

# Sentiment Analysis

Unlocking Insights from Social Conversations on X (formerly Twitter)



# THE TEAM



**Bojan Mircheski**

Marketing Ops/Rev Ops/  
Data Analyst



**Hashim Khan**

PhD Student/Data Analyst



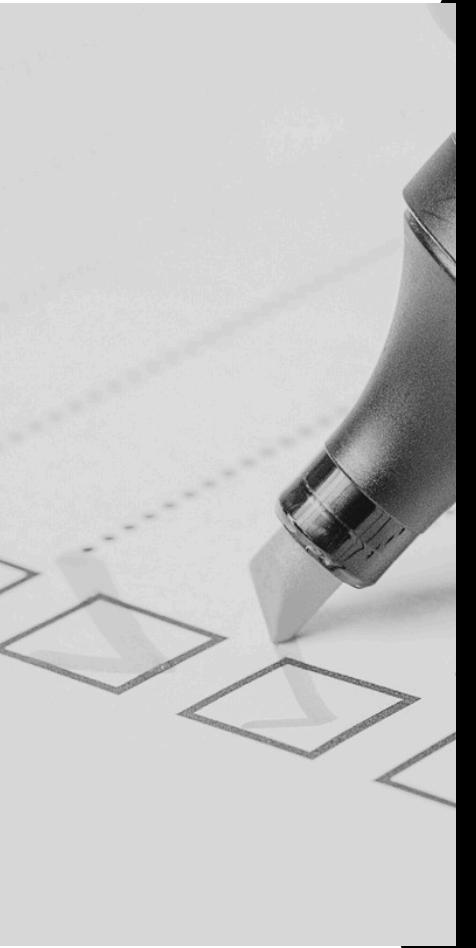
**Tristan Mertins**

Online Marketing/Advertising  
Data Analyst



# Agenda

- Motivation
- Objective and Method
- Stakeholders
- Trends
- Data Preparation
- Sentiment Analysis
- Reccomendations
- Future Work



# Five killed and emergency workers targeted as firework chaos mars Germany's New Year celebrations

By Sophie Tanno, Stephanie Halasz and Fre  
⌚ 3 minute read · Updated 11:06 AM EST,

 3 minute read · Updated 11:06 AM EST,



# **Five killed and hundreds arrested in Germany after New Year's Eve chaos**

n 2025 12:29 0

 Save  Add a comment   



# Explosive Motivation

# Public Sentiment

- **Objective:** X Public opinions
- **Method:** Examine user-generated content to gather insights
- **Focus Areas:**
  - Prevailing opinions about the fireworks
  - Relevant patterns and trends

X



# Stakeholders

## Environmentally Focused Groups:

- BUND
  - NABU
  - Berliner Klimatag
- Community Groups

## General Community Groups in Germany:

- Stadtteilvereine  
Nachbarschaftshilfe  
Deutschland
- Freiwillige  
Feuerwehren

## Animal Welfare Organizations:

- Deutscher Tierschutzbund
- Berliner Tiertafel



# Steps to be followed

Before Beginning Sentiment Analysis

TREND IDENTIFICATION	DATA COLLECTION	DATA STORAGE	CLEANING AND PREPROCESSING	TRANSLATING
-Relevant trending hashtags or keywords using Twitter trending sites	-Scraping Tweets Use APIs (e.g., Twitter API)	-Saved as a CSV-file	-Remove irrelevant content -Remove duplicates -Summarize/ adjust locations -Merge the files	-Tokenize -Translate into English

# Trend Identification



# Böller

# Kugelbombe

# Feuerwerk

# Silvester

# Data Collection



**Data Collection:** Tweets scraped via Python script

**Tools Used:** Tweepy library + X API + X Developer Account

**Free Account Limitations:**

- Max 100 tweets/month
- Data from last 7 days only

**Solution:** Multiple X Developer Accounts created

# Extracted Data

```
from tabulate import tabulate
# Convert the dataframe to a pretty table
print(tabulate(the_raw_data.tail(), headers='keys', tablefmt='fancy_grid')) # tablefmt='fancy_grid'
```

