In [3]:
```
import pandas as pd
import numpy as np
```

In [4]:
```
df = pd.read_csv(r"D:\Sem 5\CA Text Analysis\Anime Reviews.csv", encoding='ISO-8859-1'
df
```

Out[4]:

	S.no	Title	Date	User	Tag	text
0	1	Kage no Jitsuryokusha ni Naritakute!	Feb 8, 2023	Ishinashi40	Recommended	Don't drop this anime after the first episodes...
1	2	Maou Gakuin no Futekigousha: Shijou Saikyou no...	Feb 8, 2023	Crimeful	NaN	Season 1 was excellent and I loved it. However...
2	3	Pumpkin Scissors	Feb 8, 2023	GenghisAres	NaN	This is a pretty decent show that is ultimatel...
3	4	Kage no Jitsuryokusha ni Naritakute!	Feb 8, 2023	The_Alpha_King	Recommended	I will say I can see how a lot of people can n...
4	5	Tian Guan Ci Fu	Feb 8, 2023	Himei25	NaN	i just watched it this year, it's a really imp...
...	...	...	...	...	...	...
31989	31990	Peach Boy Riverside	Sep 16, 2021	TotalCadenza	NaN	Like what everyone else says, this decently-wr...
31990	31991	Black Clover	Sep 16, 2021	Dransyo	Recommended	Black Clover is an enigma, it get's dunked on ...
31991	31992	Gakuen Babysitters	Sep 16, 2021	giraffenanime	Recommended	I needed a feel good anime... and ohhh boy, di...
31992	31993	Kidou Senshi Gundam Thunderbolt: December Sky	Sep 16, 2021	ulthtwac	Recommended	Gundam Thunderbolt is the most enjoyable time ...
31993	31994	Akudama Drive	Sep 16, 2021	UltraReviewShow	NaN	At first it was a dumb action anime with a sci...

31994 rows × 6 columns

In [5]:
```
df.head()
```

Out[5]:

	S.no	Title	Date	User	Tag	text
0	1	Kage no Jitsuryokusha ni Naritakute!	Feb 8, 2023	Ishinashi40	Recommended	Don't drop this anime after the first episodes...
1	2	Maou Gakuin no Futekigousha: Shijou Saikyou no...	Feb 8, 2023	Crimeful	NaN	Season 1 was excellent and I loved it. However...
2	3	Pumpkin Scissors	Feb 8, 2023	GenghisAres	NaN	This is a pretty decent show that is ultimatel...
3	4	Kage no Jitsuryokusha ni Naritakute!	Feb 8, 2023	The_Alpha_King	Recommended	I will say I can see how a lot of people can n...
4	5	Tian Guan Ci Fu	Feb 8, 2023	Himei25	NaN	i just watched it this year, it's a really imp...

