

PRODIGY INFOTECH

TASK 4

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

-BHOOMADI LIKHITHA REDDY | Data Science Intern

```
In [2]: import numpy as np  
import pandas as pd
```

```
In [4]: data = pd.read_csv("sentimentdataset.csv")  
data
```

Out[4]:

	Unnamed: 0.1	Unnamed: 0	Text	Sentiment	Timestamp	User
0	0	0	Enjoying a beautiful day at the park! ...	Positive	2023-01-15 12:30:00	User123
1	1	1	Traffic was terrible this morning. ...	Negative	2023-01-15 08:45:00	CommuterX
2	2	2	Just finished an amazing workout! 🏋️ ...	Positive	2023-01-15 15:45:00	FitnessFan
3	3	3	Excited about the upcoming weekend getaway! ...	Positive	2023-01-15 18:20:00	AdventureX
4	4	4	Trying out a new recipe for dinner tonight. ...	Neutral	2023-01-15 19:55:00	ChefCook
...
727	728	732	Collaborating on a science project that receiv...	Happy	2017-08-18 18:20:00	ScienceProjectSuccessHighSchool
728	729	733	Attending a surprise birthday party organized ...	Happy	2018-06-22 14:15:00	BirthdayPartyJoyHighSchool
729	730	734	Successfully fundraising for a school charity ...	Happy	2019-04-05 17:30:00	CharityFundraisingTriumphHighSchool
730	731	735	Participating in a multicultural festival, cel...	Happy	2020-02-29 20:45:00	MulticulturalFestivalJoyHighSchool
731	732	736	Organizing a virtual talent show during challe...	Happy	2020-11-15 15:15:00	VirtualTalentShowSuccessHighSchool

732 rows × 15 columns



```
In [6]: data.head()
```

Out[6]:

	Unnamed: 0.1	Unnamed: 0	Text	Sentiment	Timestamp	User	Platform	Hashtags	Retw
0	0	0	Enjoying a beautiful day at the park! ...	Positive	2023-01-15 12:30:00	User123	Twitter	#Nature #Park	
1	1	1	Traffic was terrible this morning. ...	Negative	2023-01-15 08:45:00	CommuterX	Twitter	#Traffic #Morning	
2	2	2	Just finished an amazing workout! 🏋️ ...	Positive	2023-01-15 15:45:00	FitnessFan	Instagram	#Fitness #Workout	
3	3	3	Excited about the upcoming weekend getaway! ...	Positive	2023-01-15 18:20:00	AdventureX	Facebook	#Travel #Adventure	
4	4	4	Trying out a new recipe for dinner tonight. ...	Neutral	2023-01-15 19:55:00	ChefCook	Instagram	#Cooking #Food	



```
In [8]: data.tail()
```

Out[8]:

	Unnamed: 0.1	Unnamed: 0	Text	Sentiment	Timestamp	User
727	728	732	Collaborating on a science project that receiv...	Happy	2017-08-18 18:20:00	ScienceProjectSuccessHighSchool
728	729	733	Attending a surprise birthday party organized ...	Happy	2018-06-22 14:15:00	BirthdayPartyJoyHighSchool
729	730	734	Successfully fundraising for a school charity ...	Happy	2019-04-05 17:30:00	CharityFundraisingTriumphHighSchool
730	731	735	Participating in a multicultural festival, cel...	Happy	2020-02-29 20:45:00	MulticulturalFestivalJoyHighSchool
731	732	736	Organizing a virtual talent show during challe...	Happy	2020-11-15 15:15:00	VirtualTalentShowSuccessHighSchool



In [10]: data.dtypes

Out[10]: Unnamed: 0.1 int64
Unnamed: 0 int64
Text object
Sentiment object
Timestamp object
User object
Platform object
Hashtags object
Retweets float64
Likes float64
Country object
Year int64
Month int64
Day int64
Hour int64
dtype: object

In [12]: data.describe()

Out[12]:

	Unnamed: 0.1	Unnamed: 0	Retweets	Likes	Year	Month	Day	Hour
count	732.000000	732.000000	732.000000	732.000000	732.000000	732.000000	732.000000	732.000000
mean	366.464481	369.740437	21.508197	42.901639	2020.471311	6.122951	15.497268	15.52185
std	211.513936	212.428936	7.061286	14.089848	2.802285	3.411763	8.474553	4.11341
min	0.000000	0.000000	5.000000	10.000000	2010.000000	1.000000	1.000000	0.000000
25%	183.750000	185.750000	17.750000	34.750000	2019.000000	3.000000	9.000000	13.000000
50%	366.500000	370.500000	22.000000	43.000000	2021.000000	6.000000	15.000000	16.000000
75%	549.250000	553.250000	25.000000	50.000000	2023.000000	9.000000	22.000000	19.000000
max	732.000000	736.000000	40.000000	80.000000	2023.000000	12.000000	31.000000	23.000000

In [14]:

```
data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 732 entries, 0 to 731
Data columns (total 15 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Unnamed: 0.1    732 non-null   int64
1   Unnamed: 0      732 non-null   int64
2   Text            732 non-null   object
3   Sentiment       732 non-null   object
4   Timestamp       732 non-null   object
5   User            732 non-null   object
6   Platform        732 non-null   object
7   Hashtags        732 non-null   object
8   Retweets        732 non-null   float64
9   Likes           732 non-null   float64
10  Country         732 non-null   object
11  Year            732 non-null   int64
12  Month           732 non-null   int64
13  Day             732 non-null   int64
14  Hour            732 non-null   int64
dtypes: float64(2), int64(6), object(7)
memory usage: 85.9+ KB
```

