```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

df=pd.DataFrame(pd.read_csv("sentimentdataset.csv"))

df.head()
```

```
   Unnamed: 0.1  Unnamed: 0  \
0             0           0
1             1           1
2             2           2
3             3           3
4             4           4

                                          Text    Sentiment  \
0   Enjoying a beautiful day at the park!        ...   Positive
1   Traffic was terrible this morning.           ...   Negative
2   Just finished an amazing workout! □          ...   Positive
3   Excited about the upcoming weekend getaway!  ...   Positive
4   Trying out a new recipe for dinner tonight.  ...   Neutral

              Timestamp            User     Platform  \
0  2023-01-15 12:30:00   User123          Twitter
1  2023-01-15 08:45:00   CommuterX        Twitter
2  2023-01-15 15:45:00   FitnessFan      Instagram
3  2023-01-15 18:20:00   AdventureX       Facebook
4  2023-01-15 19:55:00   ChefCook        Instagram

                                    Hashtags  Retweets  Likes
Country  \
0   #Nature #Park                                 15.0   30.0     USA

1   #Traffic #Morning                              5.0   10.0
Canada
2   #Fitness #Workout                             20.0   40.0   USA

3   #Travel #Adventure                             8.0   15.0     UK

4   #Cooking #Food                                12.0   25.0
Australia

   Year  Month  Day  Hour
0  2023      1   15    12
1  2023      1   15     8
2  2023      1   15    15
3  2023      1   15    18
4  2023      1   15    19
```

```
df.tail()
```

```
     Unnamed: 0.1  Unnamed: 0  \
727           728         732
728           729         733
729           730         734
730           731         735
731           732         736

                                                  Text Sentiment  \
727  Collaborating on a science project that receiv...    Happy
728  Attending a surprise birthday party organized ...    Happy
729  Successfully fundraising for a school charity ...    Happy
730  Participating in a multicultural festival, cel...    Happy
731  Organizing a virtual talent show during challe...    Happy

                Timestamp                                    User
Platform  \
727  2017-08-18 18:20:00       ScienceProjectSuccessHighSchool
Facebook
728  2018-06-22 14:15:00            BirthdayPartyJoyHighSchool
Instagram
729  2019-04-05 17:30:00   CharityFundraisingTriumphHighSchool
Twitter
730  2020-02-29 20:45:00    MulticulturalFestivalJoyHighSchool
Facebook
731  2020-11-15 15:15:00    VirtualTalentShowSuccessHighSchool
Instagram

                                          Hashtags  Retweets  Likes
Country  \
727         #ScienceFairWinner #HighSchoolScience       20.0   39.0
UK
728    #SurpriseCelebration #HighSchoolFriendship       25.0   48.0
USA
729      #CommunityGiving #HighSchoolPhilanthropy       22.0   42.0
Canada
730         #CulturalCelebration #HighSchoolUnity       21.0   43.0
UK
731   #VirtualEntertainment #HighSchoolPositivity       24.0   47.0
USA

     Year  Month  Day  Hour
727  2017      8   18    18
728  2018      6   22    14
729  2019      4    5    17
730  2020      2   29    20
731  2020     11   15    15

df.shape

(732, 15)
```

```
df.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

df.isnull()

     Unnamed: 0.1  Unnamed: 0   Text  Sentiment  Timestamp   User
Platform  \
0           False       False  False      False      False  False
False
1           False       False  False      False      False  False
False
2           False       False  False      False      False  False
False
3           False       False  False      False      False  False
False
4           False       False  False      False      False  False
False
..            ...         ...    ...        ...        ...    ...
...
727         False       False  False      False      False  False
False
728         False       False  False      False      False  False
False
729         False       False  False      False      False  False
False
730         False       False  False      False      False  False
False
731         False       False  False      False      False  False
```

```
False
```

	Hashtags	Retweets	Likes	Country	Year	Month	Day	Hour
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
..	...	...	...	...	...	...	...	...
