Arnav Tumbde
RCOEM Nagpur

Domain : Data Science

#Task 4

- Analyze and visualize sentiment patterns in social media data to understand public opinion and attitudes towards specific topics or brands.

Screenshots :

Arnav Tumbde

Domain : Data Science

Task : 4

In [3]:
```
import pandas as pd
import numpy as np
from nltk.tokenize import sent_tokenize, word_tokenize
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
from sklearn.datasets import fetch_20newsgroups
from nltk.corpus import stopwords
import string
from nltk import pos_tag
from nltk.stem import WordNetLemmatizer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn import preprocessing
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

Out[4]: True

```
In [6]: data = pd.read_csv("D:\\Semester - IV\\Prodigy Internship\\Task 4\\archive (1)\\twitter_training.csv")
        v_data = pd.read_csv("D:\\Semester - IV\\Prodigy Internship\\Task 4\\archive (1)\\twitter_validation.csv")
```

```
In [7]: data
```

Out[7]:

	2401	Borderlands	Positive	im getting on borderlands and i will murder you all ,
0	2401	Borderlands	Positive	I am coming to the borders and I will kill you...
1	2401	Borderlands	Positive	im getting on borderlands and i will kill you ...
2	2401	Borderlands	Positive	im coming on borderlands and i will murder you...
3	2401	Borderlands	Positive	im getting on borderlands 2 and i will murder ...
4	2401	Borderlands	Positive	im getting into borderlands and i can murder y...
...	...	...	...	...
74676	9200	Nvidia	Positive	Just realized that the Windows partition of my...
74677	9200	Nvidia	Positive	Just realized that my Mac window partition is ...
74678	9200	Nvidia	Positive	Just realized the windows partition of my Mac ...
74679	9200	Nvidia	Positive	Just realized between the windows partition of...
74680	9200	Nvidia	Positive	Just like the windows partition of my Mac is l...

74681 rows × 4 columns

	3364	Facebook	Irrelevant	I mentioned on Facebook that I was struggling for motivation to go for a run the other day, which has been translated by Tom’s great auntie as ‘Hayley can’t get out of bed’ and told to his grandma, who now thinks I’m a lazy, terrible person 🤣
0	352	Amazon	Neutral	BBC News - Amazon boss Jeff Bezos rejects clai...
