```python
In [1]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn as sns
        %matplotlib inline

        import warnings
        warnings.filterwarnings('ignore')
```

```python
In [2]: # Reading Data

        df = pd.read_csv('spam.csv',encoding='ISO-8859-1')
        df.head()
```

Out[2]:

	v1	v2	Unnamed: 2	Unnamed: 3	Unnamed: 4
0	ham	Go until jurong point, crazy.. Available only ...	NaN	NaN	NaN
1	ham	Ok lar... Joking wif u oni...	NaN	NaN	NaN
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...	NaN	NaN	NaN
3	ham	U dun say so early hor... U c already then say...	NaN	NaN	NaN
4	ham	Nah I don't think he goes to usf, he lives aro...	NaN	NaN	NaN

```python
In [3]: df.shape
```

Out[3]: (5572, 5)

```python
In [4]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5572 entries, 0 to 5571
Data columns (total 5 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   v1          5572 non-null   object
 1   v2          5572 non-null   object
 2   Unnamed: 2  50 non-null     object
 3   Unnamed: 3  12 non-null     object
 4   Unnamed: 4  6 non-null      object
dtypes: object(5)
memory usage: 217.8+ KB
```

```python
In [5]: df.drop(columns=['Unnamed: 2','Unnamed: 3','Unnamed: 4'],inplace=True)
```

```python
In [6]: df.head()
```

Out[6]:

	v1	v2
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...

```python
In [7]: #renaming the columns
        df.rename(columns={'v1':'Target','v2':'Text'},inplace=True)
        df.head()
```

Out[7]:

	Target	Text
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...

```python
In [8]: #checking null values

        df.isna().sum()
```

```
Out[8]:  Target    0
         Text      0
         dtype: int64
```

```
In [9]:  #checking duplicate values

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

```
Out[9]:  403
```

```
In [10]:  df.shape
```

```
Out[10]:  (5572, 2)
```

```
In [11]:  df=df.drop_duplicates(keep='first')
```

```
In [12]:  df.duplicated().sum()
```

```
Out[12]:  0
```

```
In [13]:  df['Target'].value_counts()
```

```
Out[13]:  Target
          ham     4516
          spam     653
          Name: count, dtype: int64
```

```
In [14]:  from sklearn.preprocessing import LabelEncoder
```

```
In [15]:  le = LabelEncoder()
```

```
In [16]:  df['Target'] = le.fit_transform(df['Target'])
```

```
In [17]:  df.head
```

```
Out[17]:  <bound method NDFrame.head of        Target                                      Text
          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...
          ...       ...                                                ...
          5567        1  This is the 2nd time we have tried 2 contact u...
          5568        0              Will Ì_ b going to esplanade fr home?
          5569        0  Pity, * was in mood for that. So...any other s...
          5570        0  The guy did some bitching but I acted like i'd...
          5571        0                          Rofl. Its true to its name

          [5169 rows x 2 columns]>
```
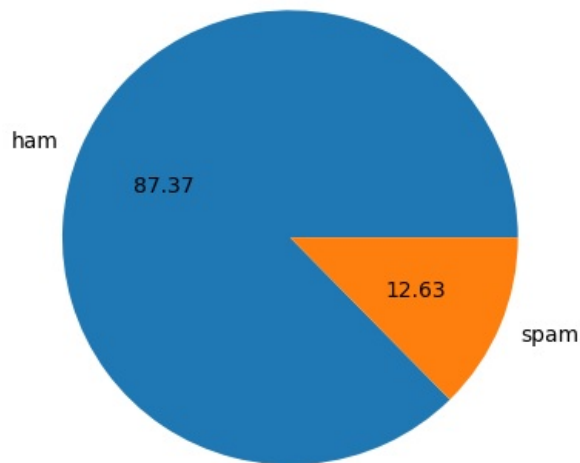
```
In [18]:  df['Target'].value_counts()
```

```
Out[18]:  Target
          0    4516
          1     653
          Name: count, dtype: int64
```

```
In [19]:  plt.pie(df['Target'].value_counts(),labels=['ham','spam'],autopct='%.2f')
          plt.show()
```

In [20]:
```
pip install nltk
```

Requirement already satisfied: nltk in c:\users\purna\appdata\local\programs\python\python311\lib\site-packages (3.8.1)
Requirement already satisfied: click in c:\users\purna\appdata\local\programs\python\python311\lib\site-packages (from nltk) (8.1.7)
Requirement already satisfied: joblib in c:\users\purna\appdata\local\programs\python\python311\lib\site-packages (from nltk) (1.2.0)
Requirement already satisfied: regex>=2021.8.3 in c:\users\purna\appdata\local\programs\python\python311\lib\site-packages (from nltk) (2024.5.15)
Requirement already satisfied: tqdm in c:\users\purna\appdata\local\programs\python\python311\lib\site-packages (from nltk) (4.66.4)
Requirement already satisfied: colorama in c:\users\purna\appdata\local\programs\python\python311\lib\site-packages (from click->nltk) (0.4.6)
Note: you may need to restart the kernel to use updated packages.

[notice] A new release of pip available: 22.3.1 -> 24.0
[notice] To update, run: C:\Users\Purna\AppData\Local\Programs\Python\Python311\python.exe -m pip install --upgrade pip

In [21]:
```python
import nltk
```

In [22]:
```python
##nltk.download()
```

In [23]:
```python
nltk.download('punkt')
```

[nltk_data] Downloading package punkt to
[nltk_data]   C:\Users\Purna\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!

Out[23]: True

In [24]:
```python
df['num_characters'] = df['Text'].apply(len)
```

In [25]:
```python
df.head()
```

Out[25]:

	Target	Text	num_characters
0	0	Go until jurong point, crazy.. Available only ...	111
1	0	Ok lar... Joking wif u oni...	29
2	1	Free entry in 2 a wkly comp to win FA Cup fina...	155
3	0	U dun say so early hor... U c already then say...	49
4	0	Nah I don't think he goes to usf, he lives aro...	61

In [26]:
```python
#number of words

