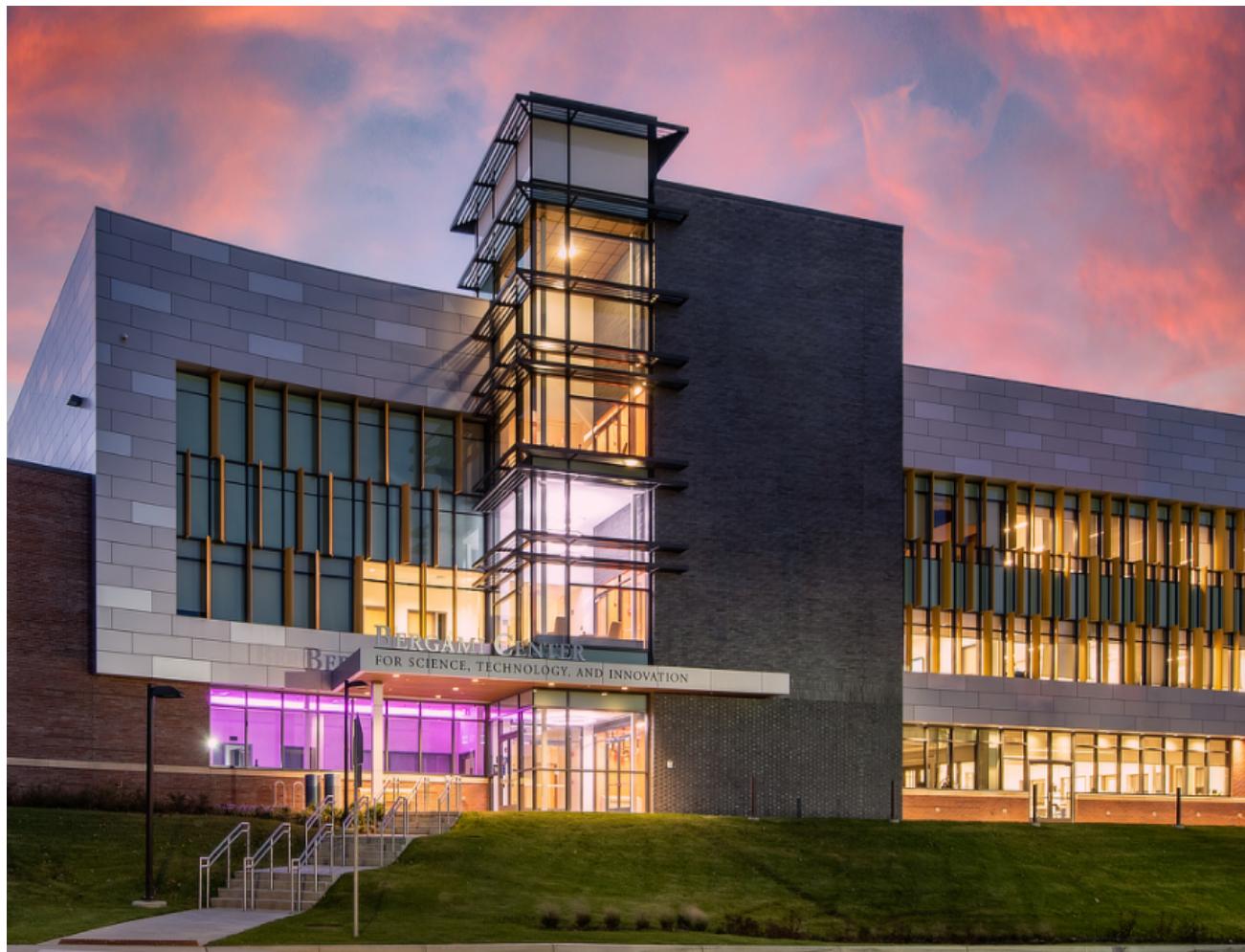


**SPRING 22**

Electrical & Computer Engineering & Computer Science (ECECS)

# TECHNICAL REPORT



# **CONTENTS**

<b>Project Name</b>	<b>1</b>
<b>Submitted on:</b>	<b>3</b>
<b>Technical Report</b>	<b>3</b>
<b>Executive Summary</b>	<b>4</b>
<b>Introductory Section</b>	<b>5</b>
<b>Methodology</b>	<b>6</b>
<b>Results Section</b>	<b>9</b>
<b>Discussion</b>	<b>12</b>
<b>Conclusion</b>	<b>13</b>