1	8312	Microsoft	Negative	@Microsoft Why do I pay for WORD when it funct...
2	4371	CS-GO	Negative	CSGO matchmaking is so full of closet hacking,...
3	4433	Google	Neutral	Now the President is slapping Americans in the...
4	6273	FIFA	Negative	Hi @EAHelp I’ve had Madeleine McCann in my cel...
...	...	...	...	...
994	4891	GrandTheftAuto(GTA)	Irrelevant	⭐️ Toronto is the arts and culture capital of ...
995	4359	CS-GO	Irrelevant	tHIS IS ACTUALLY A GOOD MOVE TOT BRING MORE VI...
996	2652	Borderlands	Positive	Today sucked so it’s time to drink wine n play...
997	8069	Microsoft	Positive	Bought a fraction of Microsoft today. Small wins.
998	6960	johnson&johnson	Neutral	Johnson & Johnson to stop selling talc baby po...

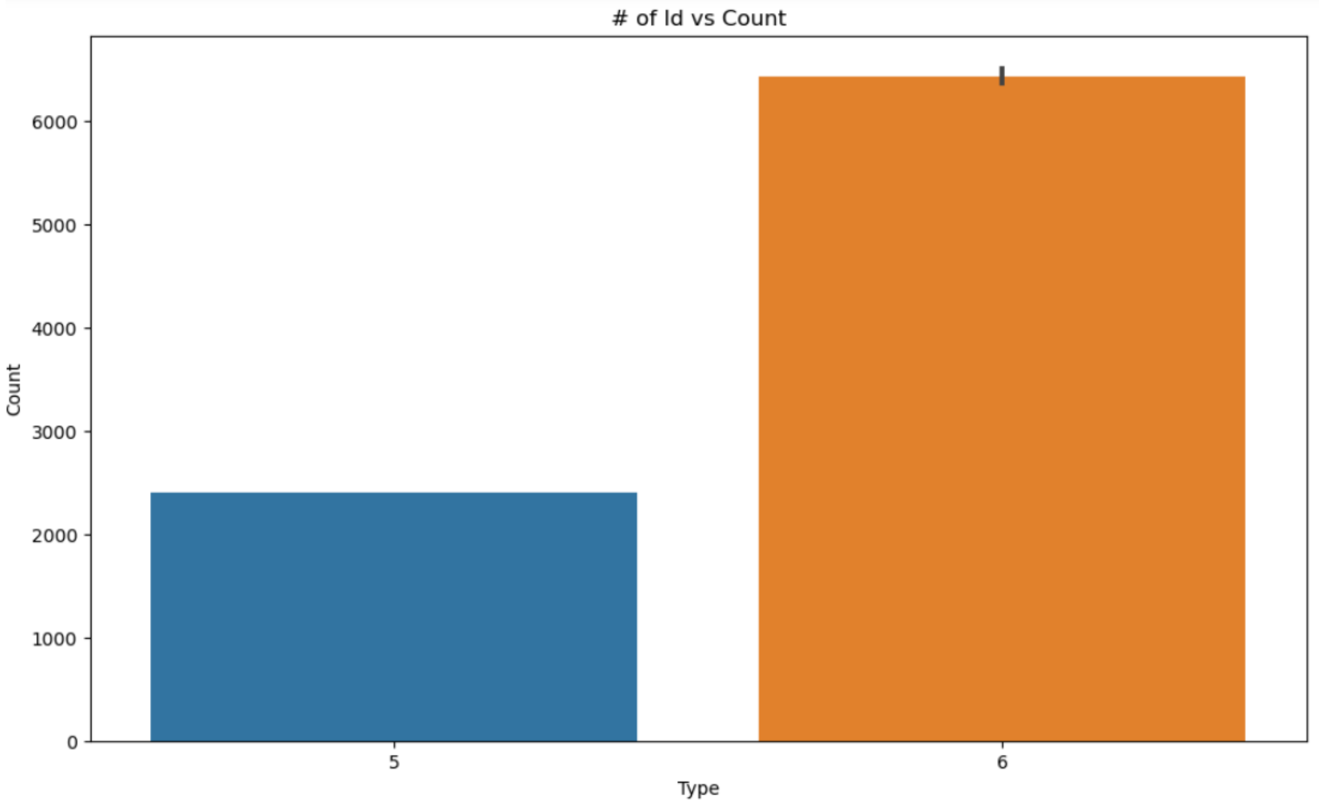
999 rows × 4 columns

In [12]:
```
data.shape
```
Out[12]:
```
(74681, 4)
```

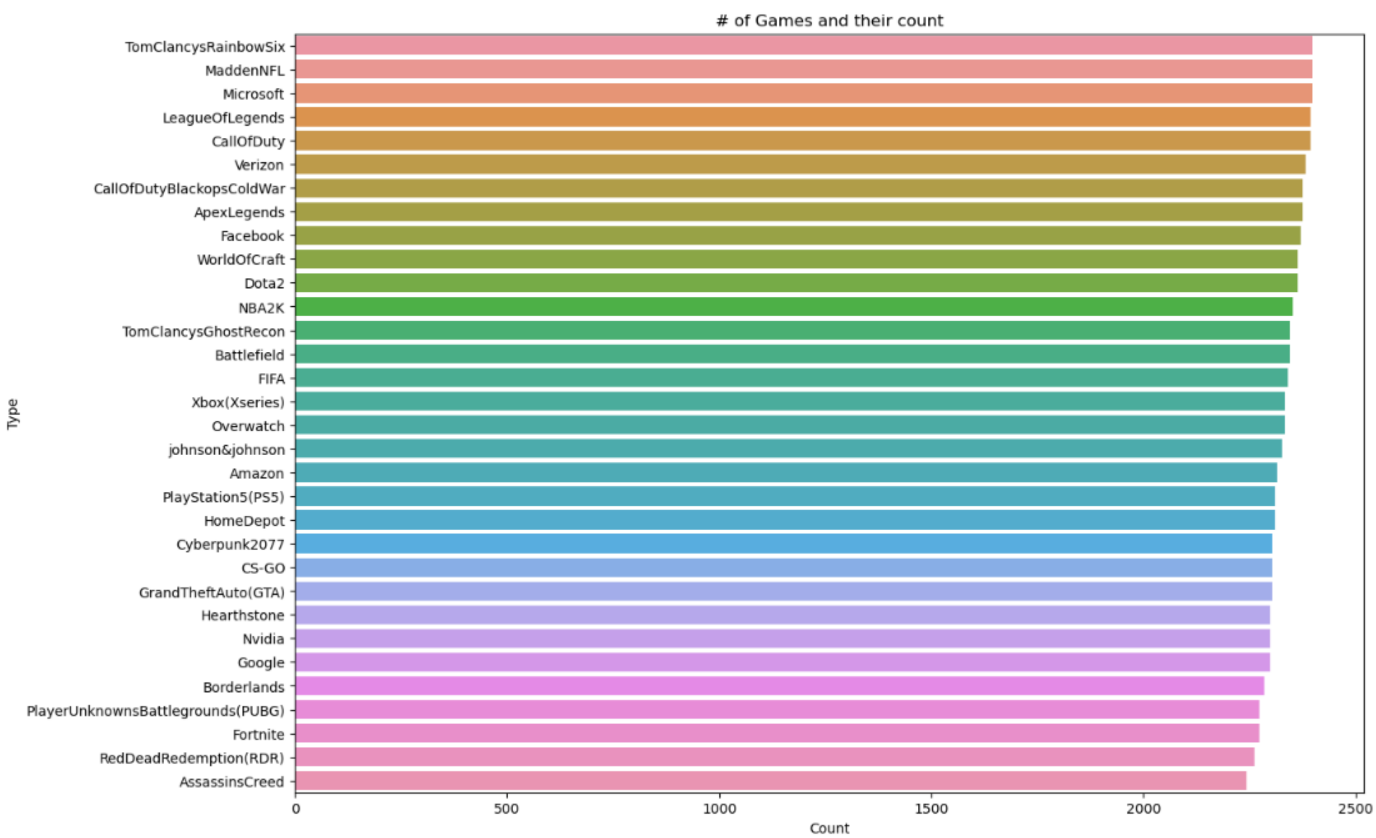
In [13]:
```
data.columns
```
Out[13]:
```
Index(['id', 'game', 'sentiment', 'text'], dtype='object')
```

In [14]:
```
data.describe(include='all')
```
Out[14]:

	id	game	sentiment	text
count	74681.000000	74681	74681	73995
unique	NaN	32	4	69490