In [6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 31994 entries, 0 to 31993
Data columns (total 6 columns):
 #   Column  Non-Null Count  Dtype 
---  ------  --------------  ----- 
 0   S.no    31994 non-null  int64 
 1   Title   31994 non-null  object
 2   Date    31994 non-null  object
 3   User    31994 non-null  object
 4   Tag     19872 non-null  object
 5   text    31994 non-null  object
dtypes: int64(1), object(5)
memory usage: 1.5+ MB
```

In [7]:
```
df.isnull().sum()
```

Out[7]:
```
S.no         0
Title        0
Date         0
User         0
Tag      12122
text         0
dtype: int64
```

In [8]:
```
df.describe()
```

Out[8]:

	S.no
count	31994.000000
mean	15997.500000
std	9236.016593
min	1.000000
25%	7999.250000
50%	15997.500000
75%	23995.750000
max	31994.000000

In [9]:
```
# Drop rows where 'Tag' is NaN
df_cleaned = df.dropna(subset=['Tag'])
```

In [10]:
```
# Check the result
df_cleaned.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 19872 entries, 0 to 31992
Data columns (total 6 columns):
 #   Column  Non-Null Count  Dtype 
---  ------  --------------  ----- 
 0   S.no    19872 non-null  int64 
 1   Title   19872 non-null  object
 2   Date    19872 non-null  object
 3   User    19872 non-null  object
 4   Tag     19872 non-null  object
 5   text    19872 non-null  object
dtypes: int64(1), object(5)
memory usage: 1.1+ MB
```

In [11]:
```
df.isnull().sum()
```

Out[11]:
```
S.no         0
Title        0
Date         0
User         0
Tag      12122
text         0
dtype: int64
```

In [12]:
```
df
```

Out[12]:

	S.no	Title	Date	User	Tag	text
0	1	Kage no Jitsuryokusha ni Naritakute!	Feb 8, 2023	Ishinashi40	Recommended	Don't drop this anime after the first episodes...
1	2	Maou Gakuin no Futekigousha: Shijou Saikyou no...	Feb 8, 2023	Crimeful	NaN	Season 1 was excellent and I loved it. However...
2	3	Pumpkin Scissors	Feb 8, 2023	GenghisAres	NaN	This is a pretty decent show that is ultimatel...
3	4	Kage no Jitsuryokusha ni Naritakute!	Feb 8, 2023	The_Alpha_King	Recommended	I will say I can see how a lot of people can n...
4	5	Tian Guan Ci Fu	Feb 8, 2023	Himei25	NaN	i just watched it this year, it's a really imp...
...	...	...	...	...	...	...
31989	31990	Peach Boy Riverside	Sep 16, 2021	TotalCadenza	NaN	Like what everyone else says, this decently-wr...
31990	31991	Black Clover	Sep 16, 2021	Dransyo	Recommended	Black Clover is an enigma, it get's dunked on ...
31991	31992	Gakuen Babysitters	Sep 16, 2021	giraffenanime	Recommended	I needed a feel good anime... and ohhh boy, di...
31992	31993	Kidou Senshi Gundam Thunderbolt: December Sky	Sep 16, 2021	ulthtwac	Recommended	Gundam Thunderbolt is the most enjoyable time ...
31993	31994	Akudama Drive	Sep 16, 2021	UltraReviewShow	NaN	At first it was a dumb action anime with a sci...

31994 rows × 6 columns

In [13]:
```
# Check unique values in 'Tag' column
df['Tag'].unique()
```

Out[13]:
```
array(['Recommended', nan], dtype=object)
```

In [14]:
```
# Replace NaN values in 'Tag' column with 'No Tag'
df['Tag'].fillna('No Tag', inplace=True)

# Check the result
df.head()
```

Out[14]:

	S.no	Title	Date	User	Tag	text
0	1	Kage no Jitsuryokusha ni Naritakute!	Feb 8, 2023	Ishinashi40	Recommended	Don't drop this anime after the first episodes...
1	2	Maou Gakuin no Futekigousha: Shijou Saikyou no...	Feb 8, 2023	Crimeful	No Tag	Season 1 was excellent and I loved it. However...
2	3	Pumpkin Scissors	Feb 8, 2023	GenghisAres	No Tag	This is a pretty decent show that is ultimatel...
3	4	Kage no Jitsuryokusha ni Naritakute!	Feb 8, 2023	The_Alpha_King	Recommended	I will say I can see how a lot of people can n...
4	5	Tian Guan Ci Fu	Feb 8, 2023	Himei25	No Tag	i just watched it this year, it's a really imp...

In [15]:
```
!pip install textblob
```