# Technical Report

## Twitter Sentiment Analysis

Submitted on: 03/09/2022

### Team Members:

Name 1 Guanyu Zhou

Name 2 Chunwang Yuan

Name 3 Preetam Soni

### Questions?

### Contact :

[cyuan3@unh.newhaven.edu](mailto:cyuan3@unh.newhaven.edu)

[gzhou2@unh.newhaven.edu](mailto:gzhou2@unh.newhaven.edu)

[psoni4@unh.newhaven.edu](mailto:psoni4@unh.newhaven.edu)



# Twitter Sentiment Analysis

## Executive Summary

Twitter sentiment analysis would provide valuable business insights that drive decisions, as it is able to analyze real-time tweets and classify the sentiment behind the message. A business brand would want to know what customers love about your brand, and what triggers their negative or positive emotions before making the next business decision.

We were implementing machine learning sentiment analysis models in order to compute public sentiment. Public sentiment is very important and valuable in business decisions such as predicting the brand opinion or stock price trend. Companies can estimate how well their products perform and decide which business move they are taking next, cutting off a negative product or promoting a positive product.

We examined sentiment analysis on Twitter data.



## Introductory Section

We are choosing twitter to work with because we believe Twitter is one of the most important social media platforms and it has a better evaluation of public sentiment than other social media platforms.

Public sentiment is very important and valuable in business decisions such as predicting the brand opinion or stock price trend. Companies can estimate how well their products perform and decide which business move they are taking next, cutting off a negative product or promoting a positive product.

This is a great project that gives us an opportunity to learn and implement machine learning.

We were implementing machine learning sentiment analysis models in order to compute public sentiment. However, due to the natural properties of human language that consist of unproductive characters, it was hard for us to implement models on the dataset.

In our project, we are using a natural language processing algorithm from the Twitter tweets dataset regarding three topics, nuclear war, Nike, and gas price through Twitter API.



# Methodology

Twitter is a social networking platform and allows users to post real-time tweets. Tweets are very short and limited to 140 characters. Besides text, a tweet can have emotions, target other users with @ symbol, and hashtags to mark topics.

we acquire tweets as our dataset from Twitter API and combine them with similar hashtags such #nuclear and #nuclearwar and etc. In the beginning, we created a Twitter API key and token in the Twitter developer portal for later access.



## Here are your keys & tokens

For security, this will be the last time we'll display these. If something happens, you can always regenerate them.

API key ⓘ



API secret key ⓘ



Bearer token ⓘ



We are using tweepy library to access twitter API in our jupyter notebook. The library tweepy is a specific library to be used to access twitter API. Our code is as shown below.

```
[21]: import tweepy as tw
import pandas as pd
import csv
import ssl
# your Twitter API key and API secret
my_api_key = "d83L0q2Cbe2[REDACTED]"
my_api_secret = "3meeaIpY0k6d7qRq3MLjncUDeAR[REDACTED]"
# authenticate
auth = tw.OAuthHandler(my_api_key, my_api_secret)
api = tw.API(auth, wait_on_rate_limit=True)
```

We access tweets through Twitter API and save all the tweet text as our dataset. There is one small problem here but we manage to solve it. As per query, our API access to Twitter allows each query to retrieve a maximum of 100 data, which means 100 rows each time when you run a query. In order to solve this, we run multiple queries with similar topics/hashtags, and combine them using pandas into a dataset. This is making our dataset comparative larger than just a single query. Therefore, our algorithm would make more sense.

```
In [50]: # Collect tweets
#test
for tweets in api.search_tweets(q="#nuclearwar ", lang="en", count=100):
    print(tweets.text)

RT @gbc2021: #GBC was listed on @coinsniper_net 🚀
Please vote here to be officially listed👉
https://t.co/eBvnztbSK4
(every 24h you can v...
RT @gbc2021: #GBC was listed on @coinsniper_net 🚀
Please vote here to be officially listed👉
https://t.co/eBvnztbSK4
(every 24h you can v...
RT @Jekova_NFT: Sad Truth 😢

#worldwar3 #WWIII #Ukraine #RussiaUkraineConflict #Russia #RussiaUkraineWar #nuclearwar https://t.co/0maNt96tQY
RT @MrMisskhanji: #UkraineRussiaWar #RussiaUkraine #nuclearwar #RussiaUkraineConflict
#Kharkiv

Hypocrisy of western world:

Freedom fighte...
RT @CarolynTGray: A very small number muted #nuclearwar as a top topic of conversation... And occasionally we chatted about w
eekends and ca...
```

