


```
import numpy as np
```




```
import pandas as pd
```

```
from google.colab import files
uploaded = files.upload()
```

 Choose Files spam\_texts.csv  
 • spam\_texts.csv(text/csv) - 503663 bytes, last modified: 3/22/2025 - 100% done


```
#here we read the uploaded file using a specific encoding known as ISO-8859-1 (also known as latin1)
df = pd.read_csv('spam_texts.csv', encoding='ISO-8859-1')
```

```
df.sample(20)
```

	v1	v2	Unnamed: 2	Unnamed: 3	Unnamed: 4
4068	ham	Hope youâ€”re not having too much fun without m...	NaN	NaN	NaN
3479	ham	What was she looking for?	NaN	NaN	NaN
2478	ham	Not yet. Just i'd like to keep in touch and it...	NaN	NaN	NaN
4511	ham	Now project pa. After that only i can come.	NaN	NaN	NaN
2850	ham	Are you going to wipro interview today?	NaN	NaN	NaN
2823	ham	No need to buy lunch for me.. I eat maggi mee..	NaN	NaN	NaN
5006	ham	Oh k. . I will come tomorrow	NaN	NaN	NaN
555	ham	O. Well uv causes mutations. Sunscreen is like...	NaN	NaN	NaN
927	ham	K:)i will give my kvb acc details:)	NaN	NaN	NaN
2051	ham	Hey darlin.. i can pick u up at college if u t...	NaN	NaN	NaN
258	spam	We tried to contact you re your reply to our o...	NaN	NaN	NaN
2510	ham	Yunny i'm walking in citylink now ħ_ faster co...	NaN	NaN	NaN
3111	ham	Just haven't decided where yet eh ?	NaN	NaN	NaN
1802	ham	Ok lor thanx... ĩĩ in school?	NaN	NaN	NaN
1156	ham	Hey.. Something came up last min.. Think i wun...	NaN	NaN	NaN
2342	spam	Double your mins & txts on Orange or 1/2 price...	NaN	NaN	NaN
3448	ham	Sorry.    mail?	NaN	NaN	NaN
1483	ham	Purity of friendship between two is not about ...	NaN	NaN	NaN
4712	ham	Big brother ħ+s really scraped the barrel with...	NaN	NaN	NaN
2439	ham	Rightio ħ 48 it is then Well arent we all u	NaN	NaN	NaN

```
df.shape
```


 (5572, 5)

```
# Drop unnecessary columns
df = df[['v1', 'v2']]
```

```
# Rename columns for clarity
df.columns = ['label', 'message']
```

## Data Cleaning and Processing

```
df.info(), df['label'].value_counts()
```

 <class 'pandas.core.frame.DataFrame'>  
 RangeIndex: 5572 entries, 0 to 5571  
 Data columns (total 2 columns):

```
# Column Non-Null Count Dtype
---
0 label 5572 non-null object
1 message 5572 non-null object
dtypes: object(2)
memory usage: 87.2+ KB
(None,
label
ham 4825
spam 747
Name: count, dtype: int64)
```

```
df.sample(10)
```

	label	message
4176	ham	Ok lor then we go tog lor...
3203	ham	She's good. How are you. Where r u working now
5228	spam	PRIVATE! Your 2003 Account Statement for <fone...
4945	ham	I'm already back home so no probably not
2354	ham	R we going with the &lt;#&gt; bus?
5268	ham	\ER
4501	ham	Wot is u up 2 then bitch?
1254	ham	What your plan for pongal?
4607	ham	We live in the next &lt;#&gt; mins
3293	ham	Beautiful tomorrow never comes. When it comes

## Data Preprocessing

```
import re
import nltk
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from nltk.tokenize import word_tokenize # Changed 'word_tokeniz' to 'word_tokenize'
```

```
# Download necessary NLTK datasets
nltk.download('punkt')
nltk.download('stopwords')
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
True
```

```
# Initialize stemmer and stop words
stemmer = PorterStemmer()
stop_words = set(stopwords.words('english'))
```

```
# Convert labels to binary values: ham = 0, spam = 1
df['label'] = df['label'].map({'ham': 0, 'spam': 1})
```

```
# Function to process text without using NLTK's word_tokenize
def process_text_manually(text):
    # Lowercase the text
    text = text.lower()

    # Tokenization: Split text into words (using split())
    tokens = text.split()

    # Remove punctuation and special characters, keep only alphabetic tokens
    tokens = [word for word in tokens if word.isalpha()]


    # Remove stopwords
    # Changed 'basic_stopwords' to 'stop_words'
    tokens = [word for word in tokens if word not in stop_words]
```


```
# Apply stemming
tokens = [stemmer.stem(word) for word in tokens]

# Join the tokens back into a single string
return ' '.join(tokens)


# Apply the new processing function to the 'message' column
df['cleaned_message'] = df['message'].apply(process_text_manually)


# Display cleaned messages and labels
df[['message', 'cleaned_message']].head()
```