727	False	False	False	False	False	False	False	False
728	False	False	False	False	False	False	False	False
729	False	False	False	False	False	False	False	False
730	False	False	False	False	False	False	False	False
731	False	False	False	False	False	False	False	False

```
[732 rows x 15 columns]
```

```
df.isnull().sum()
```

```
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
```

```
df.isnull().sum()/len(df)*100
```

```
Unnamed: 0.1    0.0
Unnamed: 0      0.0
Text            0.0
Sentiment       0.0
Timestamp       0.0
User            0.0
Platform        0.0
Hashtags        0.0
Retweets        0.0
Likes           0.0
Country         0.0
Year            0.0
```

```
Month           0.0
Day             0.0
Hour            0.0
dtype: float64
```

```
pip install missingno
```

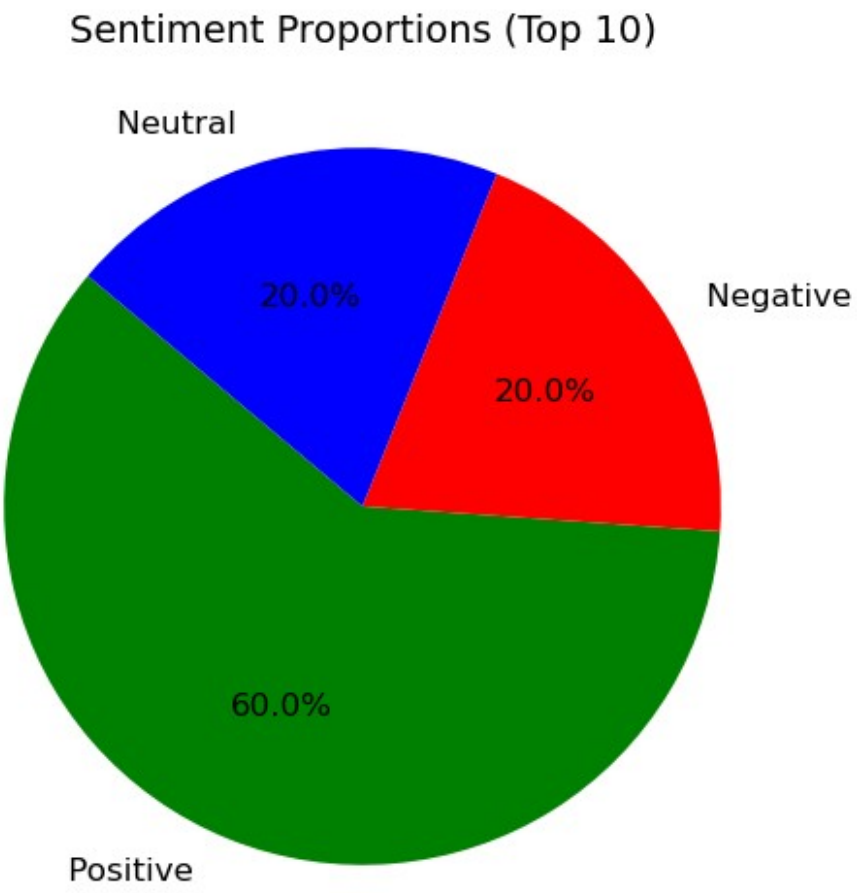
```
Requirement already satisfied: missingno in c:\users\nivas\anaconda3\
lib\site-packages (0.5.2)
Requirement already satisfied: numpy in c:\users\nivas\anaconda3\lib\