top	NaN	TomClancysRainbowSix	Negative	
freq	NaN	2400	22542	172
mean	6432.640149	NaN	NaN	NaN
std	3740.423819	NaN	NaN	NaN
min	1.000000	NaN	NaN	NaN
25%	3195.000000	NaN	NaN	NaN
50%	6422.000000	NaN	NaN	NaN
75%	9601.000000	NaN	NaN	NaN
max	13200.000000	NaN	NaN	NaN



# of Id vs Count
Count
0
1000
2000
3000
4000
5000
6000
5
6
Type

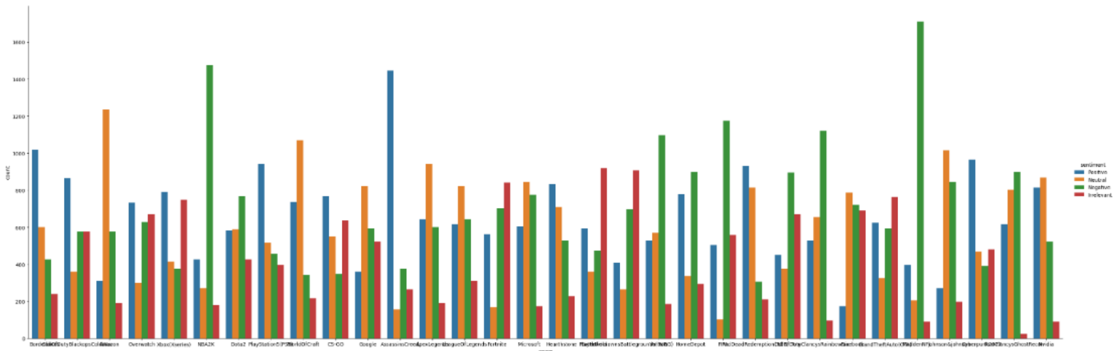


# of Games and their count
Type
TomClancysRainbowSix
MaddenNFL
Microsoft
LeagueOfLegends
CallOfDuty
Verizon
CallOfDutyBlackopsColdWar
ApexLegends
Facebook
WorldOfCraft
Dota2
NBA2K
TomClancysGhostRecon
Battlefield
FIFA
Xbox(Xseries)
Overwatch
johnson&johnson
Amazon
PlayStation5(PS5)
HomeDepot
Cyberpunk2077
CS-GO
GrandTheftAuto(GTA)
Hearthstone
Nvidia
Google
Borderlands
PlayerUnknownsBattlegrounds(PUBG)
Fortnite
RedDeadRedemption(RDR)
AssassinsCreed
0
500
1000
1500
2000
2500
Count

In [19]:
```
sns.catplot(x="game",hue="sentiment", kind="count",height=10, aspect=3, data=data)
```

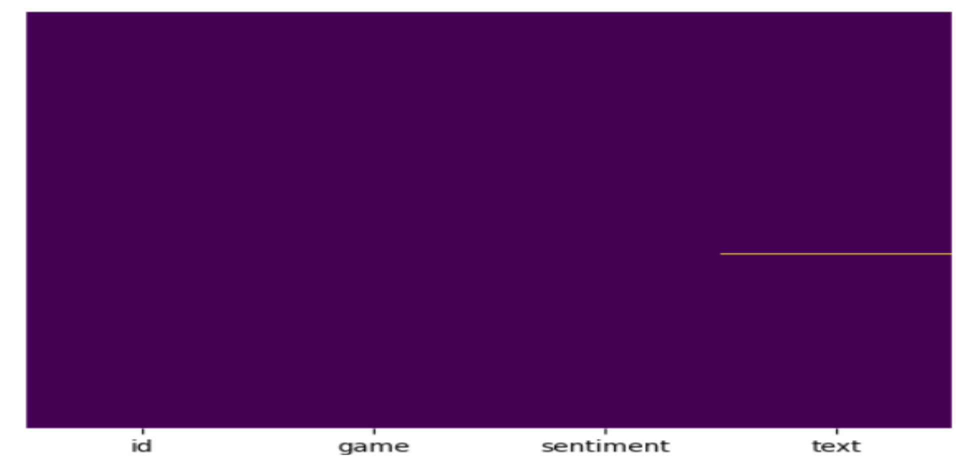
```
C:\Users\Arnav\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserWarning: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```

