



CIS600
**APPLIED NATURAL
LANGUAGE PROCESSING**

TERM PROJECT PRESENTATION
SPRING - 2024

TITLE OF THE PROJECT

SENTIMENT ANALYSIS ON SYRIAN CIVIL WAR USING REDDIT DATA



TEAM MEMBERS

VENKATA SRI SIVA RAMAKRISHNA PALAPARTHY

- 433976193

NIKHILESH GUNNAM

- 502430168

PRATHIK REDDY SANNAPUREDDY

- 344477369

GIRISH KANJIYANI

- 587443048

RUIYING CHEN

- 255618384

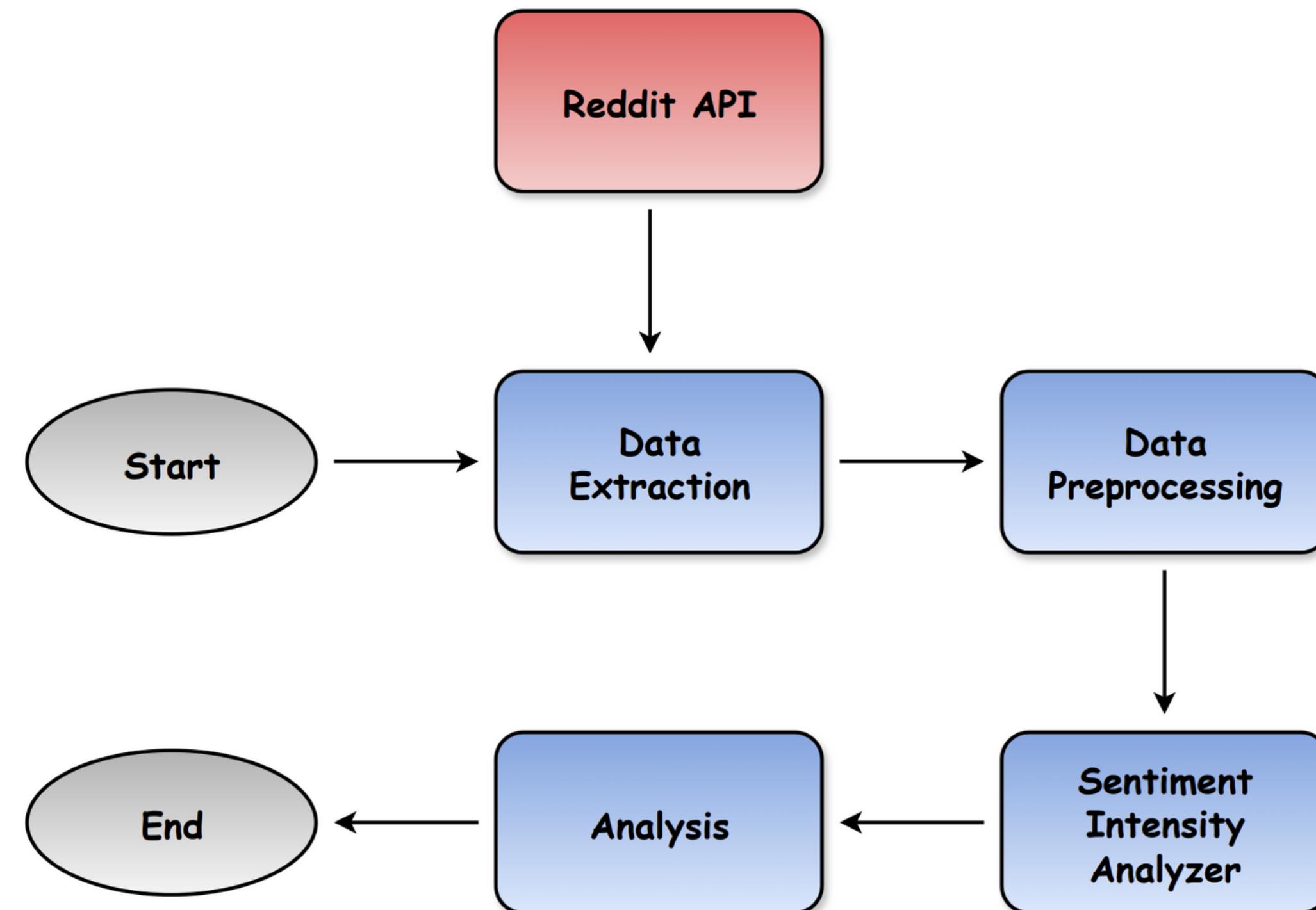


INTRODUCTION

The Syrian Civil War, which began in 2011, has had significant repercussions within the region as well as worldwide. This led to substantial loss of life, massive relocation, and instability in politics, as well as heated debate and commentary over several mediums, namely social media. Reddit, being one of the most popular discussion sites on the internet, is an invaluable resource of data on general public sentiment, views, and debates on a variety of issues, particularly global crises such as the Syrian Civil War. Studying Reddit data offers essential understanding into the opinions, feelings, and attitudes concerning the ongoing issue.

In this project, our goal is to perform sentiment analysis on the Reddit debates associated with the Syrian Civil War. Using natural language processing (NLP) tools, we hope to extract, evaluate, and understand views expressed by Reddit users about various parts of the war. Using this study, we hope to acquire an improved knowledge of the dominant sentiments, popular narratives, and fluctuations in public perceptions as time passes.