✓ 0.0s

on Order				
	username	name	location	tweet_text
94	blaulichtzecke	Blaulichtzecke ;	nan	Der hoffentlich letzte Nachklapp zum Thema #Böller. Insbesondere im Interesse der Verletzten ist zu hoffen, dass die E <a href="https://t.co/0x9c2Ud0Qi">https://t.co/0x9c2Ud0Qi</a>
95	KaiserfrontFan	Kaiserfront Extra Fan	Hamburg	@berlinerzeitung @GdPHauptstadt Ich glaube nicht das die #Böller da Das Problem ist ein Anderes und ein gewisser #MartinSellner hat au
96	KH82865104	K.Heinz 	Deutschland	Ich bin für ein generelles #Böllerverbot in Großstädten, vor allem : #Silvester2024 #Boeller  <a href="https://t.co/newF5CNw3k">https://t.co/newF5CNw3k</a> via @WEBDENews
97	Joffrey77855104	Rebel Radio	Next to them `ol barn	Letztes Jahr deutete sich schon an, daß das Imperium in Neukölln gew
98	citius_fortius1	Citius, Altius, Fortius	DACHI	@Karl_Lauterbach 🇩🇪 Solange die 60'000 ausländischen Ärzte . mit #Böller keine Wohnungen in Brand setzen . nicht mit Autos in #Weihnachtsmärkte fahren . oder Menschen umbringen:  Alles bestens, @Karl_Lauterbach.



# Cleaning the Data

- Removed irrelevant content
- Removed Emojis
- Removed duplicates
- Summarize/ adjust location data
- Merged the files

# Cleaned Data

```
### the steps for display in the presentation is here.  
justDisplay = pd.read_csv("/Users/hashimkhan/Documents/greenbootcamps/sentiment_analysis_X/cleaned_data/Silvester  
from tabulate import tabulate  
# Convert the dataframe to a pretty table  
print(tabulate(justDisplay.head(6), headers='keys', tablefmt='fancy_grid'))  
✓ 0.0s
```

	username	location	tweet_text
0	MiGAZIN	Germany	Polizei soll an rechtes Medium durchgestochen haben Ausschreitungen und Bl
1	M4xPeen	no_location	Welten zwischen und Schlimmster Datenleak seit der Bekanntgabe der Vorname
2	gabonn	Bonn	Die Polizeigewerkschaft fordert mit einer ein bundesweites Knapp 15 Millio
3	endthesealhunt	Germany	please RT EUwide ban on the sale of pyrotechnics for private purposes
4	PeterSchwarz11	Beckingen/ Saar	ist seit anderthalb Wochen vorbei doch der politische Streit zndet erst je
5	bellarocca21	berlin	es knallt seit stunden ich male bin allein mezzanotte knips ich nach gegen

# Merged Data Summary

```
▶ 
### the steps for display in the presentation is here.
justDisplay = pd.read_csv("/Users/hashimkhan/Documents/greenbootcamps/sentiment_analysis_X/merged_data/merged_data.csv")
# Create a summary of the DataFrame
summary = {
    "Column Name": justDisplay.columns,
    "Non-Null Count": justDisplay.count().values,
    "Data Type": justDisplay.dtypes.values,
}

# Convert to DataFrame
summary_df = pd.DataFrame(summary)

from tabulate import tabulate
# Convert the dataframe to a pretty table
print(tabulate(summary_df, headers='keys', tablefmt='fancy_grid'))
[4] ✓ 0.0s
...