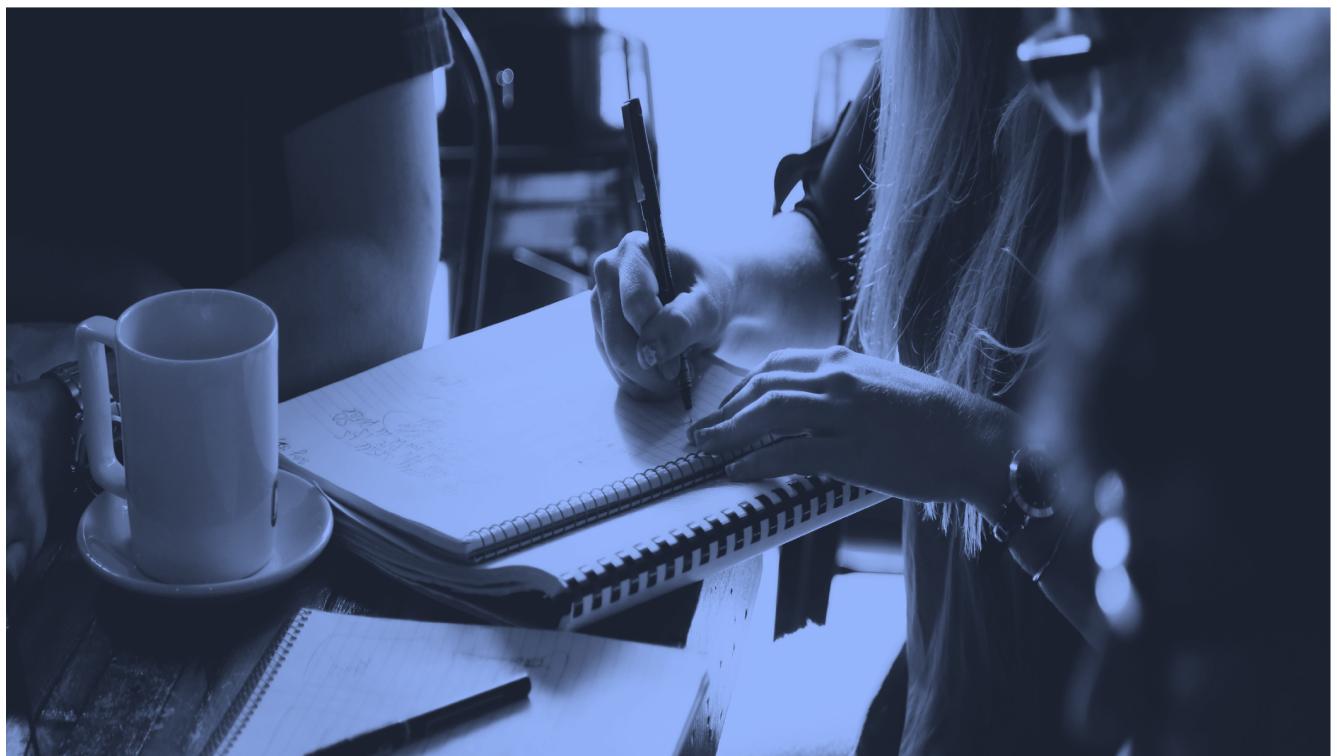
The dataset we retrieved looked like this.

```
In [5]: df.head()

Out[5]: 0
          1
          2
          3
          4
```

0	#NuclearWar: Americans today seem to take this...
1	RT @eliasamare: When something is repeated wit...
2	RT @subyroy: @cvneuvres @MarkAmesExiled #nuclea...
3	RT @subyroy: @JohnHemmings2 Hence the regime c...
4	RT @subyroy: @ZelenskyyUa Well a #nuclearwar #...