WORKFLOW



DATA COLLECTION - REDDIT API

```
merged_df.sample(5)
```

		title	score	id	url	comms_num	created	body	timestamp	subreddit
16358	The curse strikes again	175	sbyrz0		https://i.redd.it/pjaytobvrpd81.png	3	1.643064e+09	NaN	2022-01-24 22:43:10	SyrianCirclejerkWar
30200	A new blast leaves more than 25 people dead or...	1	eeoj71		http://www.syriahr.com/en/?p=151248	0	1.577126e+09	NaN	2019-12-23 18:36:59	syrianconflict
409	Are East Asian men seen as cute in the Middle ...	3	1brq1eb		https://www.reddit.com/r/AskMiddleEast/comment...	13	1.711827e+09	Because we have baby face and no beard, do you...	2024-03-30 19:26:57	AskMiddleEast
11828	Turkish Space Agency sponsored by Mountain ker...	58	pajnu8		https://i.redd.it/7l3o7hu8v9j71.jpg	0	1.629796e+09	NaN	2021-08-24 09:09:58	SyrianCirclejerkWar
17248	"Just minor lossed comrade"	10	t6gpya		https://streamable.com/16l49r	3	1.646392e+09	NaN	2022-03-04 11:14:40	SyrianCirclejerkWar

DATA COLLECTION

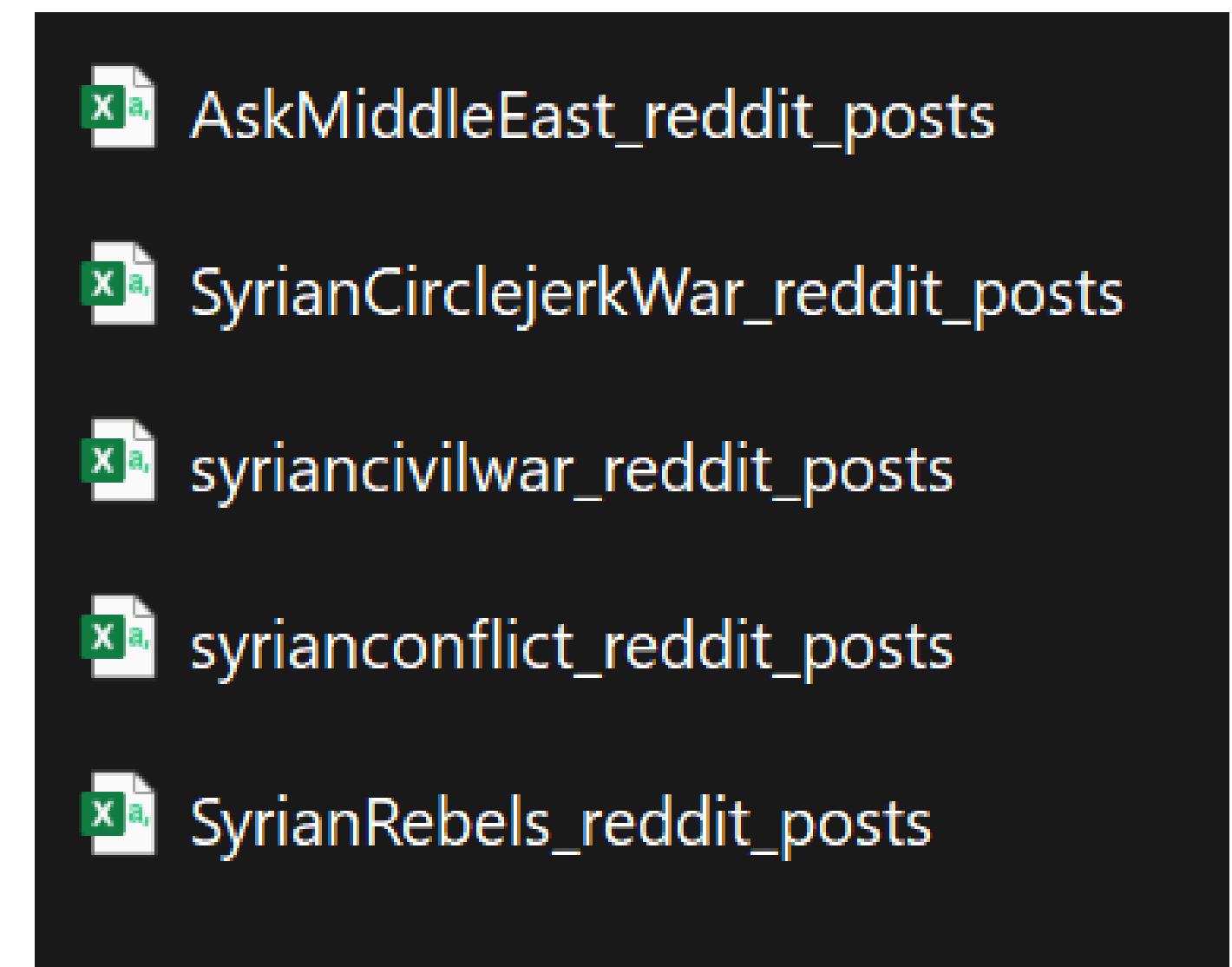
- PRAW

- PRAW (Python Reddit API Wrapper) is a Python library that serves as a wrapper for the Reddit API.
- With PRAW, developers can use Python to interact with Reddit programmatically without dealing with the low-level details of the API.
- It provides a way for external applications, scripts, or services to access and manipulate Reddit data, such as posts, comments, user information, and subreddit details.