```

	Column Name	Non-Null Count	Data Type
0	username	384	object
1	location	384	object
2	tweet_text	384	object
3	created_at	384	object
4	retweets	384	int64

# Translation

## Components:

### 1. MarianMTModel:

- A pretrained sequence-to-sequence model for translating text between languages.
- Optimized for specific language pairs, e.g., German to English (de-en).
- 

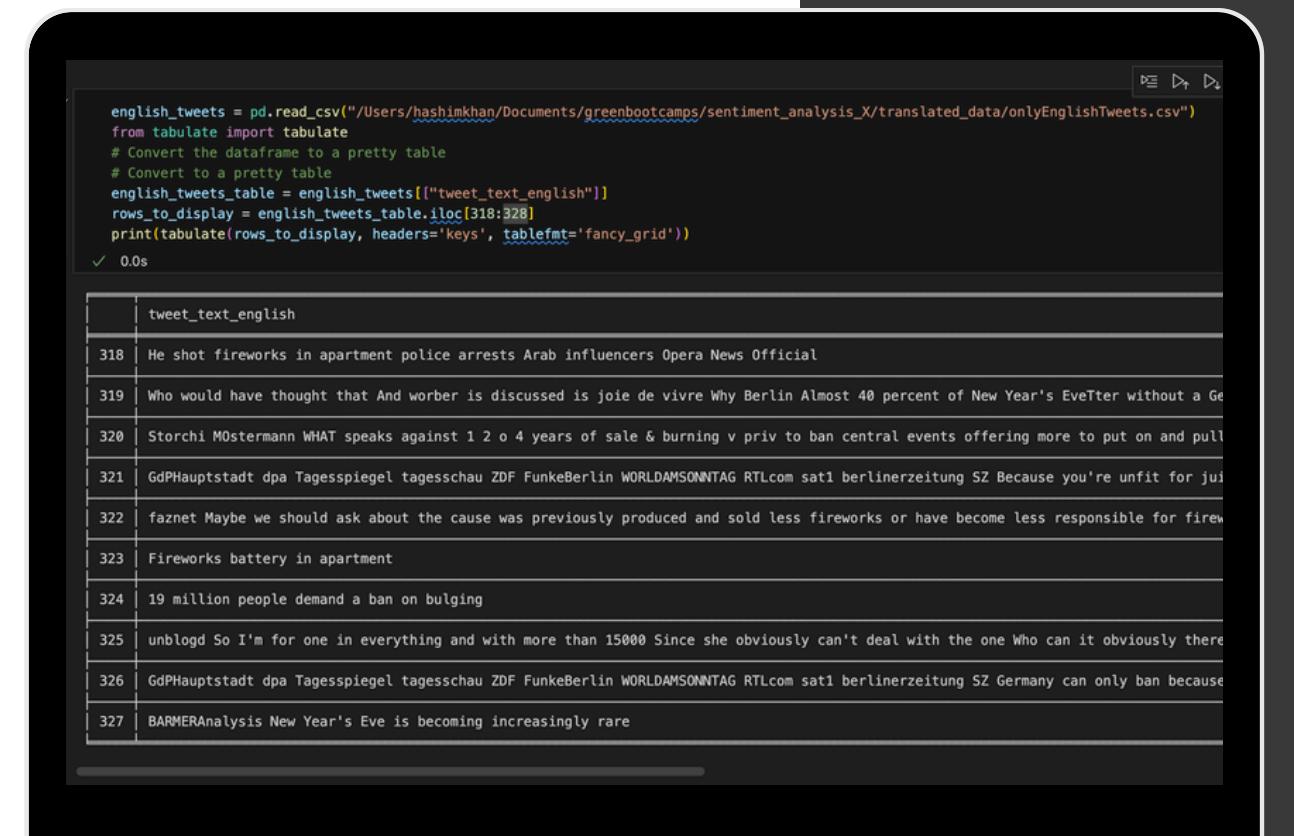
### 2. MarianTokenizer:

- Prepares input text for the MarianMT model by:
  - Tokenizing text into subword units.
  - Adding special tokens.
  - Converting text into numerical format for model processing.