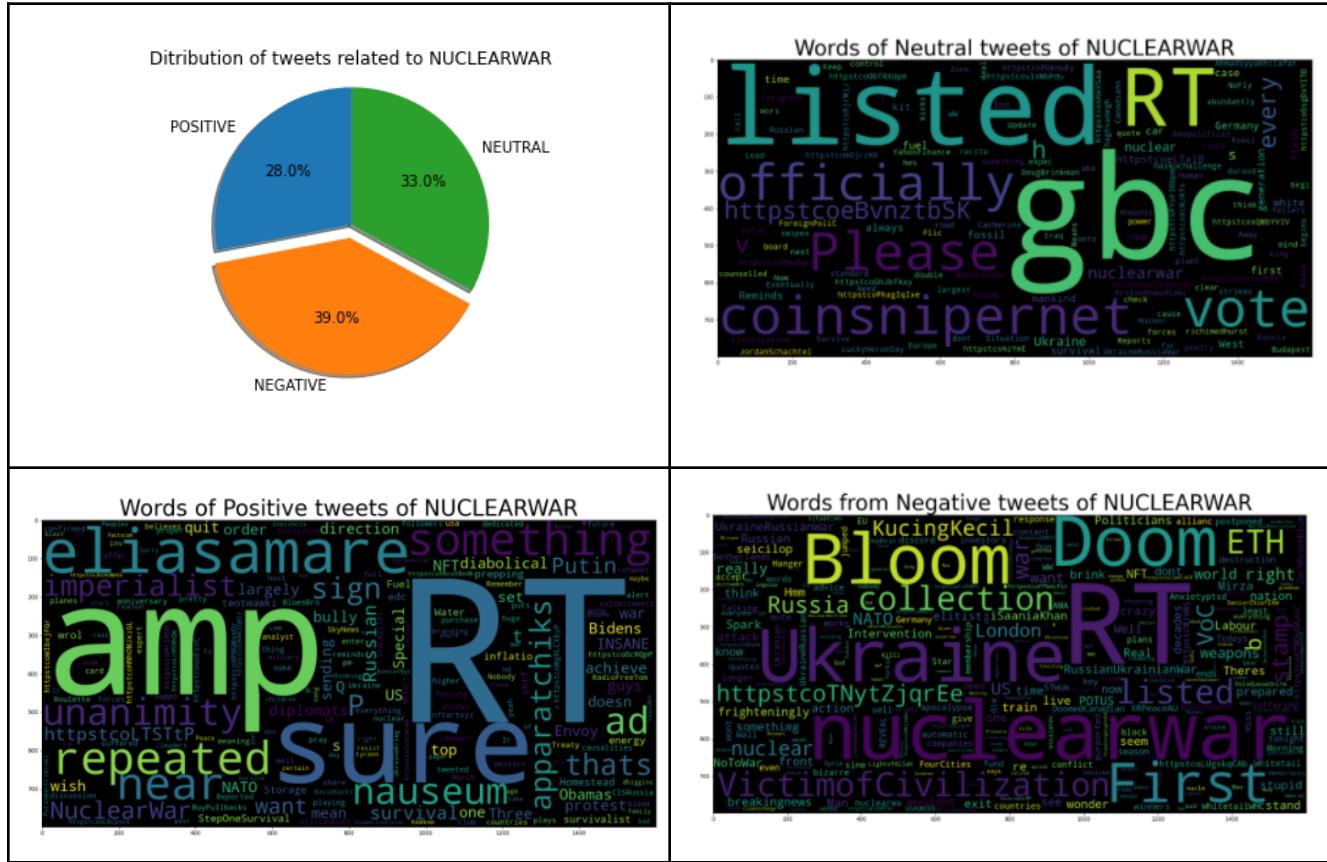
Libraries	Processing
string.punctuation	Cleaning and removing punctuations
re.sub	Cleaning and removing repeating characters
	Cleaning and removing URL's
	Cleaning and removing Numeric numbers
vaderSentiment.vaderSentiment.SentimentIntensityAnalyzer	<p><b>VADER</b> (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media.</p> <p>Using VADER to determine whether a sentence is positive, negative or neutral.</p>
matplotlib.pyplot	Visualizing the pie chart of positive negative and neutral tweets
WordCloud	Plot a cloud of words for positive negative and neutral tweets