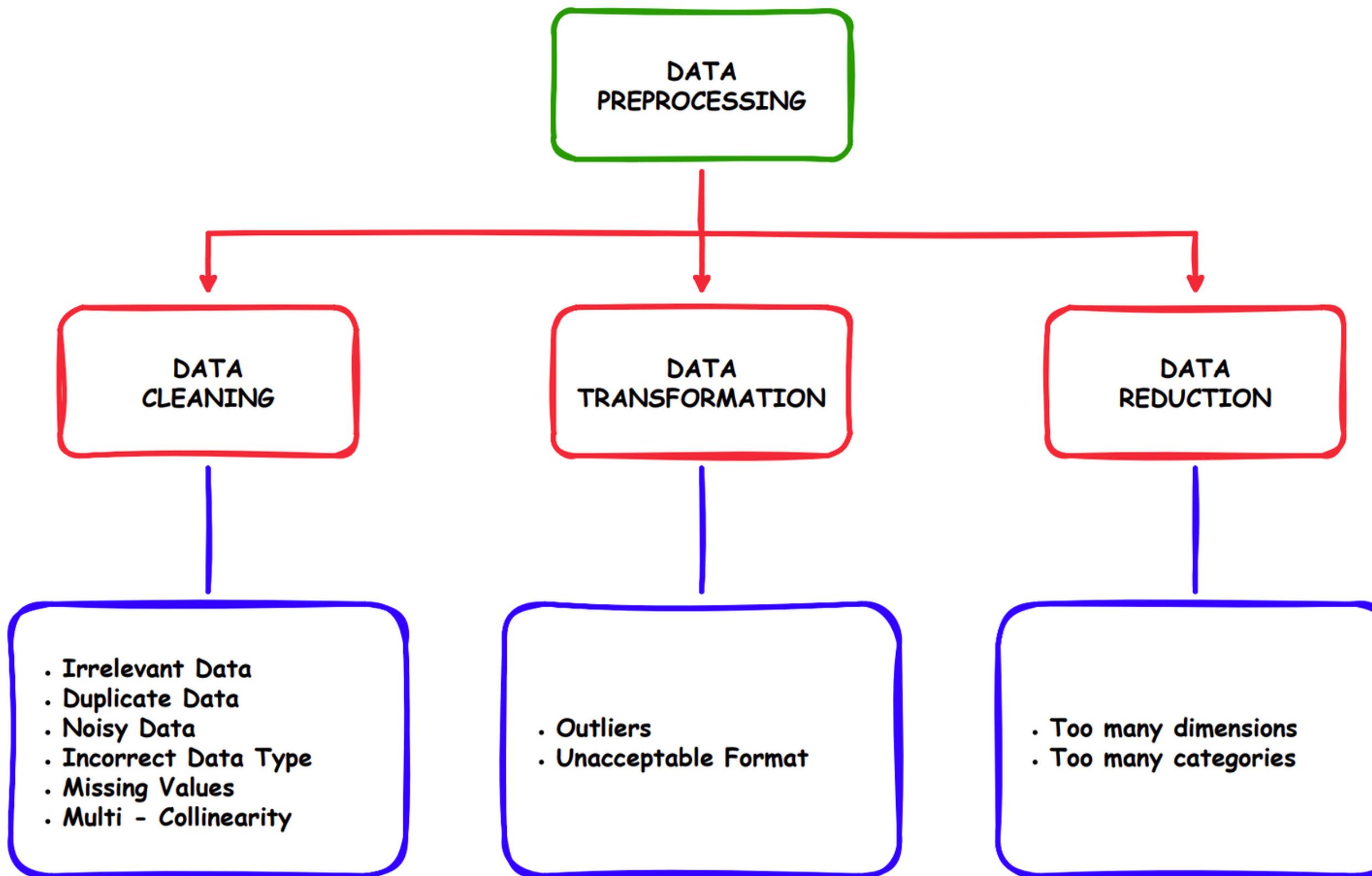
```
reddit = praw.Reddit(  
    client_id='your_client_id',  
    client_secret='your_client_secret',  
    user_agent='your_user_agent'  
)  
  
subreddit = reddit.subreddit('your_subreddit')
```

STEPS INVOLVED IN DATA EXTRACTION FROM REDDIT

- Importing Libraries like PRAW, pandas
- Setting Up Reddit API Credentials
- Accessing Subreddit - “SyriaWar”, “IsraelPalestineWar_23” etc ..
- Setting up parameters such as limit and number of requests
- Fetching posts in chunks
- Collecting post information in the form of python objects
- Creating a dataframe
- Converting the dataframe into csv
- Merging data into dataframe before data pre processing