# Translation Function

```
def translate_to_english(text):
    """
    Detect language and translate text to English if it's in
    German.
    """
    try:
        lang = detect(text)
        if lang == "de":
            # Prepare input for the model
            inputs = tokenizer.encode(text, return_tensors="pt",
max_length=512, truncation=True)
            # Generate translation
            outputs = model.generate(inputs, max_length=512,
num_beams=4, early_stopping=True)
            # Decode and return translated text
            translated_text = tokenizer.decode(outputs[0],
skip_special_tokens=True)
            return translated_text
        else:
            return text
    except Exception as e:
        return f"Error: {e}"
```



```
english_tweets = pd.read_csv("/Users/hashimkhan/Documents/greenbootcamps/sentiment_analysis_X/translated_data/onlyEnglishTweets.csv")
from tabulate import tabulate
# Convert the dataframe to a pretty table
# Convert to a pretty table
english_tweets_table = english_tweets[["tweet_text_english"]]
rows_to_display = english_tweets_table.iloc[318:328]
print(tabulate(rows_to_display, headers='keys', tablefmt='fancy_grid'))
```

0.0s

tweet_text_english
318   He shot fireworks in apartment police arrests Arab influencers Opera News Official
319   Who would have thought that And worker is discussed is joie de vivre Why Berlin Almost 40 percent of New Year's EveTter without a Ge
320   Storch! MÖstermann WHAT speaks against 1 2 o 4 years of sale & burning v priv to ban central events offering more to put on and pull
321   GdPHauptstadt dpa Tagesspiegel tagesschau ZDF FunkeBerlin WORLDAMSONNTAG RTLcom sat1 berlinerzeitung SZ Because you're unfit for ju
322   faznet Maybe we should ask about the cause was previously produced and sold less fireworks or have become less responsible for firew
323   Fireworks battery in apartment
324   19 million people demand a ban on bulging
325   unblog So I'm for one in everything and with more than 15000 Since she obviously can't deal with the one Who can it obviously there
326   GdPHauptstadt dpa Tagesspiegel tagesschau ZDF FunkeBerlin WORLDAMSONNTAG RTLcom sat1 berlinerzeitung SZ Germany can only ban because
327   BARMERanalysis New Year's Eve is becoming increasingly rare



```
english_tweets = pd.read_csv("/Users/hashimkhan/Documents/greenbootcamps/sentiment_analysis_X/translated_data/onlyEnglishTweets.csv")
from tabulate import tabulate
# Convert the dataframe to a pretty table
# Convert to a pretty table
english_tweets_table = english_tweets[["tweet_text_english"]]
rows_to_display = english_tweets_table.iloc[318:328]
print(tabulate(rows_to_display, headers='keys', tablefmt='fancy_grid'))
```

✓ 0.0s

	tweet_text_english
318	He shot fireworks in apartment police arrests Arab influencers Opera News Official
319	Who would have thought that And worber is discussed is joie de vivre Why Berlin Almost 40 percent of New Year's EveTter without a Ge
320	Storchi M0stermann WHAT speaks against 1 2 o 4 years of sale & burning v priv to ban central events offering more to put on and pull
321	GdPHauptstadt dpa Tagesspiegel tagesschau ZDF FunkeBerlin WORLDAMSONNTAG RTLcom sat1 berlinerzeitung SZ Because you're unfit for ju
322	faznet Maybe we should ask about the cause was previously produced and sold less fireworks or have become less responsible for firew
323	Fireworks battery in apartment
324	19 million people demand a ban on bulging
325	unblogd So I'm for one in everything and with more than 15000 Since she obviously can't deal with the one Who can it obviously there
326	GdPHauptstadt dpa Tagesspiegel tagesschau ZDF FunkeBerlin WORLDAMSONNTAG RTLcom sat1 berlinerzeitung SZ Germany can only ban because
327	BARMERanalysis New Year's Eve is becoming increasingly rare