In [16]:

```
data.isnull().sum()
```

```
Out[16]: Unnamed: 0.1    0
          Unnamed: 0    0
          Text          0
          Sentiment     0
          Timestamp     0
          User          0
          Platform      0
          Hashtags      0
          Retweets      0
          Likes         0
          Country       0
          Year          0
          Month         0
          Day           0
          Hour          0
          dtype: int64
```

```
In [18]: data.duplicated().sum()
```

```
Out[18]: 0
```

```
In [20]: import matplotlib.pyplot as plt
          import seaborn as sns
```

```
In [22]: temp = data.groupby('Sentiment').count()['Text'].reset_index().sort_values(by='Text', ascending=False)
          temp.style.background_gradient(cmap='Purples')
```

Out[22]:

	Sentiment	Text
214	Positive	44
172	Joy	42
110	Excitement	32
139	Happy	14
56	Contentment	14
196	Neutral	14
128	Gratitude	9
239	Sad	9
152	Hopeful	9
64	Curiosity	8
91	Embarrassed	8
178	Loneliness	7
89	Elation	6
211	Playful	6
143	Hate	6
28	Bad	6
72	Despair	6
187	Melancholy	5
158	Indifference	5
163	Inspired	5
32	Bitterness	5
121	Frustrated	5
1	Acceptance	5
68	Curiosity	5
132	Gratitude	5
53	Confusion	5
201	Nostalgia	5
203	Numbness	5
102	Enthusiasm	5
12	Ambivalence	5
57	Contentment	5
95	Empowerment	5
245	Serenity	5
246	Serenity	5

	Sentiment	Text
78	Determination	5
48	Compassionate	4
111	Excitement	4
133	Grief	4
122	Frustration	4
127	Grateful	4
109	Euphoria	4
160	Inspiration	4
71	Desolation	4
24	Awe	4
151	Hope	4
198	Nostalgia	4
197	Neutral	4
22	Arousal	4
195	Negative	4
243	Serenity	4
29	Betrayal	4
261	Tenderness	4
220	Proud	4
79	Devastated	3
94	Empathetic	3
256	Surprise	3
85	Dismissive	3
231	Resentment	3
225	Regret	3
135	Grief	3
218	Pride	3
217	Pride	3
103	Envious	3
116	Fearful	3
118	Free-spirited	3
176	Kind	3
168	Jealous	3
208	Overwhelmed	3

	Sentiment	Text
0	Acceptance	3
31	Bitter	3
2	Accomplishment	3
50	Confident	3
47	Compassion	3
8	Adventure	3
36	Boredom	3
38	Calmness	3
137	Happiness	2
136	Happiness	2
40	Captivation	2
134	Grief	2
240	Sadness	2
278	Zest	2
241	Satisfaction	2
6	Adoration	2
129	Gratitude	2
4	Admiration	2
125	Fulfillment	2
124	Fulfillment	2
123	Frustration	2
248	Shame	2
115	Fear	2
235	Reverence	2
145	Heartbreak	2
234	Reverence	2
228	Rejuvenation	2
192	Mischievous	2
25	Awe	2
184	Love	2
181	Loss	2
26	Awe	2
16	Anticipation	2
173	Joy	2

	Sentiment	Text
15	Anger	2
170	Jealousy	2
167	Isolation	2
13	Amusement	2
223	Reflection	2
224	Reflection	2
9	Affection	2
227	Regret	2
114	Exploration	2
21	Apprehensive	2
58	Coziness	2
73	Despair	2
77	Determination	2
75	Despair	2
96	Enchantment	2
97	Enchantment	2
262	Thrill	2
100	Enjoyment	2
101	Enthusiasm	2
55	Contemplation	2
80	Disappointed	2
92	Emotion	2
81	Disappointment	2
267	Tranquility	2
105	Envy	2
257	Surprise	2
270	Whimsy	2
84	Disgust	2
83	Disgust	2
52	Confusion	2
277	Yearning	2
60	Creativity	2
14	Amusement	1
215	Positivity	1

	Sentiment	Text
213	Positive	1
216	Pressure	1
271	Whispers of the Past	1
272	Winter Magic	1
212	PlayfulJoy	1
273	Wonder	1
254	Success	1
210	Pensive	1
209	Overwhelmed	1
20	Appreciation	1
19	Anxiety	1
199	Nostalgia	1
200	Nostalgia	1
276	Wonderment	1
18	Anxiety	1
202	Numbness	1
17	Anticipation	1
204	Obstacle	1
275	Wonder	1
205	Ocean's Freedom	1
219	Pride	1
206	Optimism	1
207	Overjoyed	1
274	Wonder	1
11	Ambivalence	1
269	Vibrancy	1
10	Amazement	1
221	Radiance	1
255	Suffering	1
252	Sorrow	1
251	Sorrow	1
3	Admiration	1
250	Solitude	1
249	Solace	1