DATA PREPROCESSING



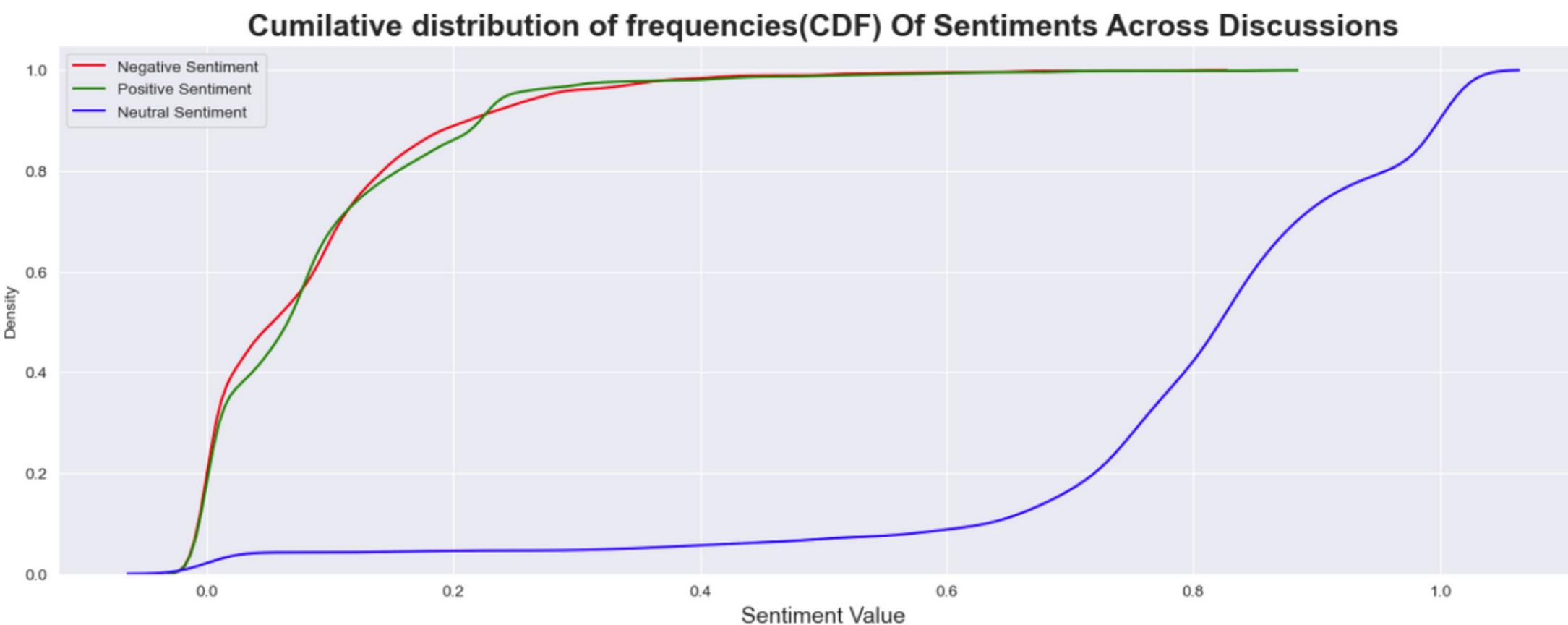
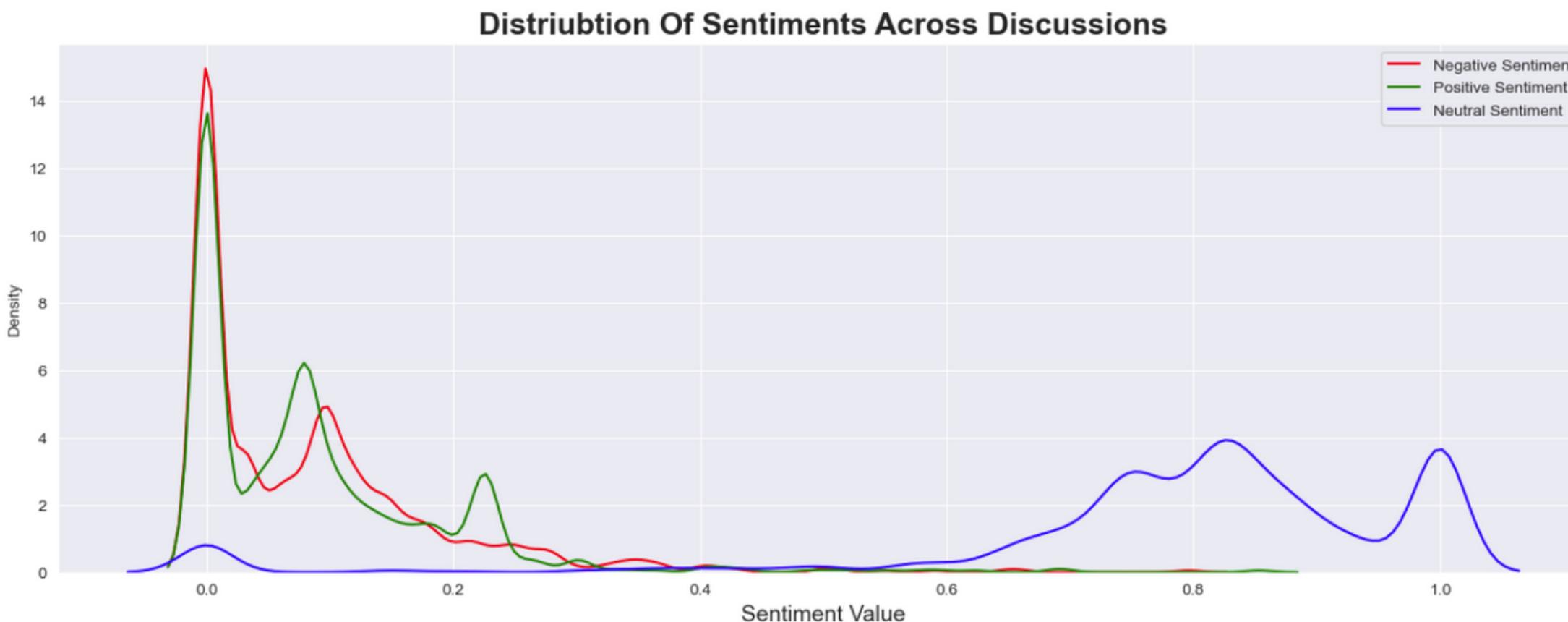
DATA PREPROCESSING INCLUDES...

- Copying relevant columns
- Dropping NaN values
- Converting text to lower case to ensure consistency
- Removing Handlers
- Removing URLs
- Removing Special Characters
- Removing Single Characters
- Substituting Multiple Spaces with Single space
- Converting timestamp to a datetime format



SENTIMENT ANALYSIS

- “VADER” model for Sentiment Analysis
- Sentiment scores using VADER
- Classification based on compound score



FEATURE ENGINEERING

- *Naïve Feature Extraction*
- *Named Entity Extraction*
- *Time Based Sentiment Analysis*
- *Text Decomposition Analysis*

