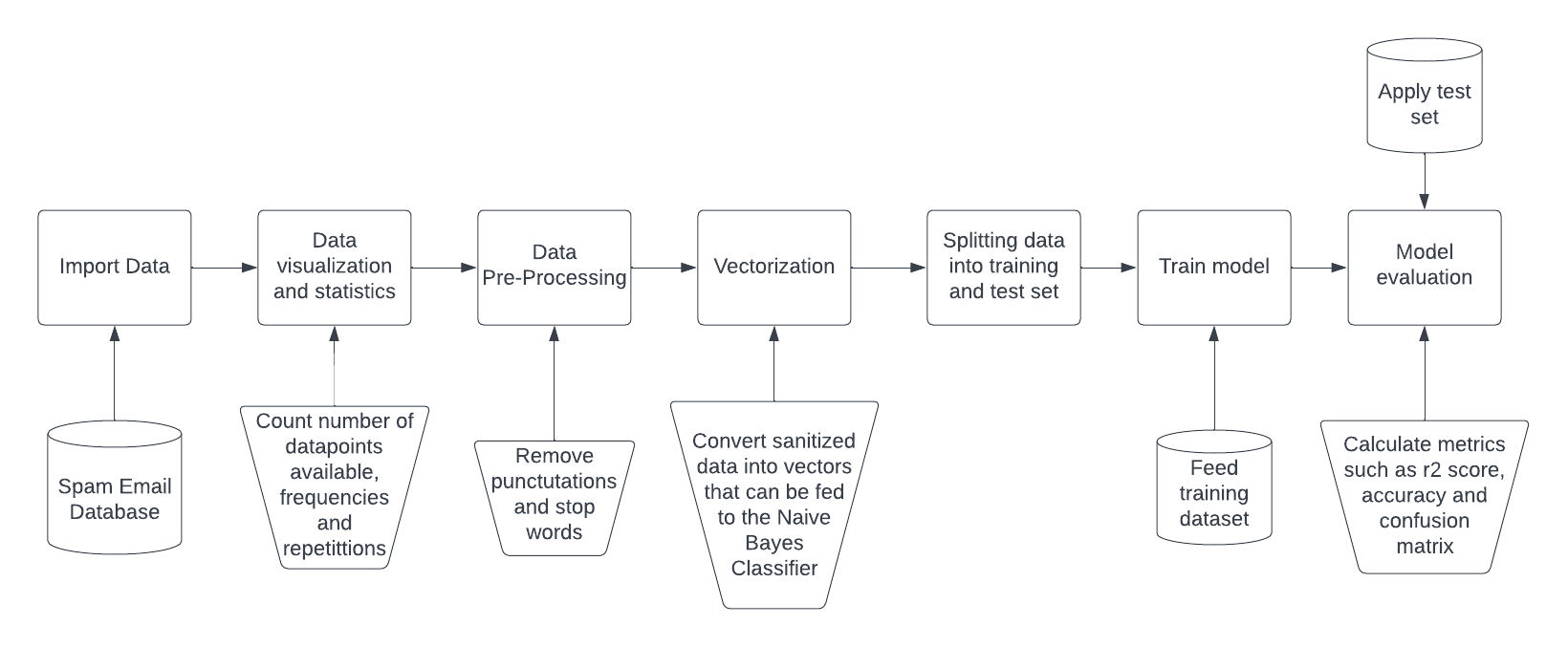
**BLOCK DIAGRAM:**

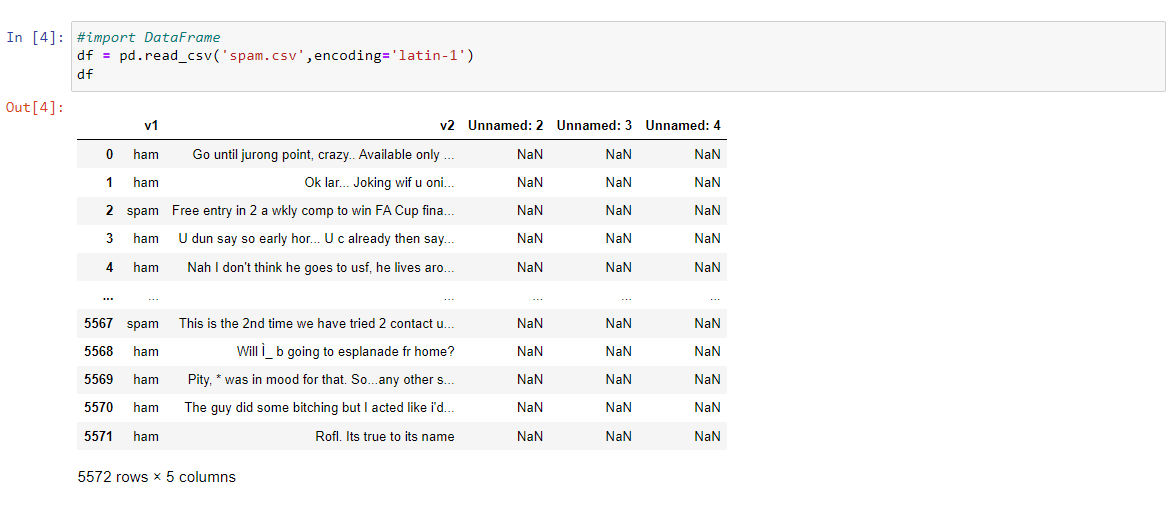


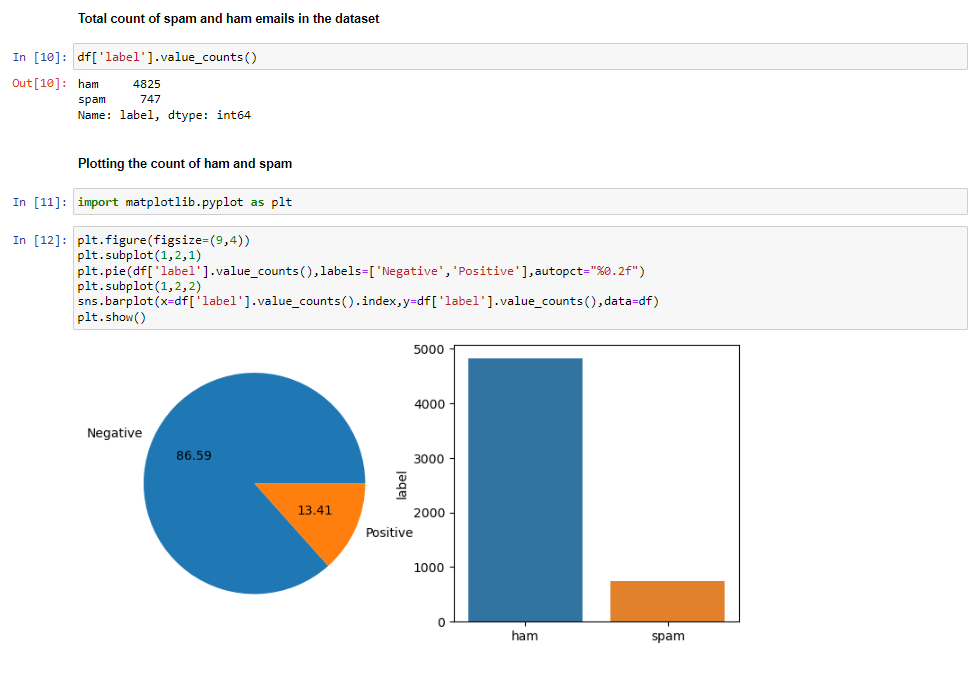
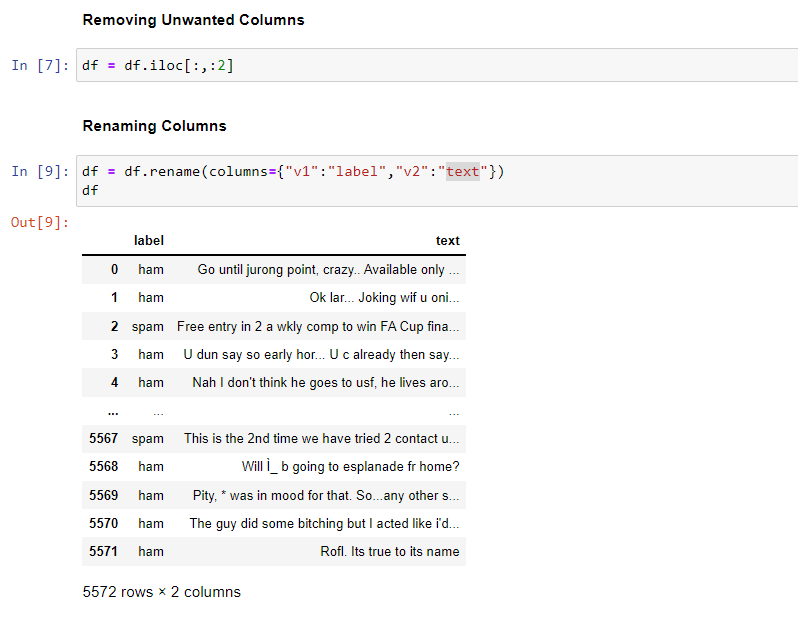
**SCREENSHOTS:**

Importing the libraries



Performing Exploratory Data Analysis (EDA)

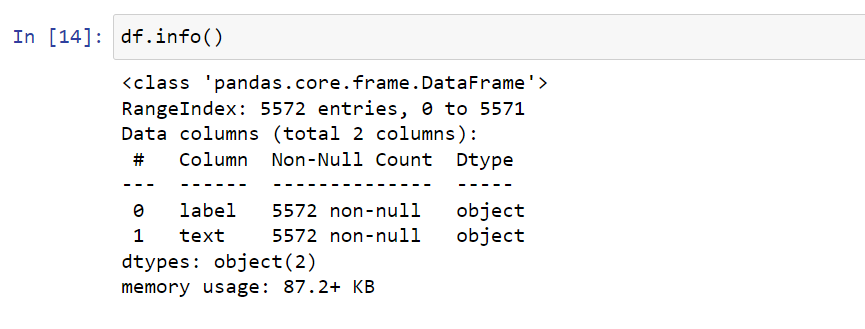




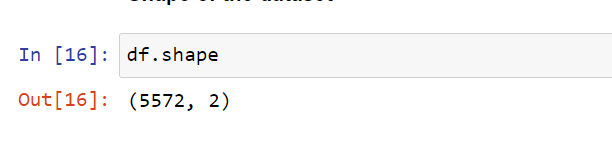
Conclusion:

We can infer that the dataset is imbalanced. Thus, we first process the dataset

