

```
# Importing necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings

# Ignoring warnings
warnings.filterwarnings('ignore')

# Ensuring plots are displayed inline
%matplotlib inline

tw_train = pd.read_csv("training.csv")
tw_valid = pd.read_csv("validation.csv")

column_name=['TweetID', 'Entity', 'Sentiment', 'Tweet_Content']
tw_train.columns=column_name
tw_valid.columns=column_name
# Combine 2 dataframes to 1 dataframe
tw=pd.concat([tw_train,tw_valid],ignore_index=False)
tw.head()
```

	TweetID	Entity	Sentiment	Tweet_Content
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...

Next steps:

Generate code with tw

☒ View recommended plots

```
tw.columns.tolist()

['TweetID', 'Entity', 'Sentiment', 'Tweet_Content']
```

```
tw.info()

<class 'pandas.core.frame.DataFrame'>
Index: 75680 entries, 0 to 998
Data columns (total 4 columns):
#   Column          Non-Null Count  Dtype
---  -
0   TweetID         75680 non-null  int64
1   Entity          75680 non-null  object
2   Sentiment       75680 non-null  object
3   Tweet_Content   74994 non-null  object
dtypes: int64(1), object(3)
memory usage: 2.9+ MB
```

```
tw.isnull().sum()

TweetID      0
Entity       0
Sentiment    0
Tweet_Content 686
dtype: int64
```

```
tw.duplicated().sum()

3216
```

```
tw.dropna(inplace=True)
tw.drop_duplicates(inplace=True)
print(tw.isnull().sum())
print("Duplicate Values:",tw.duplicated().sum())

TweetID      0
Entity       0
Sentiment    0
Tweet_Content 0
dtype: int64
Duplicate Values: 0
```

```
tw.drop(columns=['TweetID','Tweet_Content'],inplace=True)
tw.head()
```

	Entity	Sentiment	
0	Borderlands	Positive	
1	Borderlands	Positive	
2	Borderlands	Positive	
3	Borderlands	Positive	
4	Borderlands	Positive	

Next steps:

Generate code with tw

☒ View recommended plots

```
tw.info()

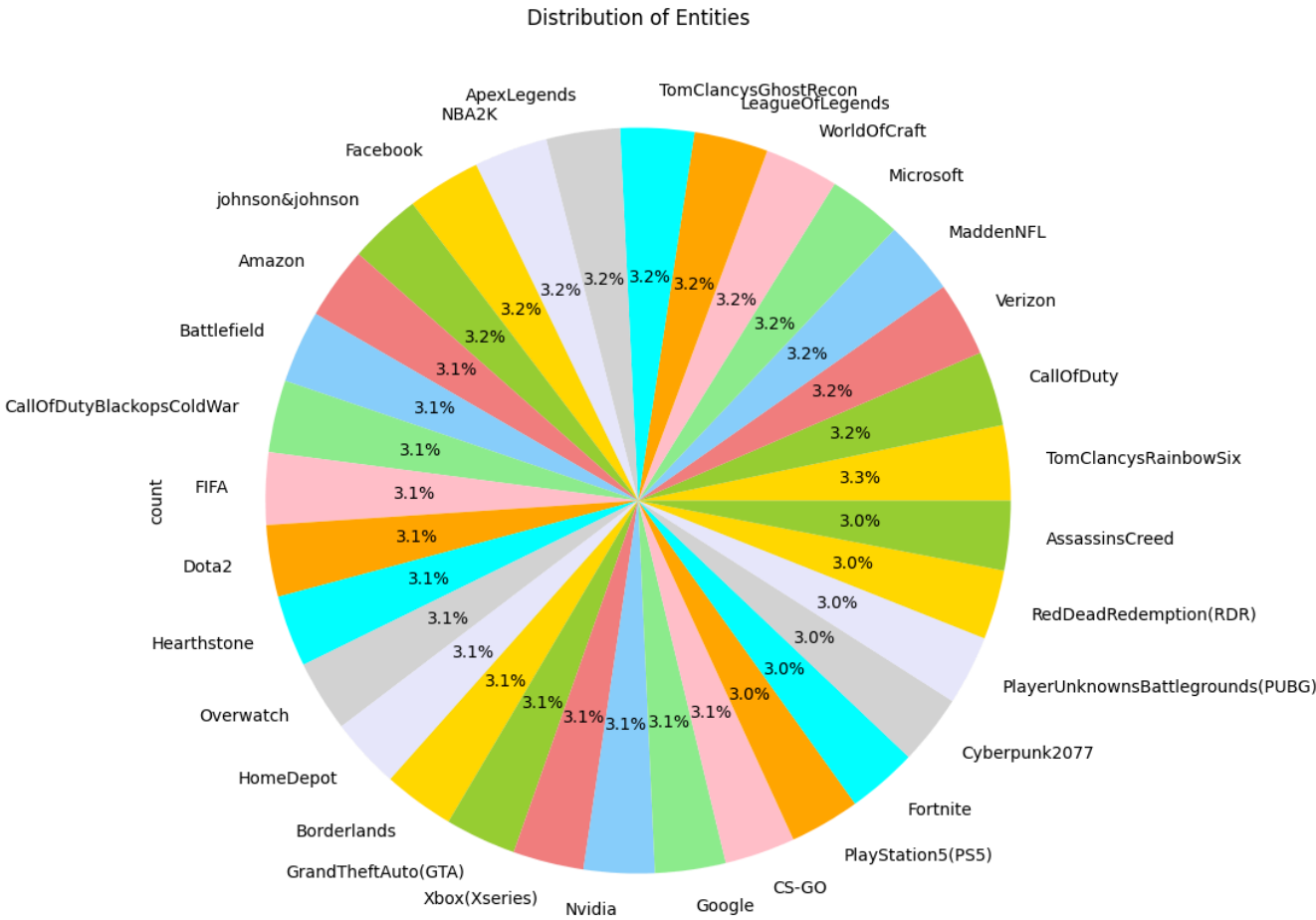
<class 'pandas.core.frame.DataFrame'>
Index: 72138 entries, 0 to 995
Data columns (total 2 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Entity      72138 non-null  object
1   Sentiment   72138 non-null  object
dtypes: object(2)
memory usage: 1.7+ MB

import matplotlib.pyplot as plt

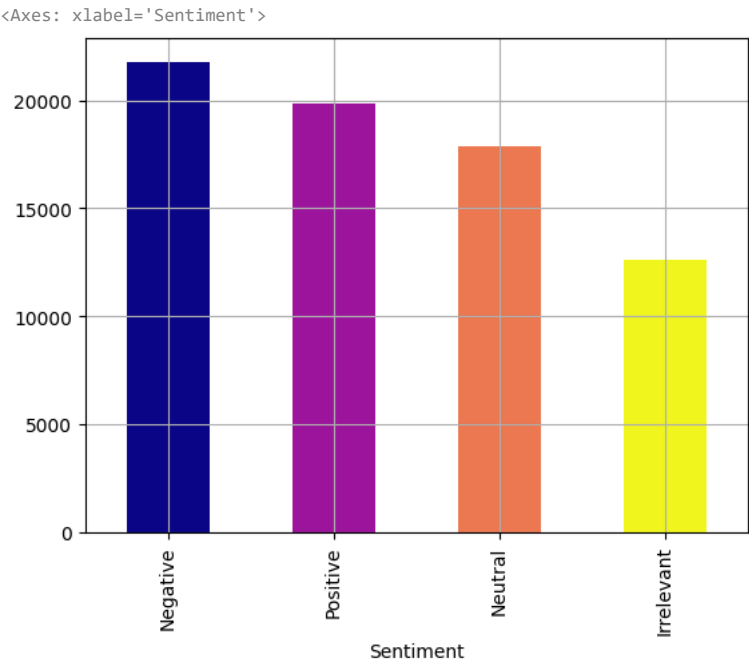
# Assuming you already have entity_content defined

# Define colors
colors = ['gold', 'yellowgreen', 'lightcoral', 'lightskyblue', 'lightgreen', 'pink', 'orange', 'cyan', 'lightgray', 'lavender']

# Plot the pie chart with specified colors
entity_content.plot(kind='pie', autopct='%1.1f%%', figsize=(10, 12), colors=colors)
plt.title('Distribution of Entities')
plt.show()
```



```
sentiment_content=tw['Sentiment'].value_counts()
color=plt.get_cmap('plasma')
colors = [color(i) for i in np.linspace(0, 1, len(sentiment_content))]
sentiment_content.plot(kind='bar',color=colors,grid=True)
```



```
reactions_entities = pd.crosstab(tw['Entity'],tw['Sentiment'])

colors = ['red', 'green', 'blue']
```