site-packages (from missingno) (1.26.4)
Requirement already satisfied: matplotlib in c:\users\nivas\anaconda3\
lib\site-packages (from missingno) (3.8.0)
Requirement already satisfied: scipy in c:\users\nivas\anaconda3\lib\
site-packages (from missingno) (1.11.4)
Requirement already satisfied: seaborn in c:\users\nivas\anaconda3\
lib\site-packages (from missingno) (0.12.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\nivas\
anaconda3\lib\site-packages (from matplotlib->missingno) (1.2.0)
Requirement already satisfied: cycler>=0.10 in c:\users\nivas\
anaconda3\lib\site-packages (from matplotlib->missingno) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\nivas\
anaconda3\lib\site-packages (from matplotlib->missingno) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\nivas\
anaconda3\lib\site-packages (from matplotlib->missingno) (1.4.4)
Requirement already satisfied: packaging>=20.0 in c:\users\nivas\
anaconda3\lib\site-packages (from matplotlib->missingno) (23.1)
Requirement already satisfied: pillow>=6.2.0 in c:\users\nivas\
anaconda3\lib\site-packages (from matplotlib->missingno) (10.2.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\nivas\
anaconda3\lib\site-packages (from matplotlib->missingno) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\nivas\
anaconda3\lib\site-packages (from matplotlib->missingno) (2.8.2)
Requirement already satisfied: pandas>=0.25 in c:\users\nivas\
anaconda3\lib\site-packages (from seaborn->missingno) (2.1.4)
Requirement already satisfied: pytz>=2020.1 in c:\users\nivas\
anaconda3\lib\site-packages (from pandas>=0.25->seaborn->missingno)
(2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in c:\users\nivas\
anaconda3\lib\site-packages (from pandas>=0.25->seaborn->missingno)
(2023.3)
Requirement already satisfied: six>=1.5 in c:\users\nivas\anaconda3\
lib\site-packages (from python-dateutil>=2.7->matplotlib->missingno)
(1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

```
top_10_data = df.head(10)
```

```