df['num_words'] = df['Text'].apply(lambda x:len(nltk.word_tokenize(x)))
```

In [27]:
```python
df.head()
```

	Target	Text	num_characters	num_words
0	0	Go until jurong point, crazy.. Available only ...	111	24
1	0	Ok lar... Joking wif u oni...	29	8
2	1	Free entry in 2 a wkly comp to win FA Cup fina...	155	37
3	0	U dun say so early hor... U c already then say...	49	13
4	0	Nah I don't think he goes to usf, he lives aro...	61	15

In [28]:
```
# number of words

df['num_sentences'] = df['Text'].apply(lambda x:len(nltk.sent_tokenize(x)))
```

In [29]:
```
df.head()
```

Out[29]:

	Target	Text	num_characters	num_words	num_sentences
0	0	Go until jurong point, crazy.. Available only ...	111	24	2
1	0	Ok lar... Joking wif u oni...	29	8	2
2	1	Free entry in 2 a wkly comp to win FA Cup fina...	155	37	2
3	0	U dun say so early hor... U c already then say...	49	13	1
4	0	Nah I don't think he goes to usf, he lives aro...	61	15	1

In [30]:
```
df[['num_characters','num_words','num_sentences']].describe()
```

Out[30]:

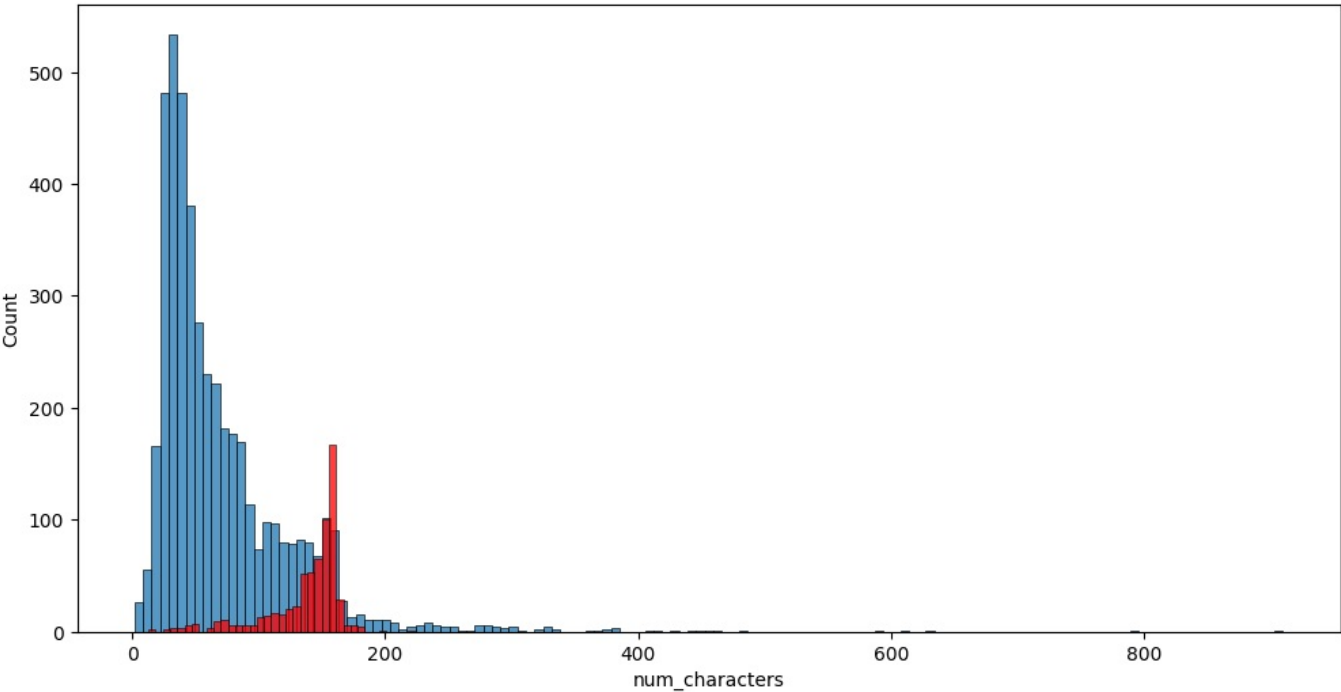|	num_characters	num_words	num_sentences
count	5169.000000	5169.000000	5169.000000
mean	78.977945	18.455794	1.965564
std	58.236293	13.324758	1.448541
min	2.000000	1.000000	1.000000
25%	36.000000	9.000000	1.000000
50%	60.000000	15.000000	1.000000
75%	117.000000	26.000000	2.000000
max	910.000000	220.000000	38.000000

In [31]:
```
# ham
df[df['Target']==0][['num_characters','num_words','num_sentences']].describe()
```

Out[31]:

	num_characters	num_words	num_sentences
count	4516.000000	4516.000000	4516.000000
mean	70.459256	17.123782	1.820195
std	56.358207	13.493970	1.383657
min	2.000000	1.000000	1.000000
25%	34.000000	8.000000	1.000000
50%	52.000000	13.000000	1.000000
75%	90.000000	22.000000	2.000000
max	910.000000	220.000000	38.000000

In [32]:
```
# spam
df[df['Target']==1][['num_characters','num_words','num_sentences']].describe()
```

Out[32]:

	num_characters	num_words	num_sentences
count	653.000000	653.000000	653.000000
mean	137.891271	27.667688	2.970904
std	30.137753	7.008418	1.488425
min	13.000000	2.000000	1.000000
25%	132.000000	25.000000	2.000000
50%	149.000000	29.000000	3.000000
75%	157.000000	32.000000	4.000000
max	224.000000	46.000000	9.000000