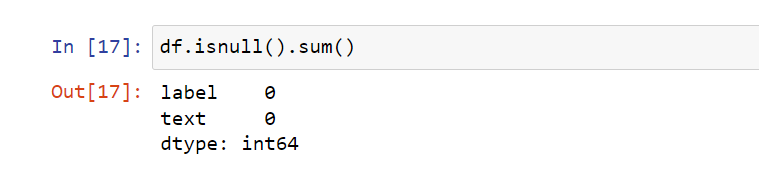
More information about the dataset



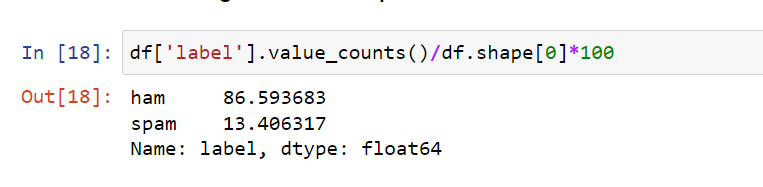
Shape of the Dataset



Checking for Null Values



Percentage of ham and spam in dataset with non-null values

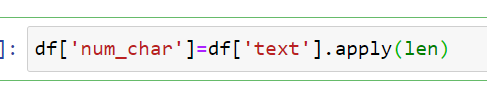


**Feature Engineering:**

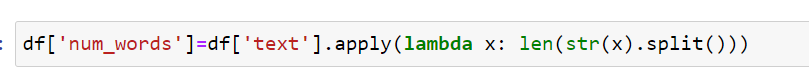
We extract certain important information from the dataset such as

