equity, jobs, and justice! , - nearly 50 legislators won’t vote for an energy bill without a #FossilFreeFuture. #CleanEnergy #twill’ |
| ‘@energy\_said Presumably, inflation also causes fossil fuel prices to rise. So, inflation should be unimportant when comparing a new FF power plant to a new renewable plant.\nThey must be talking about comparing new renewable to existing FF.’ | ‘Presumably, inflation also causes fossil fuel prices to rise. So, inflation should be unimportant when comparing a new FF power plant to a new renewable plant. They must be talking about comparing new renewable to existing FF.’ |

Table 7: Tweet content before and after cleaning

# Chapter 4: Proposed Modelling and Evaluation

The proposed method for sentiment analysis is to use an unsupervised, lexicon-based model to classify tweet sentiments. The proposed model to be applied is “Valence Aware Dictionary for Sentiment Reasoning (VADER)”. According to the study by Ballestar et al. (2020), VADER is a robust rule-based model which does not need to be trained in advance and is computationally efficient – which is ideal in our study. One limitation of using VADER is that the dictionary used only contains English words and will score foreign language tweets a sentiment score of 0 (Dahal et al., 2019). However, this is not a major limitation in this study as only tweets in English were collected. Moreover, VADER is specifically attuned to handle sentiments expressed in social media content and it was found in prior studies to outperform even individual human scoring (Ballestar et al., 2020). However, as with all models, there is still the likelihood of erroneous classifications as found in the study by Dahal et al. (2019).

VADER works by mapping lexical features which are labelled according to their semantic orientation to return a metric score for the sentiment of the tweet. It not only returns if a tweet is positive or negative, but also how positive or how negative the tweet’s sentiment is. The score returned by VADER ranges from 1 (strongly positive) to –1 (strongly negative). In both studies by Dahal et al. (2019) and Ballestar et al. (2020), manual classification was involved to reduce the number of false negatives and false positives. This was done by expanding the ‘neutral’ threshold – for instance, Dahal et al. (2019) classified tweets that had scores between –0.25 to 0.25 as neutral. The same methodology for reducing the occurrences of wrong classifications can be adopted in this study.

After sentiment analysis is complete, tweets will be analysed based on geographical location and over time. It is also proposed that tweets be analysed based on topics – by using hashtags or common keywords as an indication, hence, the ‘hashtag’ field in the dataset. The objective of this is to discover if certain sentiments are attributed to a certain aspect of low-carbon energy sources, for example, cost or inadequate policies. Geospatial analysis will also aid in identifying if a certain geographical area is more likely to possess certain sentiments. Using VADER, we will also be able to better understand how extreme the sentiments are. These findings can empower decision makers to better understand the developments and obstacles of the low-carbon energy transition faced by the general public.

# Chapter 5: Proposed Schedule



# References

Ballestar, M. T., Cuerdo-Mir, M., & Freire-Rubio, M. T. (2020). The Concept of Sustainability on Social Media: A Social Listening Approach. *Sustainability*, *12*(5), 2122. https://doi.org/10.3390/su12052122

Dahal, B., Kumar, S.A.P., & Li, Z. (2019). Topic modeling and sentiment analysis of global climate change tweets. *Social Network Analysis and Mining*, *9*(1), 24. https://doi.org/10.1007/s13278-019-0568-8

International Energy Agency. (2022). *Russia’s War on Ukraine*. Retrieved August 10, 2022, from https://www.iea.org/topics/russia-s-war-on-ukraine

Kim, S. Y., Ganesan, K., Dickens, P., & Panda, S. (2021). Public Sentiment toward Solar Energy—Opinion Mining of Twitter Using a Transformer-Based Language Model. *Sustainability 2021, 13*(5), 2673. https://doi.org/10.3390/su13052673

Reyes-Menendez, A., Saura, J. R., & Alvarez-Alonso, C. (2018). Understanding #WorldEnvironmentDay User Opinions in Twitter: A Topic-Based Sentiment Analysis Approach. *International Journal of Environmental Research and Public Health*, *15*(11), 2537. https://doi.org/10.3390/ijerph15112537

United Nations. (2020). *The UN Secretary-General speaks on the state of the planet*. Retrieved July 29, 2022, from https://www.un.org/en/climatechange/un-secretary-general-speaks-state-planet

United Nations. (n.d.). *What is renewable energy?* Retrieved July 29, 2022, from https://www.un.org/en/climatechange/what-is-renewable-energy

United States Environmental Protection Agency. (2022). *Sources of Greenhouse Gas Emissions*. Retrieved July 29, 2022, from https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions

Zhang, Y., Abbas, M., & Iqbal, W. (2022). Perceptions of GHG emissions and renewable energy sources in Europe, Australia and the USA. *Environmental Science and Pollution Research 29*(4), 5971–5987. https://doi.org/10.1007/s11356-021-15935-7