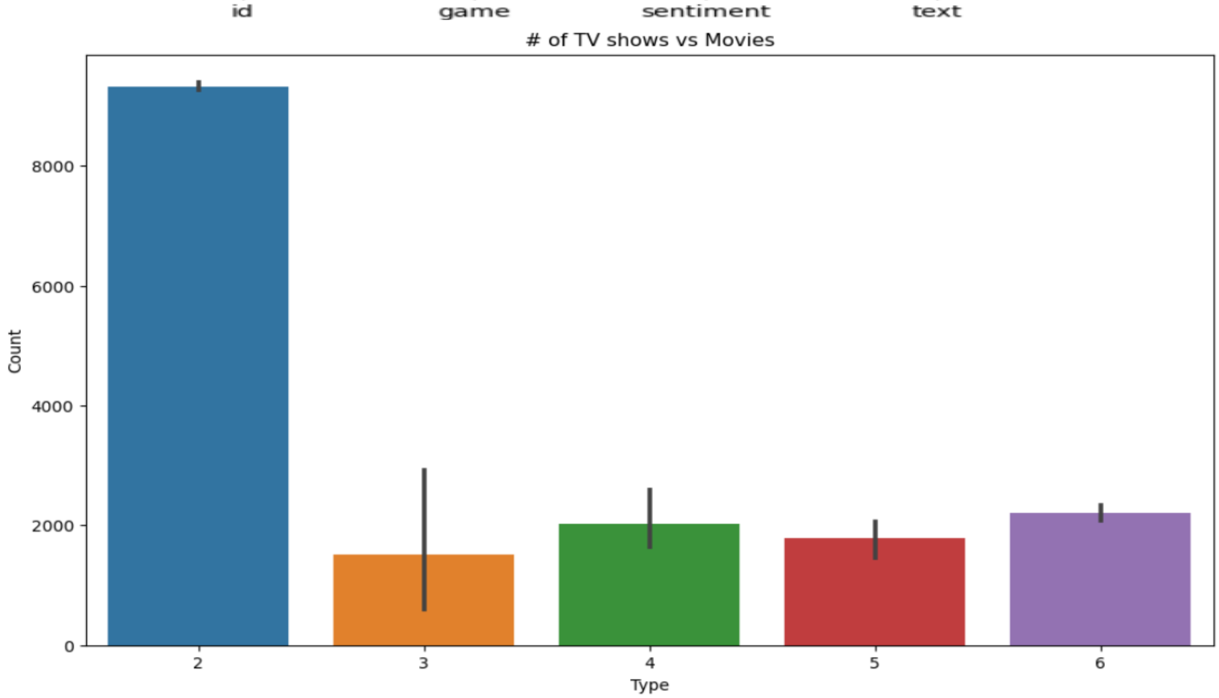
Out[19]:
```
<seaborn.axisgrid.FacetGrid at 0x1bed5e3eed0>
```