# Sentiment Analysis

## Sentiment Analysis Process

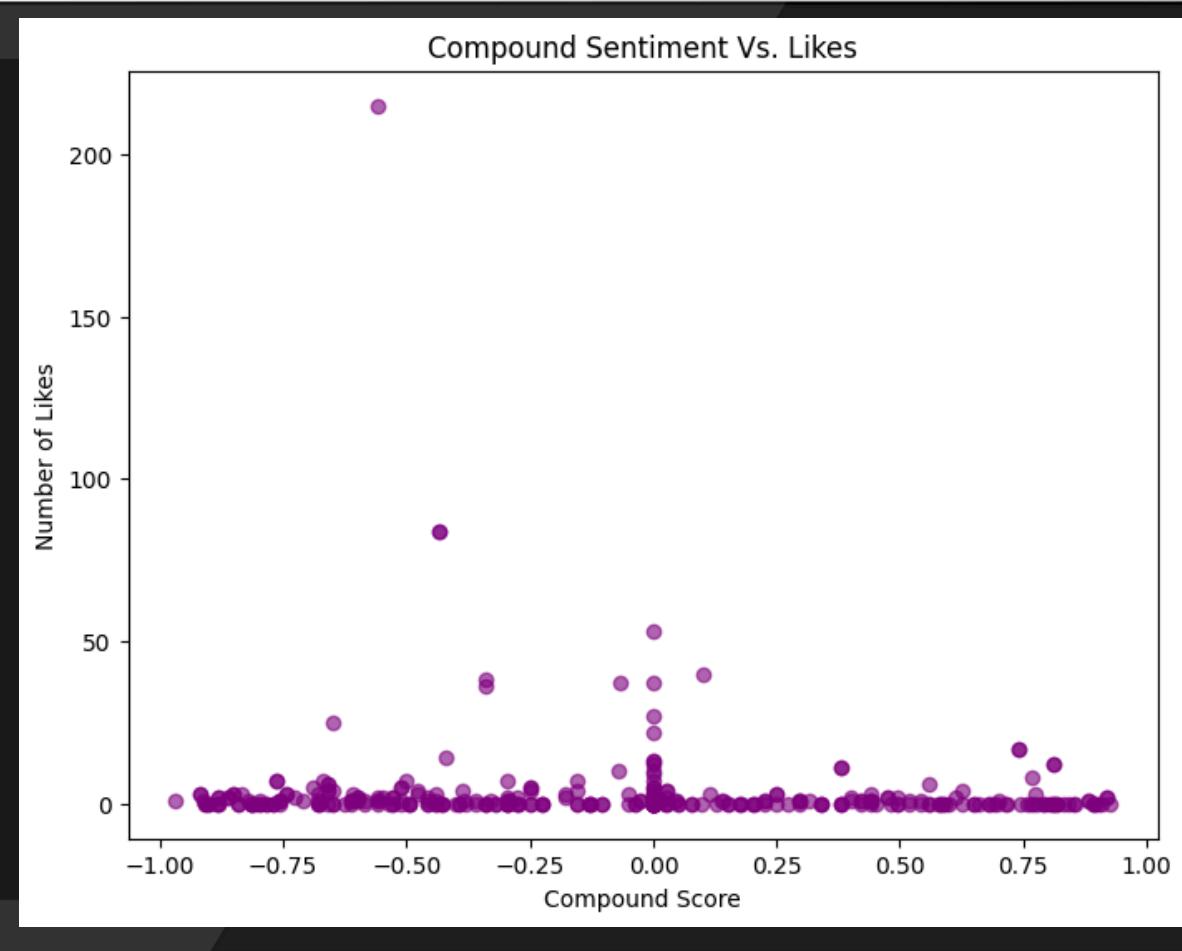
- **Library Used:** `vaderSentiment` for sentiment scoring.
- **Steps Taken:**
  - a. **Initialization:** Created a `SentimentIntensityAnalyzer` instance.
  - b. **Sentiment Scoring:**
    - Applied `analyzer.polarity_scores` on each English-translated tweet to calculate the compound score.
  - c. **Sentiment Classification:**
    - Labeled tweets as:
      - *Positive*: Compound score  $> 0.05$ .
      - *Negative*: Compound score  $< -0.05$ .
      - *Neutral*: Compound score between  $-0.05$  and  $0.05$ .
  - d. **Data Integration:** Added the calculated sentiment scores and labels as new columns in the dataset (Sentiment and Sentiment Label).