	Sentiment	Text
247	Shame	1
258	Surprise	1
259	Suspense	1
260	Sympathy	1
5	Admiration	1
244	Serenity	1
23	ArtisticBurst	1
263	Thrill	1
7	Adrenaline	1
238	Runway Creativity	1
237	Ruins	1
236	Romance	1
264	Thrill	1
265	Thrilling Journey	1
233	Resilience	1
232	Resilience	1
266	Touched	1
230	Renewed Effort	1
229	Relief	1
268	Triumph	1
253	Spark	1
226	Regret	1
222	Radiance	1
242	Satisfaction	1
69	Darkness	1
194	Nature's Beauty	1
46	Compassion	1
108	Euphoria	1
51	Confusion	1
112	Excitement	1
113	Exhaustion	1
49	Confidence	1
117	FestiveJoy	1
119	Freedom	1

	Sentiment	Text
120	Friendship	1
45	Colorful	1
193	Motivation	1
126	Grandeur	1
44	Charm	1
43	Challenge	1
130	Gratitude	1
131	Gratitude	1
42	Celestial Wonder	1
41	Celebration	1
138	Happiness	1
107	Euphoria	1
106	Euphoria	1
104	Envisioning History	1
54	Connection	1
67	Curiosity	1
66	Curiosity	1
74	Despair	1
76	Desperation	1
65	Curiosity	1
63	CulinaryOdyssey	1
82	Disgust	1
62	Culinary Adventure	1
86	DreamChaser	1
87	Ecstasy	1
88	Elation	1
61	Creativity	1
90	Elegance	1
93	EmotionalStorm	1
59	Creative Inspiration	1
98	Energy	1
99	Engagement	1
39	Calmness	1
140	Harmony	1

	Sentiment	Text
141	Harmony	1
165	Intimidation	1
169	Jealousy	1
171	Journey	1
30	Betrayal	1
70	Dazzle	1
175	JoyfulReunion	1
27	Awe	1
177	Kindness	1
179	Loneliness	1
180	Loneliness	1
182	LostLove	1
183	Love	1
185	Marvel	1
186	Melancholy	1
188	Melodic	1
189	Mesmerizing	1
190	Mindfulness	1
191	Miscalculation	1
166	Intrigue	1
164	Intimidation	1
142	Harmony	1
33	Bittersweet	1
144	Heartache	1
146	Heartbreak	1
147	Heartwarming	1
148	Helplessness	1
149	Helplessness	1
150	Hope	1
37	Breakthrough	1
153	Hypnotic	1
154	Iconic	1
155	Imagination	1
156	Immersion	1

	Sentiment	Text
157	Indifference	1
35	Boredom	1
159	InnerJourney	1
34	Blessed	1
161	Inspiration	1
162	Inspiration	1
174	Joy in Baking	1

```
In [24]: #Common words used on Text
from collections import Counter
from sklearn.model_selection import train_test_split

data['temp_list'] = data['Text'].apply(lambda x:str(x).split())
top = Counter([item for sublist in data['temp_list'] for item in sublist])
temp = pd.DataFrame(top.most_common(20))
temp.columns = ['Common_words', 'count']
temp.style.background_gradient(cmap='Blues')
```

Out[24]:

	Common_words	count
0	the	808
1	of	623
2	a	621
3	in	259
4	to	133
5	and	111
6	with	107
7	for	99
8	on	91
9	by	69
10	through	51
11	an	49
12	my	41
13	at	40
14	new	39
15	each	39
16	In	39
17	that	37
18	as	36
19	A	33

```
In [26]: #Bar plot for Common words on Text
import plotly.express as px
fig = px.bar(temp, x="count", y="Common_words", title='Common Words in Selected Text',
             width=500, height=500,color='Common_words')
fig.show()
```



```
In [28]: top = Counter([item for sublist in data['temp_list'] for item in sublist])
temp = pd.DataFrame(top.most_common(20))
temp = temp.iloc[1:,:]
temp.columns = ['Common_words', 'count']
temp.style.background_gradient(cmap='Purples')
```

Out[28]:

	Common_words	count
1	of	623
2	a	621
3	in	259
4	to	133
5	and	111
6	with	107
7	for	99
8	on	91
9	by	69
10	through	51
11	an	49
12	my	41
13	at	40
14	new	39
15	each	39
16	In	39
17	that	37
18	as	36
19	A	33

```
In [30]: #TreeMap of Most Common Words
import plotly.express as px
fig = px.treemap(temp, path=['Common_words'], values='count',width=500, height=500,tit
fig.show()
```

In [32]: `!pip install textblob`

Collecting textblob

Obtaining dependency information for textblob from <https://files.pythonhosted.org/packages/02/07/5fd2945356dd839974d3a25de8a142dc37293c21315729a41e775b5f3569/textblob-0.18.0.post0-py3-none-any.whl.metadata>

Downloading textblob-0.18.0.post0-py3-none-any.whl.metadata (4.5 kB)

Requirement already satisfied: nltk>=3.8 in c:\users\dell\anaconda3\lib\site-packages (from textblob) (3.8.1)

Requirement already satisfied: click in c:\users\dell\anaconda3\lib\site-packages (from nltk>=3.8->textblob) (8.0.4)

Requirement already satisfied: joblib in c:\users\dell\anaconda3\lib\site-packages (from nltk>=3.8->textblob) (1.2.0)

Requirement already satisfied: regex>=2021.8.3 in c:\users\dell\anaconda3\lib\site-packages (from nltk>=3.8->textblob) (2022.7.9)

Requirement already satisfied: tqdm in c:\users\dell\anaconda3\lib\site-packages (from nltk>=3.8->textblob) (4.65.0)

Requirement already satisfied: colorama in c:\users\dell\anaconda3\lib\site-packages (from click->nltk>=3.8->textblob) (0.4.6)

