Task4

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

df = pd.read_csv("C:\\Users\\hp\\Internship\\twitter_validation.csv")

df.head()

   3364   Facebook Irrelevant  \
0   352     Amazon    Neutral
1  8312  Microsoft   Negative
2  4371      CS-GO   Negative
3  4433     Google    Neutral
4  6273       FIFA   Negative

  I mentioned on Facebook that I was struggling for motivation to go
for a run the other day, which has been translated by Tom’s great
auntie as ‘Hayley can’t get out of bed’ and told to his grandma, who
now thinks I’m a lazy, terrible person 
0  BBC News - Amazon boss Jeff Bezos rejects clai...

1  @Microsoft Why do I pay for WORD when it funct...

2  CSGO matchmaking is so full of closet hacking,...

3  Now the President is slapping Americans in the...

4  Hi @EAHelp I’ve had Madeleine McCann in my cel...


df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 999 entries, 0 to 998
Data columns (total 4 columns):
 #   Column
Non-Null Count  Dtype
---  ------
--------------  -----
 0   3364
999 non-null    int64
 1   Facebook
999 non-null    object
```

```
 2   Irrelevant
999 non-null    object
 3   I mentioned on Facebook that I was struggling for motivation to
go for a run the other day, which has been translated by Tom’s great
auntie as ‘Hayley can’t get out of bed’ and told to his grandma, who
now thinks I’m a lazy, terrible person □  999 non-null    object
dtypes: int64(1), object(3)
memory usage: 31.3+ KB
```

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

```
3364
0
Facebook
0
Irrelevant
0
I mentioned on Facebook that I was struggling for motivation to go for
a run the other day, which has been translated by Tom’s great auntie
as ‘Hayley can’t get out of bed’ and told to his grandma, who now
thinks I’m a lazy, terrible person □    0
dtype: int64
```

```
df.columns
```

```
Index(['3364', 'Facebook', 'Irrelevant',
       'I mentioned on Facebook that I was struggling for motivation
to go for a run the other day, which has been translated by Tom’s
great auntie as ‘Hayley can’t get out of bed’ and told to his grandma,
who now thinks I’m a lazy, terrible person □'],
      dtype='object')
```

```
import re

def clean_text(text):
    text = re.sub(r'http\S+|www\S+|https\S+', '', str(text))
    text = re.sub(r'\@\w+|\#','', text)
    text = re.sub(r'[^A-Za-z\s]', '', text)
    text = text.lower()
    return text

df['clean_text'] = df['Facebook'].apply(clean_text)
print(df)
```

```
     3364             Facebook  Irrelevant  \
0     352               Amazon     Neutral
1    8312            Microsoft    Negative
2    4371                CS-GO    Negative
3    4433               Google     Neutral
4    6273                 FIFA    Negative
..    ...                  ...         ...
```

```
994  4891  GrandTheftAuto(GTA)  Irrelevant
995  4359                CS-GO  Irrelevant
996  2652           Borderlands    Positive
997  8069             Microsoft    Positive
998  6960      johnson&johnson     Neutral

    I mentioned on Facebook that I was struggling for motivation to go
for a run the other day, which has been translated by Tom’s great
auntie as ‘Hayley can’t get out of bed’ and told to his grandma, who
now thinks I’m a lazy, terrible person □  \
0    BBC News - Amazon boss Jeff Bezos rejects clai...

1    @Microsoft Why do I pay for WORD when it funct...

2    CSGO matchmaking is so full of closet hacking,...

3    Now the President is slapping Americans in the...

4    Hi @EAHelp I’ve had Madeleine McCann in my cel...

..                                                 ...

994   Toronto is the arts and culture capital of ...

995  tHIS IS ACTUALLY A GOOD MOVE TOT BRING MORE VI...