```
Requirement already satisfied: textblob in c:\users\prince kumar\.ipython\anaconda\li
b\site-packages (0.18.0.post0)
Requirement already satisfied: nltk>=3.8 in c:\users\prince kumar\.ipython\anaconda\l
ib\site-packages (from textblob) (3.9.1)
Requirement already satisfied: joblib in c:\users\prince kumar\.ipython\anaconda\lib
\site-packages (from nltk>=3.8->textblob) (1.1.1)
Requirement already satisfied: regex>=2021.8.3 in c:\users\prince kumar\.ipython\anac
onda\lib\site-packages (from nltk>=3.8->textblob) (2022.7.9)
Requirement already satisfied: tqdm in c:\users\prince kumar\.ipython\anaconda\lib\si
te-packages (from nltk>=3.8->textblob) (4.64.1)
Requirement already satisfied: click in c:\users\prince kumar\.ipython\anaconda\lib\s
ite-packages (from nltk>=3.8->textblob) (8.0.4)
Requirement already satisfied: colorama in c:\users\prince kumar\.ipython\anaconda\li
b\site-packages (from click->nltk>=3.8->textblob) (0.4.6)
```

In [16]:
```
import nltk
nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to C:\Users\PRINCE
[nltk_data]     KUMAR\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```

Out[16]:
```
True
```

In [17]:
```
from textblob import TextBlob

# Function to calculate sentiment polarity
def get_sentiment(text):
    analysis = TextBlob(text)
    return analysis.sentiment.polarity  # Returns a value between -1 (negative) and 1

# Apply the function to the 'text' column
df['Sentiment_Polarity'] = df['text'].apply(get_sentiment)

# Display the first few rows with sentiment polarity
df[['text', 'Sentiment_Polarity']].head()
```

Out[17]:

	text	Sentiment_Polarity
0	Don't drop this anime after the first episodes...	0.316204
1	Season 1 was excellent and I loved it. However...	0.198958
2	This is a pretty decent show that is ultimatel...	0.047853
3	I will say I can see how a lot of people can n...	0.149423
4	i just watched it this year, it's a really imp...	0.488636

In [18]:

```
def categorize_sentiment(polarity):
    if polarity > 0:
        return 'Positive'
    elif polarity < 0:
        return 'Negative'
    else:
        return 'Neutral'

# Apply the categorization function
df['Sentiment_Category'] = df['Sentiment_Polarity'].apply(categorize_sentiment)

# Check the distribution of sentiment categories
df['Sentiment_Category'].value_counts()
```

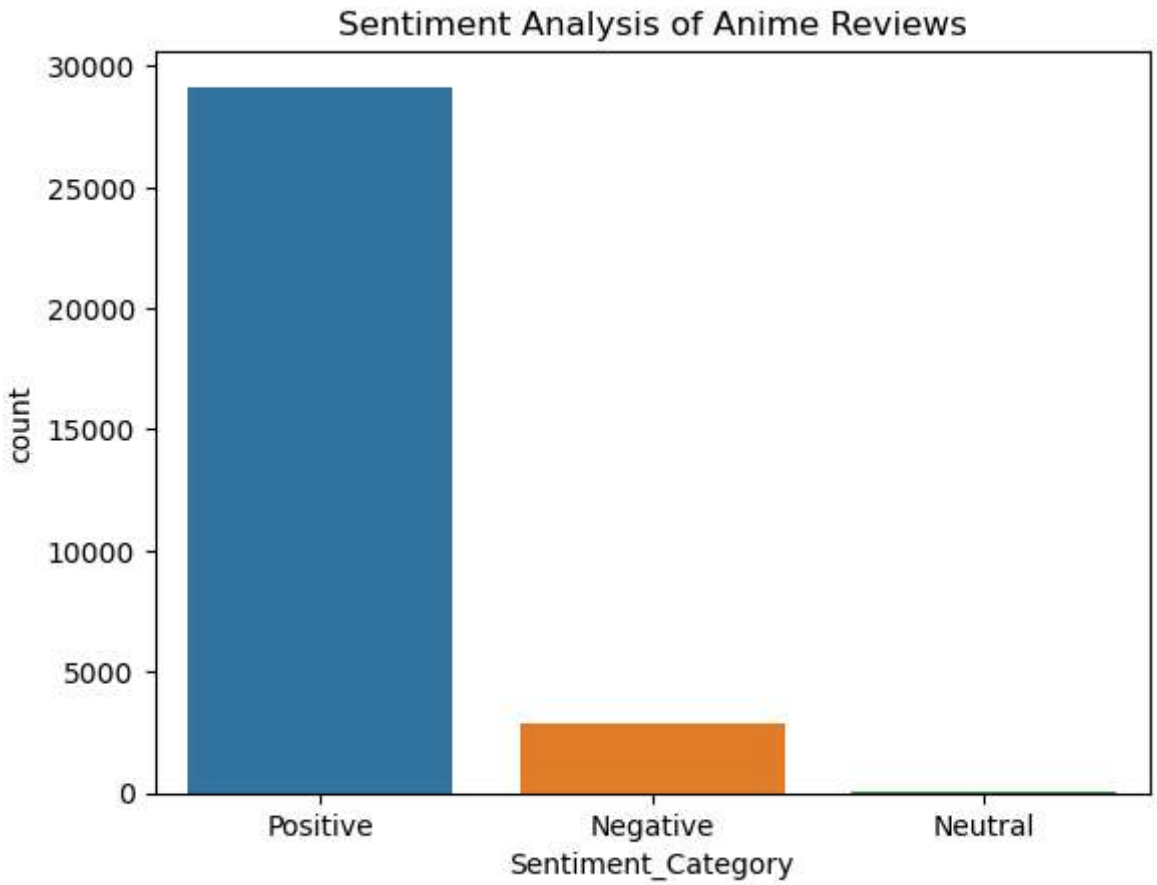
Out[18]:

```
Positive    29148
Negative     2835
Neutral        11
Name: Sentiment_Category, dtype: int64
```