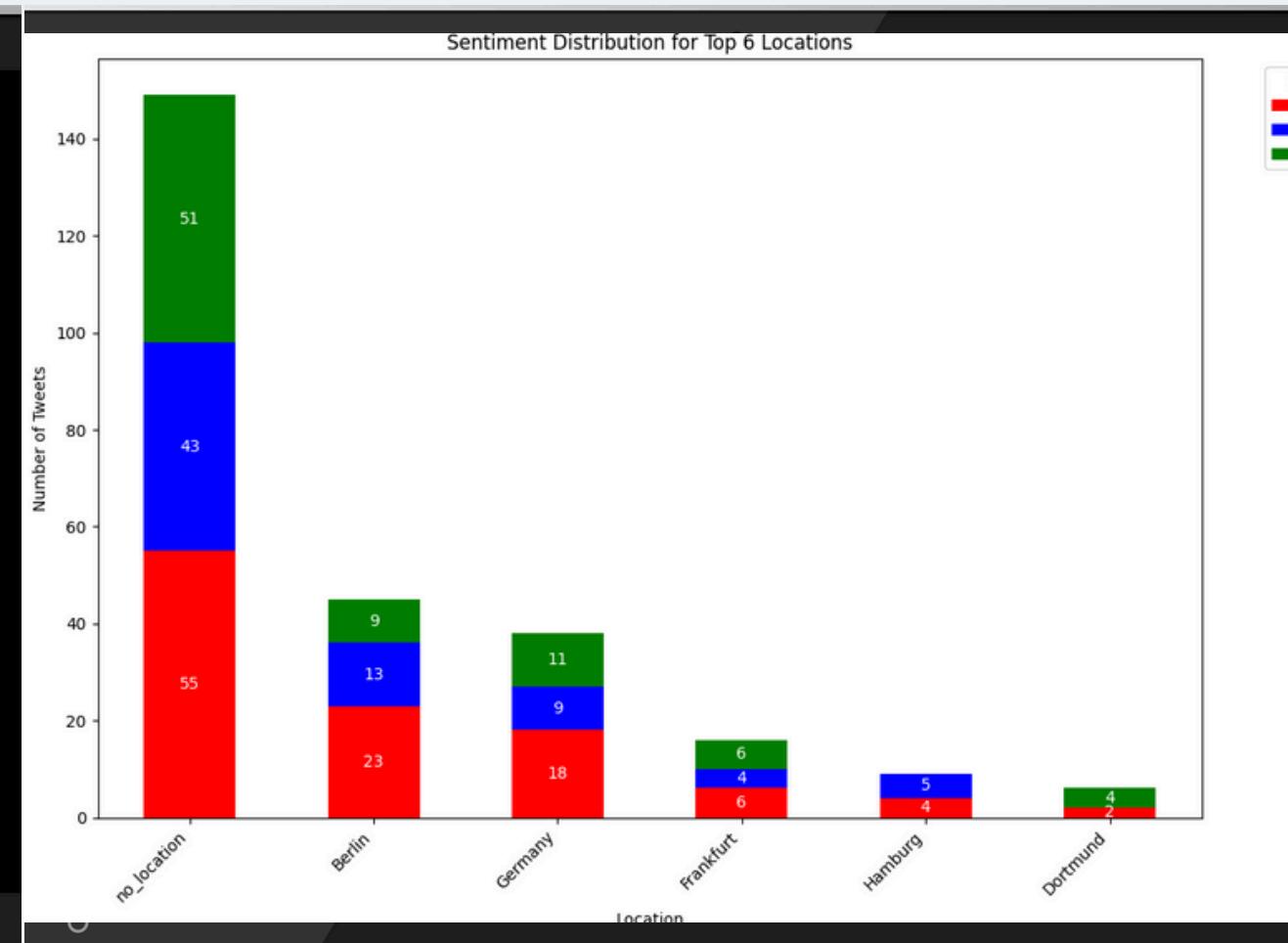
```
# View the results
print(df[['tweet_text_english', 'Sentiment_Label']])
```