	message	cleaned_message	
0	Go until jurong point, crazy.. Available only ...	go jurong avail bugi n great world la e cine g...	
1	Ok lar... Joking wif u oni...	ok joke wif u	
2	Free entry in 2 a wkly comp to win FA Cup fina...	free entri wkli comp win fa cup final tkt may ...	
3	U dun say so early hor... U c already then say...	u dun say earli u c already	
4	Nah I don't think he goes to usf, he lives ar...	nah think gae live around though	

```
df.sample(20)
```



	label	message	cleaned_message	
1547	0	Bishan lar nearer... No need buy so early cos ...	bishan lar need buy earli co buy gotta park	
817	0	Also are you bringing galileo or dobby	also bring galileo dobby	
2654	0	Its sarcasm.. nt scarcasim	scarcasim	
4451	0	I've told you everything will stop. Just dont ...	told everyth dont let get	
4938	0	Tomarrow i want to got to court. At &DECIM...	tomorrow want got come bu stand	
3002	0	I will see in half an hour	see half hour	
3509	0	I just saw ron burgundy captaining a party boa...	saw ron burgundi captain parti boat yeah	
650	0	Thats cool! Sometimes slow and gentle. Sonetim...	that sometim slow sonetim rough hard	
3484	0	No:-)i got rumour that you going to buy apartm...	got rumour go buy apart	
149	0	Sindu got job in birla soft ..	sindu got job birla soft	
2483	0	Mm have some kanji dont eat anything heavy ok	mm kanji dont eat anyth heavi ok	
3626	0	Still chance there. If you search hard you wil...	still chanc search hard get tri	
2849	0	She's fine. Good to hear from you. How are you...	good hear happi new year	
4977	0	JADE ITS PAUL. Y DIDN&Ouml;T U TXT ME? DO U REMEMB...	jade didn&Ouml;t u txt u rememb want talk txt	
4595	0	Carlos is down but I have to pick it up from h...	carlo pick swing usf littl bit	
3091	0	Dear, take care. I am just reaching home.love ...	take reach u	
3273	0	Thanx a lot...	thanx	
661	0	Then mum's repent how?	repent	
203	0	Goodmorning sleeping ga.	goodmorn sleep	
5440	0	Thank you. do you generallv date the brothas?	thank gener date	

```
# missing values
df.isnull().sum()
```



	0
label	0
message	0
cleaned_message	0

```
# checking for duplicate values
df.duplicated().sum()
```

```
np.int64(403)
```

```
# remove duplicates
df = df.drop_duplicates(keep='first')
```

```
df.duplicated().sum()
```

```
np.int64(0)
```

```
df.shape
```

```
(5169, 3)
```

## Text Vectorization

```
from sklearn.model_selection import train_test_split
```

```
# Split into train and test sets
X_train, X_test, y_train, y_test = train_test_split(
    df['cleaned_message'], df['label'], test_size=0.2, random_state=42)
```

```
# Show number of samples in each set
len(X_train), len(X_test)
```

```
(4135, 1034)
```

```
# Display first 3 samples from training and test sets
```

```
train_samples = pd.DataFrame({'message': X_train[:3].values, 'label': y_train[:3].values})
test_samples = pd.DataFrame({'message': X_test[:3].values, 'label': y_test[:3].values})
```

```
train_samples, test_samples
```

```
(
      message  label
0  exact intent      0
1             one      0
2  lololo ok next time      0,
      message  label
0             u download fring      0
1  pass di ur contact n see wat u luv wid put smi...      0
2                                     0)
```

```
from sklearn.feature_extraction.text import TfidfVectorizer
```

```
# Initialize TF-IDF Vectorizer
vectorizer = TfidfVectorizer(max_features=5000)
```

```
# Fit and transform training data, transform test data
X_train_vec = vectorizer.fit_transform(X_train)
X_test_vec = vectorizer.transform(X_test)
```

```
# Show the shape of transformed data
X_train_vec.shape, X_test_vec.shape
```

```
((4135, 4376), (1034, 4376))
```