top_10_sentiment_counts = top_10_data['Sentiment'].value_counts()
```

TOP 10 SENTIMENT PROPORTIONS

```
plt.figure(figsize=(6, 6))
top_10_sentiment_counts.plot(kind='pie', autopct='%1.1f%%',
startangle=140, colors=['green', 'red', 'blue'],
textprops={'fontsize': 12})
plt.title('Sentiment Proportions (Top 10)', fontsize=14)
plt.ylabel('')
plt.show()
```

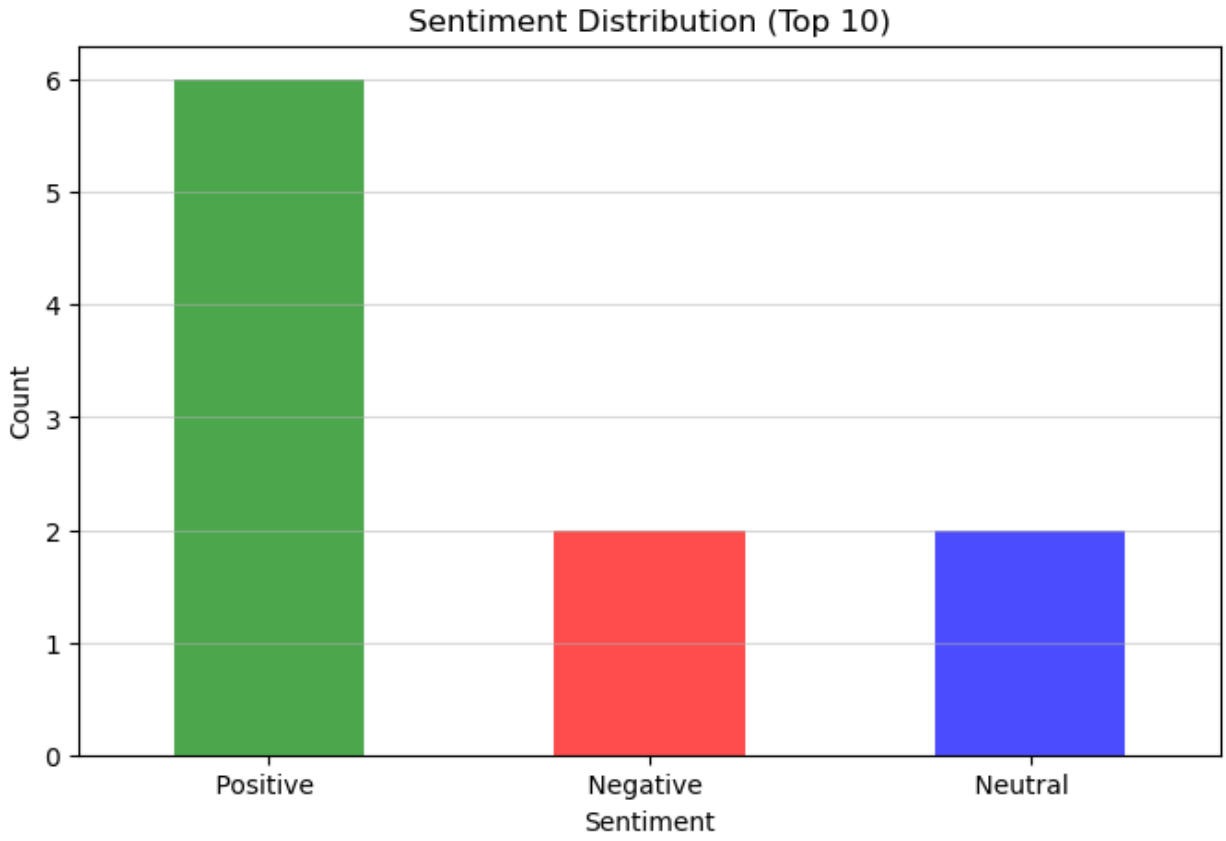



```
top_data = df.head(10)
```

SENTIMENT DISTRIBUTION

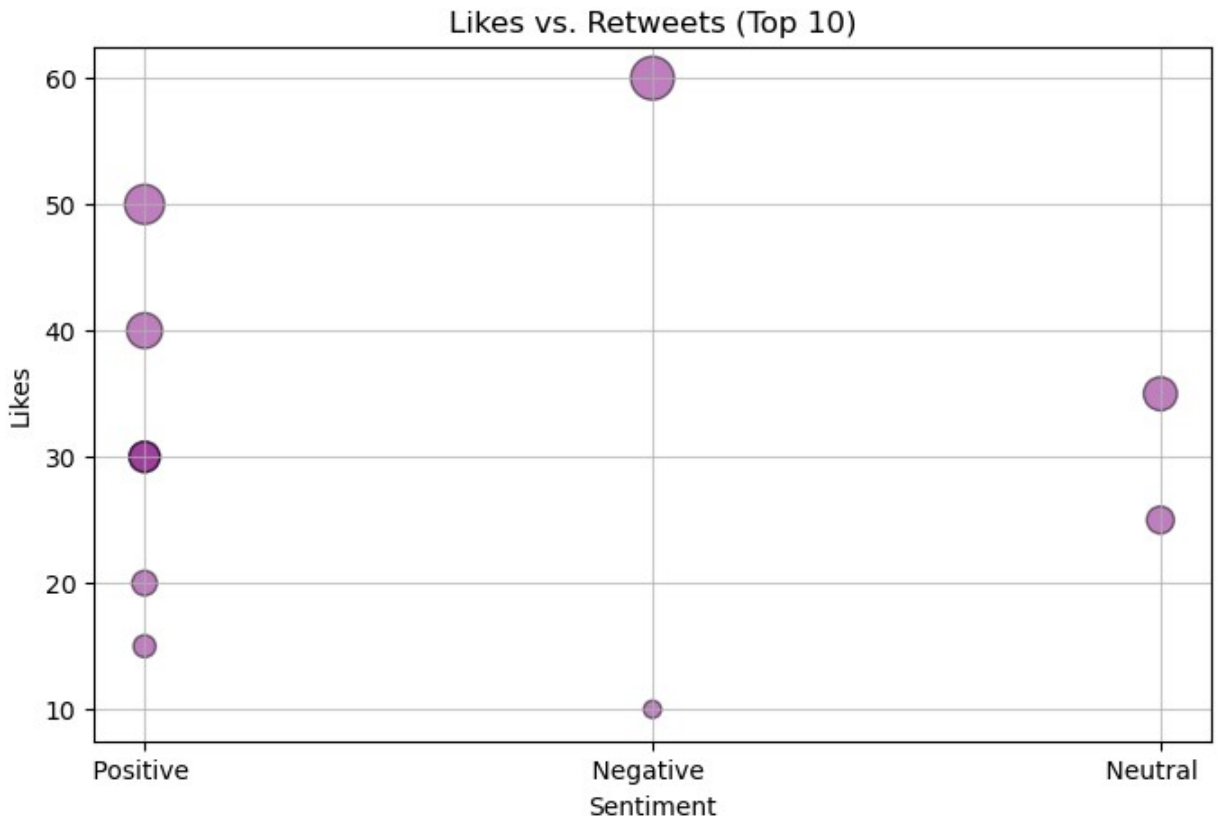
```
top_sentiment_counts = top_data['Sentiment'].value_counts()
plt.figure(figsize=(8, 5))
top_sentiment_counts.plot(kind='bar', color=['green', 'red', 'blue'],
alpha=0.7)
plt.title('Sentiment Distribution (Top 10)')
```

```
plt.xlabel('Sentiment')
plt.ylabel('Count')
plt.xticks(rotation=0)
plt.grid(axis='y', alpha=0.5)
plt.show()
```



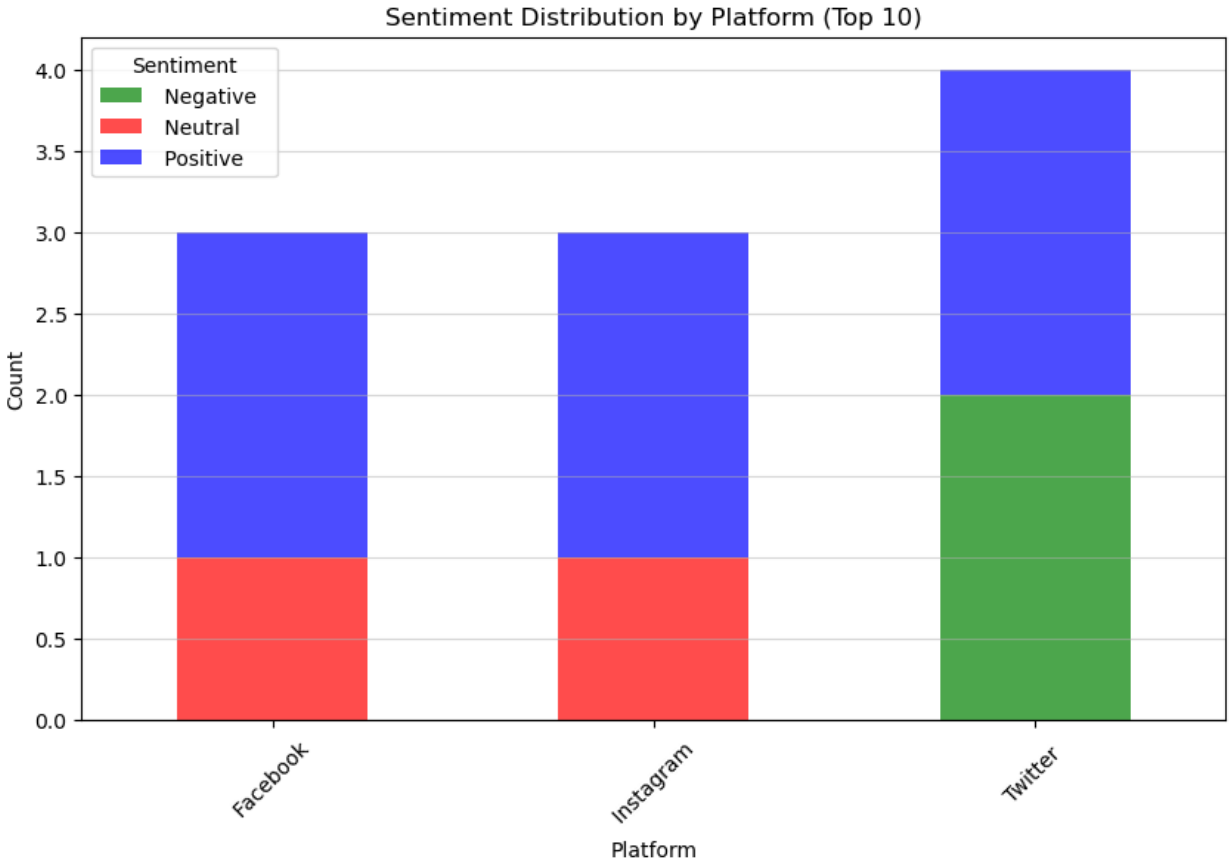
LIKES VS RETWEETS

```
plt.figure(figsize=(8, 5))
plt.scatter(top_data['Sentiment'], top_data['Likes'],
s=top_data['Retweets'] * 10, alpha=0.5, color='purple',