## Results Section

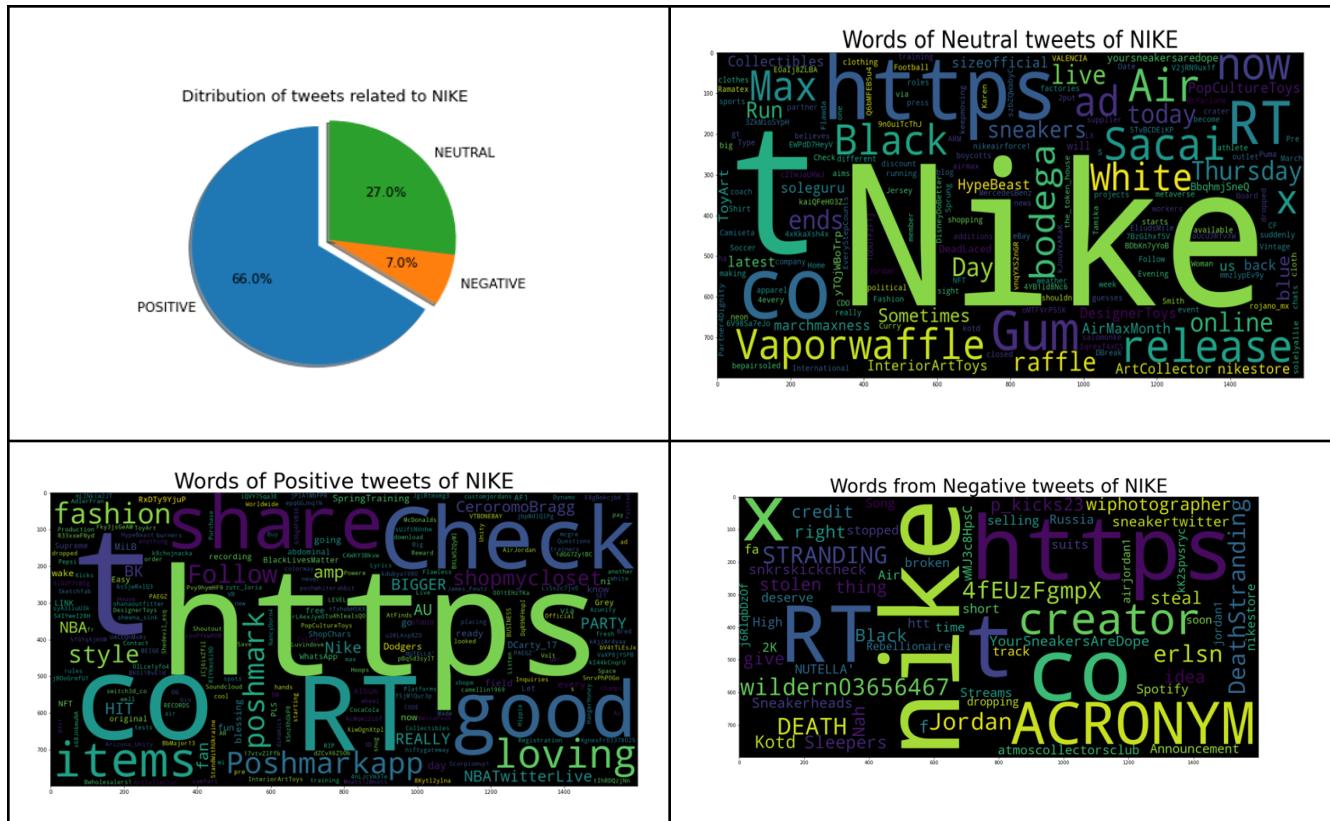
As mentioned above, we were using three datasets for the public sentiment which are nuclear war, Nike, and gas prices. Below are our empirical findings regarding the three topics. Each of them will be classified into three kinds of sentiments, positive, negative, and neutral.

The nuclear war finding has been narrowed down into four graphs below:



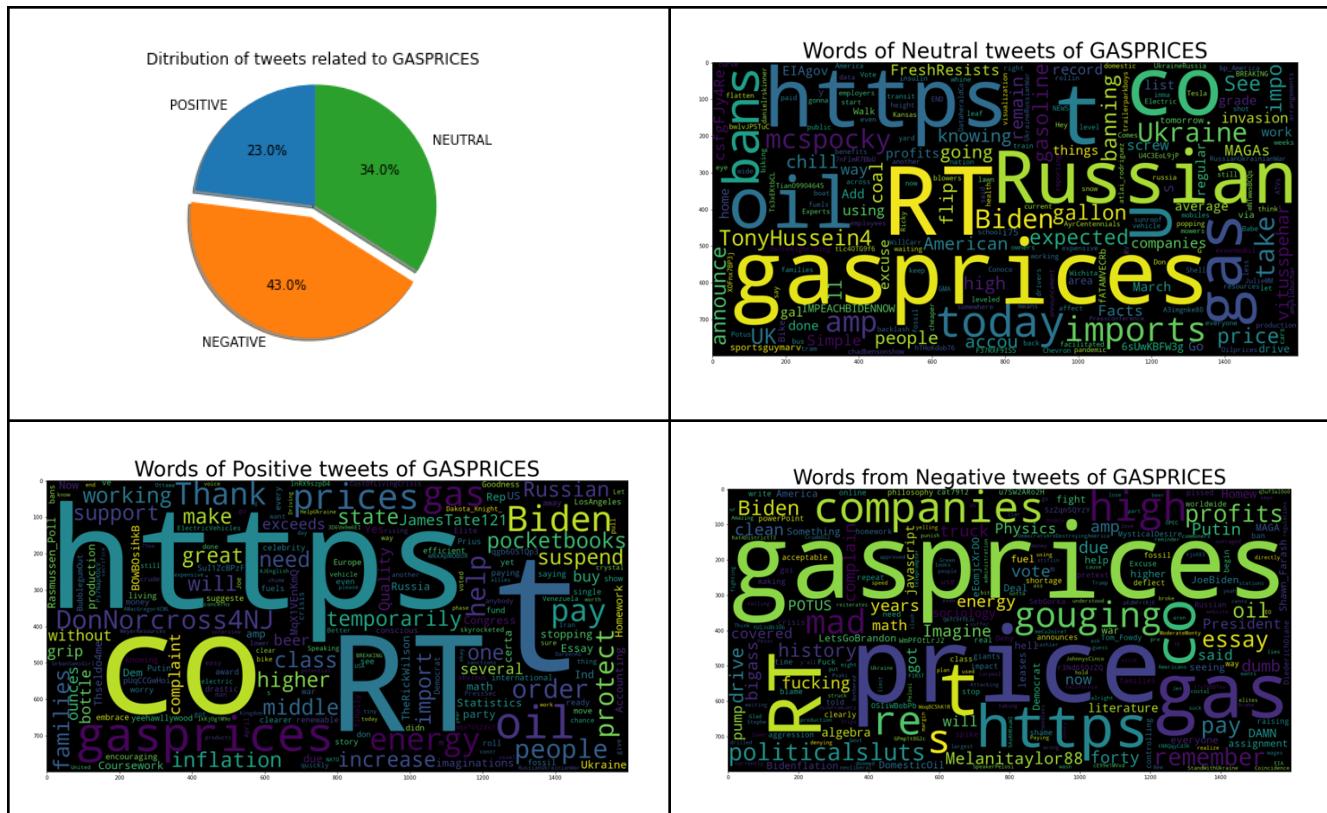
We originally assumed that nuclear war is a scary topic and the public sentiment regarding it would be mostly negative. However, the empirical finding using natural language processing algorithms is giving us a surprising result, with neutral sentiments taking up to 33% and positive sentiments taking up to 28% positive, while negative is only 39%. It is possible that most people might know nuclear war is scary and causes extinction of earth races, but since they have never experienced it, they might just worry about the possibility that it could happen, however, they do not have fear when talking about nuclear war. This might explain that the negative sentiments are only taking up to 39%.

The Nike brand finding has been narrowed down into four graphs below:



The hypothesis of public sentiment regarding the Nike brand falls under our assumption. Up to 66 percent, the majority of sentiments are positive. This is consistent with the impression that Nike propaganda has been advertising in the long term, sport and positive energy. Nevertheless, as a famous brand lasting many long years, its excellent product quality reputation must be also another factor that increases positive sentiment percentage.