## Exploratory Data Analysis

```
df.sample(10)
```

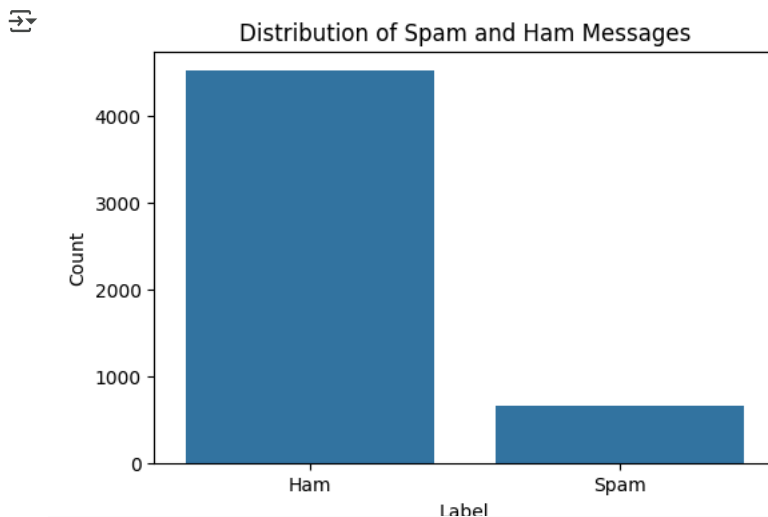
	label	message	cleaned_message
5474	0	Where's mummy's boy ? Is he being good or bad ...	boy good bad posit neg mummi made
3856	0	No! But we found a diff farm shop to buy some ...	found diff farm shop buy way back call
3609	0	Call me. I m unable to cal. Lets meet bhaskar,...	call unabl let meet deep
1925	0	We don call like &#x26; times oh. No give...	call like time give us hypertens
366	1	Update_Now - Xmas Offer! Latest Motorola, Sony...	xma latest sonyericsson nokia free doubl min t...
3516	0	So you think i should actually talk to him? No...	think actual talk call boss went place last ye...
2012	0	Beautiful Truth against Gravity.. Read careful...	beauti truth read heart feel light someone feel...
4048	0	Yeah that's the impression I got	yeah impress got
130	0	I'm home.	
1427	0	THING R GOOD THANX GOT EXAMS IN MARCH IVE DONE	thing r good thanx got exam march ive done fra

```
df['label'].value_counts()
```

	count
label	
0	4516
1	653

```
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# Plot distribution of spam vs ham messages
plt.figure(figsize=(6, 4))
sns.countplot(x='label', data=df)
plt.title('Distribution of Spam and Ham Messages')
plt.xlabel('Label')
plt.ylabel('Count')
plt.xticks([0, 1], ['Ham', 'Spam'])
plt.show()
```



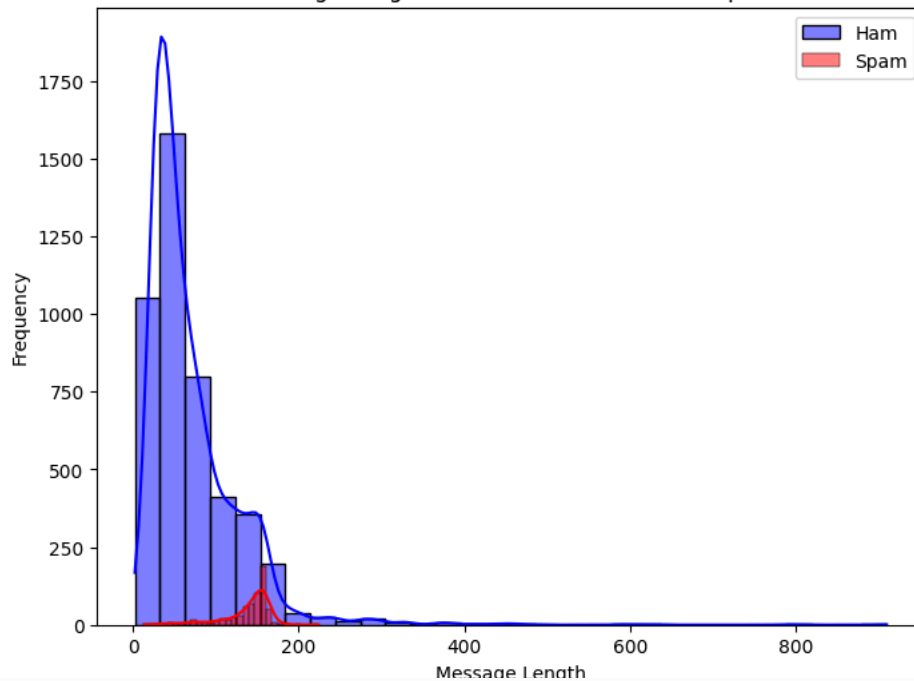
```
# Display the length of messages for spam vs ham
df['message_length'] = df['message'].apply(len)
```

```
plt.figure(figsize=(8, 6))
sns.histplot(df[df['label'] == 0]['message_length'], color='blue', label='Ham', kde=True, bins=30)
sns.histplot(df[df['label'] == 1]['message_length'], color='red', label='Spam', kde=True, bins=30)
plt.title('Message Length Distribution for Ham and Spam')
plt.xlabel('Message Length')
plt.ylabel('Frequency')
```

```
plt.legend()
plt.show()
```



Message Length Distribution for Ham and Spam



```
import nltk
```

```
!pip install nltk
```



Requirement already satisfied: nltk in /usr/local/lib/python3.11/dist-packages (3.9.1)  
 Requirement already satisfied: click in /usr/local/lib/python3.11/dist-packages (from nltk) (8.1.8)  
 Requirement already satisfied: joblib in /usr/local/lib/python3.11/dist-packages (from nltk) (1.4.2)  
 Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.11/dist-packages (from nltk) (2024.11.6)  
 Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-packages (from nltk) (4.67.1)

```
df.head()
```



	label	message	cleaned_message	message_length
0	0	Go until jurong point, crazy.. Available only ...	go jurong avail bugi n great world la e cine g...	111
1	0	Ok lar... Joking wif u oni...	ok joke wif u	29
2	1	Free entry in 2 a wkly comp to win FA Cup fina...	free entri wkli comp win fa cup final tkt may ...	155
3	0	U dun say so early hor... U c already then say...	u dun say earli u c already	49
4	0	Nah I don't think he goes to usf he lives aro...	nah think one live around though	61

Next steps:

[Generate code with df](#)
[View recommended plots](#)
[New interactive sheet](#)

```
# Add a new column for the number of words in each message
df['num_words'] = df['message'].apply(lambda x: len(x.split()))
```

```
# Display the first few rows with the new column
df[['message', 'num_words']].head()
```

	message	num_words	
0	Go until jurong point, crazy.. Available only ...	20	
1	Ok lar... Joking wif u oni...	6	
2	Free entry in 2 a wkly comp to win FA Cup fina...	28	
3	U dun say so early hor... U c already then say...	11	
4	Nah I don't think he goes to usf he lives aro	13	

```
df.sample(12)
```

	label	message	cleaned_message	message_length	num_words	
372	0	I'm going for bath will msg you next &lt;#&gt;...	go bath msg next	54	10	
4533	0	I have no money 4 steve mate! !	money steve	31	8	
5278	1	URGENT! Your Mobile number has been awarded wi...	mobil number award prize call land claim valid	144	23	
3177	0	K k :-)) then watch some films.	k k watch	33	7	
2497	0	Dai what this da.. Can i send my resume to thi...	dai send resum	51	12	
1731	0	Lol. Well quality aint bad at all so i aint co...	well qualiti aint bad aint complain	55	11	
3138	0	Is ur lecture over?	ur lectur	19	4	
3105	0	I had been hoping i would not have to send you...	hope would send rent due dont enough reserv co...	297	64	
5325	0	Gal n boy walking in d park. gal-can i hold ur...	gal n boy walk hold ur u think would run jst w...	160	33	
2816	1	Gr8 Poly tones 4 ALL mobs direct 2u rply with ...	poli tone mob direct rpli poli titl eg poli ymca	160	25	
687	0	Dear,Me at cherthala.in case u r coming cochin...	case u r come cochin pl call bfore u shall als...	169	30	
2538	0	The monthly amount is not that terrible and vo	monthli amount terribl nav anvth till finish	106	18	

```
df[['cleaned_message', 'message_length', 'num_words']].describe()
```

	message_length	num_words	
count	5169.000000	5169.000000	
mean	78.977945	15.340685	
std	58.236293	11.068488	
min	2.000000	1.000000	
25%	36.000000	7.000000	
50%	60.000000	12.000000	
75%	117.000000	22.000000	
max	910.000000	171.000000	

```
# ham
df[df['label'] == 0][['message_length', 'num_words']].describe()
```

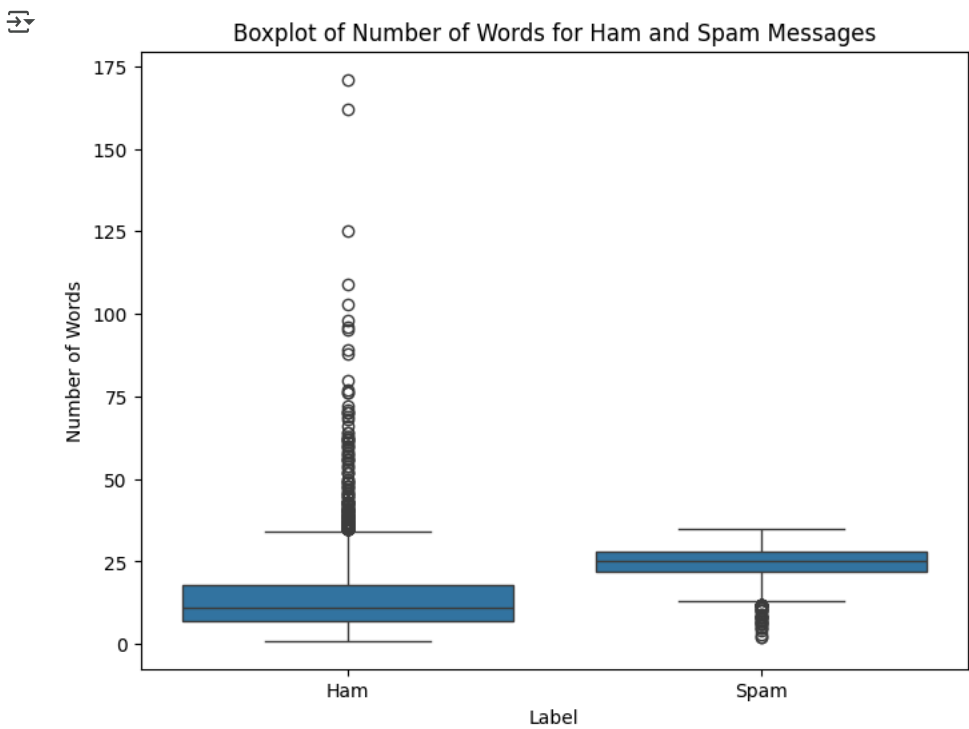
	message_length	num_words	
count	4516.000000	4516.000000	
mean	70.459256	14.134632	
std	56.358207	11.116240	
min	2.000000	1.000000	
25%	34.000000	7.000000	
50%	52.000000	11.000000	
75%	90.000000	18.000000	
max	910.000000	171.000000	

```
# spam
df[df['label'] == 1][['message_length', 'num_words']].describe()
```

	message_length	num_words	
count	653.000000	653.000000	
mean	137.891271	23.681470	
std	30.137753	5.967672	
min	13.000000	2.000000	
25%	132.000000	22.000000	
50%	149.000000	25.000000	
75%	157.000000	28.000000	
max	224 000000	35 000000	

```
# Check for outliers in num_words using a boxplot
plt.figure(figsize=(8, 6))
sns.boxplot(x='label', y='num_words', data=df)
plt.title('Boxplot of Number of Words for Ham and Spam Messages')
plt.xlabel('Label')
plt.ylabel('Number of Words')
plt.xticks([0, 1], ['Ham', 'Spam'])
plt.show()

# Show the top 5 messages with the most words
top_5_longest_messages = df[['message', 'num_words']].sort_values(by='num_words', ascending=False).head(5)
top_5_longest_messages
```



	message	num_words	
1084	For me the love should start with attraction.i...	171	
1862	The last thing i ever wanted to do was hurt yo...	162	
2157	Sad story of a Man - Last week was my b'day. M...	125	
2433	Indians r poor but India is not a poor country...	109	
1578	How to Make a girl Hannv? It's not at all diff	103	

Next steps:

Generate code with top\_5\_longest\_messages

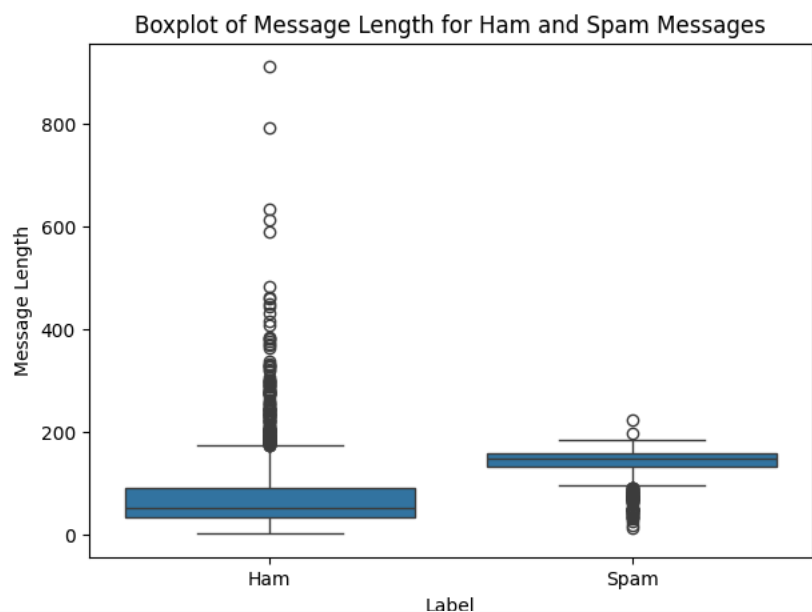
View recommended plots

New interactive sheet

```
#Plotting message_length
# Check for outliers in message length using a boxplot
```



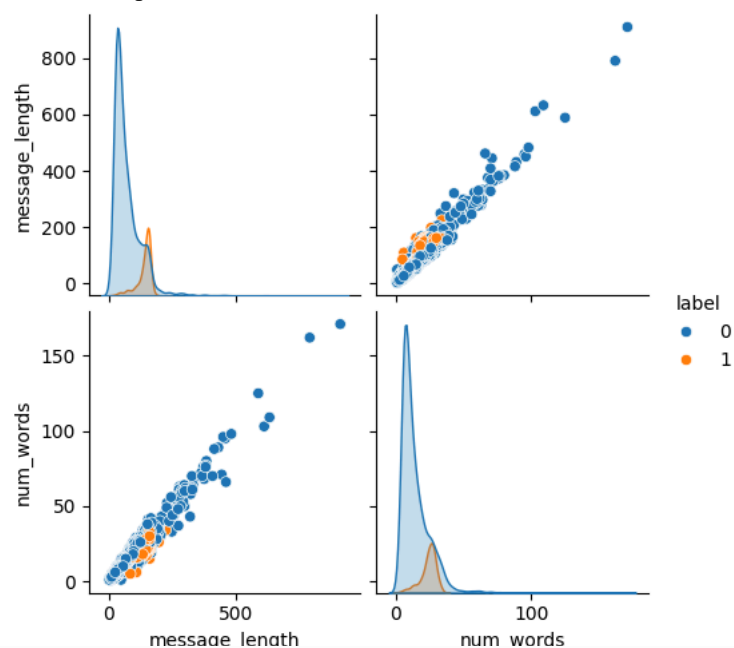
```
plt.figure(figsize=(7, 5))
sns.boxplot(x='label', y='message_length', data=df)
plt.title('Boxplot of Message Length for Ham and Spam Messages')
plt.xlabel('Label')
plt.ylabel('Message Length')
plt.xticks([0, 1], ['Ham', 'Spam'])
plt.show()
```



```
sns.pairplot(df, hue='label')
```

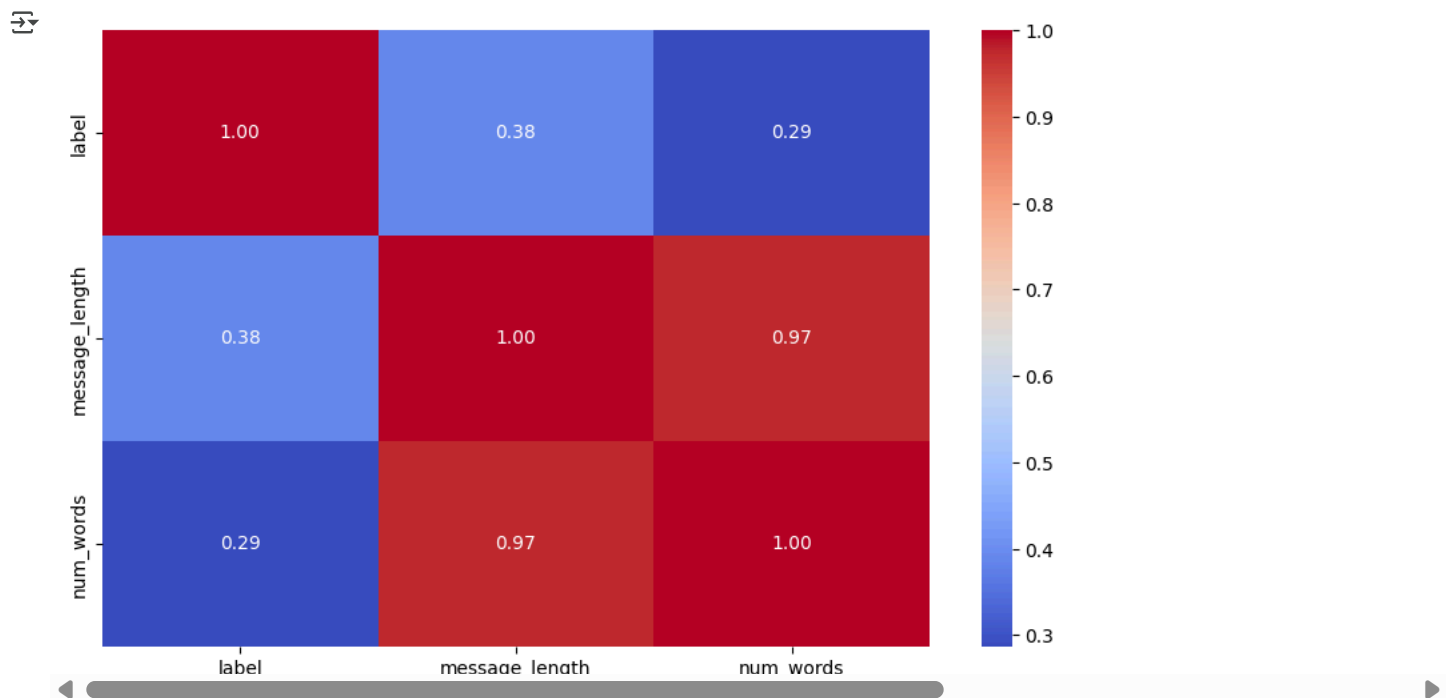


<seaborn.axisgrid.PairGrid at 0x7dcd4385b890>



```
# Select only numeric columns
numeric_df = df.select_dtypes(include=['number'])

# Compute and plot correlation matrix
plt.figure(figsize=(10,6))
sns.heatmap(numeric_df.corr(), annot=True, cmap='coolwarm', fmt='.2f')
plt.show()
```



```
# Display the number of rows and columns
```

```
shape = df.shape
```

```
# Display the column names
```

```
columns = df.columns
```

```
# Display the first 5 rows of the dataset
```

```
head = df.head()
```

```
shape, columns, head
```

```
((5169, 5),
Index(['label', 'message', 'cleaned_message', 'message_length', 'num_words'], dtype='object'),
label      message \
0      0  Go until jurong point, crazy.. Available only ...
1      0      Ok lar... Joking wif u oni...
2      1  Free entry in 2 a wkly comp to win FA Cup fina...
3      0  U dun say so early hor... U c already then say...
4      0  Nah I don't think he goes to usf, he lives aro...

      cleaned_message  message_length \
0  go jurong avail bugi n great world la e cine g...      111
1      ok joke wif u      29
2  free entri wkli comp win fa cup final tkt may ...      155
3      u dun say earli u c already      49
4      nah think goe live around though      61

num_words
0      20
1      6
2      28
3      11
4      13 )
```

```
current_shape = df.shape
```

```
current_shape
```

```
((5169, 5))
```

## Model Training

```
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.svm import SVC
from sklearn.ensemble import RandomForestClassifier
```

```
import xgboost as xgb
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
from sklearn.feature_extraction.text import TfidfVectorizer
```

```
df.sample(15)
```

	label	message	cleaned_message	message_length	num_words
3595	0	Good morning princess! Happy New Year!	good morn happi new	38	6
344	0	My sister cleared two round in birla soft yest...	sister clear two round birla soft	52	9
2672	0	Super msg da:)nalla timing.	super msg	27	4
4044	0	If You mean the website. Yes.	mean	29	6
865	1	Congratulations ur awarded either a yrs supply...	congratul ur award either yr suppli cd virgin ...	161	24
3274	0	Hurry home u big butt. Hang up on your last ca...	hurri home u big hang last caller u food done ask	120	26
5549	0	You know, wot people wear. T shirts, jumpers, ...	wot peopl r cribb	87	18
3167	0	Ugh hopefully the asus ppl dont randomly do a ...	ugh hope asu ppl dont randomli	55	10
5411	0	I ask if u meeting da ge tmr nite...	ask u meet da ge tmr	36	9
4548	0	Haha, my friend tyler literally just asked if ...	friend tyler liter ask could get dubsack	73	14
3140	0	Customer place i will call you	custom place call	30	6
4914	0	We took hooch for a walk toaday and i fell ove...	took hooch walk toaday fell graze knee stay see	134	25
2436	0	Btw regarding that we should really try to see...	btw regard realli tri see anyon els guy commit...	114	24
1265	0	Im in inperialmusic listening2the weirdest tra...	im inperialmus weirdest track ever byãóleafcut...	156	19
302	0	Oh and by the way you do have more food in you	oh wav food want no meal	91	21

```
df.head()
```

	label	message	cleaned_message	message_length	num_words
0	0	Go until jurong point, crazy.. Available only ...	go jurong avail bugi n great world la e cine g...	111	20
1	0	Ok lar... Joking wif u oni...	ok joke wif u	29	6
2	1	Free entry in 2 a wkly comp to win FA Cup fina...	free entri wkli comp win fa cup final tkt may ...	155	28
3	0	U dun say so early hor... U c already then say...	u dun say earli u c already	49	11
4	0	Nah I don't think he goes to usf he lives aro	nah think one live around though	61	13

Next steps: [Generate code with df](#) [View recommended plots](#) [New interactive sheet](#)

```
# Split the data into features and target
X = df['cleaned_message']
y = df['label']
```

```
# Vectorize the data
vectorizer = TfidfVectorizer(max_features=5000)
X_vec = vectorizer.fit_transform(X)
```

```
# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X_vec, y, test_size=0.2, random_state=42)
```

```
# Initialize the models
svm_model = SVC()
rf_model = RandomForestClassifier(n_estimators=100)
xgb_model = xgb.XGBClassifier()
```

```
# Train the models
svm_model.fit(X_train, y_train)
rf_model.fit(X_train, y_train)
xgb_model.fit(X_train, y_train)
```

XGBClassifier ⓘ

```
# Make predictions
svm_pred = svm_model.predict(X_test)
rf_pred = rf_model.predict(X_test)
xgb_pred = xgb_model.predict(X_test)

# Evaluate the models
svm_accuracy = accuracy_score(y_test, svm_pred)
rf_accuracy = accuracy_score(y_test, rf_pred)
xgb_accuracy = accuracy_score(y_test, xgb_pred)

svm_report = classification_report(y_test, svm_pred)
rf_report = classification_report(y_test, rf_pred)
xgb_report = classification_report(y_test, xgb_pred)

svm_conf_matrix = confusion_matrix(y_test, svm_pred)
rf_conf_matrix = confusion_matrix(y_test, rf_pred)
xgb_conf_matrix = confusion_matrix(y_test, xgb_pred)

# Cross-validation scores
svm_cv_scores = cross_val_score(svm_model, X_vec, y, cv=5)
rf_cv_scores = cross_val_score(rf_model, X_vec, y, cv=5)
xgb_cv_scores = cross_val_score(xgb_model, X_vec, y, cv=5)

svm_accuracy, rf_accuracy, xgb_accuracy, svm_report, rf_report, xgb_report, svm_conf_matrix, rf_conf_matrix, xgb_conf_matrix, svm_cv_scores.
```

```
(0.9661508704061895,
0.9690522243713733,
0.9613152804642167,
',
precision recall f1-score support\n\n
0.97 0.78 0.87 145\n\n accuracy
0.92 1034\nweighted avg 0.97 0.97 0.96 1034\n',
',
precision recall f1-score support\n\n
0.99 0.79 0.88 145\n\n accuracy
0.93 1034\nweighted avg 0.97 0.97 0.97 1034\n',
',
precision recall f1-score support\n\n
0.91 0.81 0.85 145\n\n accuracy
0.92 1034\nweighted avg 0.96 0.96 0.96 1034\n',
array([[886, 3],
[ 32, 113]]),
array([[888, 1],
[ 31, 114]]),
array([[877, 12],
[ 28, 117]]),
np.float64(0.9651757009030802),
np.float64(0.9653702479679287),
np.float64(0.9620812978292742))
```

```
# Gather results
results = {
    "svm_accuracy": svm_accuracy,
    "rf_accuracy": rf_accuracy,
    "xgb_accuracy": xgb_accuracy,
    "svm_report": svm_report,
    "rf_report": rf_report,
    "xgb_report": xgb_report,
    "svm_conf_matrix": svm_conf_matrix,
    "rf_conf_matrix": rf_conf_matrix,
    "xgb_conf_matrix": xgb_conf_matrix,
    "svm_cv_mean": svm_cv_scores.mean(),
    "rf_cv_mean": rf_cv_scores.mean(),
    "xgb_cv_mean": xgb_cv_scores.mean()
}
```

results

```
{'svm_accuracy': 0.9661508704061895,
'rf_accuracy': 0.9690522243713733,
```

```

'xgb_accuracy': 0.9613152804642167,
'svm_report': '
precision recall f1-score support\n\n
1 0.97 0.78 0.87 145\n\n accuracy
0.92 1034\nweighted avg 0.97 0.97 0.96 1034\n',
'rf_report': '
precision recall f1-score support\n\n
1 0.99 0.79 0.88 145\n\n accuracy
0.93 1034\nweighted avg 0.97 0.97 0.97 1034\n',
'xgb_report': '
precision recall f1-score support\n\n
1 0.91 0.81 0.85 145\n\n accuracy
0.92 1034\nweighted avg 0.96 0.96 0.96 1034\n',
'svm_conf_matrix': array([[886, 3],
[ 32, 113]]),
'rf_conf_matrix': array([[888, 1],
[ 31, 114]]),

```

	0	0.97	1.00	0.98	889\n	
0.97	1034\n	macro avg	0.97	0.89		

	0	0.97	1.00	0.98	889\n	
0.97	1034\n	macro avg	0.98	0.89		

	0	0.97	0.99	0.98	889\n	
0.96	1034\n	macro avg	0.94	0.90		