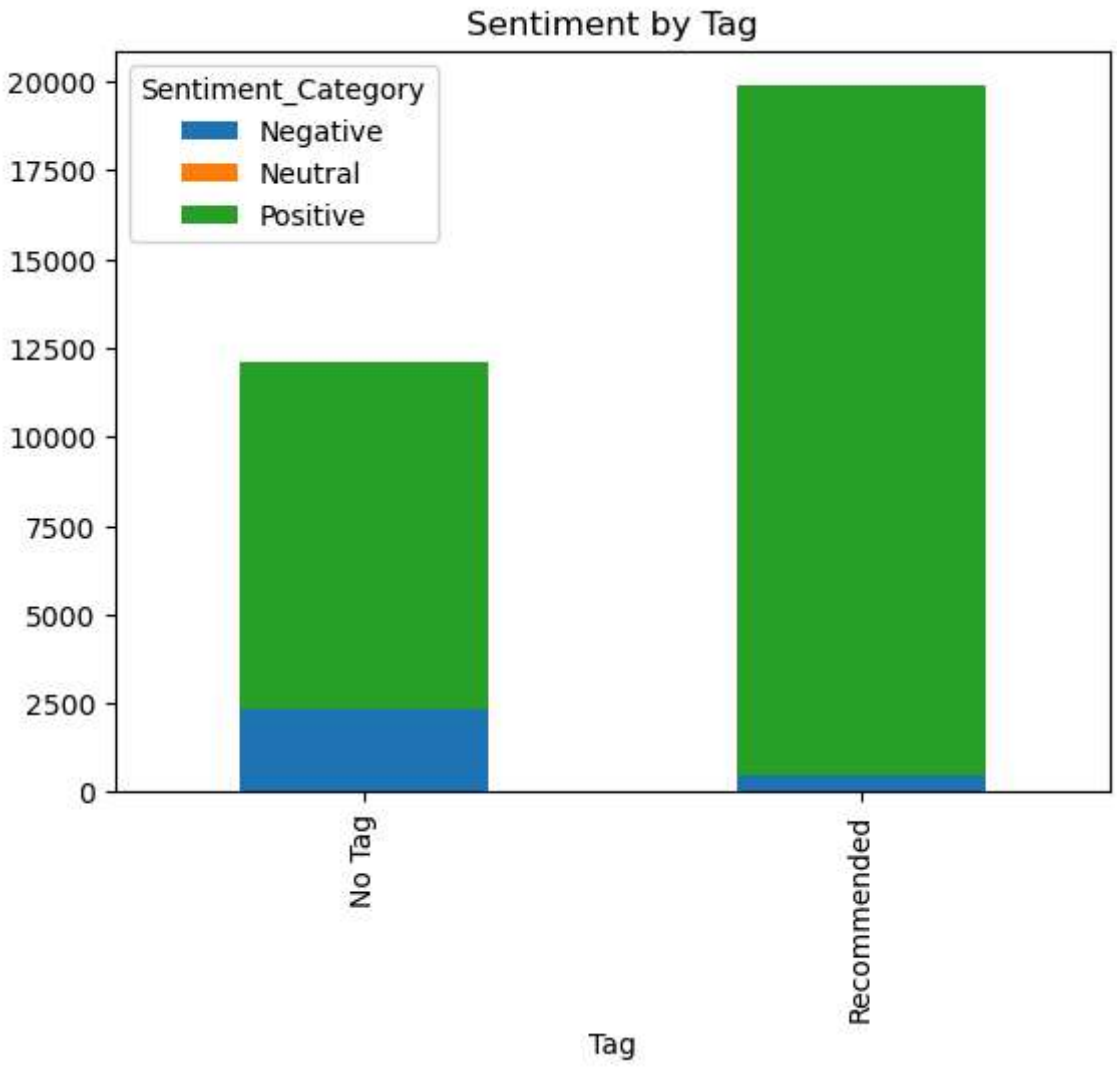
In [19]:

```
import matplotlib.pyplot as plt
import seaborn as sns

# Plot the sentiment category distribution
sns.countplot(x='Sentiment_Category', data=df)
plt.title('Sentiment Analysis of Anime Reviews')
plt.show()
```