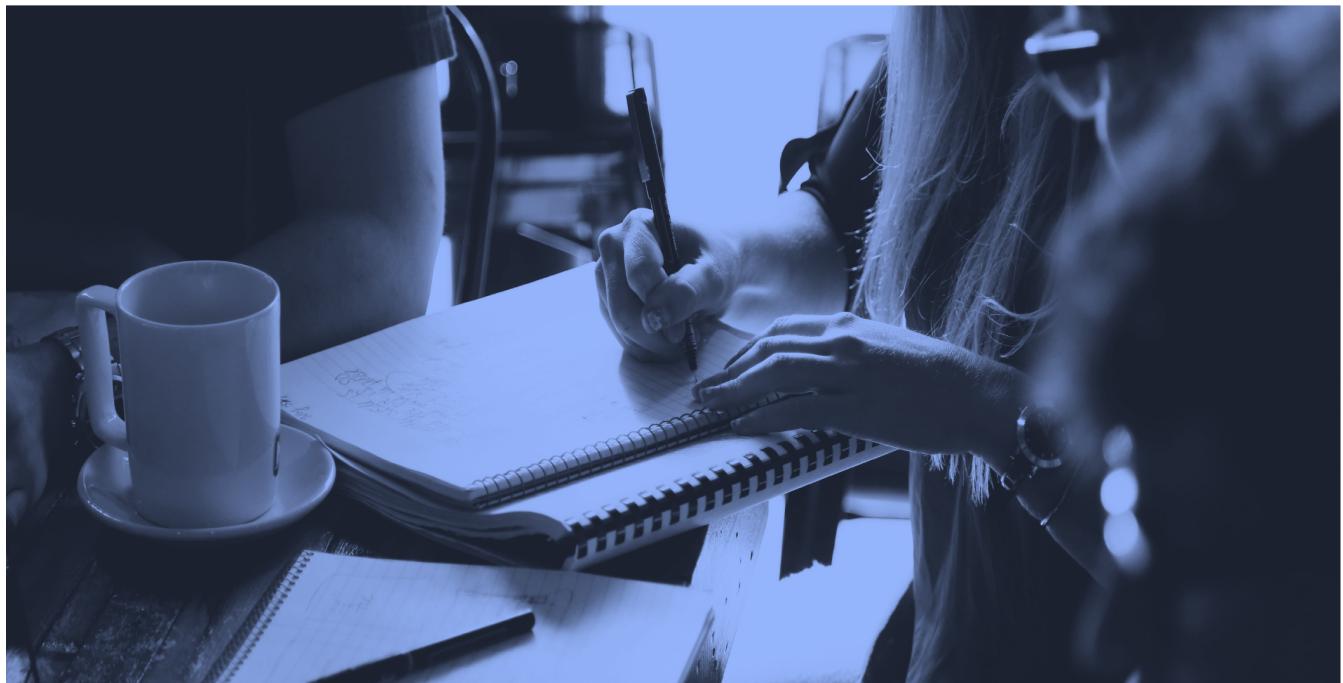
The gas price finding has been narrowed down into four graphs below:



The global gas price finding is exactly the same as our assumption. With 43% negative public sentiments and 34% neutral sentiments, combined with 77% in total that is not positive. This is explainable based on the nuclear war result, due to most people have never experienced the real nuclear war, therefore they might think the nuclear war is scary but they do not fear what they have never seen before. On the contrary, gas price is related to your daily life, the commodity price in shops, and the commute cost of going to work. Therefore, the war ongoing between Russia and Ukraine, two of the important petroleum countries in the world, has caused panic global wide and worries about the shortage of gas supplies in the market.

## Discussion

1. The Nike brand has mostly 66% positive sentiments. Does this mean the Nike advertising team is making great propaganda and people are fond of this product?
2. The news regarding the war between two important petroleum countries, Russia and Ukraine, triggers the nuclear topic accessing the global Twitter platform. Top regarding the nuclear war has only 39% public negative sentiment. Does this mean people are not worried about nuclear war?
3. While the war is ongoing between 2 important petroleum countries, the topic regarding gas price has 43% negative sentiment. Will oil and gold-related stock prices plummet because global investors are pessimistic?



## Conclusion

In our task of public sentiment analysis, our findings of three topics, nuclear war, Nike, and gas price, each of them can help businesses to decide the next move of the business. For example, as the nuclear public negative public sentiments skyrocket in recent days, gas and other commodities prices can fluctuate sharply. Similarly, Nike has increased negative public sentiment percentage, business would question what steps have been made wrong recently, such as a public criticized advertising video.

The difficulty that we faced was data size acquired through public access to Twitter API. Twitter limits the number of data each time people query down to a maximum of 100. The solution of dealing with small data was to combine similar hashtags data, however, this increased the number of queries times and workloads, and increased the uncertainty and inaccuracy of particular interest that we pursued.

Our approach is still in its developing stage and our models will need further improvement by adding more information. In the future, we will be exploring more linguistic analysis and using a higher level of Twitter API access to acquire bigger and integrated data.