Downloading textblob-0.18.0.post0-py3-none-any.whl (626 kB)

```
----- 0.0/626.3 kB ? eta -:-:--
----- 10.2/626.3 kB ? eta -:-:--
-- ----- 41.0/626.3 kB 487.6 kB/s eta 0:00:02
--- ----- 61.4/626.3 kB 656.4 kB/s eta 0:00:01
----- 92.2/626.3 kB 655.4 kB/s eta 0:00:01
----- 122.9/626.3 kB 654.9 kB/s eta 0:00:01
----- 153.6/626.3 kB 654.6 kB/s eta 0:00:01
----- 204.8/626.3 kB 692.4 kB/s eta 0:00:01
----- 204.8/626.3 kB 692.4 kB/s eta 0:00:01
----- 256.0/626.3 kB 714.4 kB/s eta 0:00:01
----- 286.7/626.3 kB 707.1 kB/s eta 0:00:01
----- 317.4/626.3 kB 678.1 kB/s eta 0:00:01
----- 358.4/626.3 kB 719.3 kB/s eta 0:00:01
----- 399.4/626.3 kB 732.8 kB/s eta 0:00:01
----- 430.1/626.3 kB 726.4 kB/s eta 0:00:01
----- 460.8/626.3 kB 720.9 kB/s eta 0:00:01
----- 491.5/626.3 kB 716.2 kB/s eta 0:00:01
----- 512.0/626.3 kB 713.7 kB/s eta 0:00:01
----- 542.7/626.3 kB 725.4 kB/s eta 0:00:01
----- 593.9/626.3 kB 717.9 kB/s eta 0:00:01
----- 626.3/626.3 kB 717.2 kB/s eta 0:00:00
```

Installing collected packages: textblob

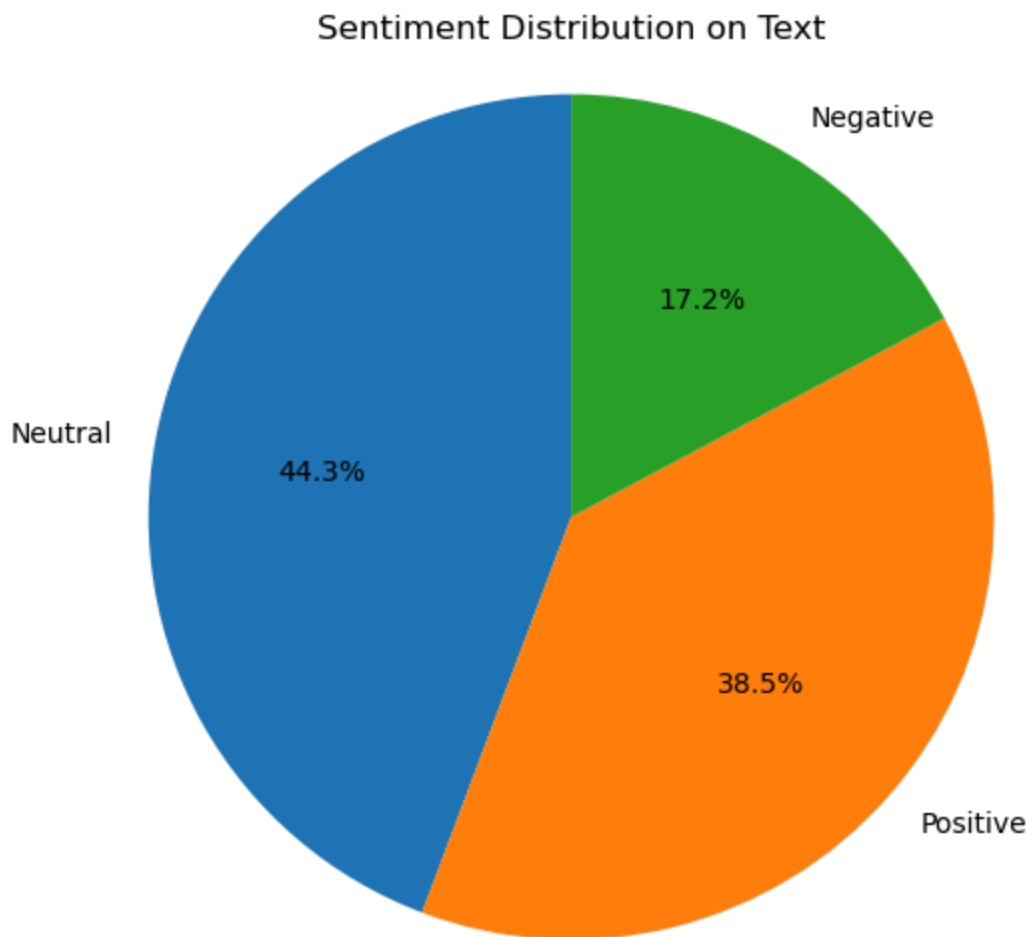
Successfully installed textblob-0.18.0.post0

```
In [34]: # Perform sentiment analysis on text
import sys
from textblob import TextBlob
data['Sentiment'] = data['Text'].apply(lambda x: TextBlob(x).sentiment.polarity)

# Categorize sentiment into positive, negative, and neutral
data['Sentiment Category'] = data['Sentiment'].apply(lambda x: 'Positive' if x > 0 else 'Negative' if x < 0 else 'Neutral')

# Calculate the count of each sentiment category
sentiment_counts = data['Sentiment Category'].value_counts()

# Plot a pie chart of sentiment distribution
plt.figure(figsize=(6, 6))
plt.pie(sentiment_counts, labels=sentiment_counts.index, autopct='%1.1f%%', startangle=90)
plt.axis('equal')
plt.title('Sentiment Distribution on Text')
plt.show()
```