In [20]:
```
# Sentiment distribution by 'Tag'
df.groupby('Tag')['Sentiment_Category'].value_counts().unstack().plot(kind='bar', stac
plt.title('Sentiment by Tag')
plt.show()
```


Sentiment by Tag
Sentiment_Category
Negative
Neutral
Positive
20000
17500
15000
12500
10000
7500
5000
2500
0
No Tag
Recommended
Tag

In [21]:
```
df
```

Out[21]:

	S.no	Title	Date	User	Tag	text	Sentiment_Polarity	S
0	1	Kage no Jitsuryokusha ni Naritakute!	Feb 8, 2023	Ishinashi40	Recommended	Don't drop this anime after the first episodes...	0.316204	
1	2	Maou Gakuin no Futekigousha: Shijou Saikyou no...	Feb 8, 2023	Crimeful	No Tag	Season 1 was excellent and I loved it. However...	0.198958	
2	3	Pumpkin Scissors	Feb 8, 2023	GenghisAres	No Tag	This is a pretty decent show that is ultimatel...	0.047853	
3	4	Kage no Jitsuryokusha ni Naritakute!	Feb 8, 2023	The_Alpha_King	Recommended	I will say I can see how a lot of people can n...	0.149423	
4	5	Tian Guan Ci Fu	Feb 8, 2023	Himei25	No Tag	i just watched it this year, it's a really imp...	0.488636	
...	...	...	...	...	...	...	...	
31989	31990	Peach Boy Riverside	Sep 16, 2021	TotalCadenza	No Tag	Like what everyone else says, this decently-wr...	0.087698	
31990	31991	Black Clover	Sep 16, 2021	Dransyo	Recommended	Black Clover is an enigma, it get's dunked on ...	0.134350	
31991	31992	Gakuen Babysitters	Sep 16, 2021	giraffenanime	Recommended	I needed a feel good anime... and ohhh boy, di...	0.242782	
31992	31993	Kidou Senshi Gundam Thunderbolt: December Sky	Sep 16, 2021	ulthtwac	Recommended	Gundam Thunderbolt is the most enjoyable time ...	0.144464	
31993	31994	Akudama Drive	Sep 16,	UltraReviewShow	No Tag	At first it was a dumb	0.035417	

S.no	Title	Date	User	Tag	text	Sentiment_Polarity	S
		2021			action anime with a sci...		

