

```
In [ ]: from sklearn.feature_extraction.text import ENGLISH_STOP_WORDS
from pandas import DataFrame, read_csv, Series, set_option
import matplotlib.pyplot as plt

from string import punctuation
from nltk.stem import PorterStemmer
import pandas as pd
from langcodes import Language
from textblob import TextBlob
import seaborn as sns
import string
import matplotlib.pyplot as plt
import spacy
```

```
In [ ]: # Test 1. A.Text Preprocessing (10 points):
# Choose a set of text documents (e.g., articles or reviews).
# Implement a Python script or Jupyter Notebook to preprocess the text data (tokeni
```

```
In [ ]: comments_csv = read_csv("./_spotify_short.csv")
comments_csv.head()
```

Out [ ]:

	Time_submitted	Review	Rating	Total_thumbsup	Reply
0	2022-07-09 15:00:00	Great music service, the audio is high quality...	5	2	NaN
1	2022-07-09 14:21:22	Please ignore previous negative rating. This a...	5	1	NaN
2	2022-07-09 13:27:32	This pop-up "Get the best Spotify experience o...	4	0	NaN
3	2022-07-09 13:26:45	Really buggy and terrible to use as of recently	1	1	NaN
4	2022-07-09 13:20:49	Dear Spotify why do I get songs that I didn't ...	1	1	NaN

```
In [ ]: columns_to_drop = ["Time_submitted", "Rating", "Total_thumbsup", "Reply"]
comments_csv.drop(columns=columns_to_drop, inplace = True)
set_option('display.max_colwidth', 180)
comments_csv.head()
```

Out [ ]:

	Review
0	Great music service, the audio is high quality and the app is easy to use. Also very quick and friendly support.
1	Please ignore previous negative rating. This app is super great. I give it five stars+
2	This pop-up "Get the best Spotify experience on Android 12" is too annoying. Please let's get rid of this.
3	Really buggy and terrible to use as of recently
4	Dear Spotify why do I get songs that I didn't put on my playlist??? And why do we have shuffle play?

```
In [ ]: # for lower_case
def convert_to_lower_case(text):
    return text.str.lower()

def apply_lower_case_conversion(dataframe, column_name):
    dataframe[column_name] = convert_to_lower_case(dataframe[column_name])
    return dataframe

columns_to_convert_to_lower = ['Review']

for column in columns_to_convert_to_lower:
```

```
comments_csv = apply_lower_case_conversion(comments_csv, column)
comments_csv.head(10)
```

Out[ ]:

	Review
0	great music service, the audio is high quality and the app is easy to use. also very quick and friendly support.
1	please ignore previous negative rating. this app is super great. i give it five stars+
2	this pop-up "get the best spotify experience on android 12" is too annoying. please let's get rid of this.
3	really buggy and terrible to use as of recently
4	dear spotify why do i get songs that i didn't put on my playlist??? and why do we have shuffle play?
5	the player controls sometimes disappear for no reason. app restart forgets what i was playing but fixes the issue.
6	i love the selection and the lyrics are provided with the song you're listening to!
7	still extremely slow when changing storage to external sd card.. i'm convinced this is done on purpose, spotify knows of this issue and has done nothing to solve it! over time ...
8	it's a great app and the best mp3 music app i have ever used but there is one problem that, why can't we play some songs or find some songs? despite this the app is wonderful i...
9	i'm deleting this app, for the following reasons: this app now has a failing business model. whether streaming services like it, or not: the consumer doesn't want to pay for mu...

In [ ]:

```
# To remove all punctuation in reviews
def remove_punctuation_from_text(text):
    translation_table = str.maketrans('', '',punctuation)
    return text.translate(translation_table)

def apply_punctuation_removal_to_column(dataframe, column_name):
    dataframe[column_name] = dataframe[column_name].apply(remove_punctuation_from_t
    return dataframe
columns_to_remove_punctuation = ['Review']

for column in columns_to_remove_punctuation:
    comments_csv = apply_punctuation_removal_to_column(comments_csv, column)

comments_csv.head(10)
```

Out[ ]:

	Review
0	great music service the audio is high quality and the app is easy to use also very quick and friendly support
1	please ignore previous negative rating this app is super great i give it five stars
2	this popup get the best spotify experience on android 12 is too annoying please lets get rid of this
3	really buggy and terrible to use as of recently
4	dear spotify why do i get songs that i didnt put on my playlist and why do we have shuffle play
5	the player controls sometimes disappear for no reason app restart forgets what i was playing but fixes the issue
6	i love the selection and the lyrics are provided with the song youre listening to
7	still extremely slow when changing storage to external sd card im convinced this is done on purpose spotify knows of this issue and has done nothing to solve it over time i hav...
8	its a great app and the best mp3 music app i have ever used but there is one problem that why cant we play some songs or find some songs despite this the app is wonderful i rec...
9	im deleting this app for the following reasons this app now has a failing business model whether streaming services like it or not the consumer doesnt want to pay for music the...

```
In [ ]: # To remove stop words in reviews
def remove_stop_words(text):
    stop_words = set(ENGLISH_STOP_WORDS)
    return ' '.join([word for word in text.split() if word.lower() not in stop_words])

def apply_stop_words_removal(dataframe, columns):
    for column in columns:
        dataframe[column] = dataframe[column].apply(remove_stop_words)
    return dataframe

columns_to_remove_stop_words = ['Review']
comments_csv = apply_stop_words_removal(comments_csv, columns_to_remove_stop_words)

comments_csv.head(10)
```

Out [ ]:

	Review
0	great music service audio high quality app easy use quick friendly support
1	ignore previous negative rating app super great stars
2	popup best spotify experience android 12 annoying lets rid
3	really buggy terrible use recently
4	dear spotify songs didnt playlist shuffle play
5	player controls disappear reason app restart forgets playing fixes issue
6	love selection lyrics provided song youre listening
7	extremely slow changing storage external sd card im convinced purpose spotify knows issue solve time changed sd cards faster read write speedsall samsung brand add dont like so...
8	great app best mp3 music app used problem play songs songs despite app wonderful recommend just best
9	im deleting app following reasons app failing business model streaming services like consumer doesnt want pay music fully 6 ads successively logging single song closed app ad n...

```
In [ ]: # Stemming the text for the review
def stemming(text):
    stemmer = PorterStemmer()
    return ' '.join([stemmer.stem(word) for word in text.split()])

def apply_stemming(dataframe, column):
    dataframe[column] = dataframe[column].apply(stemming)
    return dataframe

columns_to_stem = ['Review']
comments_csv = apply_stemming(comments_csv, 'Review')

comments_csv.head(10)
```

Out [ ]:

	Review
0	great music servic audio high qualiti app easi use quick friendli support
1	ignor previou neg rate app super great star
2	popup best spotifi experi android 12 annoy let rid
3	realli buggi terribl use recent
4	dear spotifi song didnt playlist shuffl play
5	player control disappear reason app restart forget play fix issu
6	love select lyric provid song your listen
7	extrem slow chang storag extern sd card im convinc purpos spotifi know issu solv time chang sd card faster read write speedsal samsung brand add dont like song appear search pl...
8	great app best mp3 music app use problem play song song despit app wonder recommend just best
9	im delet app follow reason app fail busi model stream servic like consum doesnt want pay music fulli 6 ad success log singl song close app ad number 6 im patient way profit you...

In [ ]:

```
# TEST 1. B. Sentiment Analysis (10 points):  
# Apply sentiment analysis to determine the sentiment of each document.  
# Use a pre-trained model or implement a simple model using libraries like NLTK or  
# Present the sentiment distribution and key findings.
```

In [ ]:

```
# Sentiment value of review  
comments_csv["Sentiment"] = comments_csv["Review"].apply(lambda text_review: TextBl  
comments_csv.head(10)
```

Out [ ]:

	Review	Sentiment
0	great music servic audio high qualiti app easi use quick friendli support	0.431111
1	ignor previou neg rate app super great star	0.566667
2	popup best spotifi experi android 12 annoy let rid	1.000000
3	realli buggi terribl use recent	0.000000
4	dear spotifi song didnt playlist shuffl play	0.000000
5	player control disappear reason app restart forget play fix issu	0.000000
6	love select lyric provid song your listen	0.375000
7	extrem slow chang storag extern sd card im convinc purpos spotifi know issu solv time chang sd card faster read write speedsal samsung brand add dont like song appear search pl...	-0.300000
8	great app best mp3 music app use problem play song song despit app wonder recommend just best	0.933333
9	im delet app follow reason app fail busi model stream servic like consum doesnt want pay music fulli 6 ad success log singl song close app ad number 6 im patient way profit you...	-0.066667

In [ ]:

```
def get_emotion(sentiment):  
    if sentiment > 0:  
        return "Positive"  
    if sentiment < 0:  
        return "Negative"  
  
    return "Neutral"  
  
comments_csv["Emotion"] = comments_csv["Sentiment"].apply(lambda sentiment: get_emo  
comments_csv.head(10)
```

Out[ ]:

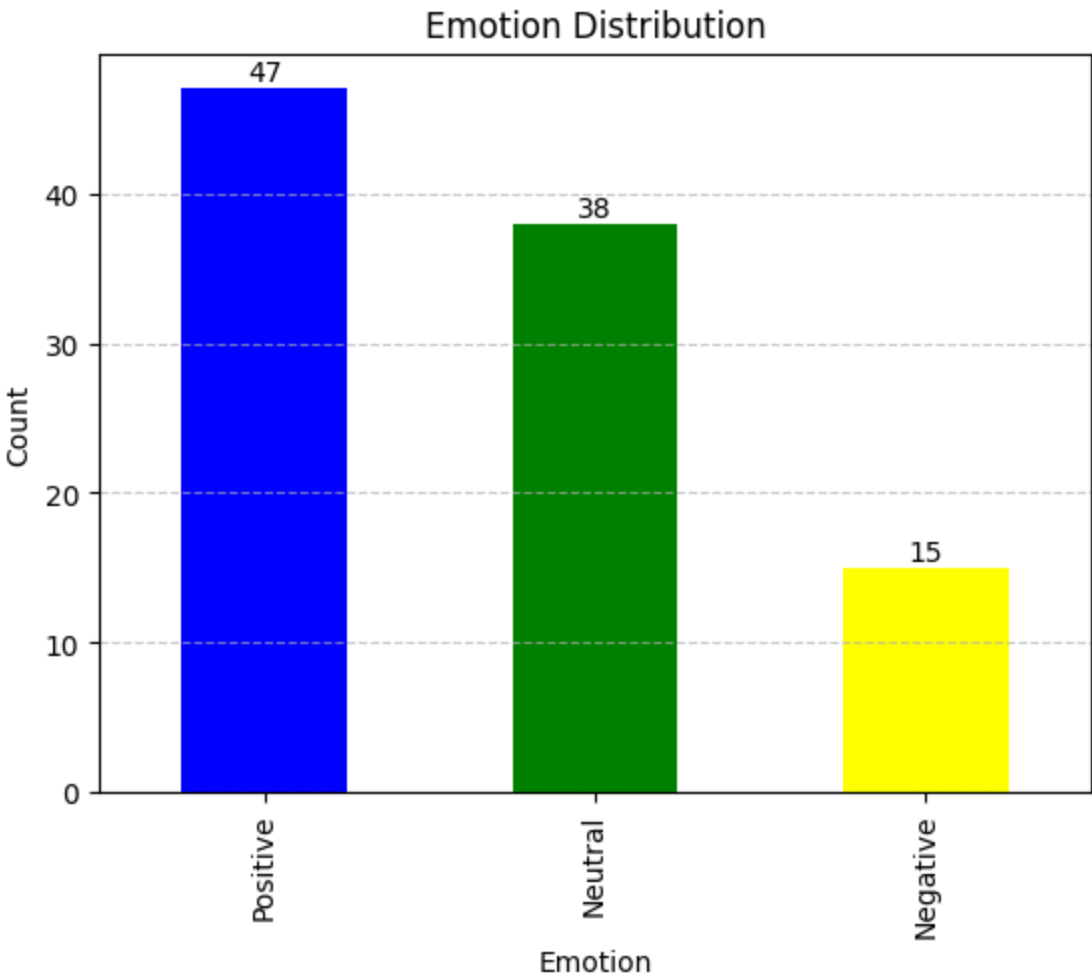
	Review	Sentiment	Emotion
0	great music servic audio high qualiti app easi use quick friendli support	0.431111	Positive
1	ignor previou neg rate app super great star	0.566667	Positive
2	popup best spotifi experi android 12 annoy let rid	1.000000	Positive
3	realli buggi terribl use recent	0.000000	Neutral
4	dear spotifi song didnt playlist shuffl play	0.000000	Neutral
5	player control disappear reason app restart forget play fix issu	0.000000	Neutral
6	love select lyric provid song your listen	0.375000	Positive
7	extrem slow chang storag extern sd card im convinc purpos spotifi know issu solv time chang sd card faster read write speedsal samsung brand add dont like song appear search pl...	-0.300000	Negative
8	great app best mp3 music app use problem play song song despit app wonder recommend just best	0.933333	Positive
9	im delet app follow reason app fail busi model stream servic like consum doesnt want pay music fulli 6 ad success log singl song close app ad number 6 im patient way profit you...	-0.066667	Negative

```
In [ ]: emotion_counts = comments_csv["Emotion"].value_counts()
bar_colors = ['blue', 'green', 'yellow']
emotion_counts.plot(kind='bar', color=bar_colors)

plt.xlabel("Emotion")
plt.ylabel("Count")
plt.title("Emotion Distribution")

for i, count in enumerate(emotion_counts):
    plt.text(i, count + 0.1, str(count), ha='center', va='bottom')

plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.show()
```



```
In [ ]: # Test 1.
# C.Named Entity Recognition (NER) (10 points):
# Implement a named entity recognition model using spaCy or a similar library.
# Identify and visualize named entities in the provided text data.
```

```
In [ ]: nlp = spacy.load("en_core_web_sm")
```

```
In [ ]: # Entities of Reviews
def classify_review(review):

    document = nlp(review)
    entities = [(entity.text, entity.label_) for entity in document.ents]
    return entities

comments_csv["Entities-NER"] = comments_csv["Review"].apply(lambda review: classify_review(review))
comments_csv.head(10)
```

Out[ ]:

	Review	Sentiment	Emotion	Entities-NER
0	great music servic audio high qualiti app easi use quick friendli support	0.431111	Positive	[(qualiti, PERSON)]
1	ignor previou neg rate app super great star	0.566667	Positive	[]
2	popup best spotifi experi android 12 annoy let rid	1.000000	Positive	[(12, CARDINAL)]
3	realli buggi terribl use recent	0.000000	Neutral	[(realli buggi terribl, PERSON)]
4	dear spotifi song didnt playlist shuffl play	0.000000	Neutral	[]
5	player control disappear reason app restart forget play fix issu	0.000000	Neutral	[(issu, ORG)]
6	love select lyric provid song your listen	0.375000	Positive	[]
7	extrem slow chang storag extern sd card im convinc purpos spotifi know issu solv time chang sd card faster read write speedsal samsung brand add dont like song appear search pl...	-0.300000	Negative	[(extrem slow chang storag extern, PERSON), (issu, ORG), (chang, PERSON), (samsung, ORG)]
8	great app best mp3 music app use problem play song song despit app wonder recommend just best	0.933333	Positive	[]
9	im delet app follow reason app fail busi model stream servic like consum doesnt want pay music fulli 6 ad success log singl song close app ad number 6 im patient way profit you...	-0.066667	Negative	[(6, CARDINAL), (6, CARDINAL), (declin, GPE)]

```
In [ ]: # NER bar graph
from collections import Counter
flat_entities = [entity[1] for entities in comments_csv["Entities-NER"] for entity in entities]
entity_counts = Counter(flat_entities)

sns.barplot(x=list(entity_counts.values()), y=list(entity_counts.keys()), color='green')
plt.title('NER Entity Counts')
plt.ylabel('Entity Type')
plt.xlabel('Count')
plt.show()
```

