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

```
import nltk
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('wordnet')
from nltk.corpus import stopwords
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
[nltk_data] Downloading package wordnet to /root/nltk_data...
```

```
from google.colab import files
uploaded = files.upload()
```

Choose Files No file chosen Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable

```
data = pd.read_csv('NLP-activity.csv')
print(data.head())
```

```
                                             REVIEWS  RATING
0                                          must buy        5
1  Amazing purchase at this prise as sound qualit...       3
2                                very good product         2
3  This buds have very good and stylish design.Th...       1
4  What I found in this earbuds is the clear voic...       4
```

```
#stemming
from nltk.stem import PorterStemmer
from nltk.tokenize import word_tokenize
stemmer = PorterStemmer()
```

```
target_column = 'REVIEWS'
```

```
def stem_text(text):
    words = word_tokenize(text)
    stemmed_words = [stemmer.stem(word) for word in words]
    return ' '.join(stemmed_words)
```

```
data[target_column] = data[target_column].fillna('')
```

```
def stem_text(text):
    words = word_tokenize(text)
```

```
    stemmed_words = [stemmer.stem(word) for word in words]
    return ' '.join(stemmed_words)
```

```
data[target_column] = data[target_column].apply(stem_text)
```

```
output_csv_file = 'output_data_stemmed.csv'
data.to_csv(output_csv_file, index=False)

print("Stemming complete.")
```

```
Stemming complete.
```

```
print(data.head())
```

```
                                          REVIEWS  RATING
0                                        must buy       5
1  amaz purchas at thi prise as sound qualiti is ...       3
2                                veri good product       2
3  thi bud have veri good and stylish design.th s...       1
4  what i found in thi earbud is the clear voic d...       4
```

```
#tokenization
from nltk.tokenize import word_tokenize
```

```
def tokenize_text(text):
    return word_tokenize(text)
```

```
data[target_column + '_tokens'] = data[target_column].apply(tokenize_text)
```

```
output_csv_file = 'output_data_tokenized.csv'
data.to_csv(output_csv_file, index=False)

print("Tokenization complete.")
```

```
Tokenization complete.
```

```
print(data.head())
```

```
                                          REVIEWS  RATING
0                                        must buy       5
1  amaz purchas at thi prise as sound qualiti is ...       3
2                                veri good product       2
3  thi bud have veri good and stylish design.th s...       1
4  what i found in thi earbud is the clear voic d...       4
```

```
#lemmatization
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
```

```
lemmatizer = WordNetLemmatizer()
```

```
def lemmatize_text(text):
    words = word_tokenize(text)
    lemmatized_words = [lemmatizer.lemmatize(word) for word in words]
    return ' '.join(lemmatized_words)
```

```
data[target_column + '_lemmatized'] = data[target_column].apply(lemmatize_text)
```

```
output_csv_file = 'output_data_lemmatized.csv'
data.to_csv(output_csv_file, index=False)

print("Lemmatization complete.")
print(data.head())
```

```
Lemmatization complete.
                                            REVIEWS  RATING
0                                          must buy       5
1  amaz purchas at thi prise as sound qualiti is ...       3
2                                veri good product       2
3  thi bud have veri good and stylish design.th s...       1
4  what i found in thi earbud is the clear voic d...       4
```

```
#stop words removal
nltk.download('stopwords')
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

```
stop_words = set(stopwords.words('english'))
```

```
def remove_stop_words(text):
    words = word_tokenize(text)
    filtered_words = [word for word in words if word.lower() not in stop_words]
    return ' '.join(filtered_words)
```

```
data[target_column + '_no_stopwords'] = data[target_column].apply(remove_stop_words)
```

```
output_csv_file = 'output_data_no_stopwords.csv'
data.to_csv(output_csv_file, index=False)

print("Stop words removal complete.")
print(data.head())
```

```
Stop words removal complete.
                                            REVIEWS  RATING
0                                          must buy       5
1  amaz purchas at thi prise as sound qualiti is ...       3
2                                veri good product       2
```

```
3  thi bud have veri good and stylish design.th s...       1
4  what i found in thi earbud is the clear voic d...       4
```

```
#lowercasing
data[target_column + '_lowercase'] = data[target_column].str.lower()
```

```
output_csv_file = 'output_data_lowercase.csv'
data.to_csv(output_csv_file, index=False)

print("Lowercasing complete.")
print(data.head())
```

```
Lowercasing complete.
                                             REVIEWS  RATING
0                                          must buy       5
1  amaz purchas at thi prise as sound qualiti is ...       3
2                                veri good product       2
3  thi bud have veri good and stylish design.th s...       1
4  what i found in thi earbud is the clear voic d...       4
```

```
from google.colab import files
uploaded = files.upload()
```

- **NLP-activity.csv**(text/csv) - 1888 bytes, last modified: 9/3/2023 - 100% done

```
Saving NLP-activity.csv to NLP-activity.csv
```

```
data = pd.read_csv('NLP-activity.csv')
print(data.head())
```

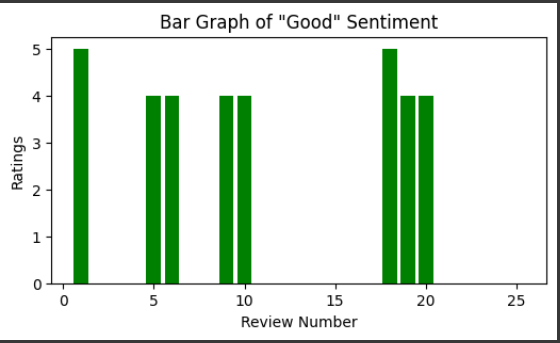
```
                                             REVIEWS  RATING
0                                          must buy       5
1  Amazing purchase at this prise as sound qualit...       3
2                               very good product          2
3  This buds have very good and stylish design.Th...       1
4  What I found in this earbuds is the clear voic...       4
```

```
csv_file = ('NLP-activity.csv')
data = pd.read_csv('NLP-activity.csv')
data['ReviewNumber'] = range(1, len(data) + 1)
print(data.head())
```

```
                                             REVIEWS  RATING  ReviewNumber
0                                          must buy       5             1
1  Amazing purchase at this prise as sound qualit...       3             2
2                               very good product          2             3
3  This buds have very good and stylish design.Th...       1             4
4  What I found in this earbuds is the clear voic...       4             5
```

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

# Sample data with 25 review numbers
```

```
data = pd.DataFrame({
    'ReviewNumber': range(1, 26),  # 25 review numbers from 1 to 25
})

# Define the custom "good" sentiment values corresponding to each review number
good_sentiment_values = [5, 0, 0, 0, 4, 4, 0, 0, 4, 4, 0, 0, 0, 0, 0, 0, 0, 5, 4, 4, 0, 0,

# Create a figure with one subplot for the "good" sentiment
plt.figure(figsize=(6, 3))

# Define the review numbers
review_numbers = data['ReviewNumber']

# Plot the custom "good" sentiment values for each review number
plt.bar(review_numbers, good_sentiment_values, color='green')
plt.xlabel('Review Number')
plt.ylabel('Ratings')
plt.title('Bar Graph of "Good" Sentiment')

# Show the plot
plt.show()
```

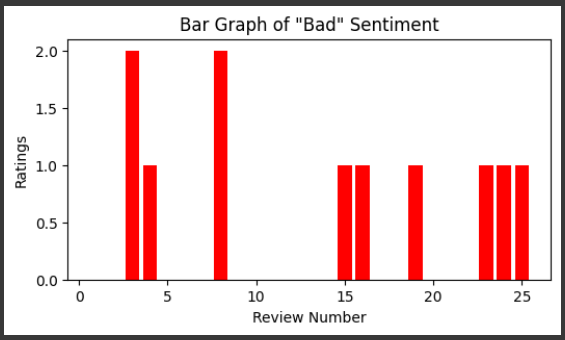



```
bad_sentiment_values = [0, 0, 2, 1, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 1
```

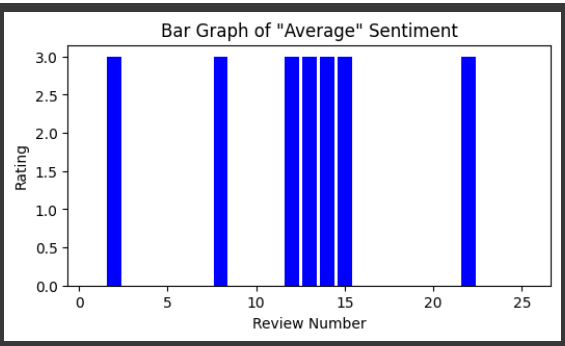
```
<Figure size 600x300 with 0 Axes>
<Figure size 600x300 with 0 Axes>
```

```
plt.figure(figsize=(6, 3))

review_numbers = data['ReviewNumber']

# Plot the custom "bad" sentiment values for each review number
plt.bar(review_numbers, bad_sentiment_values, color='red')
plt.xlabel('Review Number')
plt.ylabel('Ratings')
```

```
plt.title('Bar Graph of "Bad" Sentiment')

# Show the plot
plt.show()
```




```
average_sentiment_values = [0, 3, 0, 0, 0, 0, 0, 3, 0, 0, 0, 3, 3, 3, 3, 0, 0, 0, 0, 0, 0,


plt.figure(figsize=(6, 3))


review_numbers = data['ReviewNumber']


plt.bar(review_numbers, average_sentiment_values, color='blue')
plt.xlabel('Review Number')
plt.ylabel('Rating')
plt.title('Bar Graph of "Average" Sentiment')


plt.show()
```



✓ 0s completed at 1:57 AM