**EXPERIMENT 02- EXPLORATORY DATA ANALYSIS USING E-MAIL DATASET**

**PROGARM:**

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from wordcloud import WordCloud, STOPWORDS

data = [

{'label': 'ham', 'message': 'Hey, are you coming to the party tonight?'},

{'label': 'spam', 'message': 'Congratulations! You have won a free ticket to Bahamas. Call now!'},

{'label': 'ham', 'message': 'Don’t forget to bring your notebook tomorrow.'},

{'label': 'spam', 'message': 'You have been selected for a $1000 gift card. Click here!'},

{'label': 'ham', 'message': 'Let’s meet at 5 PM in the library.'},

{'label': 'spam', 'message': 'Free entry in a contest. Text WIN to 12345 now!'},

{'label': 'ham', 'message': 'Lunch at 1?'},

{'label': 'ham', 'message': 'Sure, I can help with your assignment.'},

{'label': 'spam', 'message': 'Exclusive offer! Buy 1 get 1 free on all products.'}

]

df = pd.DataFrame(data)

df['length'] = df['message'].apply(len)

sns.countplot(data=df, x='label')

plt.title("Spam vs Ham Distribution")

plt.show()

sns.histplot(data=df, x='length', hue='label', bins=10, kde=True)

plt.title("Message Length Distribution")

plt.show()

print("\nAverage Message Lengths:")

print(df.groupby('label')['length'].mean())

spam\_words = " ".join(df[df['label'] == 'spam']['message'])

ham\_words = " ".join(df[df['label'] == 'ham']['message'])

spam\_wc = WordCloud(width=500, height=300, background\_color='black', stopwords=STOPWORDS).generate(spam\_words)

ham\_wc = WordCloud(width=500, height=300, background\_color='white', stopwords=STOPWORDS).generate(ham\_words)

plt.figure(figsize=(10, 4))

plt.subplot(1, 2, 1)

plt.imshow(spam\_wc, interpolation='bilinear')

plt.axis('off')

plt.title("Spam Word Cloud")

plt.subplot(1, 2, 2)

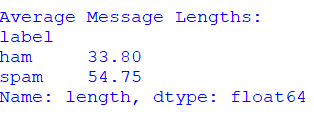
plt.imshow(ham\_wc, interpolation='bilinear')

plt.axis('off')

plt.title("Ham Word Cloud")

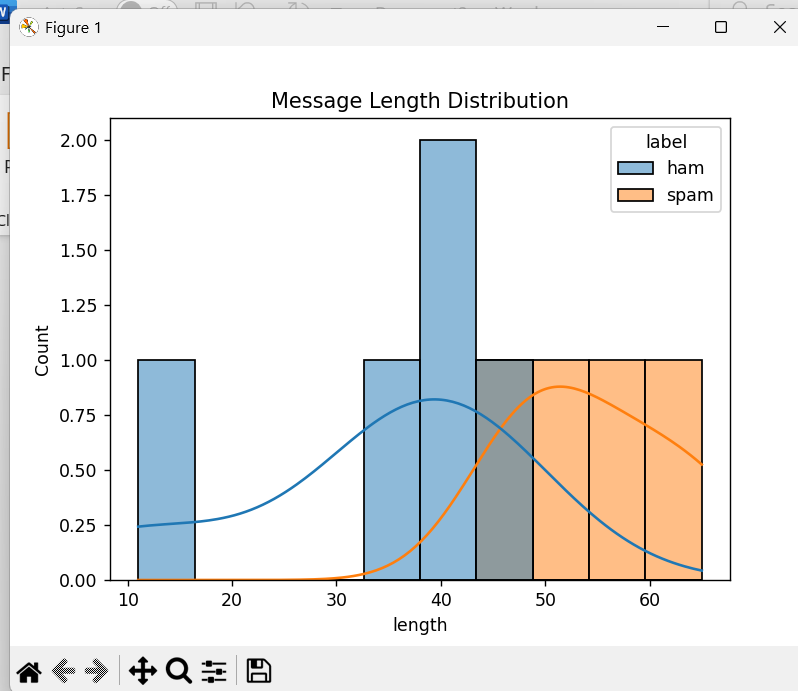
plt.tight\_layout()

plt.show()



A screenshot of a computer screen

AI-generated content may be incorrect.



A screenshot of a computer screen

AI-generated content may be incorrect.