```python
In [33]: plt.figure(figsize=(12,6))
         sns.histplot(df[df['Target']==0]['num_characters'])
         sns.histplot(df[df['Target']==1]['num_characters'],color='red')
```

Out[33]: <Axes: xlabel='num_characters', ylabel='Count'>


```python
In [34]: plt.figure(figsize=(12,6))
         sns.histplot(df[df['Target']==0]['num_words'])
         sns.histplot(df[df['Target']==1]['num_words'],color='red')
```
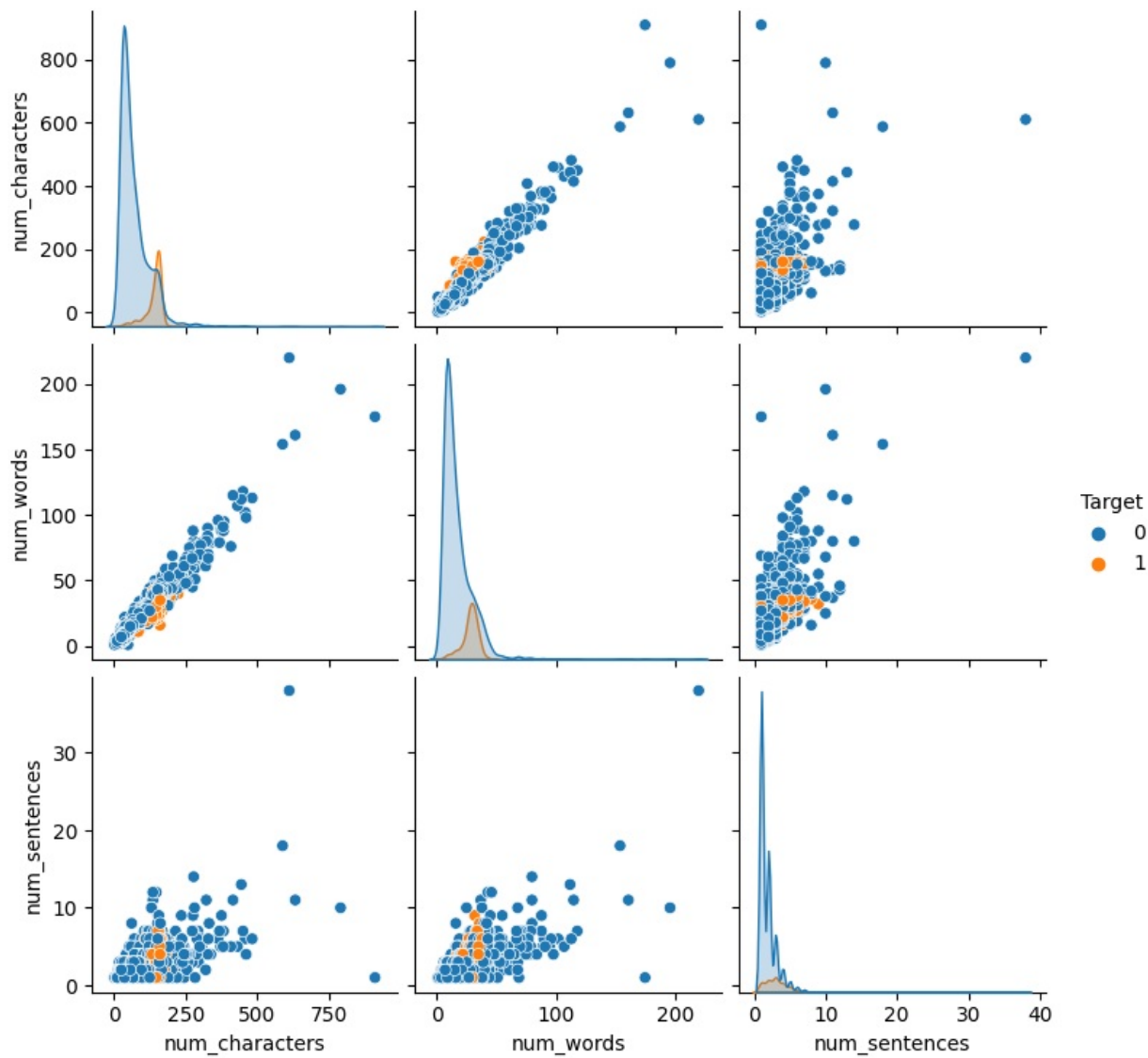
Out[34]: <Axes: xlabel='num_words', ylabel='Count'>