1. Total No. of Characters
2. Total No. of Words
3. Total No. of Sentences

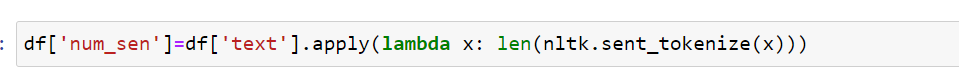
Total No. of Characters

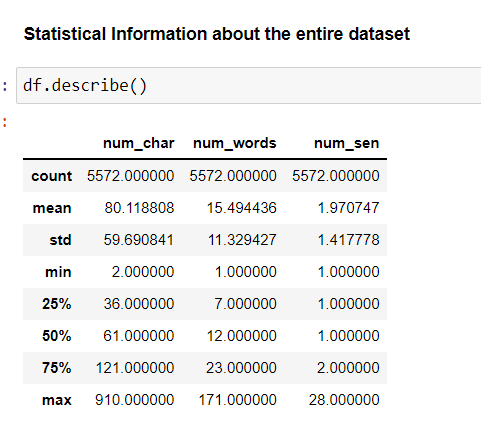


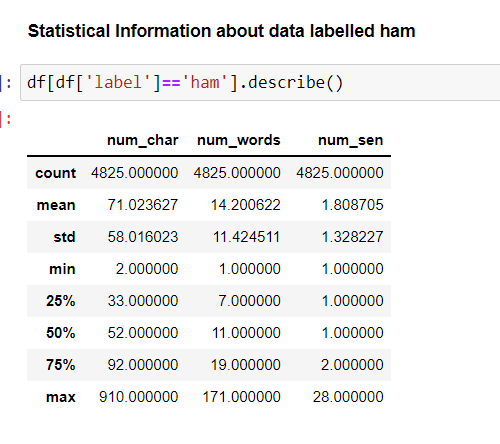
Total Number of Words

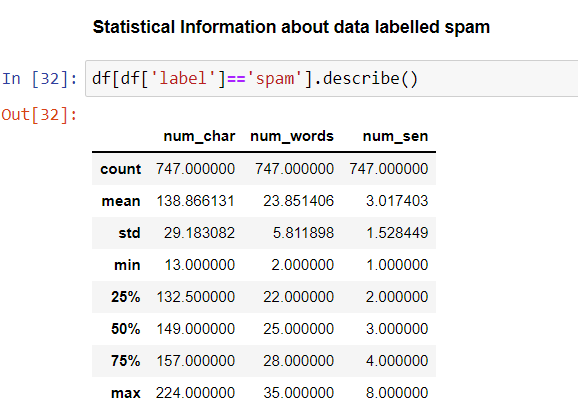


Total Number of Sentences



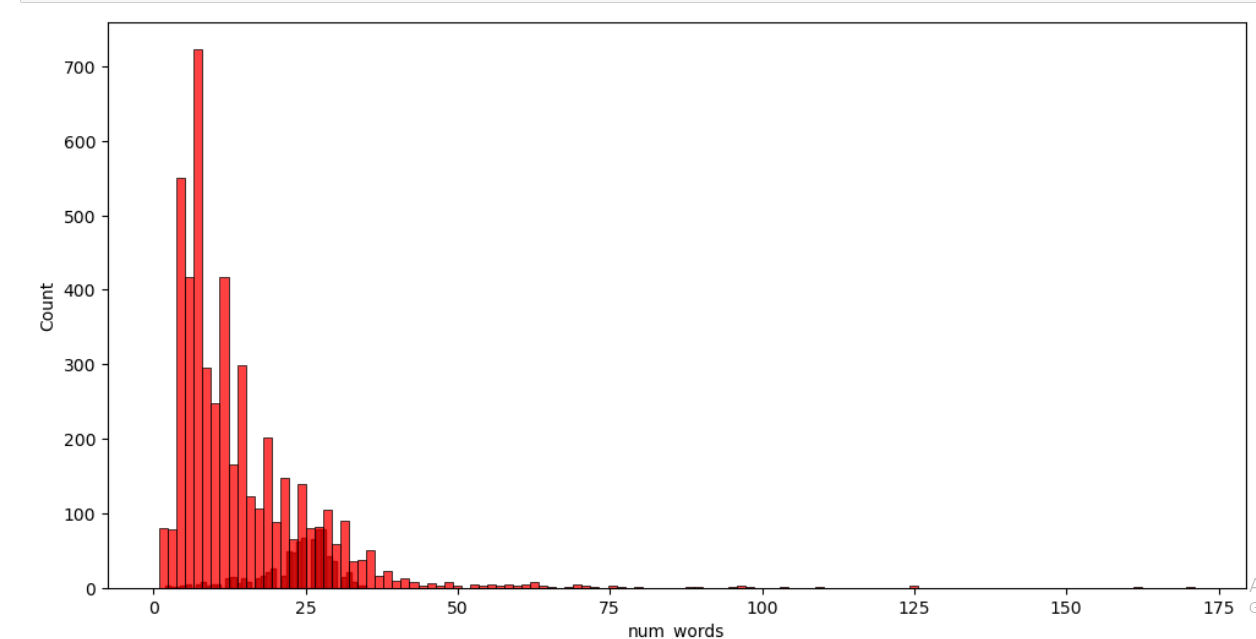




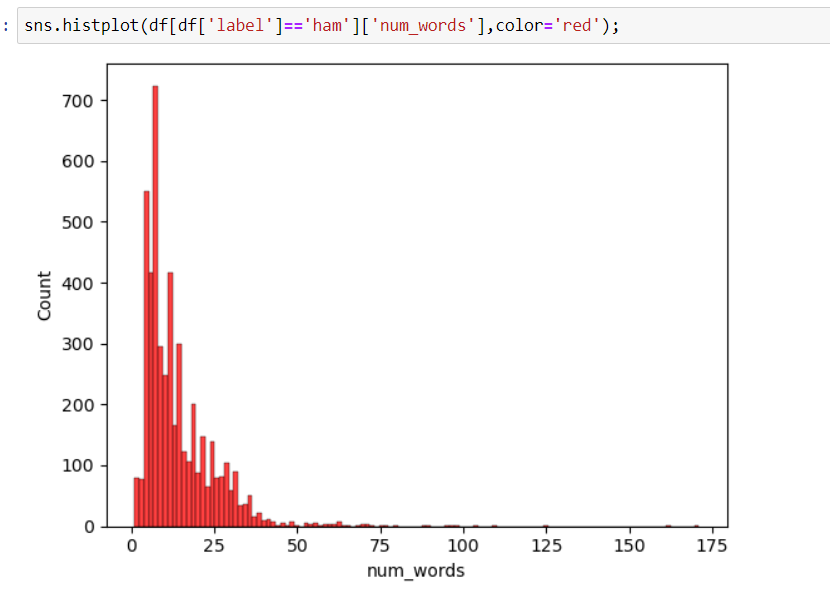


Histogram Plot For Spam and Ham

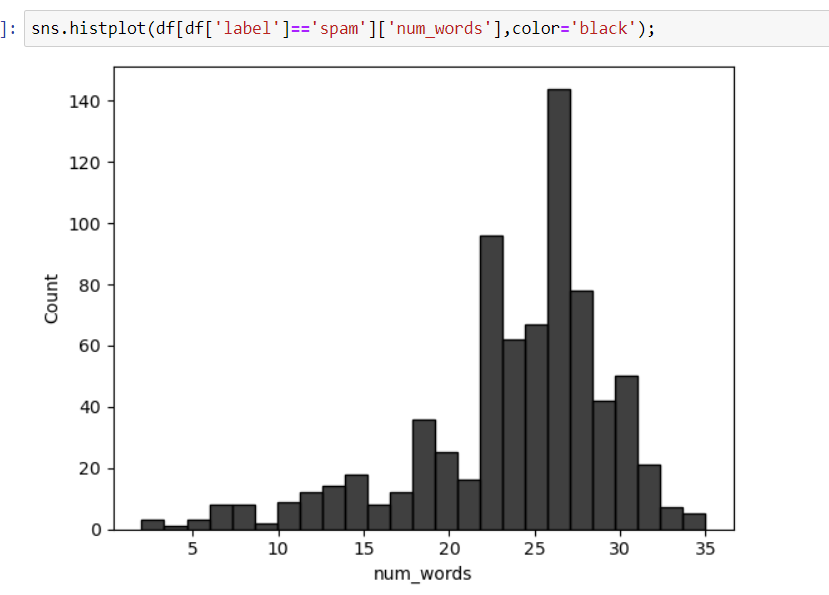




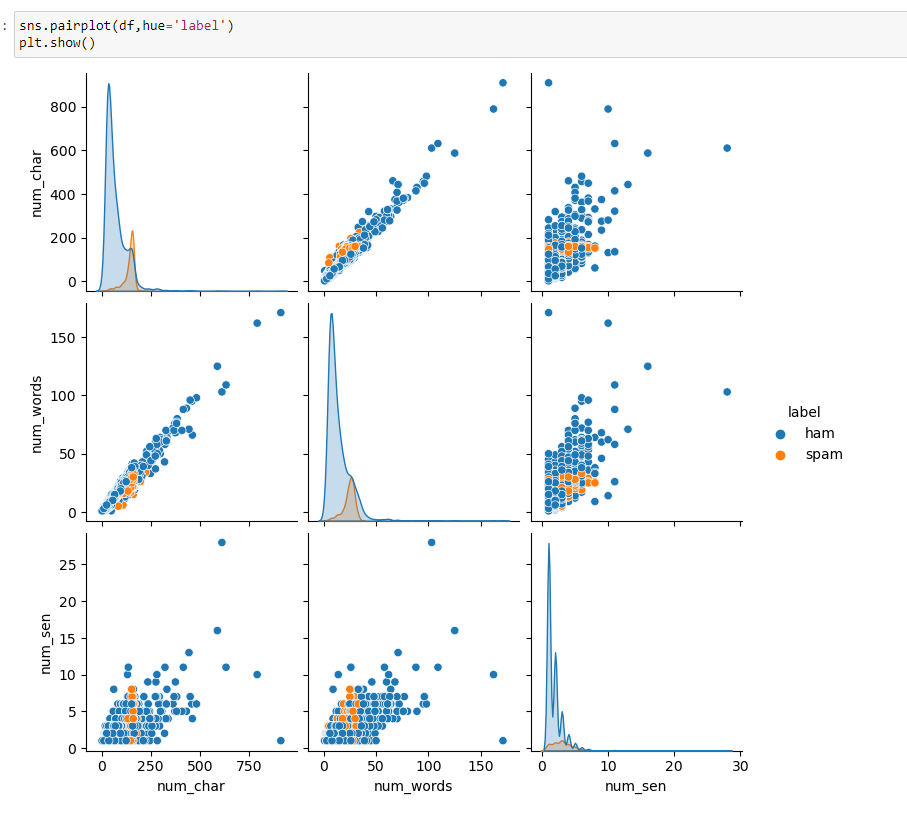
Histogram Plot for Ham



Histogram Plot for Spam



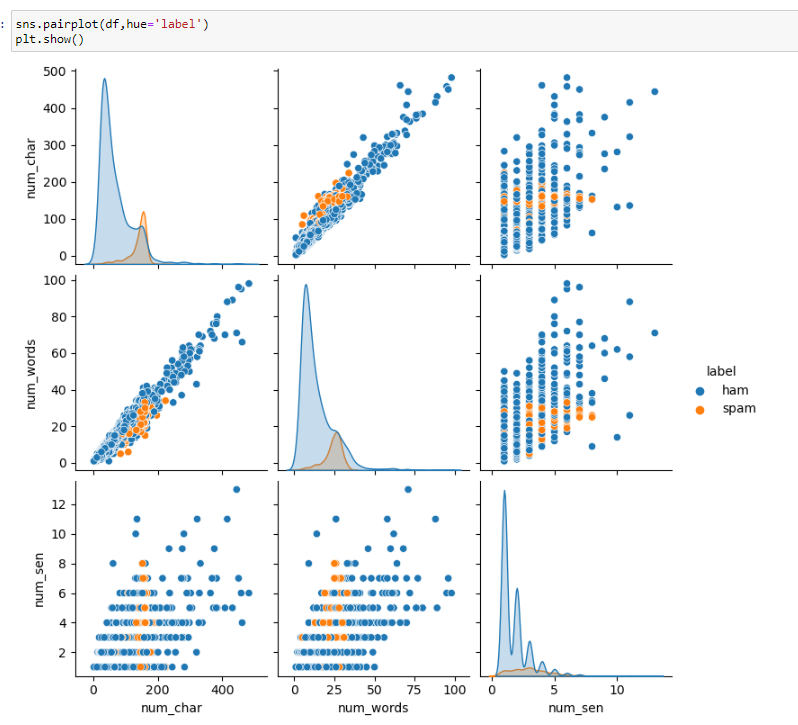
Pair Plot For Better Visualization



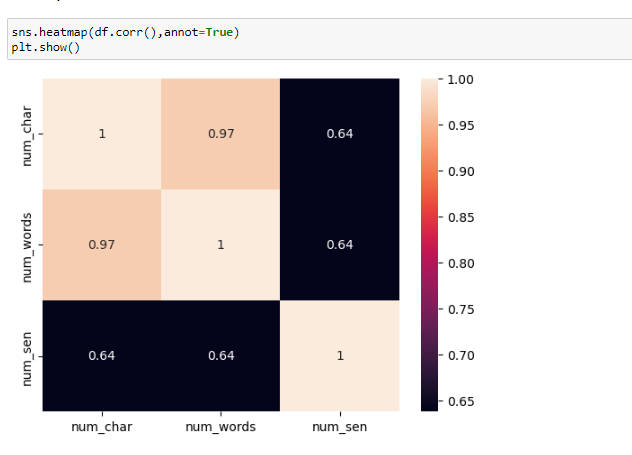
Handling Outliers



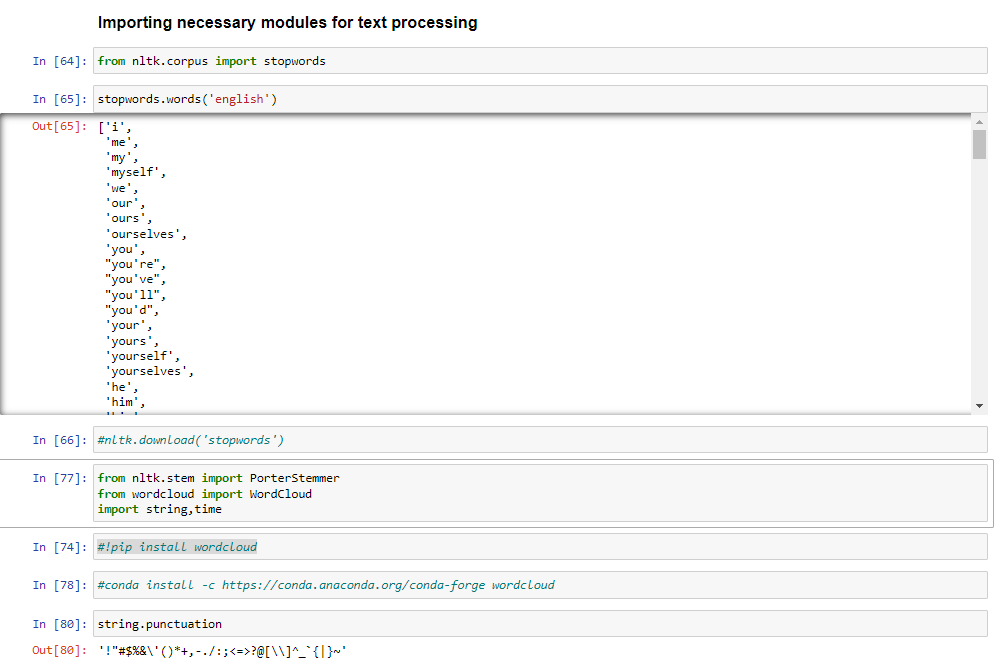
Pair Plot After Removing Emails with over 500 characters

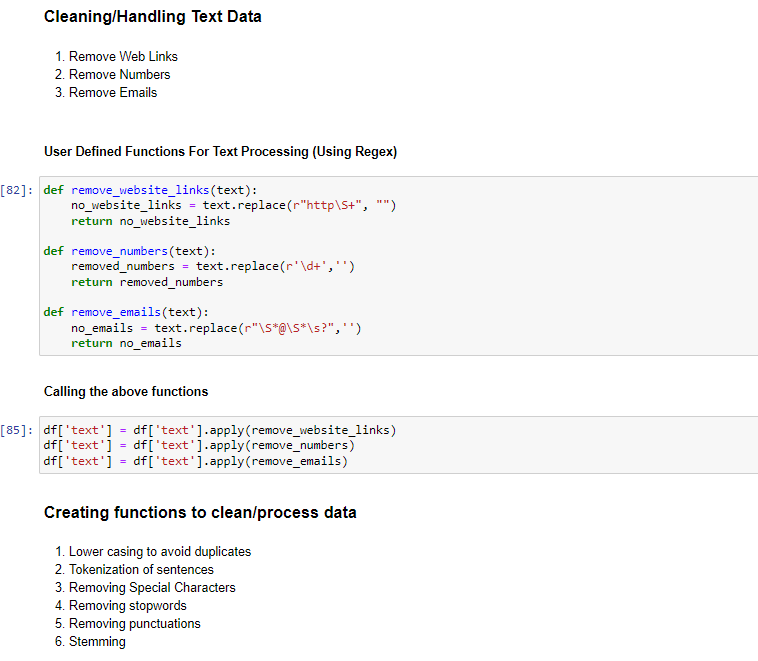


Heat Map



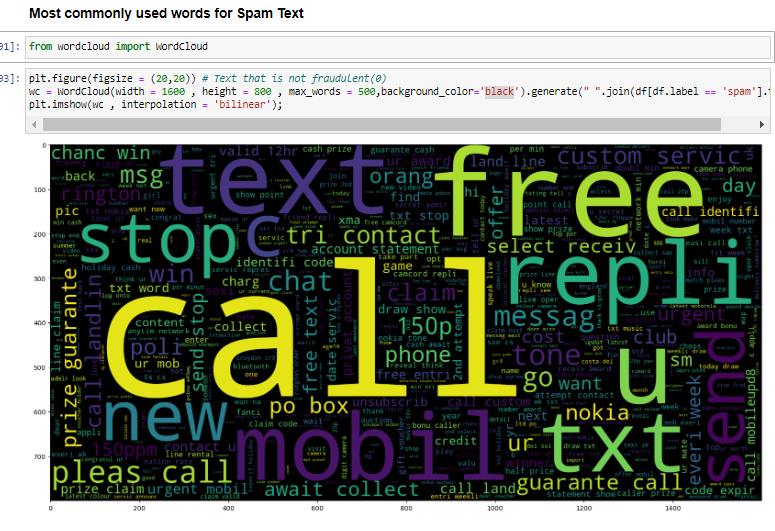
Data Pre-processing

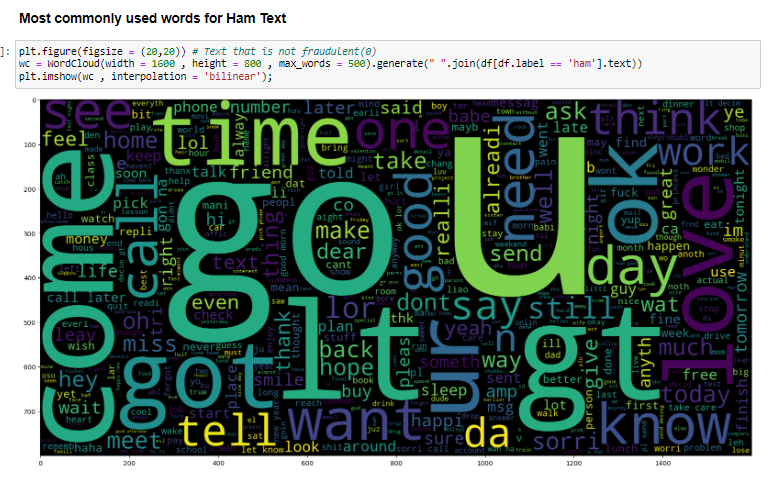




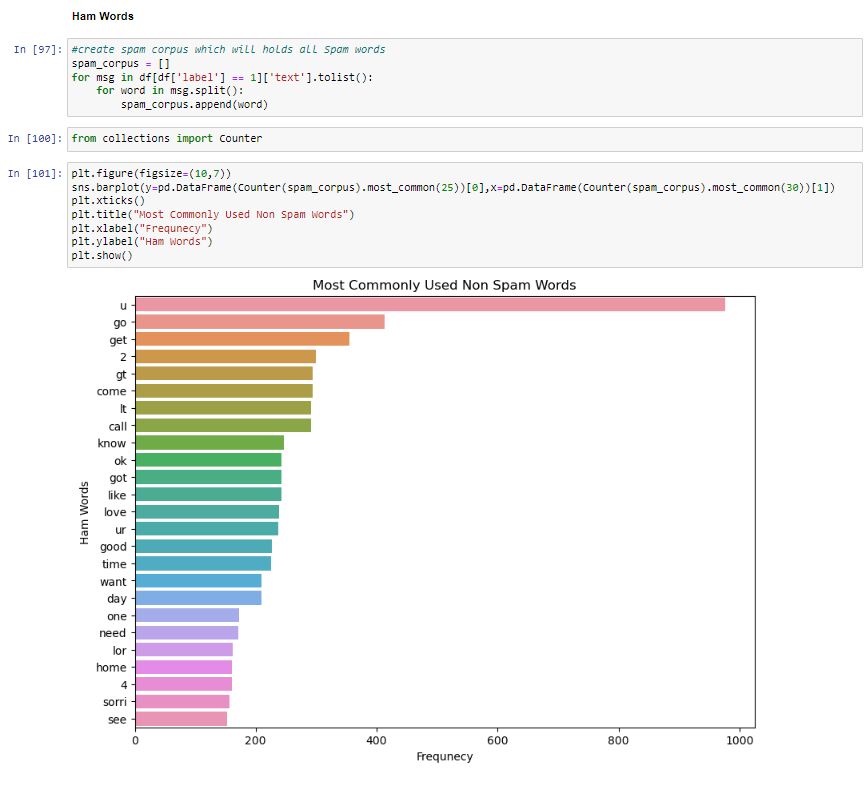


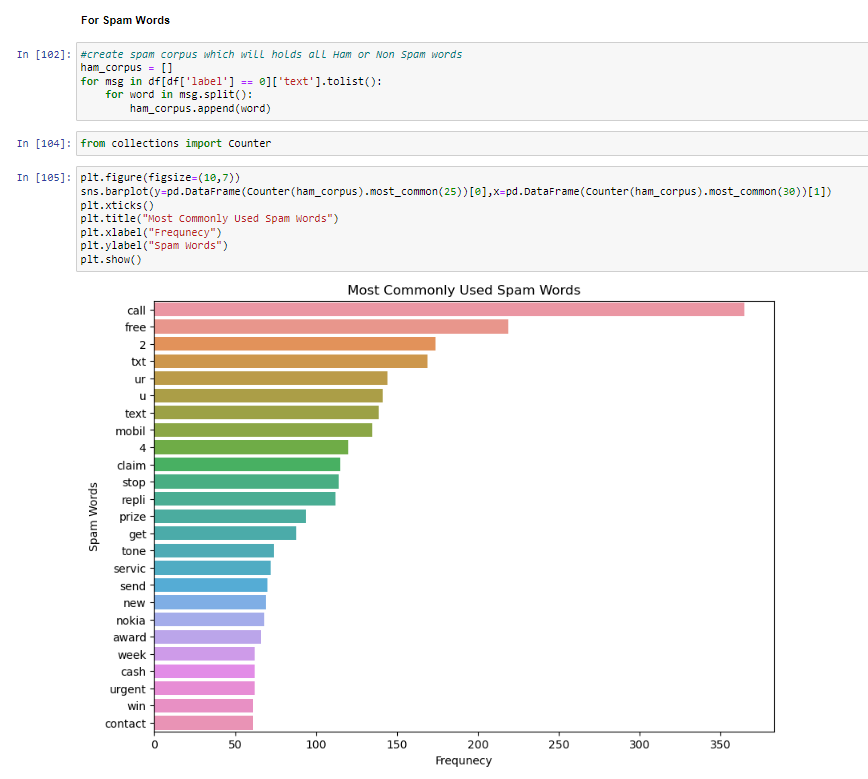
Data Visualization



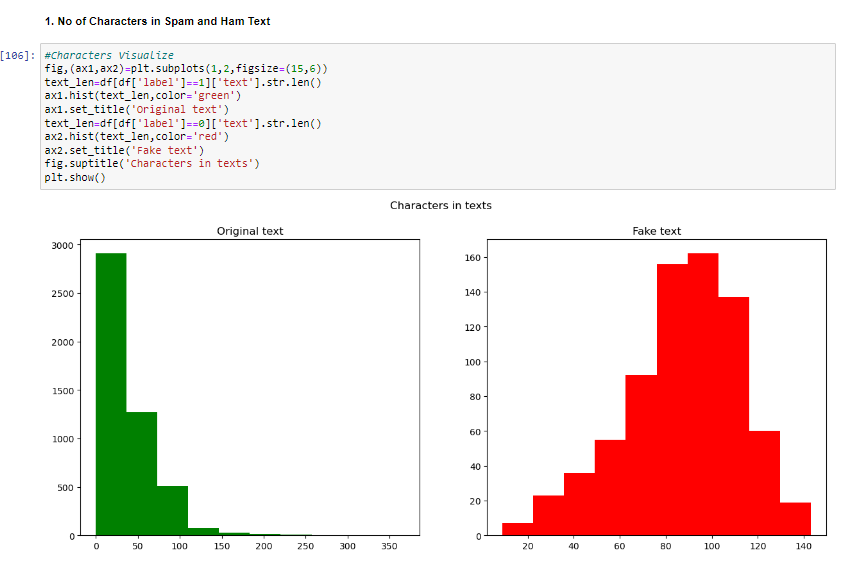


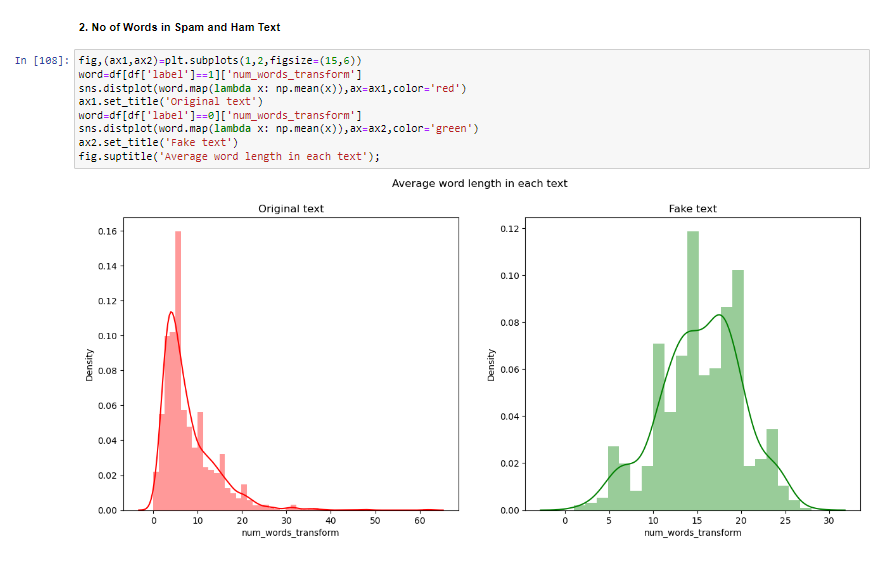






Comparing Total number of characters in ham and spam text





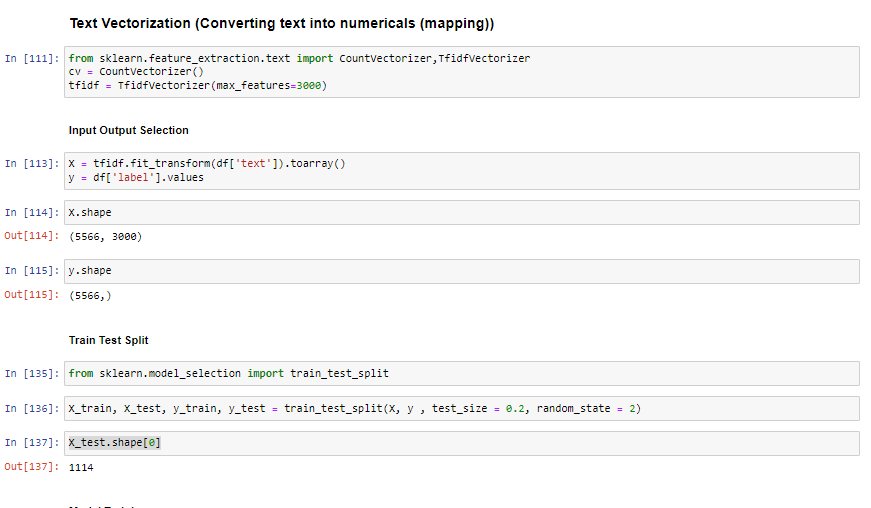
Conclusion

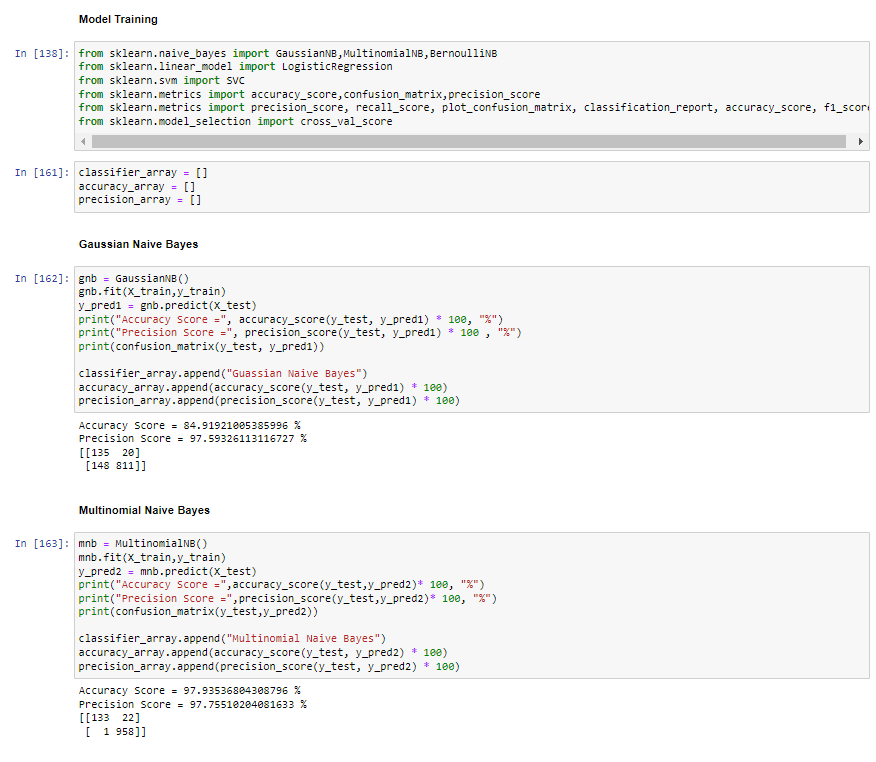
- After visualization, we can conclude that spam text has more words and characters as compare to Ham text

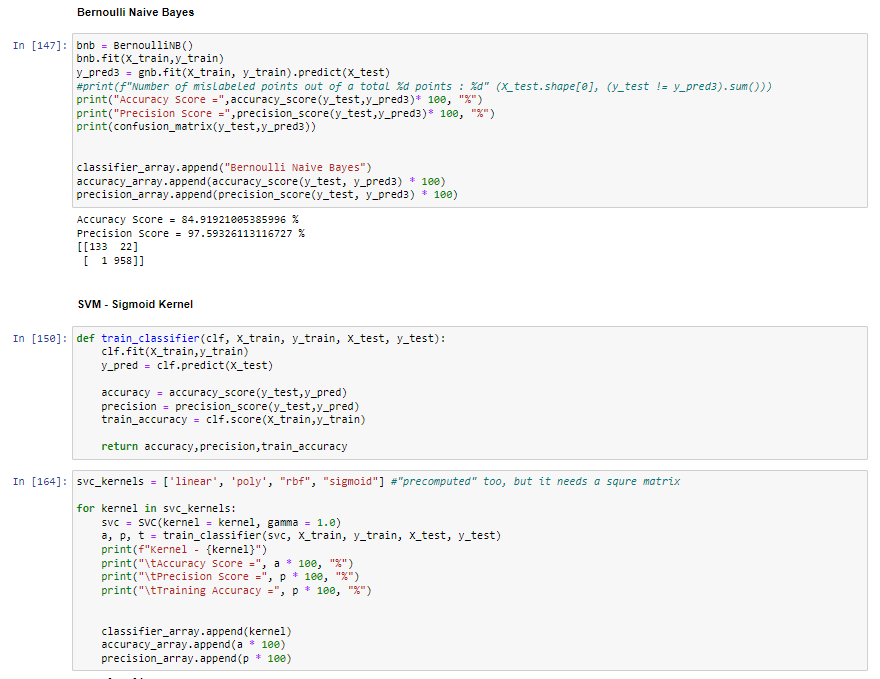
- Average characters includes in spam message is around 90 characters

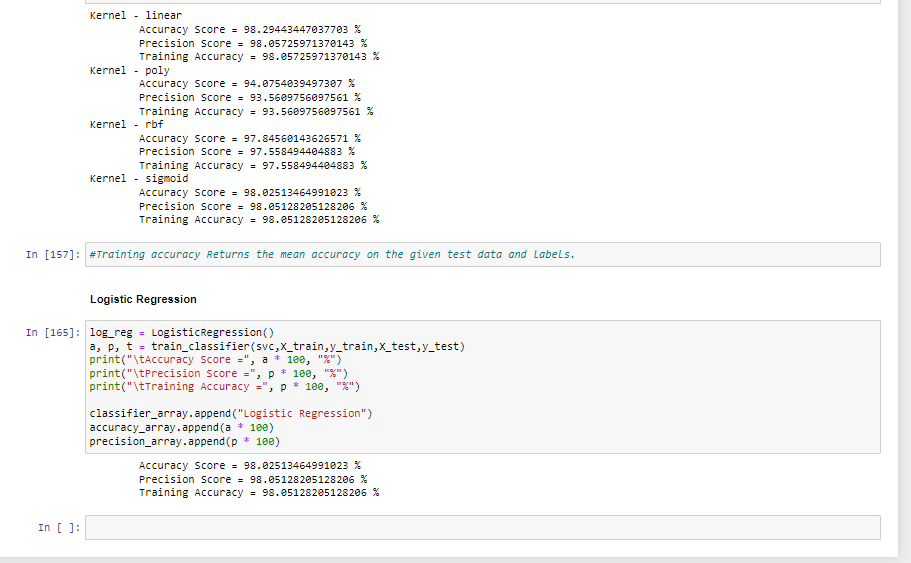
- Average words includes in spam message is around 15 word

**Model Building**

****

****

****

****