996  Today sucked so it’s time to drink wine n play...

997  Bought a fraction of Microsoft today. Small wins.

998  Johnson & Johnson to stop selling talc baby po...


             clean_text
0                amazon
1             microsoft
2                  csgo
3                google
4                  fifa
..                  ...
994  grandtheftautogta
995                csgo
996          borderlands
997            microsoft
998       johnsonjohnson

[999 rows x 5 columns]

def get_sentiment(text):
    return TextBlob(text).sentiment.polarity
```

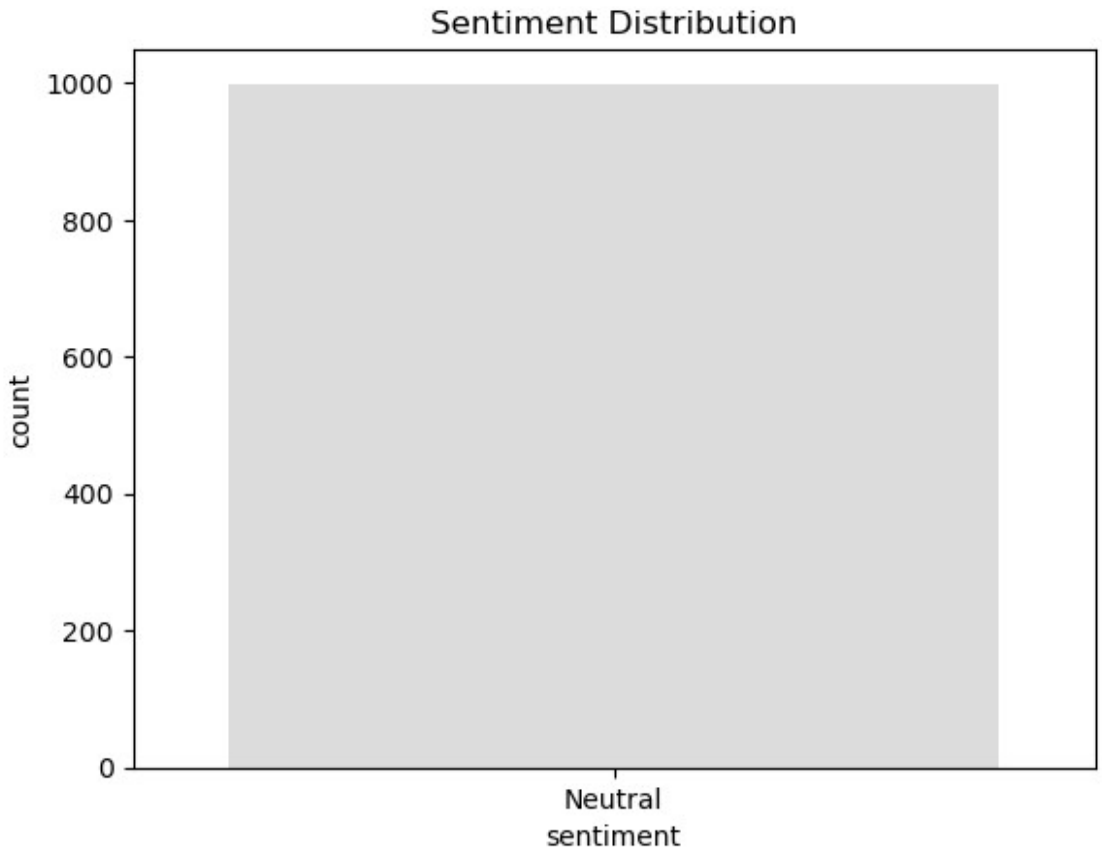
```
df['polarity'] = df['clean_text'].apply(get_sentiment)

def sentiment_label(score):
    if score > 0:
        return 'Positive'
    elif score < 0:
        return 'Negative'
    else:
        return 'Neutral'