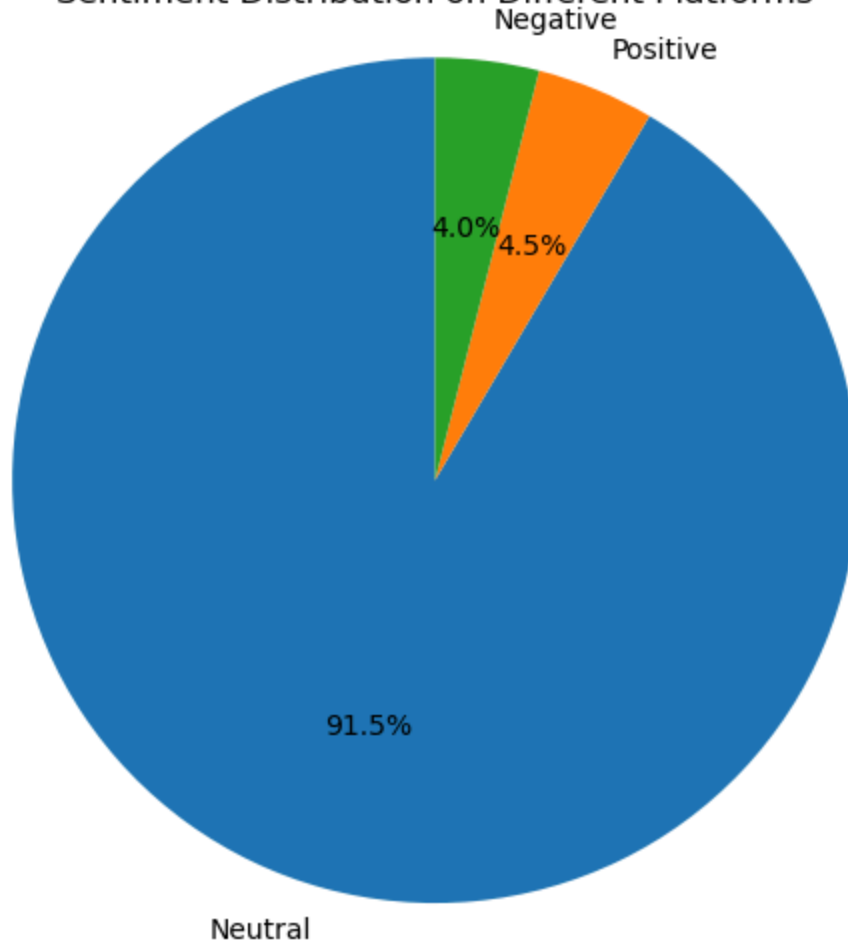
```
In [36]: # Perform sentiment analysis on platform
import sys
from textblob import TextBlob
data['Sentiment'] = data['Hashtags'].apply(lambda x: TextBlob(x).sentiment.polarity)

# Categorize sentiment into positive, negative, and neutral
data['Sentiment Category'] = data['Sentiment'].apply(lambda x: 'Positive' if x > 0 else 'Negative' if x < 0 else 'Neutral')

# Calculate the count of each sentiment category
sentiment_counts = data['Sentiment Category'].value_counts()

# Plot a pie chart of sentiment distribution
plt.figure(figsize=(6, 6))
plt.pie(sentiment_counts, labels=sentiment_counts.index, autopct='%1.1f%%', startangle=90)
plt.axis('equal')
plt.title('Sentiment Distribution on Different Platforms')
plt.show()
```

