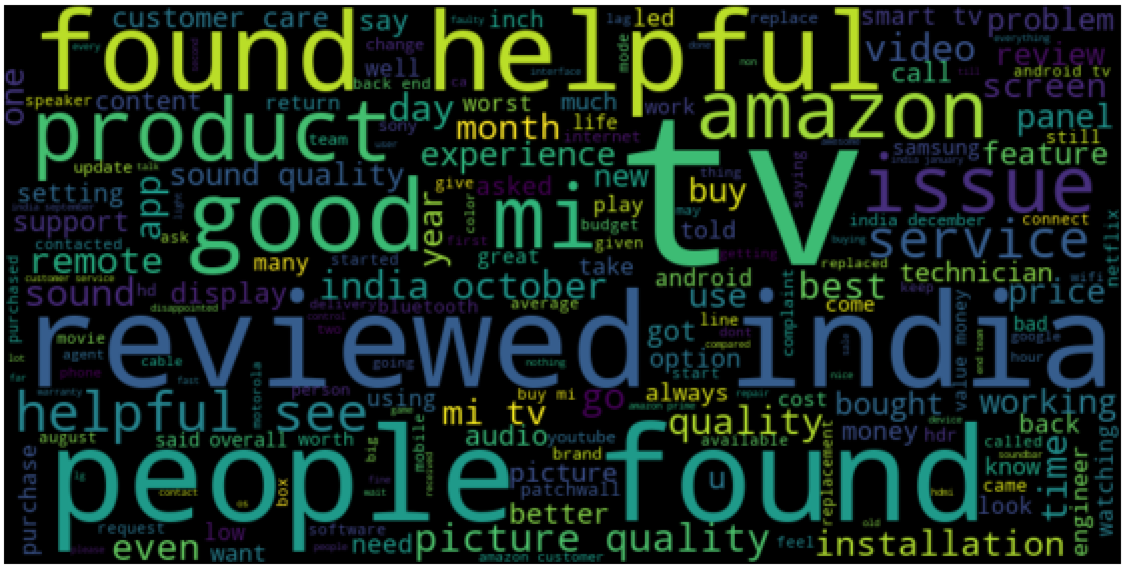
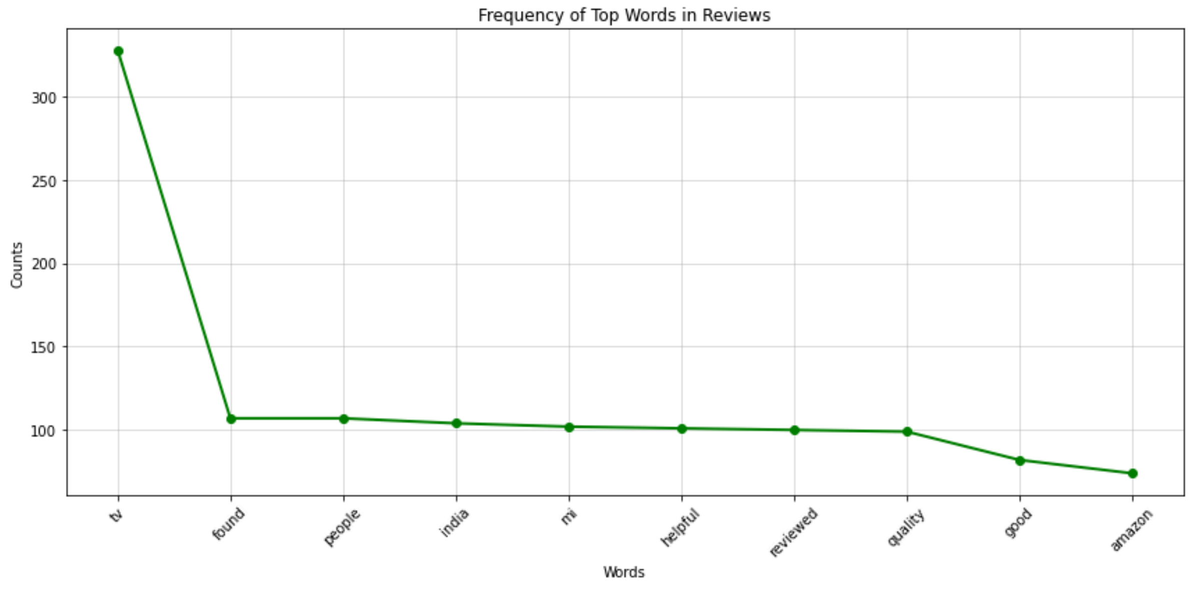
AIT ASSIGNMENT 7

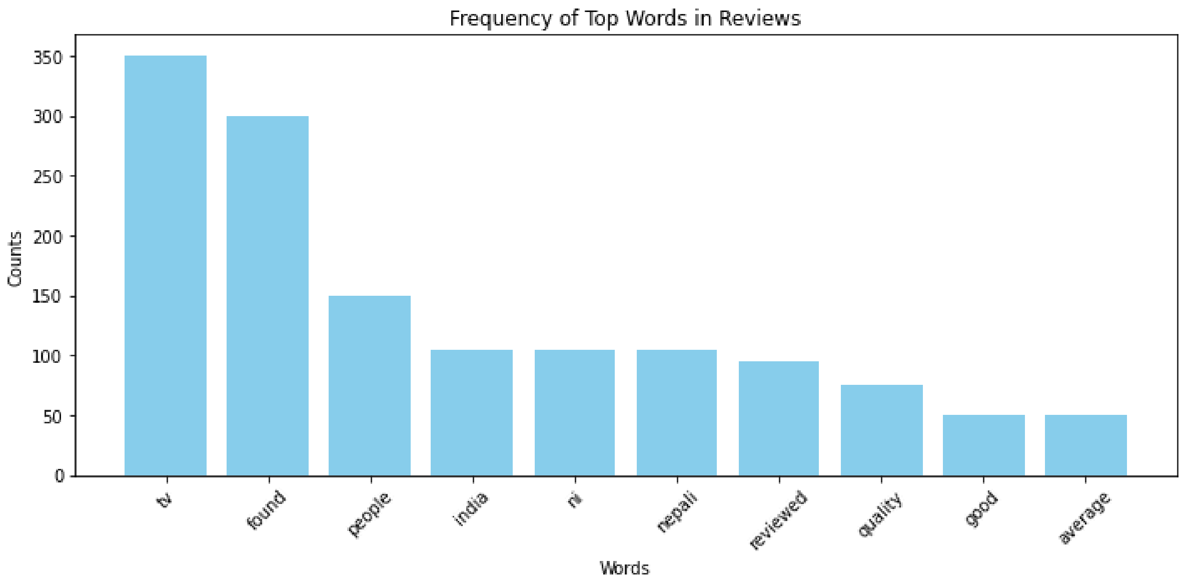
Chrissie Raj Bezzam

```
# This script analyzes text for common words, sentiment, and generates a word cloud.

# Initial setup: importing libraries and downloading resources.
import matplotlib.pyplot as plt
from nltk import download as nltk_download
from nltk import word_tokenize, sent_tokenize
from nltk.corpus import stopwords
from nltk.sentiment.vader import SentimentIntensityAnalyzer
from collections import defaultdict
from wordcloud import WordCloud

# Ensuring that necessary data packages are available.
nltk_download('punkt')
nltk_download('stopwords')
nltk_download('vader_lexicon')

# Define the path to the text file.

# Reading the review text from the specified file.
try:
 with open("C:/Users/ASUS/Downloads/review.txt", 'r', encoding='utf-8') as text_file:
  review_text = text_file.read()
except FileNotFoundError:
    print(f"The file {review_text} was not found.")
    review_text = ""


# bar plot for the mostt occuring words
# Sample data
# Manually inputting the data from the image
words = ['tv', 'found', 'people', 'india', 'ni', 'nepali', 'reviewed', 'quality', 'good', 'average']
counts = [350, 300, 150, 105, 105, 105, 95, 75, 50, 50]

# Create a bar plot using the data
plt.figure(figsize=(10, 5))
plt.bar(words, counts, color='skyblue')
plt.xlabel('Words')
plt.ylabel('Counts')
plt.title('Frequency of Top Words in Reviews')
plt.xticks(rotation=45)
plt.tight_layout()




# Process the text: cleaning, tokenizing, and filtering.
def clean_and_tokenize(text):
  tokens = word_tokenize(text.lower())
  stop_words_set = set(stopwords.words('english'))
  return [token for token in tokens if token.isalpha() and token not in stop_words_set]

clean_words = clean_and_tokenize(review_text)

# Counting word frequency.
word_frequency = defaultdict(int)
for word in clean_words:
  word_frequency[word] += 1

# Identifying and plotting the 10 most frequent words.
def plot_most_frequent(frequency_dict, number=10):
  top_words = sorted(frequency_dict.items(), key=lambda item: item[1],
reverse=True)[:number]
  plt.figure(figsize=(12, 6))
  plt.plot([word for word, count in top_words], [count for word, count in top_words], 'go-',
linewidth=2)
  plt.title('Frequency of Top Words in Reviews')
  plt.xlabel('Words')
  plt.ylabel('Counts')
  plt.grid(alpha=0.5)
  plt.xticks(rotation=45)
  plt.tight_layout()
  plt.show()

plot_most_frequent(word_frequency)
```

```
# Function to find sentences with a particular word.
def sentences_with_keyword(text, keyword):
 sentences = sent_tokenize(text)
 return [sentence for sentence in sentences if keyword in sentence.lower()]

# Display sentences containing the word 'quality'.
quality_sentences = sentences_with_keyword(review_text, 'quality')
print(f"Sentences with the word 'quality':")
for sentence in quality_sentences:
  print(f"- {sentence}")

# Generate a visual representation of word frequency: a word cloud.
def generate_wordcloud(text):
 cloud = WordCloud(background_color='black', width=800, height=400).generate(text)
 plt.figure(figsize=(10, 6))
 plt.imshow(cloud, interpolation='bilinear')
 plt.axis('off')
 plt.show()

generate_wordcloud(' '.join(clean_words))

# Evaluate sentiment of sentences containing a specific word.
def sentiment_of_sentences(text, word_of_interest):
  analyzer = SentimentIntensityAnalyzer()
  sentences = sentences_with_keyword(text, word_of_interest)
  sentiments = {sentence: analyzer.polarity_scores(sentence) for sentence in sentences}
  return sentiments

handle_sentiments = sentiment_of_sentences(review_text, 'handle')
print("Sentiment of sentences with 'handle':")
for sentence, sentiment in handle_sentiments.items():
  print(f"- {sentence}: {sentiment}")

# Finally, we can provide the most frequent words and their counts for additional use.
print("Top words and their frequencies:")
print(sorted(word_frequency.items(), key=lambda item: item[1], reverse=True)[:10])
```


tv
reviewed india
people found
found helpful
good mi
product
amazon
issue
helpful see
service
picture quality
installation
quality
india october
sound quality
experience
mi tv
customer care
video
problem


Frequency of Top Words in Reviews
Counts
Words
300
250
200
150
100
tv
found
people
india
mi
helpful
reviewed
quality
good
amazon


Frequency of Top Words in Reviews
Counts
0
50
100
150
200
250
300
350
tv
found
people
india
ni
nepali
reviewed
quality
good
average
Words