df['sentiment'] = df['polarity'].apply(sentiment_label)
df[['clean_text', 'polarity', 'sentiment']].head()
```

	clean_text	polarity	sentiment
0	amazon	0.0	Neutral
1	microsoft	0.0	Neutral
2	csgo	0.0	Neutral
3	google	0.0	Neutral
4	fifa	0.0	Neutral

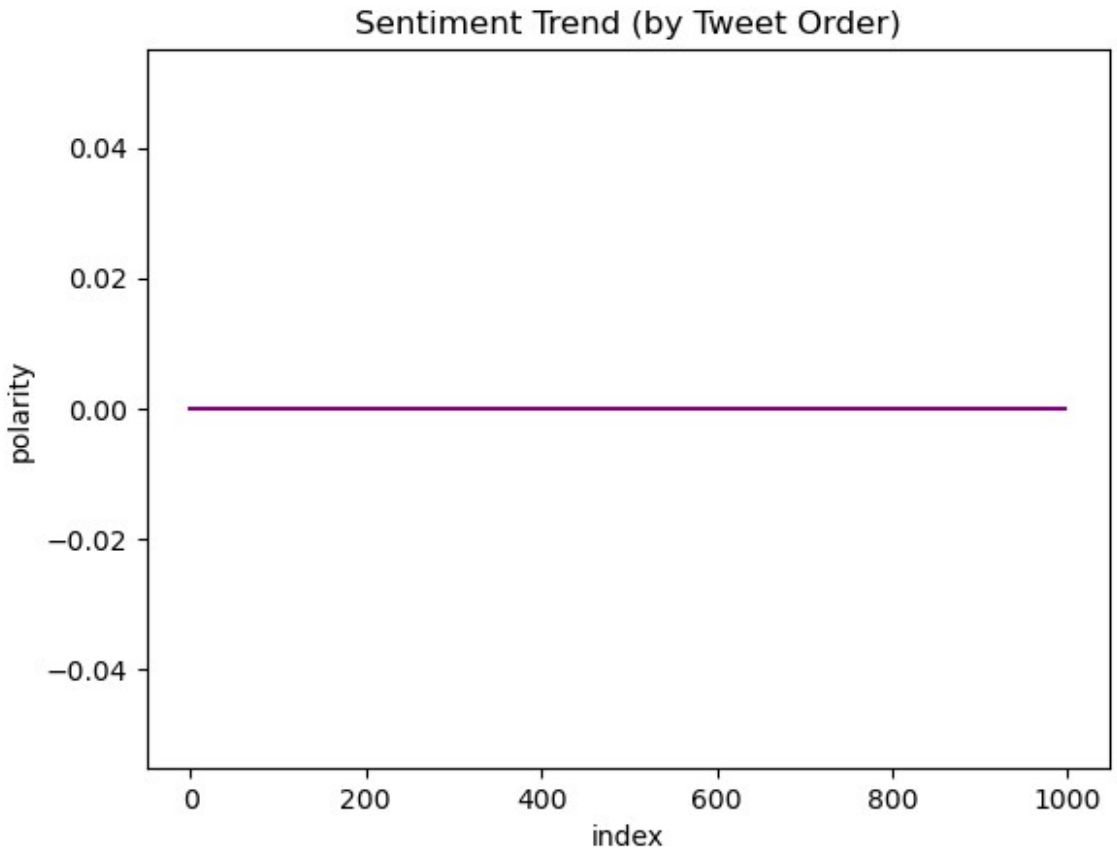
```
sns.countplot(data=df, x='sentiment',hue='sentiment',
palette='coolwarm')
plt.title('Sentiment Distribution')
plt.show()
```




```
for sentiment in df['sentiment'].unique():
    text = ' '.join(df[df['sentiment'] == sentiment]['clean_text'])
    wc = WordCloud(width=800, height=400,
background_color='white').generate(text)
    plt.figure(figsize=(8,4))
    plt.imshow(wc, interpolation='bilinear')
    plt.axis('off')
    plt.title(f'Word Cloud - {sentiment}')
    plt.show()
```


```
df['index'] = df.index
sns.lineplot(data=df, x='index', y='polarity', color='purple')
plt.title('Sentiment Trend (by Tweet Order)')
plt.show()

sentiment_counts = df['sentiment'].value_counts(normalize=True) * 100
print("\nSentiment Distribution (%):\n", sentiment_counts)
```




```
Sentiment Distribution (%):
 sentiment
Neutral    100.0
Name: proportion, dtype: float64

plt.figure(figsize=(12,4))

# Positive
plt.subplot(1,3,1)
sns.histplot(df[df['sentiment']=='Positive']['polarity'],
color='green', kde=True)
plt.title('Positive Sentiments')

# Negative
plt.subplot(1,3,2)
sns.histplot(df[df['sentiment']=='Negative']['polarity'], color='red',
kde=True)
plt.title('Negative Sentiments')

# Neutral
plt.subplot(1,3,3)
sns.histplot(df[df['sentiment']=='Neutral']['polarity'], color='gray',
```

```
kde=True)
plt.title('Neutral Sentiments')

plt.tight_layout()
plt.show()
```