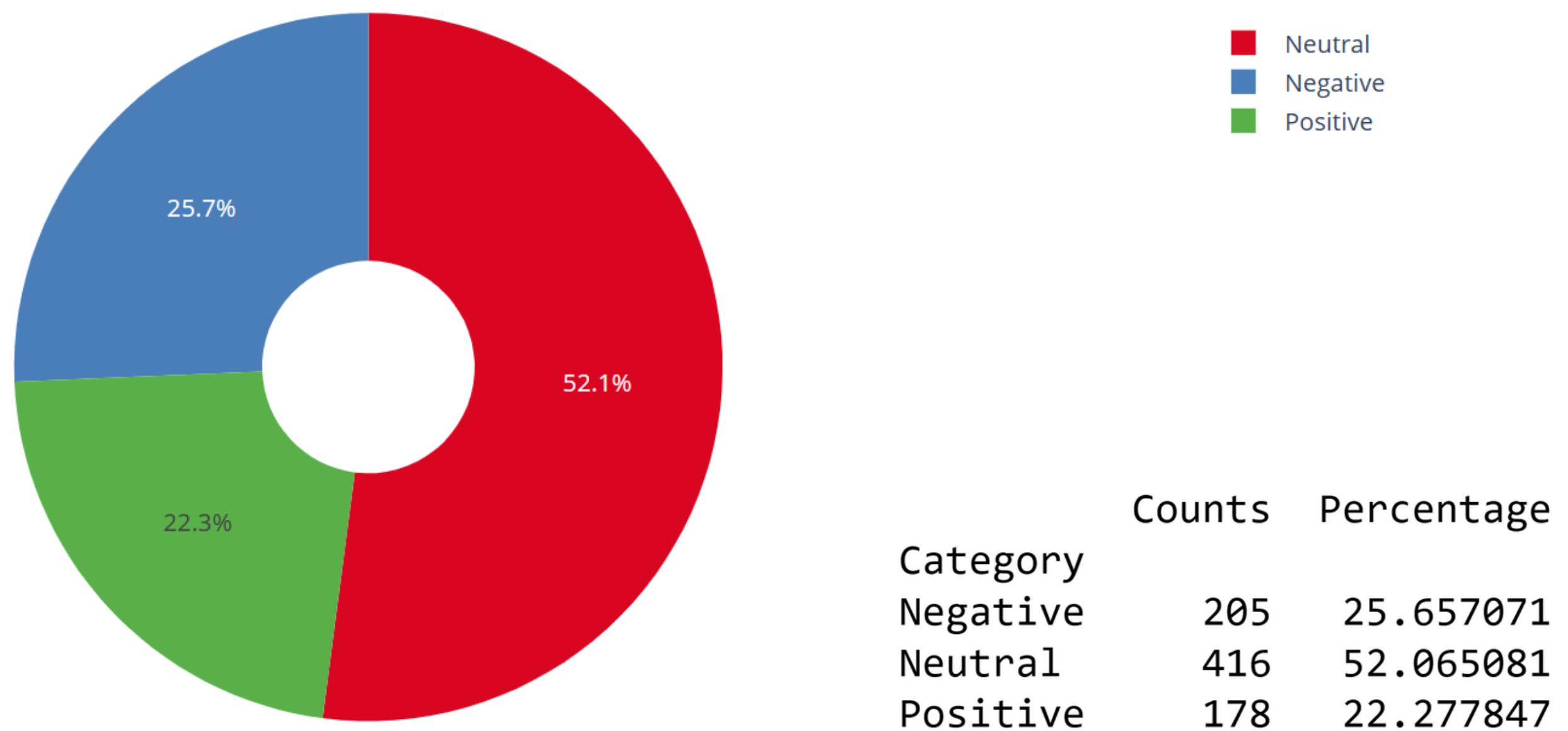
Named Entity Extraction

	body	timestamp	Positive Sentiment	Neutral Sentiment	Negative Sentiment	# Of Words	# Of StopWords	Average Word Length	# Of Times Events Was Mentioned	# Of Organizations Mentioned	# Of Political locations Mentioned	# Of Non Political locations Mentioned
0	ifrom the us historically not politically acti...	2024-04-10	0.000001	0.685001	0.315001	23	10	5.692308	0	0	1	0
