Code:

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

from wordcloud import WordCloud

import string

df = pd.read\_csv("spam.csv", encoding='latin-1')[['v1', 'v2']]

df.columns = ['label', 'message']

print("Head of dataset:")

print(df.head())

print("\nDataset Info:")

print(df.info())

print("\nStatistical Summary:")

print(df.describe())

print("\nMissing values per column:")

print(df.isnull().sum())

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

plt.title("Distribution of Spam vs Ham")

plt.xlabel("Label")

plt.ylabel("Count")

plt.show()

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

plt.figure(figsize=(10, 6))

df[df['label']=='ham']['length'].plot.hist(bins=50, alpha=0.7, label='Ham',color='red')

df[df['label']=='spam']['length'].plot.hist(bins=50, alpha=0.7, label='Spam',color='yellow')

plt.legend()

plt.title("Histogram of Message Lengths")

plt.xlabel("Length")

plt.show()

def clean\_text(msg):

msg = msg.lower()

msg = "".join([char for char in msg if char not in string.punctuation])

return msg

df['clean\_message'] = df['message'].apply(clean\_text)

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

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

spam\_wc = WordCloud(width=600, height=400, background\_color='black').generate(spam\_words)

ham\_wc = WordCloud(width=600, height=400, background\_color='white').generate(ham\_words)

plt.figure(figsize=(12, 6))

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.show()

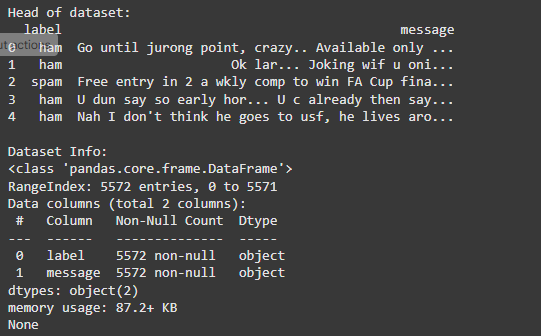
print("Average message length (ham):", df[df['label'] == 'ham']['length'].mean())

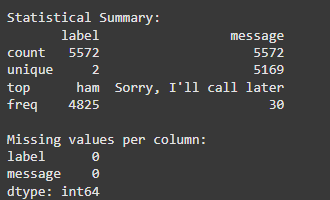
print("Average message length (spam):", df[df['label'] == 'spam']['length'].mean())

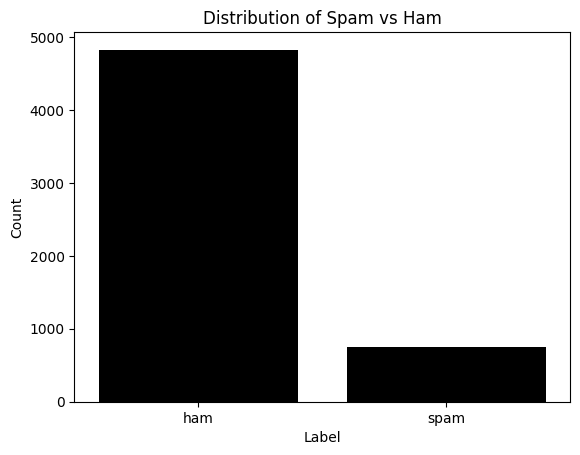
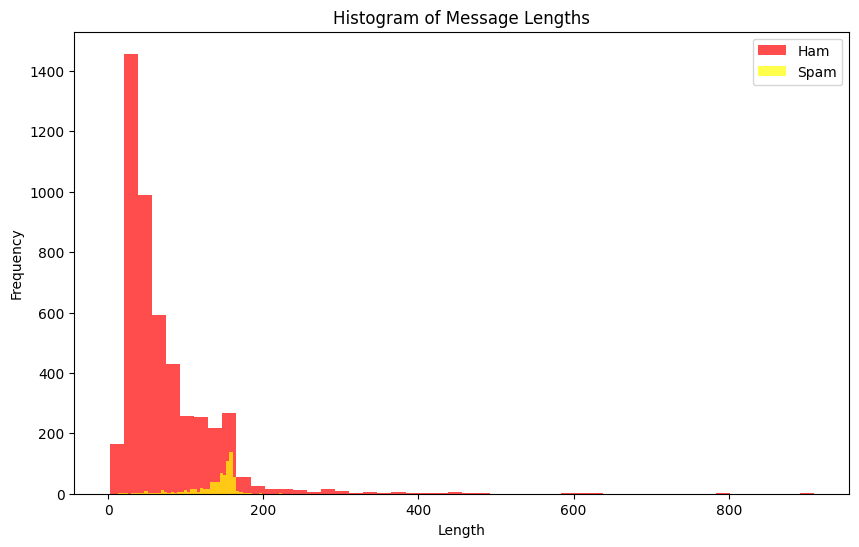
print("Max message length:", df['length'].max())

print("Min message length:", df['length'].min())

Output:





  
  
Average message length (ham): 71.02362694300518

Average message length (spam): 138.8661311914324

Max message length: 910

Min message length: 2