	tweet_text_english	Sentiment_Label
0	Massageportal24 Provider Create entry etc	positive
1	Beware of all life-spans of the WHO Swissmedic...	negative
2	Obesity of prion diseases Heat stress Deficien...	negative
3	DIOSynVax opens new advisory body to advance t...	negative
4	Health professionals Immerse yourself in the w...	negative
..	...	...
379	sent to found on updayDE for demolition work	neutral
380	In both cases, as in the case of broken buildi...	negative
381	If the people of X do not believe it, electric...	negative
382	Die SALTED die die Welt beenden knnte	negative
383	Apparently on in Loud police is on Thursday at...	negative
384	[384 rows x 2 columns]	

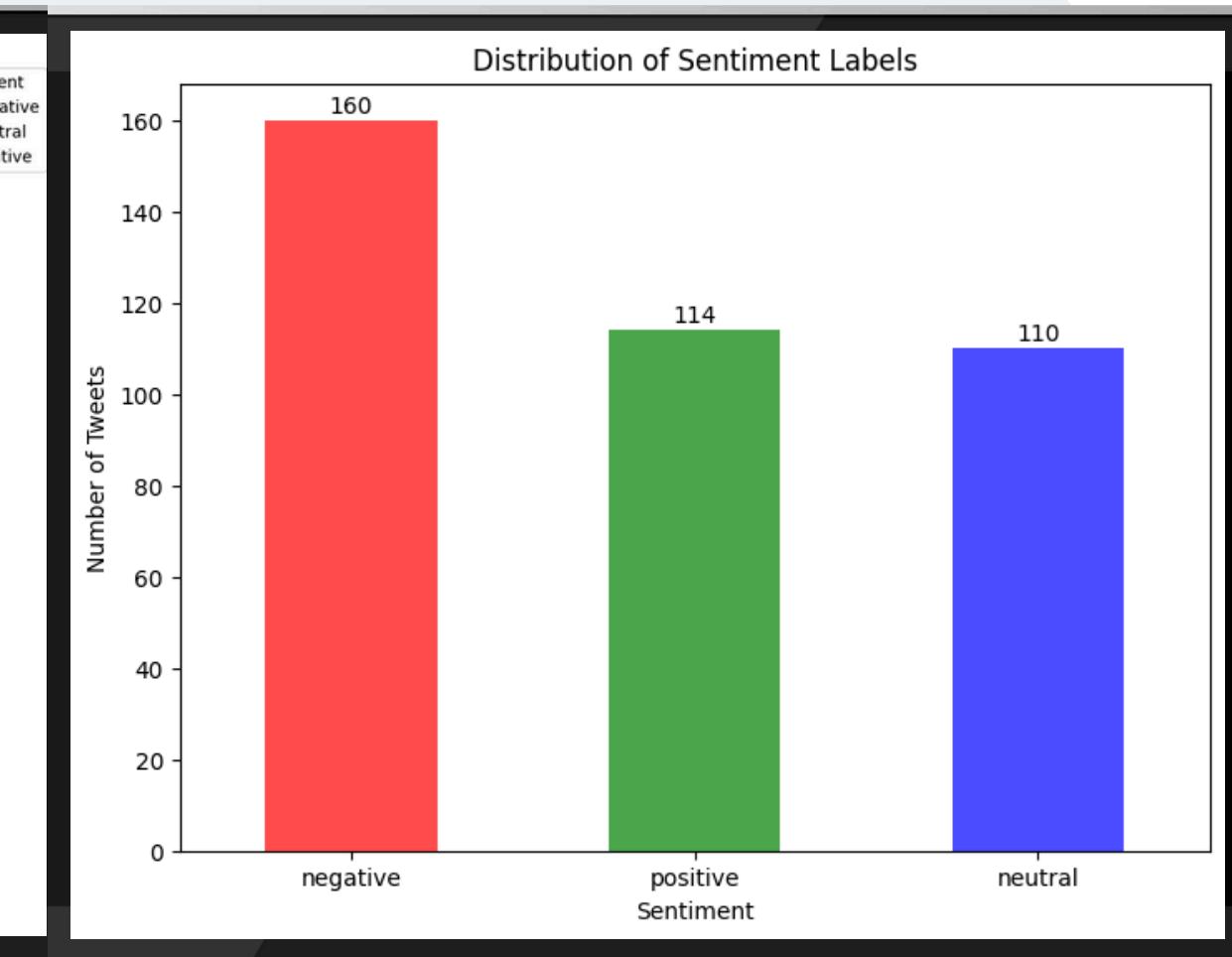
## Compound Score Vs. No. of likes



## Top 6 locations based Sentiment statistics



## Count of Sentiment Labels





## Most Negative

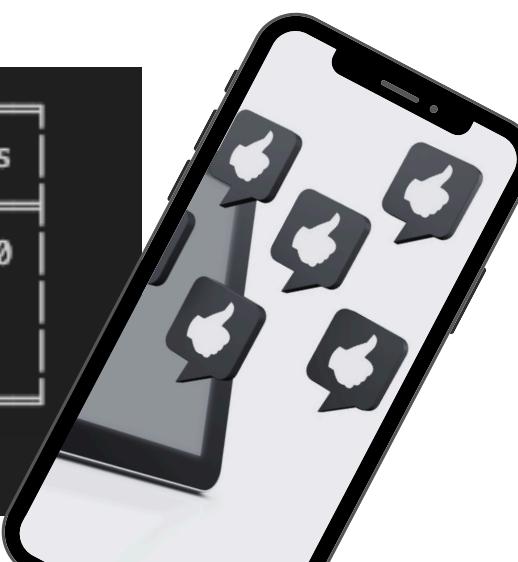
	tweet_text_english	compound
317	M0stermann It's not a solution but we can't help but protect the police for now The problem as such is a multi-ethnic conflict that can lead to war That's why it's important to have a practical government that recognizes and eliminates the danger.	-0.9698
283	Environmental aid First of all, we start with a DUH ban or worry that there are people in the DUH who can think up to wallpaper There is also a ban on weapons, however, there are undisputed pistols etc in D brigens D also has limits	-0.9186

## Most Positive

	tweet_text_english	compound
61	Discover the amazing secret against varicose veins Our innovative means surprise with effective help Try now	0.9274
235	BettinaJarasch Those who want to celebrate safely in the future and want to start the new year in a healthy way Should choose from safety grounds Not those are the problem which is the problem every healthy person with a healthy mind needs to recognize the AfD is	0.9186

## Most Likes and Retweets

	tweet_text_english	likes	compound	retweets
115	Attalah Younes 24 fired a rocket that flew through a window into a nursery and exploded Now he landed in the cell for the time being	215	-0.5574	30





A dark gray background featuring a complex network of white lines and dots forming various geometric shapes like triangles and hexagons. A large white plus sign is positioned in the center. In the bottom right corner, there is a faint, large red 'X'.

**Thank You  
For Your  
Attention!**