2	like trade on land water and air and air travel	2024-04-10	0.217001	0.783001	0.000001	10	4	4.333333	0	0	0	0
3	why is the uae so hostile towards algeria rece...	2024-04-10	0.000001	0.877001	0.123001	25	10	7.133333	0	0	3	0
5	he is currently hiding in qatar and his commen...	2024-04-10	0.000001	0.821001	0.179001	25	12	7.000000	0	0	1	0
9	x200b	2024-04-10	0.000001	1.000001	0.000001	1	0	5.000000	0	0	0	0

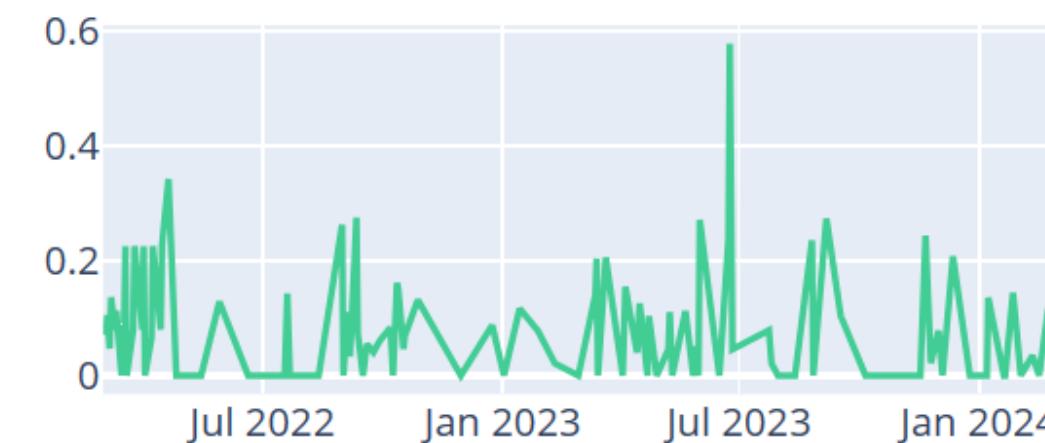
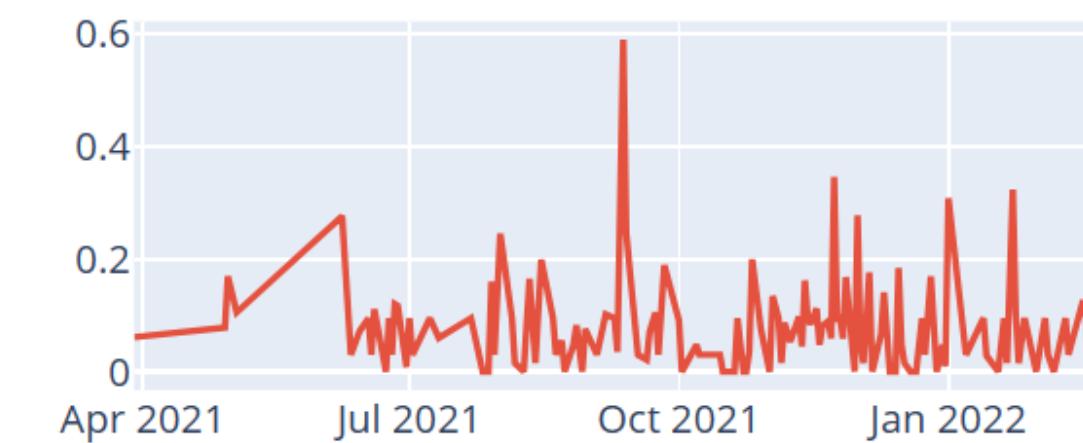
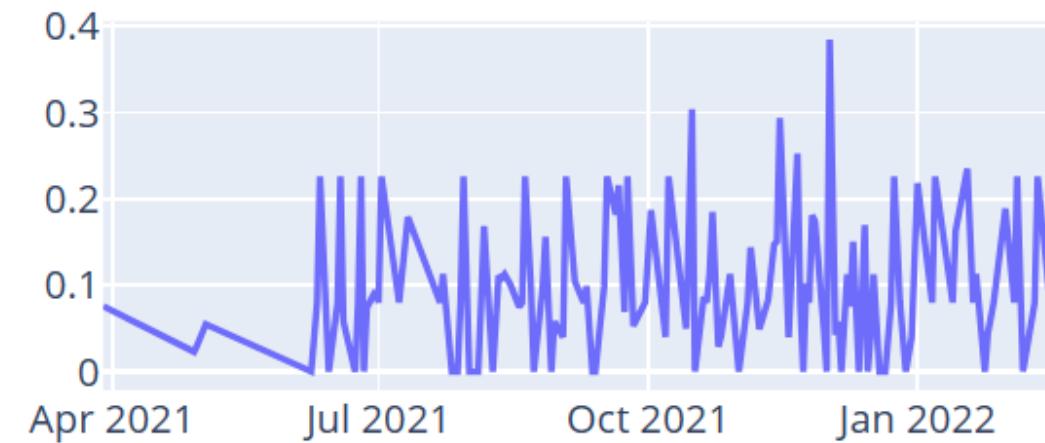
RESULTS