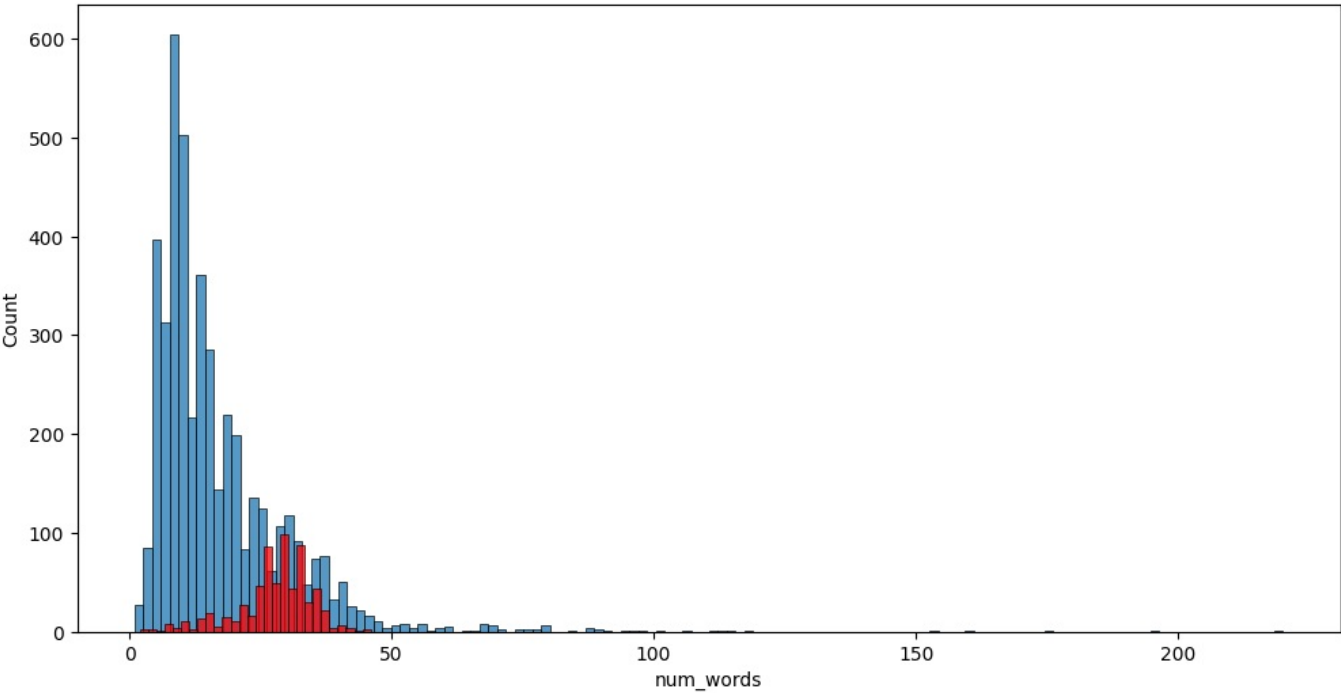
```
In [35]:  df.head()
```

Out[35]:

	Target	Text	num_characters	num_words	num_sentences
0	0	Go until jurong point, crazy.. Available only ...	111	24	2
1	0	Ok lar... Joking wif u oni...	29	8	2
2	1	Free entry in 2 a wkly comp to win FA Cup fina...	155	37	2
3	0	U dun say so early hor... U c already then say...	49	13	1
4	0	Nah I don't think he goes to usf, he lives aro...	61	15	1

```
In [36]:  sns.pairplot(df,hue='Target')
```

Out[36]: <seaborn.axisgrid.PairGrid at 0x169b04b0950>