Sentiment Distribution on Different Platforms



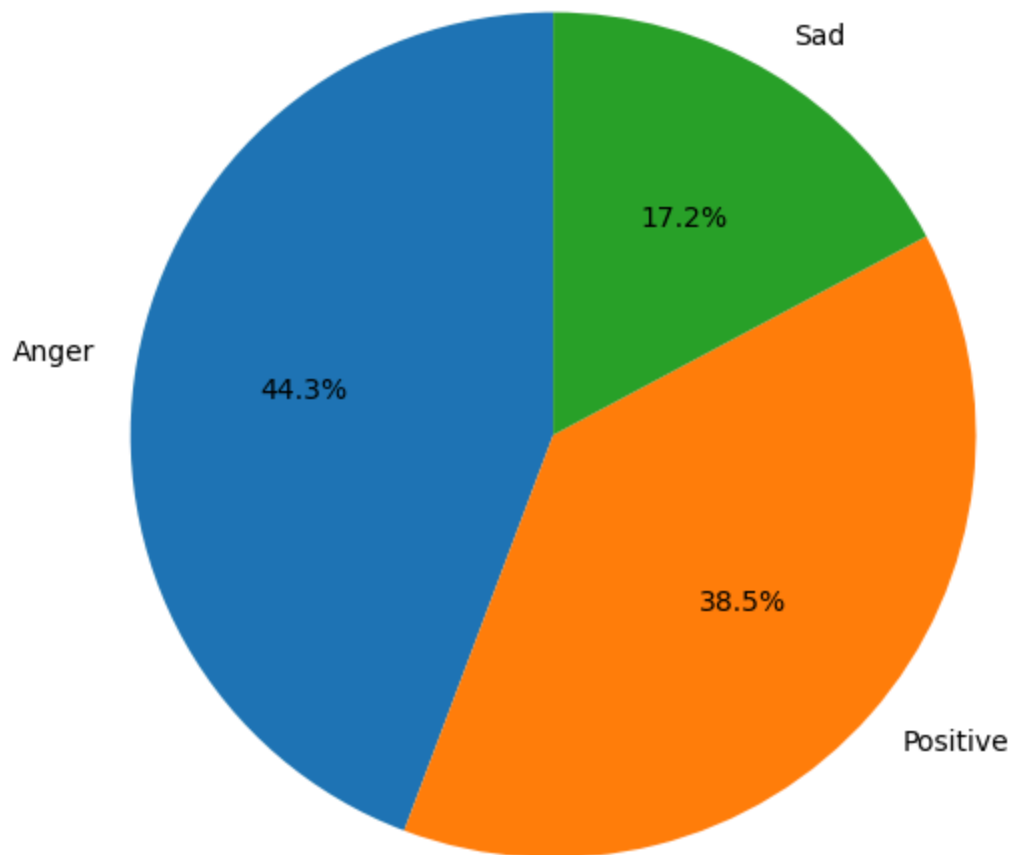
```
In [38]: # Perform sentiment analysis on text
import sys
from textblob import TextBlob
data['Sentiment'] = data['Text'].apply(lambda x: TextBlob(x).sentiment.polarity)

# Categorize sentiment into positive, sad, and anger
data['Sentiment Category'] = data['Sentiment'].apply(lambda x: 'Positive' if x > 0 else

# Calculate the count of each sentiment category
sentiment_counts = data['Sentiment Category'].value_counts()

# Plot a pie chart of sentiment distribution
plt.figure(figsize=(6, 6))
plt.pie(sentiment_counts, labels=sentiment_counts.index, autopct='%1.1f%%', startangle=
plt.axis('equal')
plt.title('Sentiment Distribution on Text')
plt.show()
```

Sentiment Distribution on Text



```
In [40]: #Distribution of various Likes
plt.figure(figsize=(8,6))
p=data.groupby(['Likes']).size().reset_index(name='counts')
piechart=px.pie(p,values='counts',names='Likes',title='Likes count')
piechart.show()
```

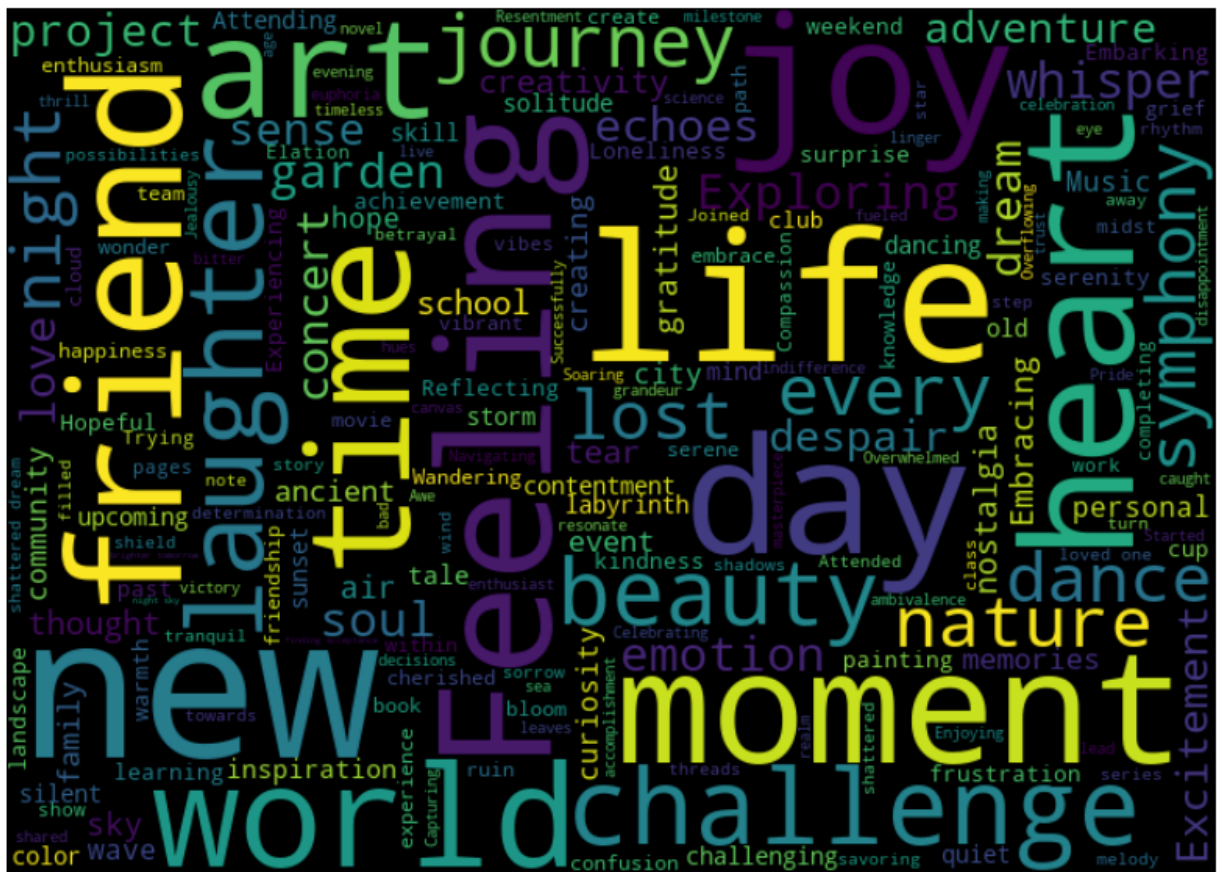
<Figure size 800x600 with 0 Axes>

```
In [42]: #Top 5 Hashtags Analysis
Hashtags_name=pd.DataFrame(data)
Hashtags_name=data['Hashtags'].str.split(',',expand=True).stack()
Hashtags_name=Hashtags_name.to_frame()
Hashtags_name.columns=['Hashtags']
Hashtags=Hashtags_name.groupby(['Hashtags']).size().reset_index(name='Total Content')
Hashtags=Hashtags[Hashtags.Hashtags !='Not specified']
Hashtags=Hashtags.sort_values(by=['Total Content'],ascending=False)
HashtagsTop5=Hashtags.head()
HashtagsTop5=HashtagsTop5.sort_values(by=['Total Content'])
fig1=px.bar(HashtagsTop5,x='Total Content',y='Hashtags',width=800, height=500,title='1')
fig1.show()
```

