Name = Yash Shinde

Class = ET2

Roll no =ET2-15

PRN = 202401070121

**20 Problem statement and its solution**

import pandas as pd

import numpy as np

from collections import Counter

**# Load the dataset**

file\_path = "ModelTrain.csv" # Replace with the actual path to your CSV file

df = pd.read\_csv(file\_path)

**# Helper function to count words in a text**

def word\_count(text):

return len(str(text).split())

**# Helper function to get all words from the reviews**

def get\_all\_words(reviews):

words = []

for review in reviews:

words.extend(str(review).lower().split())

return words

**# 1. Total number of reviews**

total\_reviews = df.shape[0] # Corrected to get the number of rows

**# 2. Count of POSITIVE and NEGATIVE reviews**

sentiment\_counts = df['Sentiment'].value\_counts()

**# 3. Percentage of POSITIVE reviews**

positive\_percentage = (sentiment\_counts['POSITIVE'] / total\_reviews) \* 100

**# 4. Percentage of NEGATIVE reviews**

negative\_percentage = (sentiment\_counts['NEGATIVE'] / total\_reviews) \* 100

**# 5. Average length of reviews (characters)**

avg\_length\_chars = df['Review'].apply(len).mean()

**# 6. Average length of positive reviews**

avg\_length\_positive = df[df['Sentiment'] == 'POSITIVE']['Review'].apply(len).mean()

**# 7. Average length of negative reviews**

avg\_length\_negative = df[df['Sentiment'] == 'NEGATIVE']['Review'].apply(len).mean()

**# 8. Review with maximum characters**

max\_length\_review = df.loc[df['Review'].apply(len).idxmax(), 'Review']

**# 9. Review with minimum characters**

min\_length\_review = df.loc[df['Review'].apply(len).idxmin(), 'Review']

**# 10. Total number of words across all reviews**

total\_words = df['Review'].apply(word\_count).sum()

**# 11. Average number of words per review**

avg\_words\_per\_review = df['Review'].apply(word\_count).mean()

**# 12. Most common word across all reviews**

all\_words\_list = get\_all\_words(df['Review'])

most\_common\_word = Counter(all\_words\_list).most\_common(1)[0][0] #Corrected to extract the word

**# Further computations (adapted from the document)**

**# 20. Reviews with more than 100 words**

reviews\_more\_than\_100\_words = df[df['Review'].apply(word\_count) > 100].shape[0]

**# 21. Average word length across all reviews**

all\_word\_lengths = [len(word) for word in all\_words\_list]

average\_word\_length = np.mean(all\_word\_lengths)

**# Displaying results**

print(f"Total Reviews: {total\_reviews}")

print(f"Sentiment Counts:\n{sentiment\_counts}")

print(f"Positive %: {positive\_percentage:.2f}")

print(f"Negative %: {negative\_percentage:.2f}")

print(f"Average Review Length (chars): {avg\_length\_chars:.2f}")

print(f"Average Positive Review Length (chars): {avg\_length\_positive:.2f}")

print(f"Average Negative Review Length (chars): {avg\_length\_negative:.2f}")

print(f"Review with Maximum Characters (preview): {max\_length\_review[:300]}...") # Truncated to 300 chars

print(f"Review with Minimum Characters: {min\_length\_review}")

print(f"Total Words: {total\_words}")

print(f"Average Words per Review: {avg\_words\_per\_review:.2f}")

print(f"Most Common Word (All Reviews): {most\_common\_word}")

print(f"Reviews with more than 100 words: {reviews\_more\_than\_100\_words}")

print(f"Average Word Length: {average\_word\_length:.2f}")

Name=Yash shinde

Class= ET2

Roll no=15

Prn=202401070121