```
In [37]:  df1=df.copy()
```

```
In [38]:  df1.head()
```

Out[38]:

	Target	Text	num_characters	num_words	num_sentences
0	0	Go until jurong point, crazy.. Available only ...	111	24	2
1	0	Ok lar... Joking wif u oni...	29	8	2
2	1	Free entry in 2 a wkly comp to win FA Cup fina...	155	37	2
3	0	U dun say so early hor... U c already then say...	49	13	1
4	0	Nah I don't think he goes to usf, he lives aro...	61	15	1

```
In [39]:  df1.drop('Text',axis=1,inplace=True)
```

```
In [40]:  df1.head()
```

	Target	num_characters	num_words	num_sentences
0	0	111	24	2
1	0	29	8	2
2	1	155	37	2
3	0	49	13	1
4	0	61	15	1

In [41]:
```python
corr = df1.corr()
```

In [42]:
```python
sns.heatmap(data=corr,annot=True,cmap='magma')
```

Out[42]: <Axes: >

In [43]:
```python
from nltk.corpus import stopwords
stopwords.words('english')
```

Out[43]:
```
['i',
 'me',
 'my',
 'myself',
 'we',
 'our',
 'ours',
 'ourselves',
 'you',
 "you're",
 "you've",
 "you'll",
 "you'd",
 'your',
 'yours',
 'yourself',
 'yourselves',
 'he',
 'him',
 'his',
 'himself',
 'she',
 "she's",
 'her',
 'hers',
 'herself',
 'it',
 "it's",
 'its',
 'itself',
 'they',
 'them',
 'their',
 'theirs',
 'themselves',
 'what',
 'which',
```

```
'who',
'whom',
'this',
'that',
"that'll",
'these',
'those',
'am',
'is',
'are',
'was',
'were',
'be',
'been',
'being',
'have',
'has',
'had',
'having',
'do',
'does',
'did',
'doing',
'a',
'an',
'the',
'and',
'but',
'if',
'or',
'because',
'as',
'until',
'while',
'of',
'at',
'by',
'for',
'with',
'about',
'against',
'between',
'into',
'through',
'during',
'before',
'after',
'above',
'below',
'to',
'from',
'up',
'down',
'in',
'out',
'on',
'off',
'over',
'under',
'again',
'further',
'then',
'once',
'here',
'there',
'when',
'where',
'why',
'how',
'all',
'any',
'both',
'each',
'few',
'more',
'most',
'other',
'some',
'such',
'no',
'nor',
'not',
'only',
```

```
        'own',
        'same',
        'so',
        'than',
        'too',
        'very',
        's',
        't',
        'can',
        'will',
        'just',
        'don',
        "don't",
        'should',
        "should've",
        'now',
        'd',
        'll',
        'm',
        'o',
        're',
        've',
        'y',
        'ain',
        'aren',
        "aren't",
        'couldn',
        "couldn't",
        'didn',
        "didn't",
        'doesn',
        "doesn't",
        'hadn',
        "hadn't",
        'hasn',
        "hasn't",
        'haven',
        "haven't",
        'isn',
        "isn't",
        'ma',
        'mightn',
        "mightn't",
        'mustn',
        "mustn't",
        'needn',
        "needn't",
        'shan',
        "shan't",
        'shouldn',
        "shouldn't",
        'wasn',
        "wasn't",
        'weren',
        "weren't",
        'won',
        "won't",
        'wouldn',
        "wouldn't"]
```

In [44]:
```python
import string
string.punctuation
```

Out[44]: `'!"#$%&\'()*+,-./:;<=>?@[\\]^_`{|}~'`

In [45]:
```python
from nltk.stem.porter import PorterStemmer
ps = PorterStemmer()
ps.stem('dancing')
```

Out[45]: `'danc'`

In [46]:
```python
def transform_text(text):
    text=text.lower()
    text = nltk.word_tokenize(text)

    y =[]
    for i in text:
        if i.isalnum():
            y.append(i)

    text = y[:]
    y.clear()
```

```
        for i in text:
            if i not in stopwords.words('english') and i not in string.punctuation:
                y.append(i)


        text = y[:]
        y.clear()

        for i in text:
            y.append(ps.stem(i))

        return " ".join(y)
```

In [47]: `df['Text'][0]`

Out[47]: `'Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got amore wat..`
`.'`

In [48]: `transform_text('Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there g`

Out[48]: `'go jurong point crazi avail bugi n great world la e buffet cine got amor wat'`

In [49]: `df['transformed_text'] = df['Text'].apply(transform_text)`

In [50]: `df.head()`

Out[50]:

	Target	Text	num_characters	num_words	num_sentences	transformed_text
0	0	Go until jurong point, crazy.. Available only ...	111	24	2	go jurong point crazi avail bugi n great world...
1	0	Ok lar... Joking wif u oni...	29	8	2	ok lar joke wif u oni
2	1	Free entry in 2 a wkly comp to win FA Cup fina...	155	37	2	free entri 2 wkli comp win fa cup final tkt 21...
3	0	U dun say so early hor... U c already then say...	49	13	1	u dun say earli hor u c alreadi say
4	0	Nah I don't think he goes to usf, he lives aro...	61	15	1	nah think goe usf live around though

In [51]: `pip install wordcloud`

```
Requirement already satisfied: wordcloud in c:\users\purna\appdata\local\programs\python\python311\lib\site-pack
ages (1.9.3)
Requirement already satisfied: numpy>=1.6.1 in c:\users\purna\appdata\local\programs\python\python311\lib\site-p
ackages (from wordcloud) (1.24.3)
Requirement already satisfied: pillow in c:\users\purna\appdata\local\programs\python\python311\lib\site-package
s (from wordcloud) (9.5.0)
Requirement already satisfied: matplotlib in c:\users\purna\appdata\local\programs\python\python311\lib\site-pac
kages (from wordcloud) (3.7.1)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\purna\appdata\local\programs\python\python311\lib\si
te-packages (from matplotlib->wordcloud) (1.0.7)
Requirement already satisfied: cycler>=0.10 in c:\users\purna\appdata\local\programs\python\python311\lib\site-p
ackages (from matplotlib->wordcloud) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\purna\appdata\local\programs\python\python311\lib\s
ite-packages (from matplotlib->wordcloud) (4.39.4)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\purna\appdata\local\programs\python\python311\lib\s
ite-packages (from matplotlib->wordcloud) (1.4.4)
Requirement already satisfied: packaging>=20.0 in c:\users\purna\appdata\local\programs\python\python311\lib\sit
e-packages (from matplotlib->wordcloud) (23.1)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\purna\appdata\local\programs\python\python311\lib\si
te-packages (from matplotlib->wordcloud) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\purna\appdata\local\programs\python\python311\li
b\site-packages (from matplotlib->wordcloud) (2.8.2)
Requirement already satisfied: six>=1.5 in c:\users\purna\appdata\local\programs\python\python311\lib\site-packa
ges (from python-dateutil>=2.7->matplotlib->wordcloud) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
[notice] A new release of pip available: 22.3.1 -> 24.0
[notice] To update, run: C:\Users\Purna\AppData\Local\Programs\Python\Python311\python.exe -m pip install --upgr
ade pip
```

In [52]:
```
from wordcloud import WordCloud
wc = WordCloud(width=500,height=500,min_font_size=10,background_color='white')
```

In [53]: `spam_wc = wc.generate(df[df['Target']==1]['transformed_text'].str.cat(sep=''))`

In [54]:
```
plt.figure(figsize=(12,6))
plt.imshow(spam_wc)
```

Out[54]: `<matplotlib.image.AxesImage at 0x169b3622910>`

```
In [55]: ham_wc = wc.generate(df[df['Target']==0]['transformed_text'].str.cat(sep=''))
```

```
In [56]: plt.figure(figsize=(12,6))
         plt.imshow(ham_wc)
```

Out[56]: <matplotlib.image.AxesImage at 0x169b36592d0>


```
In [57]: df[df['Target']==1]
```

	Target	Text	num_characters	num_words	num_sentences	transformed_text
2	1	Free entry in 2 a wkly comp to win FA Cup fina...	155	37	2	free entri 2 wkli comp win fa cup final tkt 21...
5	1	FreeMsg Hey there darling it's been 3 week's n...	148	39	4	freemsg hey darl 3 week word back like fun sti...
8	1	WINNER!! As a valued network customer you have...	158	32	5	winner valu network custom select receivea pri...
9	1	Had your mobile 11 months or more? U R entitle...	154	31	3	mobil 11 month u r entitl updat latest colour ...
11	1	SIX chances to win CASH! From 100 to 20,000 po...	136	31	3	six chanc win cash 100 pound txt csh11 send co...
...	...	...	...	...	...	...
5537	1	Want explicit SEX in 30 secs? Ring 02073162414...	90	18	3	want explicit sex 30 sec ring 02073162414 cost...
5540	1	ASKED 3MOBILE IF 0870 CHATLINES INCLU IN FREE ...	160	38	6	ask 3mobil 0870 chatlin inclu free min india c...
5547	1	Had your contract mobile 11 Mnths? Latest Moto...	160	35	5	contract mobil 11 mnth latest motorola nokia e...
5566	1	REMINDER FROM O2: To get 2.50 pounds free call...	147	30	1	remind o2 get pound free call credit detail gr...
5567	1	This is the 2nd time we have tried 2 contact u...	161	35	4	2nd time tri 2 contact u pound prize 2 claim e...

653 rows × 6 columns

```python
spam_corpus = []
for msg in df[df['Target']==1]['transformed_text'].tolist():
    for word in msg.split():
        spam_corpus.append(word)
```

```python
len(spam_corpus)
```

9939

```python
from collections import Counter
Counter(spam_corpus)
```

Counter({'call': 320,
         'free': 191,
         '2': 155,
         'txt': 141,
         'text': 122,
         'u': 119,
         'ur': 119,
         'mobil': 114,
         'stop': 104,
         'repli': 103,
         'claim': 98,
         '4': 97,
         'prize': 82,
         'get': 74,
         'new': 64,
         'servic': 64,
         'tone': 63,
         'send': 60,
         'urgent': 57,
         'nokia': 57,
         'contact': 56,
         'award': 55,
         'phone': 52,
         'cash': 51,
         'pleas': 51,
         'week': 49,
         'win': 48,
         'c': 45,
         'collect': 45,
         'min': 45,
         'custom': 42,
         'messag': 42,
         'guarante': 42,
         'per': 41,
         'chat': 38,
         'tri': 37,
         'msg': 35,

'draw': 35,
'number': 35,
'cs': 35,
'show': 33,
'today': 33,
'offer': 33,
'line': 33,
'go': 32,
'receiv': 31,
'want': 31,
'latest': 30,
'rington': 30,
'landlin': 30,
'150ppm': 29,
'video': 29,
'1': 28,
'voucher': 28,
'code': 27,
'holiday': 27,
'rate': 26,
'network': 26,
'select': 26,
'150p': 26,
'date': 26,
'box': 26,
'day': 26,
'everi': 26,
'appli': 24,
'r': 24,
'cost': 24,
'end': 24,
'po': 24,
'orang': 24,
'await': 24,
'camera': 23,
'chanc': 22,
'charg': 22,
'sm': 22,
'live': 22,
'poli': 22,
'attempt': 22,
'entri': 21,
'word': 21,
'valid': 21,
'find': 21,
'3': 20,
'back': 20,
'uk': 20,
'mob': 20,
'18': 20,
'game': 20,
'weekli': 20,
'pound': 19,
'know': 19,
'500': 19,
'club': 19,
'1st': 19,
'2nd': 19,
'8007': 19,
'pic': 19,
'credit': 18,
'http': 18,
'deliveri': 18,
'account': 18,
'join': 18,
'16': 18,
'sae': 18,
'time': 18,
'bonu': 17,
'wait': 17,
'expir': 17,
'order': 17,
'750': 17,
'unsubscrib': 17,
'wk': 17,
'next': 16,
'thank': 16,
'08000930705': 16,
'special': 16,
'12hr': 16,
'privat': 16,
'see': 16,
'take': 16,

'land': 16,
'dear': 16,
'gift': 16,
'camcord': 15,
'statement': 15,
'point': 15,
'identifi': 15,
'name': 15,
'music': 15,
'xma': 15,
'price': 15,
'enter': 15,
'hi': 15,
'auction': 15,
'final': 14,
'freemsg': 14,
'updat': 14,
'colour': 14,
'100': 14,
'play': 14,
'match': 14,
'top': 14,
'yr': 14,
'sexi': 14,
'content': 14,
'86688': 14,
'winner': 13,
'weekend': 13,
'friend': 13,
'someon': 13,
'10p': 13,
'hot': 13,
'doubl': 13,
'mobileupd8': 13,
'2003': 13,
'800': 13,
'tell': 13,
'like': 12,
'use': 12,
'ye': 12,
'last': 12,
'caller': 12,
'real': 12,
'dog': 12,
'start': 12,
'reveal': 12,
'congratul': 12,
'anytim': 12,
'minut': 12,
'fanci': 12,
'08000839402': 12,
'half': 12,
'help': 12,
'good': 12,
'enjoy': 12,
'info': 11,
'pobox': 11,
'nation': 11,
'eg': 11,
'complimentari': 11,
'oper': 11,
'ltd': 11,
'make': 11,
'think': 11,
'cd': 11,
'current': 11,
'rental': 11,
'import': 11,
'need': 11,
'wan': 11,
'xxx': 10,
'tomorrow': 10,
'follow': 10,
'player': 10,
'sent': 10,
'direct': 10,
'invit': 10,
'87066': 10,
'unlimit': 10,
'easi': 10,
'babe': 10,
'motorola': 10,
'love': 10,

'ipod': 10,
'shop': 10,
'one': 10,
'opt': 10,
'sex': 10,
'na': 10,
'b': 10,
'wkli': 9,
'question': 9,
'reward': 9,
'wap': 9,
'dont': 9,
'year': 9,
'0800': 9,
'quiz': 9,
'age': 9,
'look': 9,
'choos': 9,
'1000': 9,
'150': 9,
'ts': 9,
'subscrib': 9,
'chariti': 9,
'girl': 9,
'inform': 9,
'great': 9,