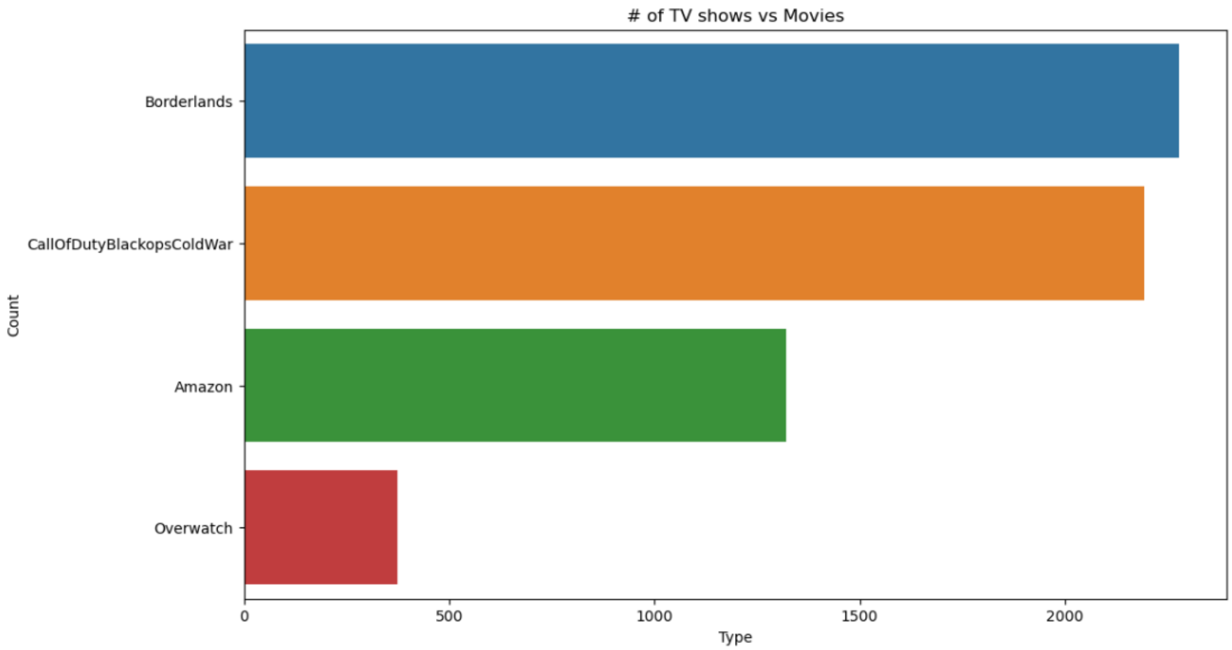


```
sns.heatmap(data.isnull(),yticklabels=False,cbar=False,cmap='viridis')
```