edgecolors='k')
plt.title('Likes vs. Retweets (Top 10)')
plt.xlabel('Sentiment')
plt.ylabel('Likes')
plt.grid(alpha=0.7)
plt.show()
```



TOP 10 SENTIMENT DISTRIBUTION BY PLATFORM

```
top10_data = df.head(10)
stacked_data_top10 = top10_data.groupby(['Platform',
'Sentiment']).size().unstack(fill_value=0)
stacked_data_top10.plot(kind='bar', stacked=True, figsize=(10, 6),
color=['green', 'red', 'blue'], alpha=0.7)
plt.title('Sentiment Distribution by Platform (Top 10)')
plt.xlabel('Platform')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.legend(title='Sentiment', fontsize=10)
plt.grid(axis='y', alpha=0.5)
plt.show()
```

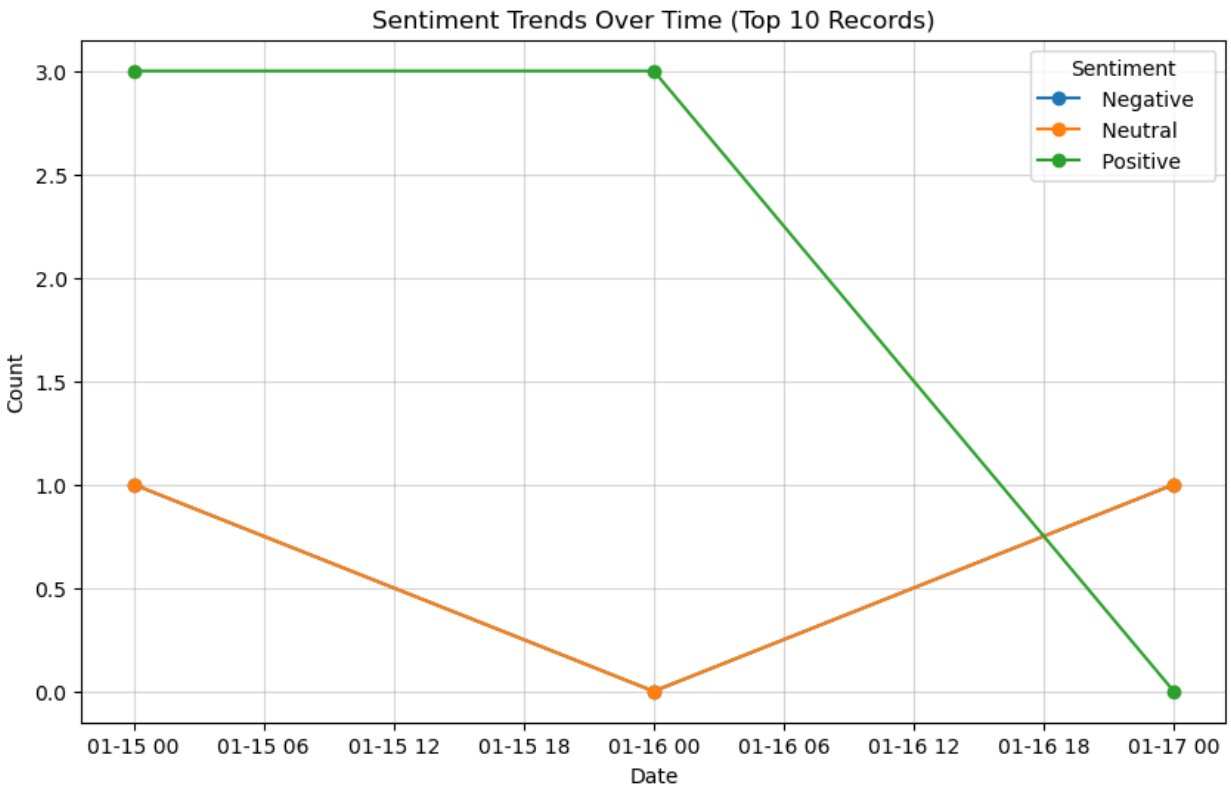


TOP 10 SENTIMENT TRENDS OVER TIME

```
top10_data = df.head(10)
top10_data['Timestamp'] = pd.to_datetime(top10_data['Timestamp'])
sentiment_over_time_top10 =
top10_data.groupby([top10_data['Timestamp'].dt.date,
'Sentiment']).size().unstack(fill_value=0)
sentiment_over_time_top10.plot(figsize=(10, 6), marker='o')