- ANALYSIS OF DATA

Distribution of Sentiment Categories in Discussions

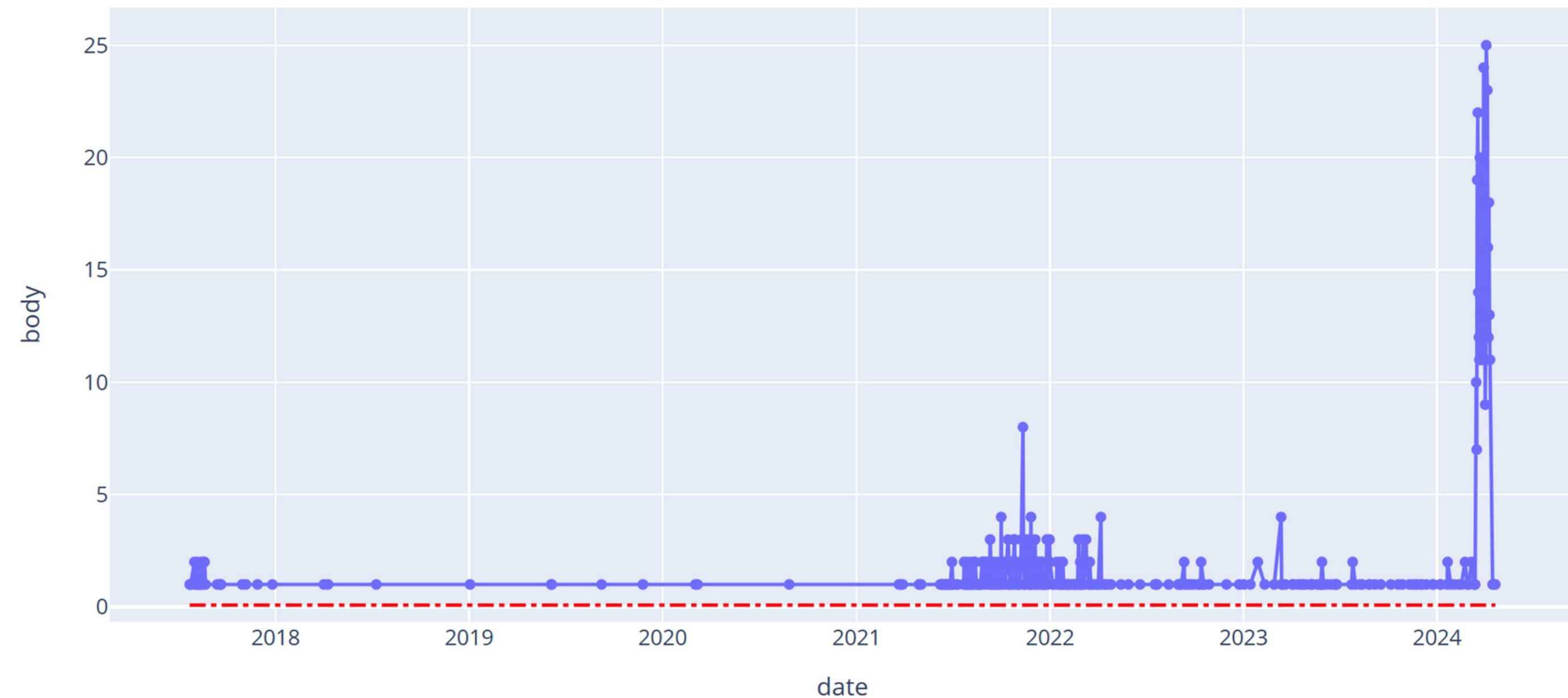


Distribution Of Daily Mean Sentiments Over Our Time Line For Each Partition

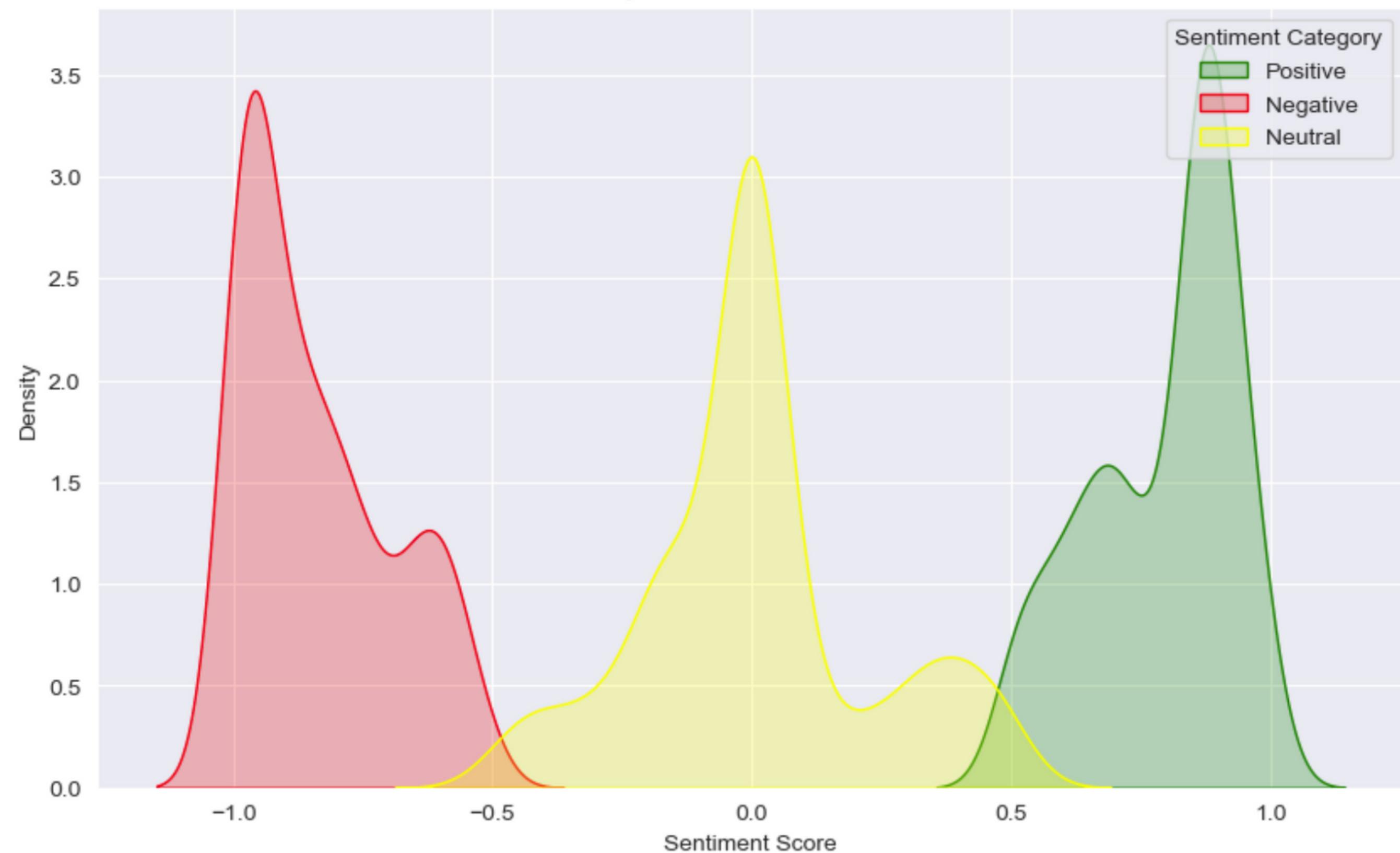


Positive Part 1
Negative Part 1
Positive Part 2
Negative Part 2
Positive Part 3
Negative Part 3