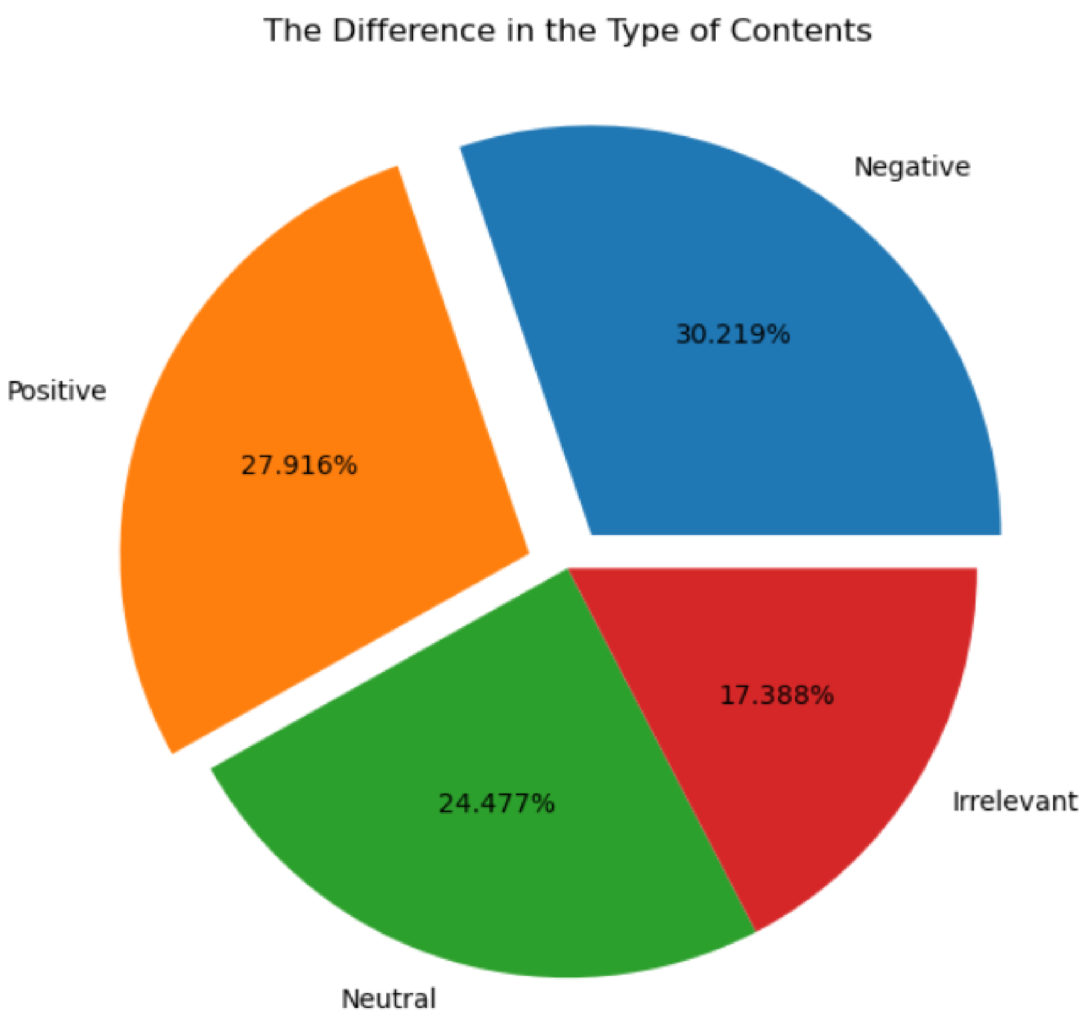
```
<Axes: >
```







# of TV shows vs Movies
Borderlands
CallOfDutyBlackopsColdWar
Amazon
Overwatch
Count
0
500
1000
1500
2000
Type



The Difference in the Type of Contents
Negative
30.219%
Positive
27.916%
17.388%
Irrelevant
24.477%
Neutral

a. About the Dataset

For this task, I used a dataset that contains social media sentiment data. The dataset includes information about public opinions and attitudes towards specific topics or brands. The dataset includes the following columns:

- **Date**: The date the social media post was made.
- **User**: The username of the person who made the post.
- **Text**: The content of the social media post.
- **Sentiment**: The sentiment score of the post (e.g., positive, negative, neutral).
- **Topic**: The specific topic or brand mentioned in the post.
- **Location**: The geographical location of the user.

b. Explanation of the Concept

The primary objective of this task was to analyze and visualize sentiment patterns in social media data to understand public opinion and attitudes towards specific topics or brands. By exploring sentiment scores and trends, we can gain insights into how different topics or brands are perceived by the public, identify emerging trends, and detect any shifts in sentiment over time.

c. Outcome of the Analysis

1. **Sentiment Distribution**: I created a pie chart showing the distribution of positive, negative, and neutral sentiments across the dataset. This visualization helped to understand the overall sentiment landscape.
2. **Sentiment Over Time**: A line chart was used to visualize the changes in sentiment scores over time. This chart revealed periods of heightened positive or negative sentiment, indicating significant events or trends.
3. **Topic-Based Sentiment**: A bar chart was created to show the sentiment distribution for different topics or brands. This highlighted which topics or brands had the most positive or negative sentiments associated with them.
4. **Geographical Sentiment Analysis**: A map was created to visualize the sentiment scores geographically. This helped to identify regions with particularly strong sentiments towards specific topics or brands.
5. **Word Cloud**: A word cloud was generated from the text content of the posts, focusing on the most frequently mentioned words. This provided a visual representation of key terms and themes in the social media conversations.

d. Conclusion

- **Overall Sentiment Landscape**: The sentiment distribution chart showed that the majority of posts were neutral, with a balanced representation of positive and negative sentiments. This gives a broad view of public opinion on social media.

- **Temporal Sentiment Trends**: The analysis of sentiment over time indicated specific periods where sentiment significantly shifted. These periods often correlated with notable events or announcements related to the topics or brands.
- **Topic-Specific Sentiment**: Certain topics or brands had more positive sentiments, while others were predominantly negative. This insight is valuable for understanding public perception and can guide marketing and communication strategies.
- **Geographical Insights**: The geographical sentiment analysis revealed regional variations in public opinion. Identifying areas with strong sentiments can help in tailoring region-specific strategies.
- **Key Themes and Terms**: The word cloud visualization provided a quick snapshot of the most discussed terms and themes, highlighting what aspects of the topics or brands are most engaging to the public.

These visualizations and analyses help in uncovering important patterns and trends, providing a deeper understanding of public sentiment and guiding strategic decisions for better engagement and reputation management.