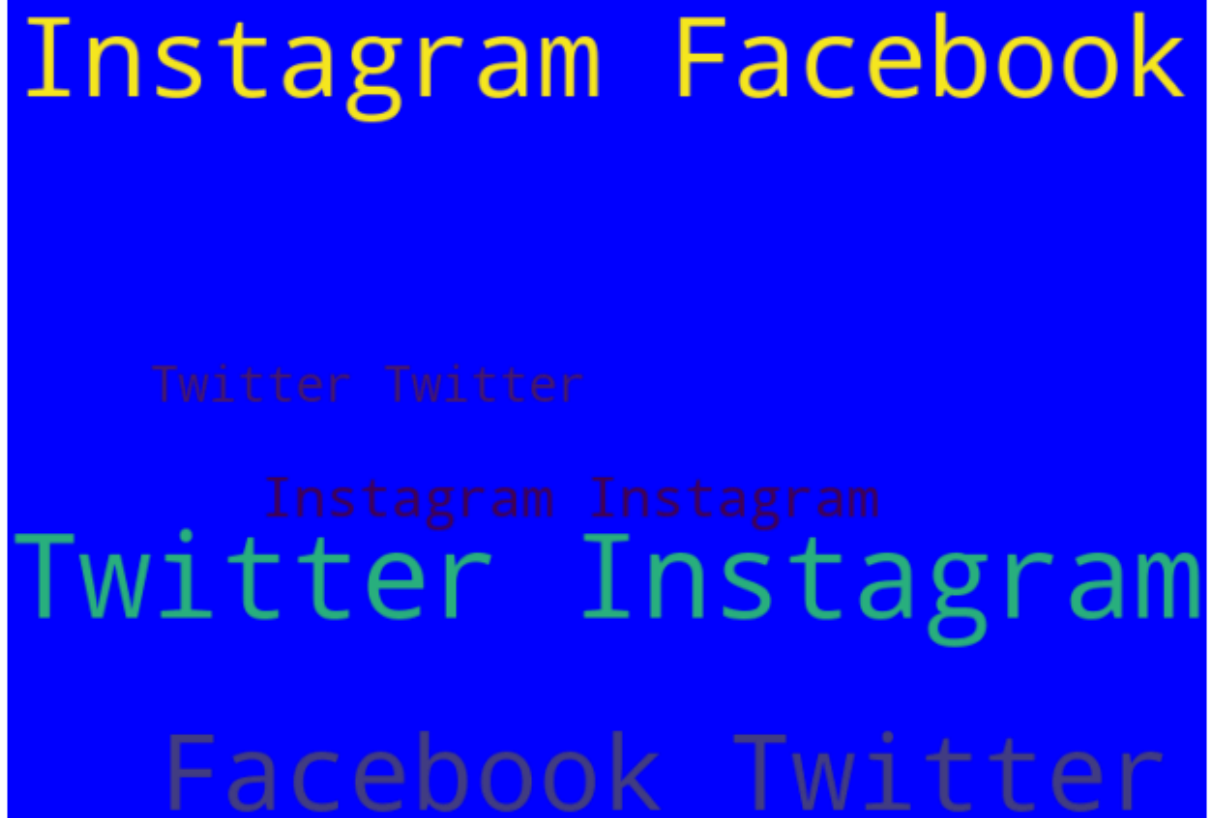


```
In [44]: #Top 5 Text Analysis
Text_name=pd.DataFrame(data)
Text_name=data['Text'].str.split(',',expand=True).stack()
Text_name=Text_name.to_frame()
Text_name.columns=['Text']
Text=Text_name.groupby(['Text']).size().reset_index(name='Total Content')
Text=Text[Text.Text != 'Not specified']
Text=Text.sort_values(by=['Total Content'],ascending=False)
TextTop5=Text.head()
TextTop5=TextTop5.sort_values(by=['Total Content'])
fig1=px.bar(TextTop5,x='Total Content',y='Text',width=800, height=500,title='Top 5 Text')
fig1.show()
```