Discussions Per Day



Density Plot of Sentiment Scores



WORD CLOUDS

COMPOUND DATA

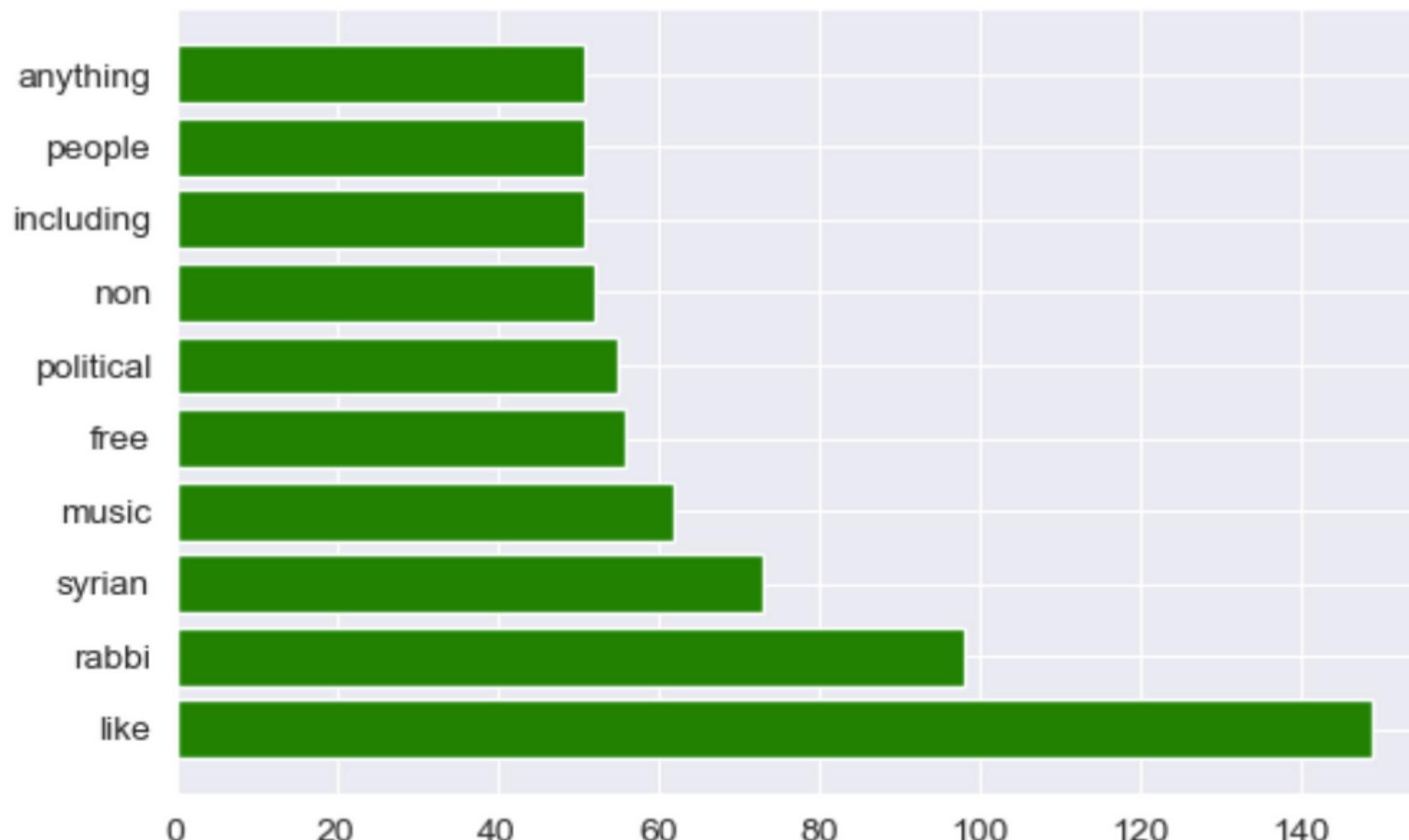
NEUTRAL DATA

POSITIVE DATA

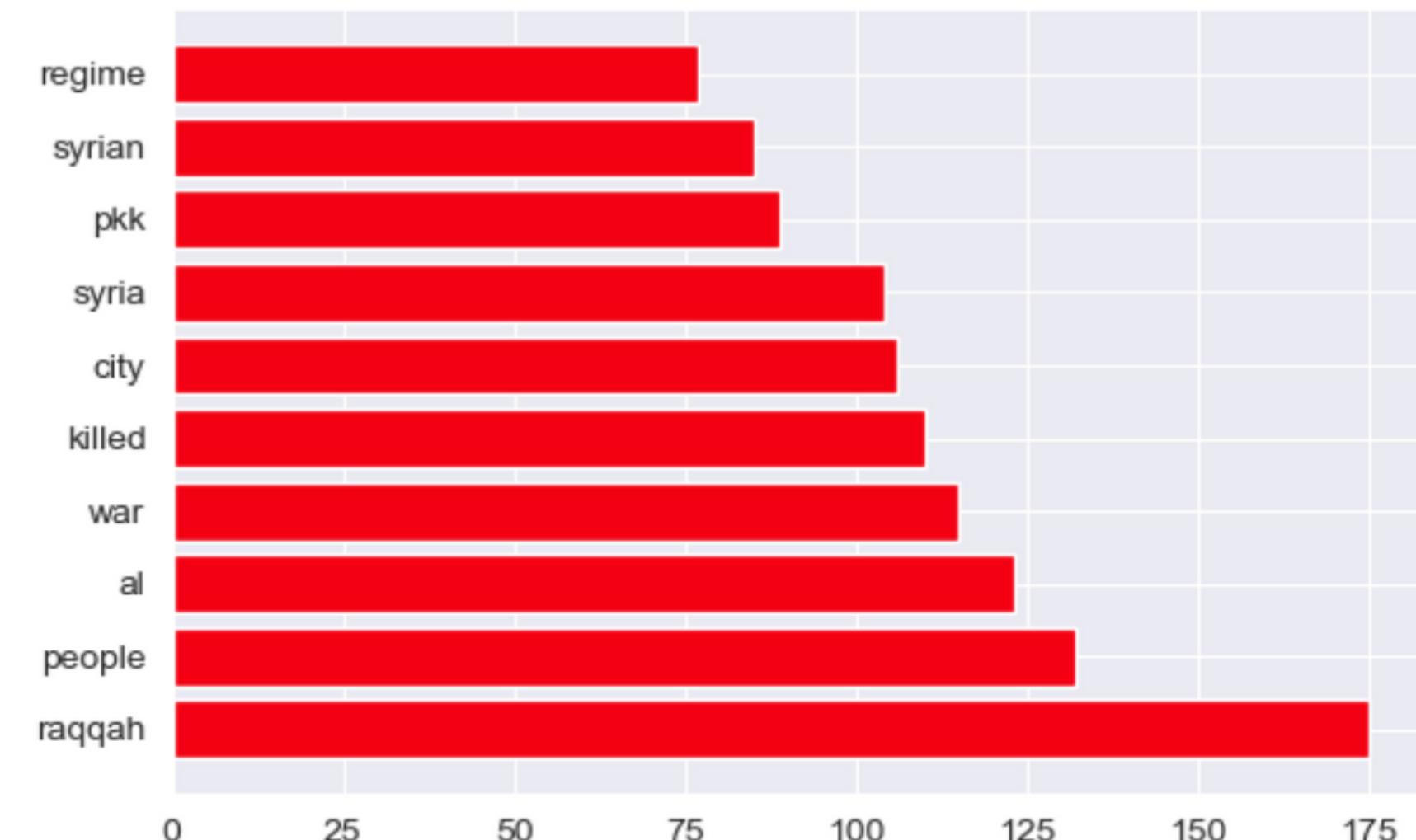
NEGATIVE DATA

FREQUENCY OF WORDS FOR SENTIMENTS

Frequent Words for Positive Sentiment



Frequent Words for Negative Sentiment



CONCLUSION

The current analysis in this project provides important insights with regard to the varied character of popular opinion towards the Syrian Civil War. By using data collected from Reddit and effective NLP techniques, it adds to an improved awareness of how people view and talk about this complicated dispute in the digital realm. In short, our project examined people's opinions and thoughts about the Syrian Civil War through Reddit posts. We discovered what most individuals were saying, what issues kept coming up, and how these conversations evolved over time. Understanding multiple user viewpoints helped us better grasp what shapes these exchanges. Our research is important for decision-makers, organizations, academics, and people working for positive change because it provides them with essential insights for making sound decisions. Throughout our project, we prioritized appropriate data processing and respect for people's privacy.





Thank you