plt.title('Sentiment Trends Over Time (Top 10 Records)')
plt.xlabel('Date')
plt.ylabel('Count')
plt.grid(alpha=0.5)
plt.legend(title='Sentiment', fontsize=10)
plt.show()
```

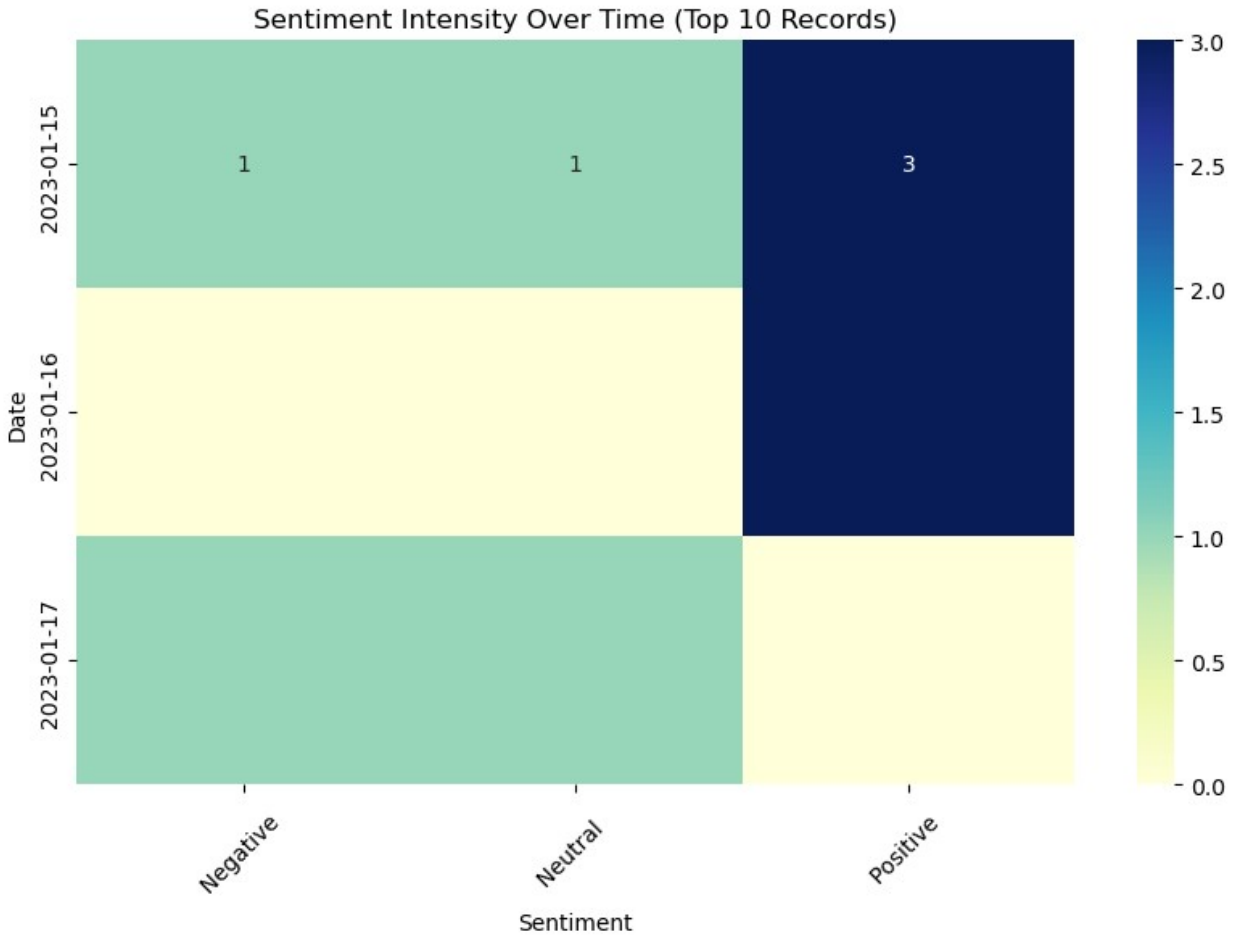
```
C:\Users\nivas\AppData\Local\Temp\ipykernel_2840\89274623.py:2:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
```

```
returning-a-view-versus-a-copy
  top10_data['Timestamp'] = pd.to_datetime(top10_data['Timestamp'])
```



TOP 10 SENTIMENT INTENSITY OVER TIME

```
import seaborn as sns
heatmap_data = top10_data.groupby([top10_data['Timestamp'].dt.date,
'Sentiment']).size().unstack(fill_value=0)
plt.figure(figsize=(10, 6))
sns.heatmap(heatmap_data, annot=True, fmt='d', cmap='YlGnBu',
cbar=True)
plt.title('Sentiment Intensity Over Time (Top 10 Records)')
plt.xlabel('Sentiment')
plt.ylabel('Date')
plt.xticks(rotation=45)
plt.show()
```


Sentiment Intensity Over Time (Top 10 Records)
Date
2023-01-15
2023-01-16
2023-01-17
1
1
3
Negative
Neutral
Positive
Sentiment
3.0
2.5
2.0
1.5
1.0
0.5
0.0