21004 rows × 8 columns

In [22]:
```
# Group the dataset by anime title and calculate average sentiment polarity
anime_sentiment = df.groupby('Title')['Sentiment_Polarity'].mean().reset_index()

# Sort by sentiment polarity to find the top-performing anime
anime_sentiment.sort_values(by='Sentiment_Polarity', ascending=False, inplace=True)

# Display the top anime by average sentiment polarity
anime_sentiment.head(10)
```

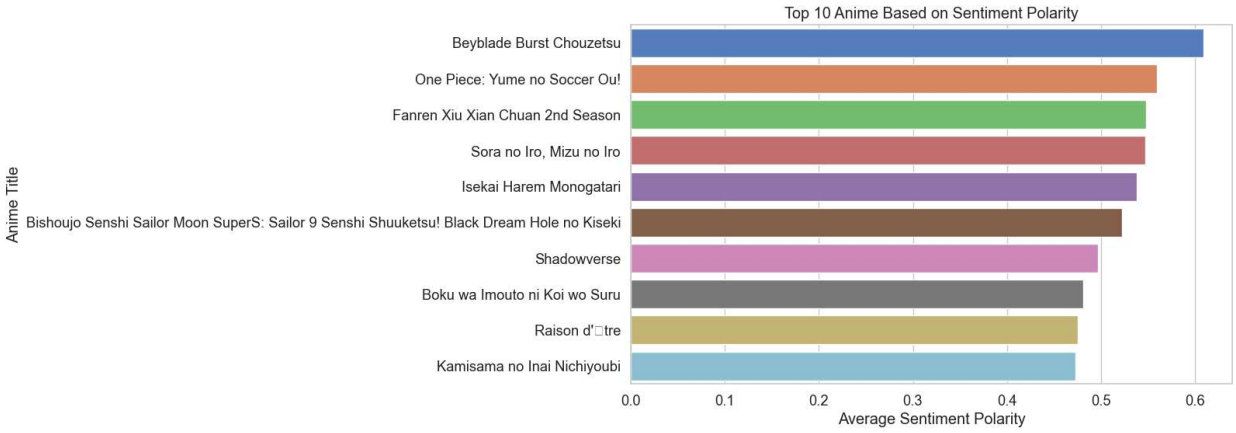
Out[22]:

	Title	Sentiment_Polarity
409	Beyblade Burst Chouzetsu	0.609524
3054	One Piece: Yume no Soccer Ou!	0.560000
1088	Fanren Xiu Xian Chuan 2nd Season	0.548214
3862	Sora no Iro, Mizu no Iro	0.547653
1800	Isekai Harem Monogatari	0.538596
436	Bishoujo Senshi Sailor Moon SuperS: Sailor 9 S...	0.522321
3643	Shadowverse	0.497222
511	Boku wa Imouto ni Koi wo Suru	0.481345
3364	Raison d'□tre	0.475694
2001	Kamisama no Inai Nichiyoubi	0.472727

In [26]:
```
import matplotlib.pyplot as plt
import seaborn as sns

# Plot the top 10 anime based on average sentiment polarity
plt.figure(figsize=(10, 6))
top_anime = anime_sentiment.head(10)
sns.barplot(x='Sentiment_Polarity', y='Title', data=top_anime)
plt.title('Top 10 Anime Based on Sentiment Polarity')
plt.xlabel('Average Sentiment Polarity')
plt.ylabel('Anime Title')
plt.show()
```