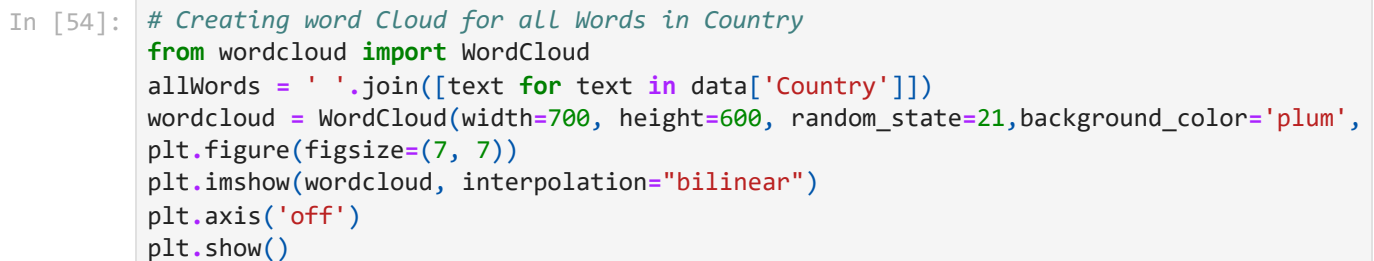
```
In [46]: # Creating word Cloud for all Words in all platform Text
from wordcloud import WordCloud
allWords = ' '.join([text for text in data['Text']])
wordcloud = WordCloud(width=700, height=500, random_state=21, max_font_size=115).generate(allWords)
plt.figure(figsize=(10, 10))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis('off')
plt.show()
```

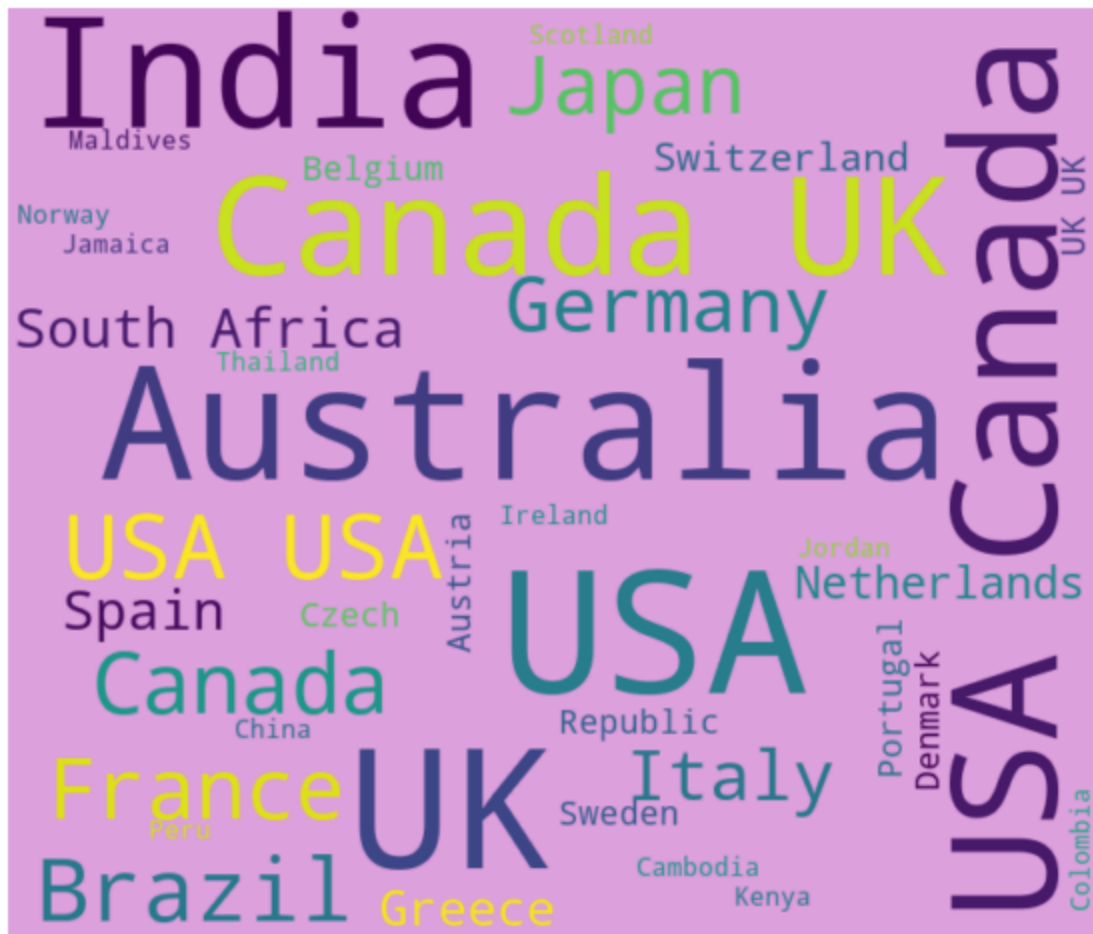


```
In [50]: # Creating word Cloud for all Words in Platform
from wordcloud import WordCloud
allWords = ' '.join([text for text in data['Platform']])
wordcloud = WordCloud(width=700, height=600, random_state=21, background_color='BLUE',
plt.figure(figsize=(8, 8))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis('off')
plt.show()
```



```
In [52]: # Creating word Cloud for all Words in Users
from wordcloud import WordCloud
allWords = ' '.join([text for text in data['User']])
wordcloud = WordCloud(width=700, height=700, random_state=21, background_color='lightblue')
plt.figure(figsize=(7, 7))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis('off')
plt.show()
```





In []: