```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from wordcloud import WordCloud
```

```
cols=['ID', 'Topic', 'Sentiment', 'Text']
train = pd.read_csv(r"twitter_training.csv",names=cols)
```

```
train.head()
```

	ID	Topic	Sentiment	Text
0	2401	Borderlands	Positive	im getting on borderlands and i will murder yo...
1	2401	Borderlands	Positive	I am coming to the borders and I will kill you...
2	2401	Borderlands	Positive	im getting on borderlands and i will kill you ...
3	2401	Borderlands	Positive	im coming on borderlands and i will murder you...
4	2401	Borderlands	Positive	im getting on borderlands 2 and i will murder ...

Next steps: Generate code with `train` View recommended plots New interactive sheet

```
train.shape
```

```
(74682, 4)
```

```
train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 74682 entries, 0 to 74681
Data columns (total 4 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   ID         74682 non-null  int64
 1   Topic      74682 non-null  object
 2   Sentiment  74682 non-null  object
 3   Text       73996 non-null  object
dtypes: int64(1), object(3)
memory usage: 2.3+ MB
```

```
train.describe(include=object)
```

	Topic	Sentiment	Text
count	74682	74682	73996
unique	32	4	69491
top	TomClancysRainbowSix	Negative	At the same time, despite the fact that there ...
freq	2400	22542	172

```
train['Sentiment'].unique()
```

```
array(['Positive', 'Neutral', 'Negative', 'Irrelevant'], dtype=object)
```

```
train.isnull().sum()
```

	0
ID	0
Topic	0
Sentiment	0
Text	686

dtype: int64

```
train.dropna(inplace=True)
```

```
train.isnull().sum()
```

	0
ID	0
Topic	0
Sentiment	0
Text	0

dtype: int64

```
train.duplicated().sum()
```

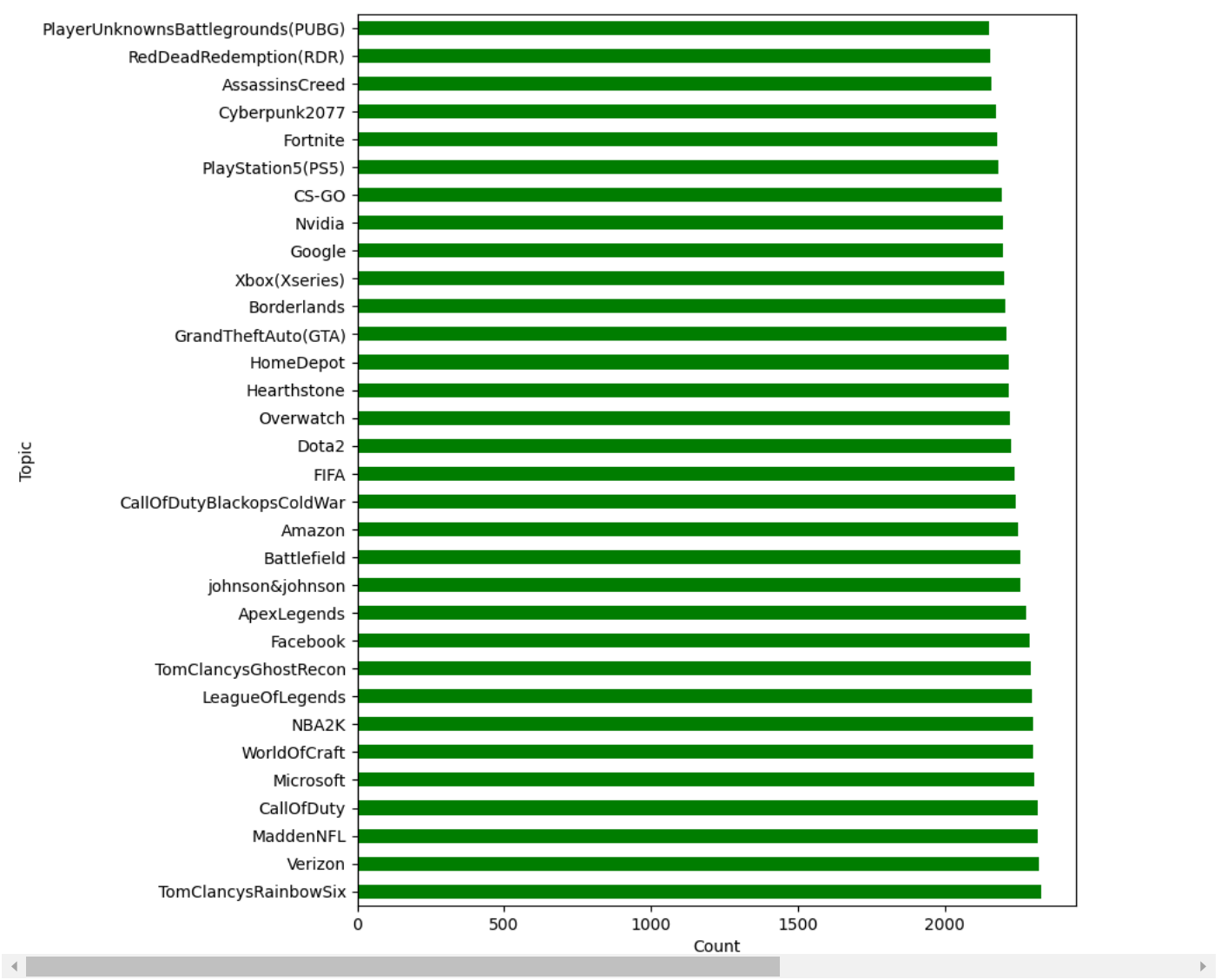
```
2340
```

```
train.drop_duplicates(inplace=True)
```

```
train.duplicated().sum()
```

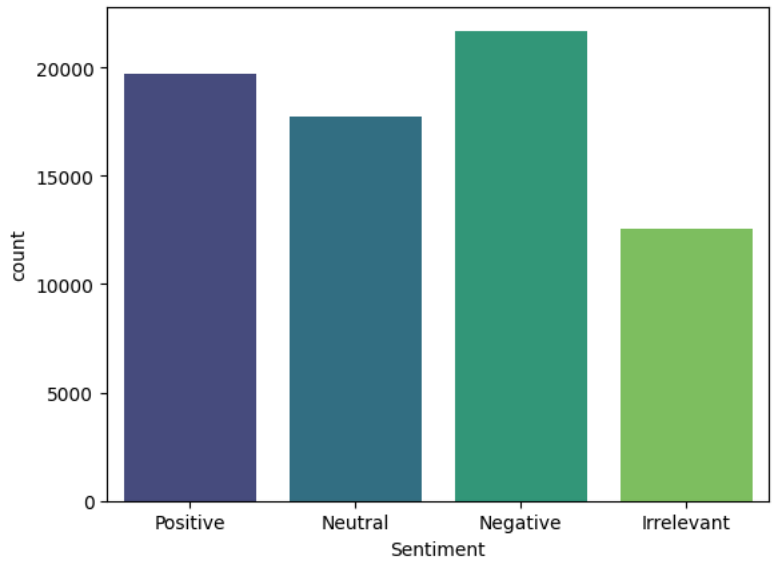
```
0
```

```
plt.figure(figsize=(8,10))
train['Topic'].value_counts().plot(kind='barh',color='g')
plt.xlabel("Count")
plt.show()
```

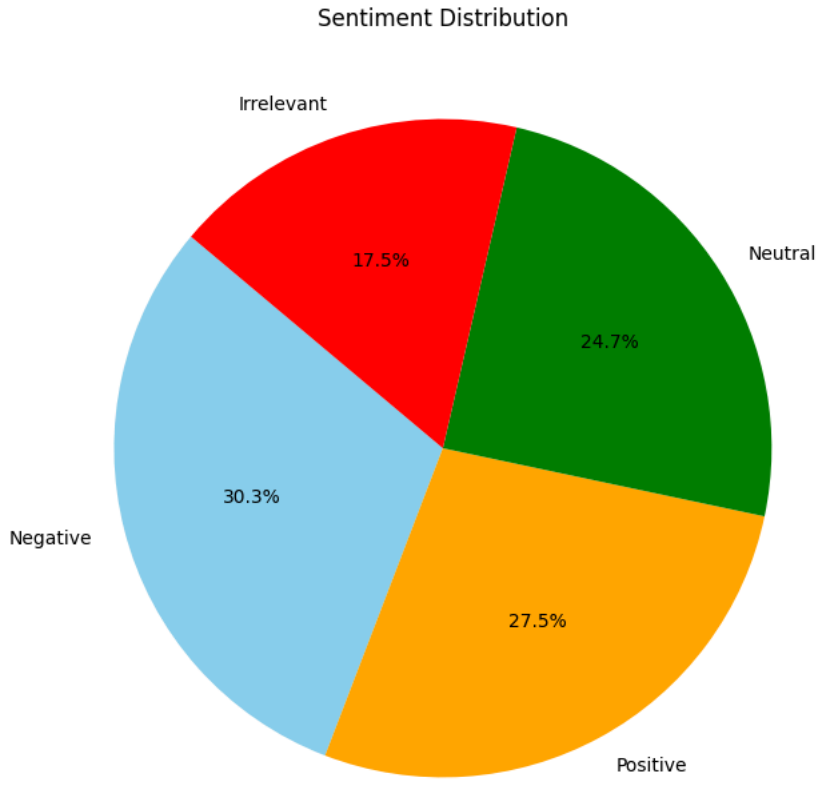



```
sns.countplot(x = 'Sentiment',data=train,palette='viridis')
plt.show()
```

```
<ipython-input-15-0f5f2096c1d5>:1: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend

  sns.countplot(x = 'Sentiment',data=train,palette='viridis')
```




```
sentiment_counts = train['Sentiment'].value_counts()

plt.figure(figsize=(8, 8))
plt.pie(sentiment_counts, labels=sentiment_counts.index, autopct="%1.1f%%", startangle=140, colors=['skyblue', 'orange', 'green', 'red', 'pur

plt.title('Sentiment Distribution')

plt.show()
```



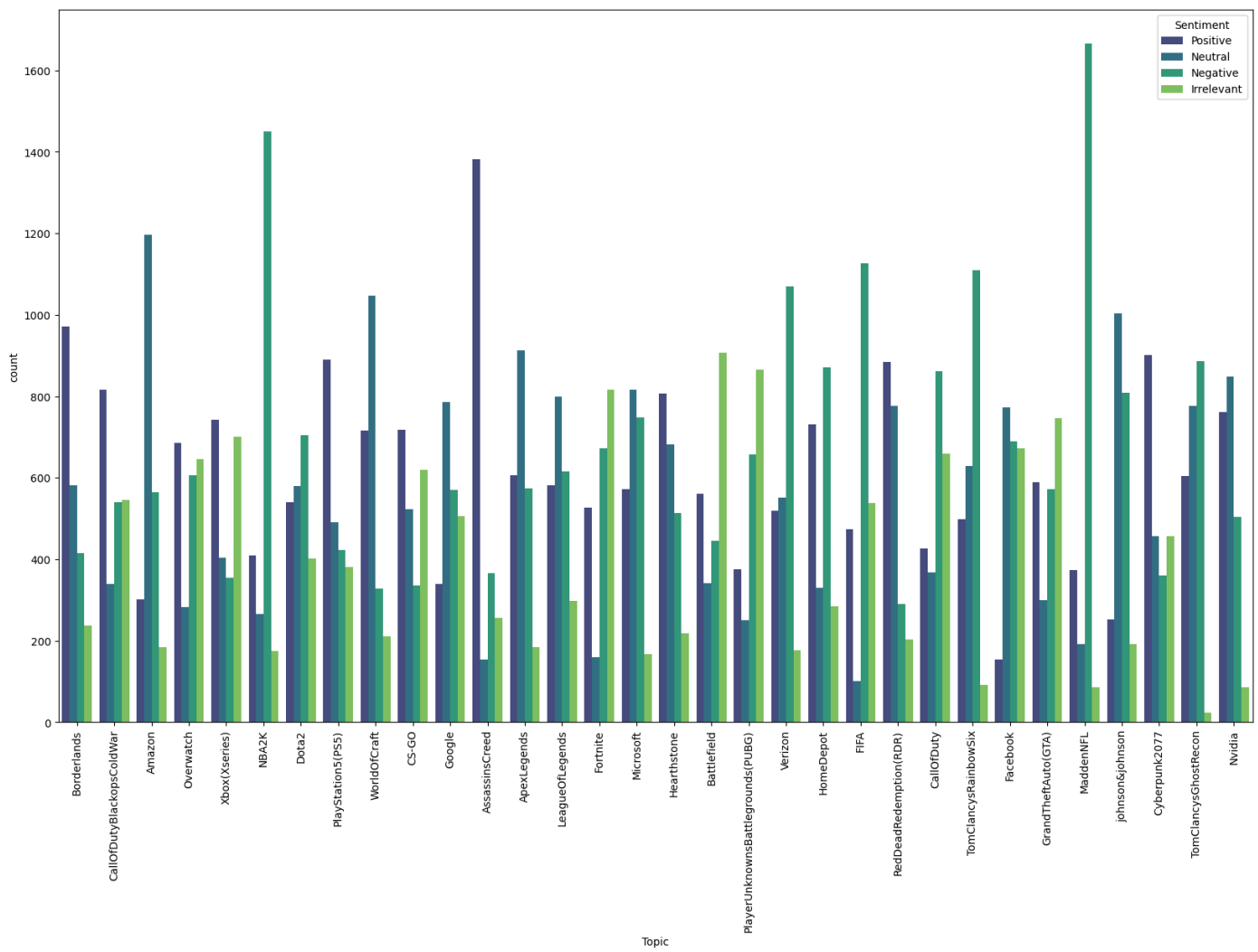

```
train
```

	ID	Topic	Sentiment	Text
0	2401	Borderlands	Positive	im getting on borderlands and i will murder yo...
1	2401	Borderlands	Positive	I am coming to the borders and I will kill you...
2	2401	Borderlands	Positive	im getting on borderlands and i will kill you ...
3	2401	Borderlands	Positive	im coming on borderlands and i will murder you...
4	2401	Borderlands	Positive	im getting on borderlands 2 and i will murder ...
...	...	...	...	...
74677	9200	Nvidia	Positive	Just realized that the Windows partition of my...
74678	9200	Nvidia	Positive	Just realized that my Mac window partition is ...
74679	9200	Nvidia	Positive	Just realized the windows partition of my Mac ...
74680	9200	Nvidia	Positive	Just realized between the windows partition of...
74681	9200	Nvidia	Positive	Just like the windows partition of my Mac is l...

71656 rows × 4 columns

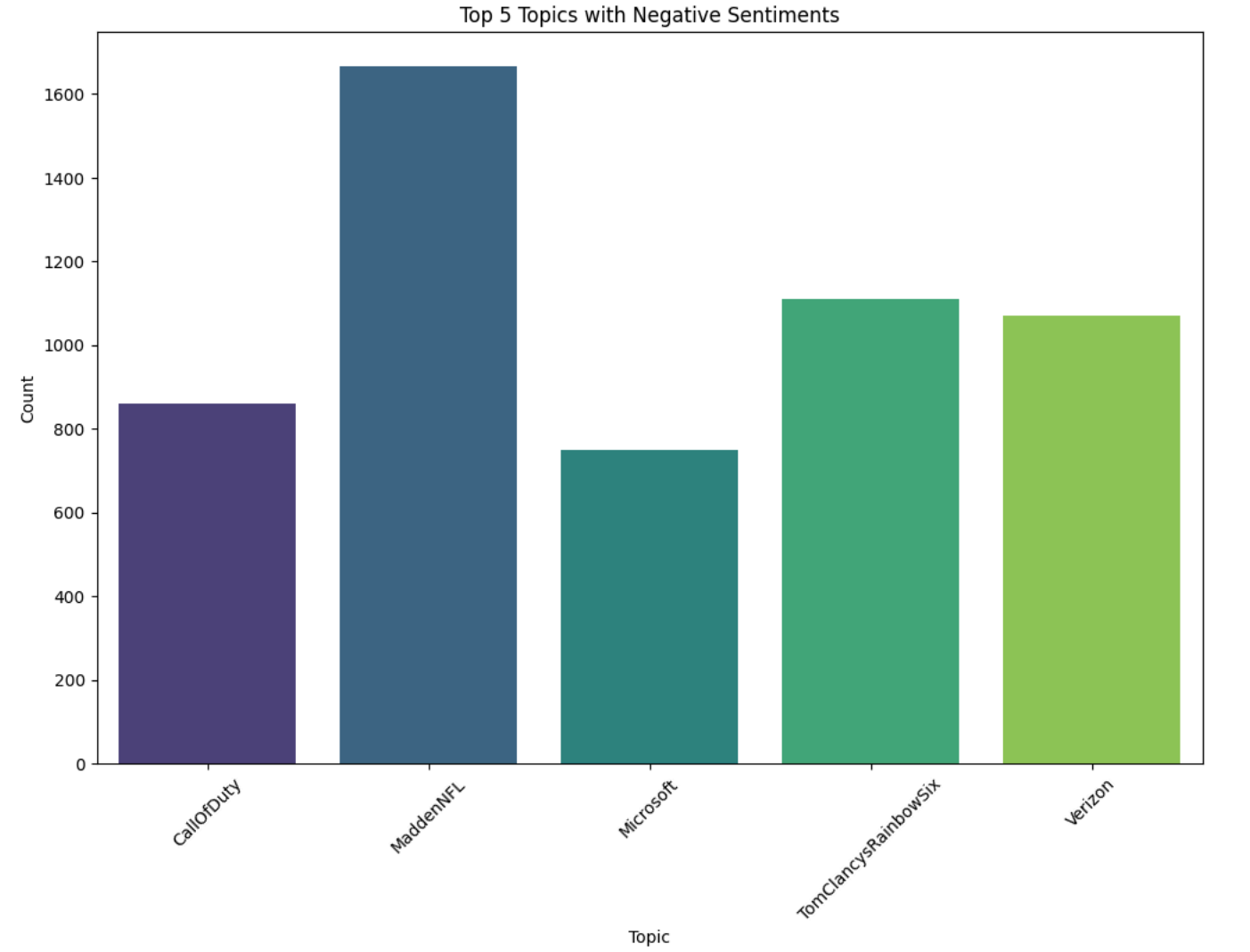
Next steps: Generate code with `train` View recommended plots New interactive sheet

```
plt.figure(figsize=(20,12))
sns.countplot(x='Topic',data=train,palette='viridis',hue='Sentiment')
plt.xticks(rotation=90)
plt.show()
```



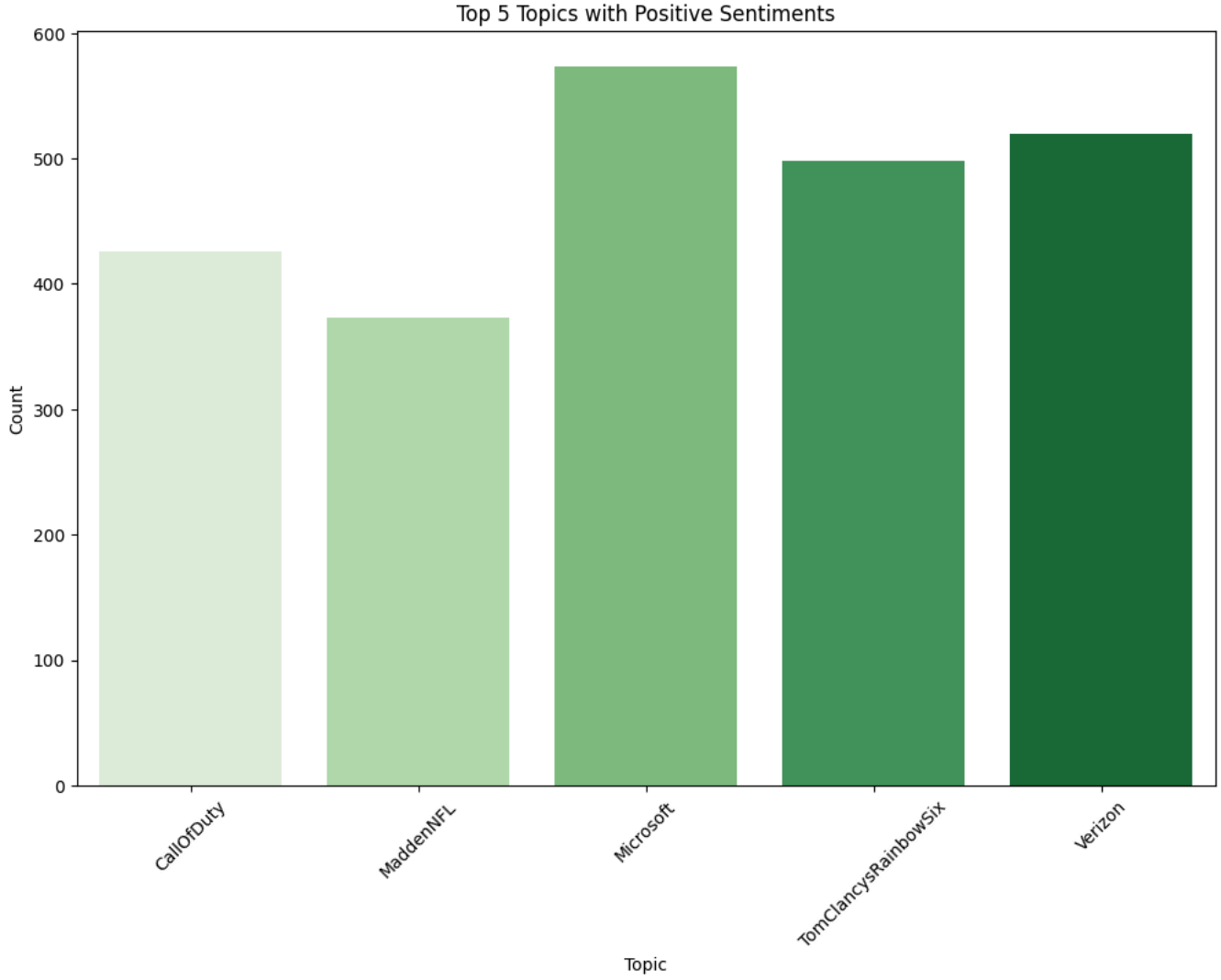
```
topic_wise_sentiment = train.groupby(["Topic", "Sentiment"]).size().reset_index(name='Count')

# Step 2: Select Top 5 Topics
topic_counts = train['Topic'].value_counts().nlargest(5).index
top_topics_sentiment = topic_wise_sentiment[topic_wise_sentiment['Topic'].isin(topic_counts)]


plt.figure(figsize=(12, 8))
sns.barplot(data=top_topics_sentiment[top_topics_sentiment['Sentiment'] == 'Negative'], x='Topic', y='Count', palette='viridis')
plt.title('Top 5 Topics with Negative Sentiments')
plt.xlabel('Topic')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.show()
```

```
<ipython-input-20-7127521535d3>:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend

  sns.barplot(data=top_topics_sentiment[top_topics_sentiment['Sentiment'] == 'Negative'], x='Topic', y='Count', palette='viridis')
```



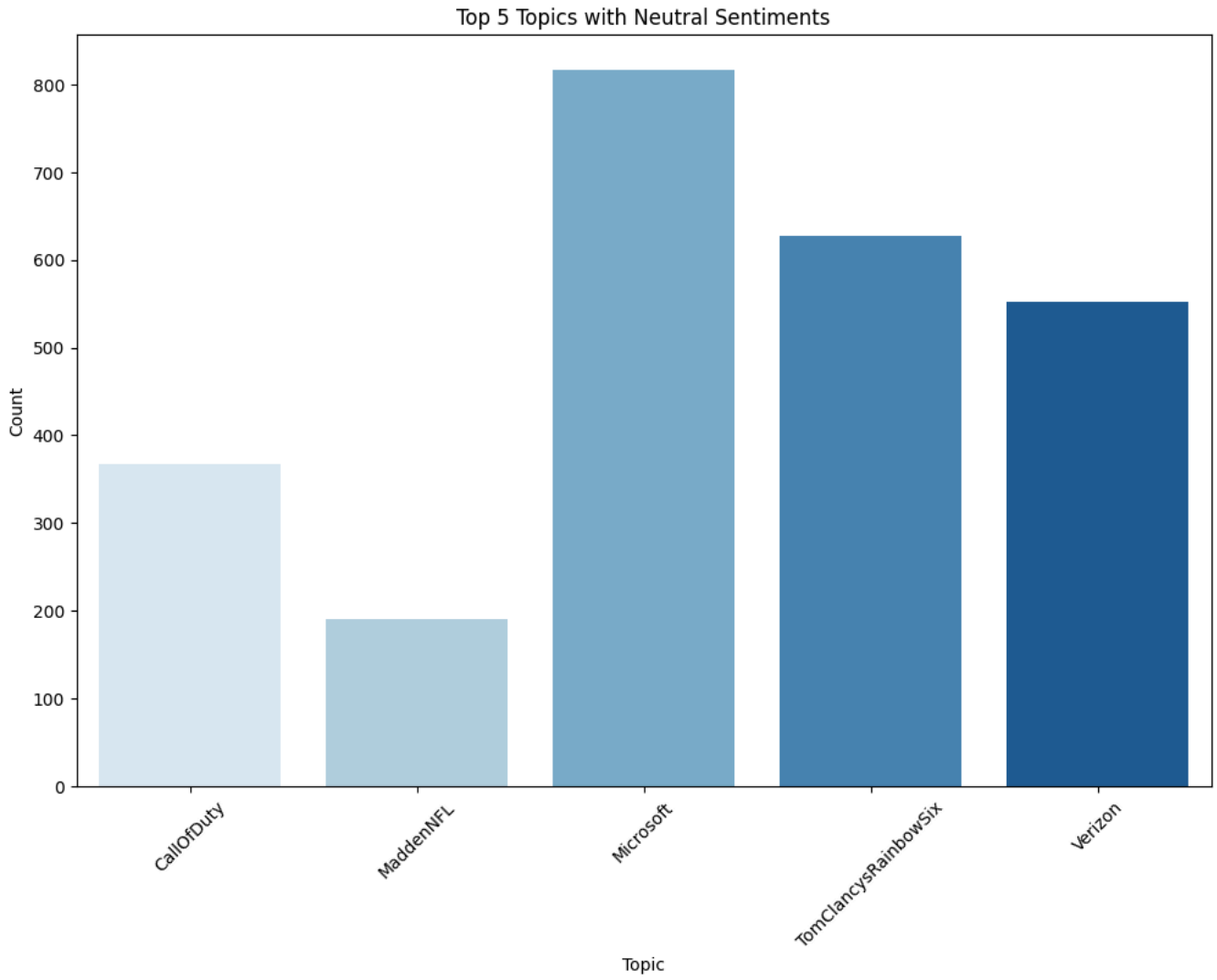

```
plt.figure(figsize=(12, 8))
sns.barplot(data=top_topics_sentiment[top_topics_sentiment['Sentiment'] == 'Positive'], x='Topic', y='Count', palette='Greens')
plt.title('Top 5 Topics with Positive Sentiments')
plt.xlabel('Topic')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.show()
```

```
<ipython-input-21-fa26222f4ed6>:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend

  sns.barplot(data=top_topics_sentiment[top_topics_sentiment['Sentiment'] == 'Positive'], x='Topic', y='Count', palette='Greens')
```



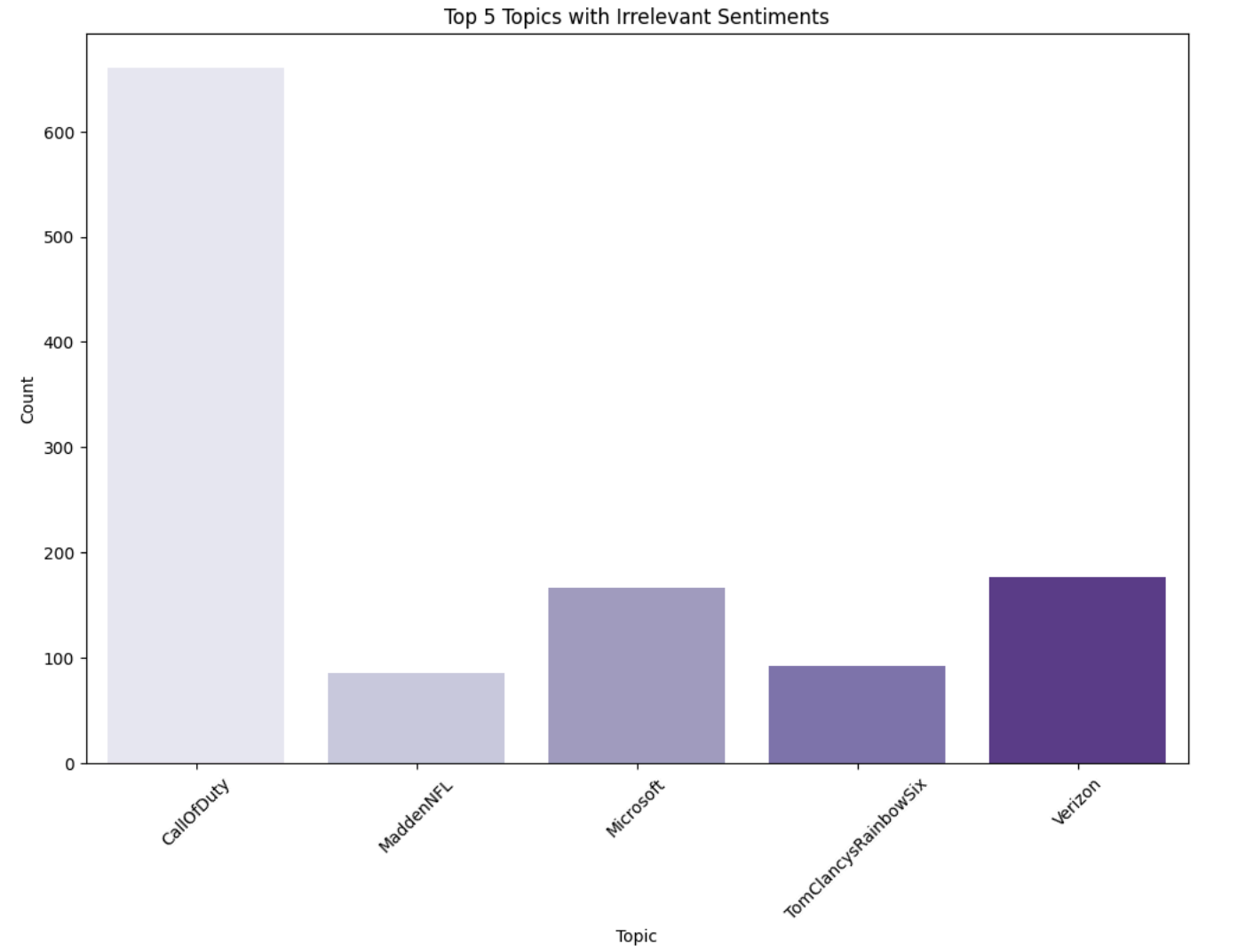

```
plt.figure(figsize=(12, 8))
sns.barplot(data=top_topics_sentiment[top_topics_sentiment['Sentiment'] == 'Neutral'], x='Topic', y='Count', palette='Blues')
plt.title('Top 5 Topics with Neutral Sentiments')
plt.xlabel('Topic')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.show()
```

```
<ipython-input-22-af01e1bcdbaa>:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend

  sns.barplot(data=top_topics_sentiment[top_topics_sentiment['Sentiment'] == 'Neutral'], x='Topic', y='Count', palette='Blues')
```

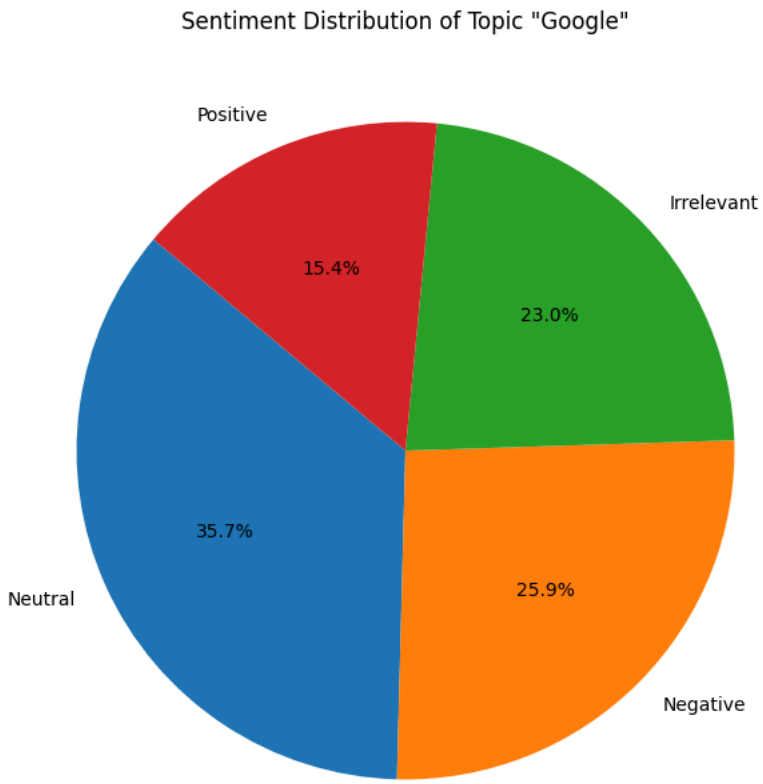



```
plt.figure(figsize=(12, 8))
sns.barplot(data=top_topics_sentiment[top_topics_sentiment['Sentiment'] == 'Irrelevant'], x='Topic', y='Count', palette='Purples')
plt.title('Top 5 Topics with Irrelevant Sentiments')
plt.xlabel('Topic')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.show()
```

```
<ipython-input-23-7662d01b7d35>:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend

  sns.barplot(data=top_topics_sentiment[top_topics_sentiment['Sentiment'] == 'Irrelevant'], x='Topic', y='Count', palette='Purples')
```




```
google_data = train[train['Topic'] == 'Google']
sentiment_counts = google_data['Sentiment'].value_counts()
plt.figure(figsize=(8, 8))
plt.pie(sentiment_counts, labels=sentiment_counts.index, autopct='%1.1f%%', startangle=140)
plt.title('Sentiment Distribution of Topic "Google"')
plt.show()
```




```
train['msg_len'] = train['Text'].apply(len)
train
```

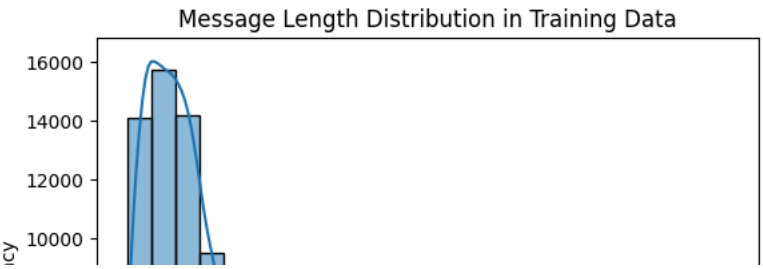
	ID	Topic	Sentiment	Text	msg_len
0	2401	Borderlands	Positive	im getting on borderlands and i will murder yo...	53
1	2401	Borderlands	Positive	I am coming to the borders and I will kill you...	51
2	2401	Borderlands	Positive	im getting on borderlands and i will kill you ...	50
3	2401	Borderlands	Positive	im coming on borderlands and i will murder you...	51
4	2401	Borderlands	Positive	im getting on borderlands 2 and i will murder ...	57
...	...	...	...	...	...
74677	9200	Nvidia	Positive	Just realized that the Windows partition of my...	128
74678	9200	Nvidia	Positive	Just realized that my Mac window partition is ...	117
74679	9200	Nvidia	Positive	Just realized the windows partition of my Mac ...	125
74680	9200	Nvidia	Positive	Just realized between the windows partition of...	159
74681	9200	Nvidia	Positive	Just like the windows partition of my Mac is l...	119

71656 rows × 5 columns

Next steps: Generate code with `train` View recommended plots New interactive sheet

```
sns.histplot(train['msg_len'], bins=25,kde=True)
plt.title('Message Length Distribution in Training Data')
plt.ylabel('Frequency')
plt.xlabel('Message Length')
plt.show()
```

[+]


Message Length Distribution in Training Data
16000
14000
12000
10000