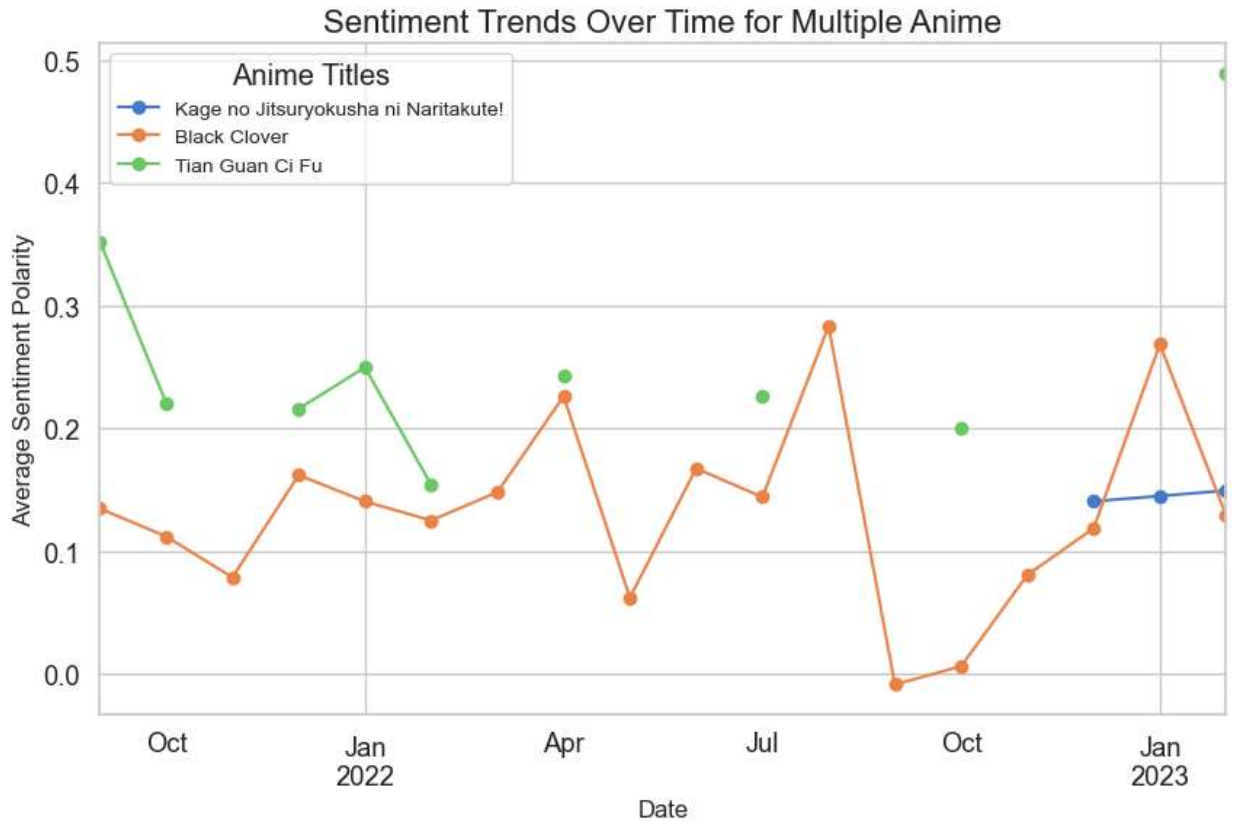
```
C:\Users\PRINCE KUMAR\.ipython\Anaconda\lib\site-packages\IPython\core\pylabtools.py:
152: UserWarning: Glyph 144 (\x90) missing from current font.
  fig.canvas.print_figure(bytes_io, **kw)
```



In [33]:
```
# Plot sentiment trends for multiple anime
anime_list = ['Kage no Jitsuryokusha ni Naritakute!', 'Black Clover', 'Tian Guan Ci Fu

# Plot each anime on the same graph
plt.figure(figsize=(10,6))
for anime in anime_list:
    anime_sentiment_over_time[anime].plot(marker='o', linestyle='-', label=anime)

# Add labels and legend
plt.title('Sentiment Trends Over Time for Multiple Anime', fontsize=16)
plt.ylabel('Average Sentiment Polarity', fontsize=12)
plt.xlabel('Date', fontsize=12)
plt.legend(title='Anime Titles', fontsize=10)
plt.grid(True)
plt.show()
```



In [36]:
```
df.shape
```

Out[36]: (31994, 8)

In [ ]: