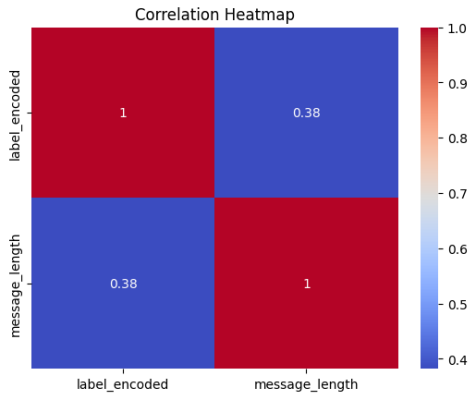
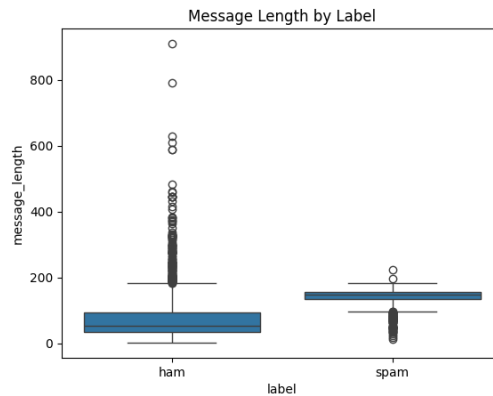


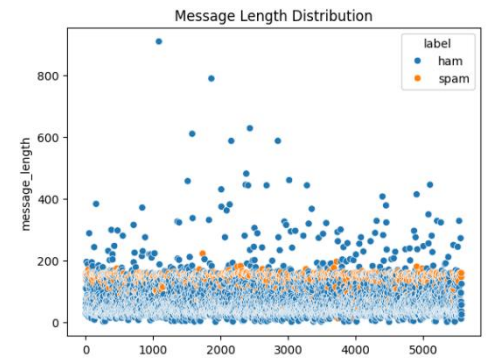
## Implementation Output:



Correlation heatmap



Boxplot of Message Length by Label



Scatter Plot of Message Lengths

## Load Dataset:


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

## Decision Tree:

Decision Tree Accuracy: 0.967713004484305

	precision	recall	f1-score	support
0	0.98	0.98	0.98	966
1	0.88	0.87	0.88	149
accuracy			0.97	1115
macro avg	0.93	0.93	0.93	1115
weighted avg	0.97	0.97	0.97	1115


## K-Nearest Neighbours:



```
KNN Accuracy: 0.9327354260089686
[[955 11]
 [ 64 85]]
```

	precision	recall	f1-score	support
0	0.94	0.99	0.96	966
1	0.89	0.57	0.69	149
accuracy			0.93	1115
macro avg	0.91	0.78	0.83	1115
weighted avg	0.93	0.93	0.93	1115

Compare the performance of DT and KNN using accuracy and F1 Score:



	Model	Accuracy	F1 Score
0	Decision Tree	0.967713	0.878378
1	KNN	0.932735	0.693878

