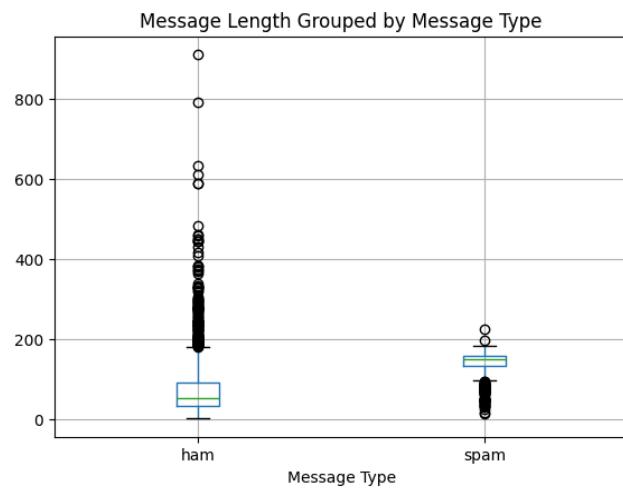
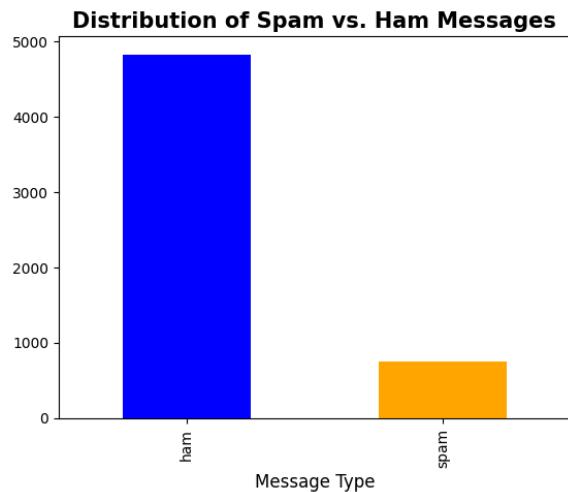


Data Dictionary

Column	Description	Potential Responses
Classification	States whether message is spam or ham	Spam, ham
Message	The text of the message	Oh k...i'm watching here:)

EDA 1



Comparison of Common Spam Words

Word/Phrase	Spam %	Ham %
Money	0.40%	1.10%
Win	13.40%	1.36%
Free	26.60%	1.40%
Immediately	0.60%	0.10%
Click here	0.10%	0%
Urgent	8.30%	0.10%
Password	0.40%	0.10%
Cash	9.10%	0.30%
Selected	3.30%	0.10%

Prize	12%	0%
-------	-----	----

EDA 2

Spam Most common Words (minus common stop words)

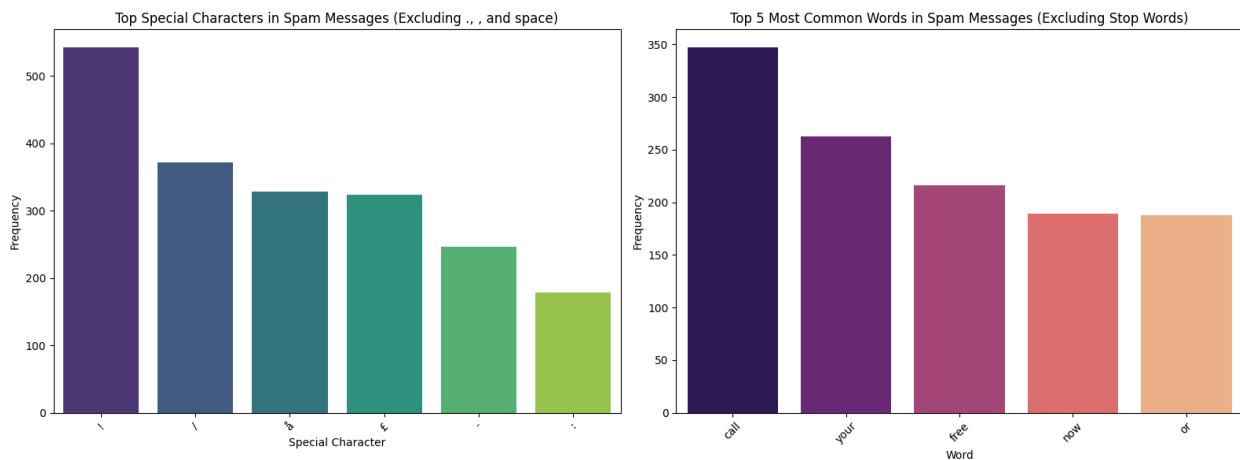
Word/Phrase	Occurrence
call	347
your	263
free	216
now	189
or	188

Ham Most common Words (minus common stop words)

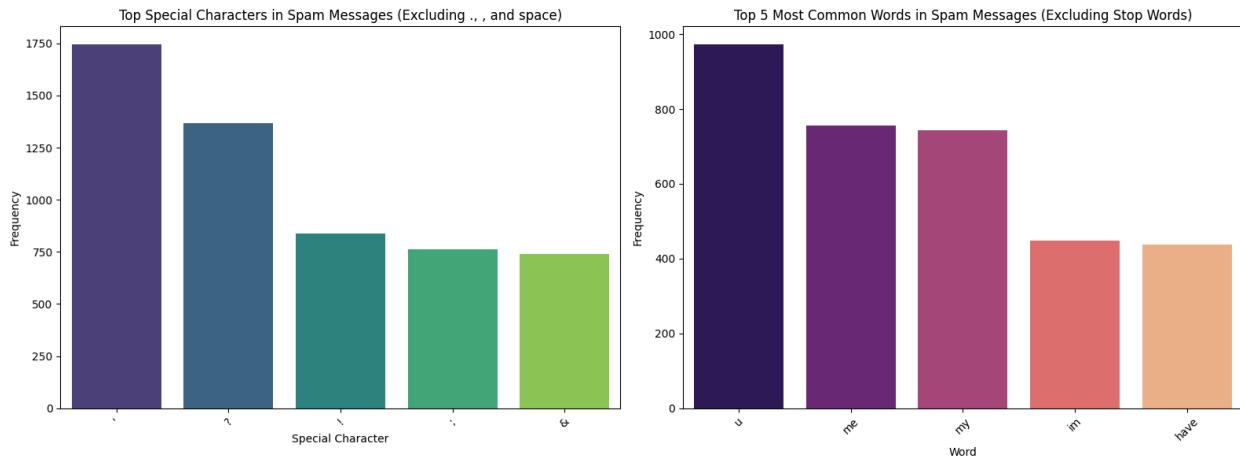
Word/Phrase	Occurrence
u	972
me	756
my	743
im	449
have	439

Average Special Characters per Message	
Spam	6.61
Ham	4.08

Top Special Characters and Words in Spam



Top Special Characters and Words in Ham



Capitalization