'comp': 8,
'fun': 8,
'valu': 8,
'miss': 8,
'news': 8,
'87077': 8,
'access': 8,
'x': 8,
'welcom': 8,
'term': 8,
'worth': 8,
'savamob': 8,
'area': 8,
'5': 8,
'mate': 8,
'either': 8,
'bid': 8,
'pl': 8,
'user': 8,
'08712460324': 8,
'lucki': 8,
'give': 8,
'part': 8,
'fantasi': 8,
'england': 7,
'10': 7,
'inc': 7,
'speak': 7,
'2004': 7,
'txting': 7,
'download': 7,
'even': 7,
'secret': 7,
'admir': 7,
'tnc': 7,
'horni': 7,
'sport': 7,
'maxim': 7,
'bluetooth': 7,
'ldn': 7,
'tenerif': 7,
'tc': 7,
'us': 7,
'custcar': 7,
'remov': 7,
'25p': 7,
'gr8': 7,
'fantast': 7,
'best': 7,
'night': 7,
'could': 7,
'summer': 7,
'store': 7,
'ntt': 7,
'log': 7,
'onto': 7,
'n': 7,

'plu': 7,
'bill': 7,
'im': 7,
'detail': 7,
'may': 6,
'std': 6,
'entitl': 6,
'link': 6,
'congrat': 6,
'etc': 6,
'repres': 6,
'flight': 6,
'08712300220': 6,
'quot': 6,
'standard': 6,
'sp': 6,
'handset': 6,
'answer': 6,
'loyalti': 6,
'cc': 6,
'discount': 6,
'first': 6,
'logo': 6,
'40gb': 6,
'member': 6,
'age16': 6,
'spree': 6,
'holder': 6,
'pc': 6,
'b4': 6,
'book': 6,
'extra': 6,
'zed': 6,
'care': 6,
'record': 6,
'asap': 6,
'gay': 6,
'guess': 6,
'digit': 6,
'an': 6,
'luck': 6,
'got': 6,
'0870': 6,
'arriv': 6,
'titl': 6,
'del': 6,
'pick': 6,
'hey': 5,
'still': 5,
'ok': 5,
'month': 5,
'click': 5,
'subscript': 5,
'pass': 5,
'come': 5,
'voda': 5,
'sunshin': 5,
'soni': 5,
'dvd': 5,
'locat': 5,
'bt': 5,
'must': 5,
'local': 5,
'luv': 5,
'frnd': 5,
'62468': 5,
'goto': 5,
'meet': 5,
'partner': 5,
'flirt': 5,
'singl': 5,
'themob': 5,
'keep': 5,
'rcvd': 5,
'sonyericsson': 5,
'eve': 5,
'pobox36504w45wq': 5,
'activ': 5,
'visit': 5,
'mp3': 5,
'sub': 5,
'ring': 5,
'announc': 5,

'guy': 5,
'vodafon': 5,
'80062': 5,
'immedi': 5,
'alert': 5,
'txtauction': 5,
'break': 5,
'no1': 5,
'talk': 5,
'08718720201': 5,
'6': 5,
'regist': 5,
'ask': 5,
'xchat': 5,
'o2': 5,
'surpris': 5,
'readi': 5,
'20p': 5,
'1327': 5,
'croydon': 5,
'cr9': 5,
'5wb': 5,
'reach': 5,
'86021': 5,
'eeri': 5,
'rpli': 5,
'comput': 5,
'fone': 5,
'flag': 5,
'spook': 5,
'tariff': 5,
'cancel': 5,
'3510i': 5,
'300': 5,
'hour': 4,
'v': 4,
'ac': 4,
'correct': 4,
'150pm': 4,
'recent': 4,
'app': 4,
'laid': 4,
'largest': 4,
'ec2a': 4,
'noth': 4,
'netcollex': 4,
'freephon': 4,
'2optout': 4,
'hmv': 4,
'upgrad': 4,
'card': 4,
'4u': 4,
'check': 4,
'decemb': 4,
'refer': 4,
'ad': 4,
'vari': 4,
'post': 4,
'5000': 4,
'cw25wx': 4,
'condit': 4,
'83355': 4,
'mono': 4,
'85023': 4,
'within': 4,
'hello': 4,
'forward': 4,
'polyphon': 4,
'pobox84': 4,
'36504': 4,
'm8': 4,
'ibiza': 4,
'ppm': 4,
'vip': 4,
'sale': 4,
'money': 4,
'away': 4,
'sk38xh': 4,
'contract': 4,
'place': 4,
'deliv': 4,
'08712405020': 4,
'cum': 4,

'xx': 4,
'08707509020': 4,
'linerent': 4,
'optout': 4,
'2nite': 4,
'hope': 4,
'hotel': 4,
'2day': 4,
'brand': 4,
'250': 4,
'london': 4,
'mobi': 4,
'action': 4,
'arcad': 4,
'08001950382': 4,
'would': 4,
'costa': 4,
'sol': 4,
'87239': 4,
'inclus': 4,
'saturday': 4,
'ntwk': 4,
'hit': 4,
'crazi': 4,
'porn': 4,
'mth': 4,
'film': 4,
'santa': 4,
'cup': 3,
'12': 3,
'11': 3,
'co': 3,
'hl': 3,
'review': 3,
'trip': 3,
'hear': 3,
'k52': 3,
'786': 3,
'unredeem': 3,
'tonight': 3,
'tv': 3,
'89545': 3,
'star': 3,
'25': 3,
'recd': 3,
'penc': 3,
'chart': 3,
'0845': 3,
'newest': 3,
'm26': 3,
'3uz': 3,
'08002986906': 3,
'germani': 3,
'via': 3,
'valentin': 3,
'83600': 3,
'11mth': 3,
'87131': 3,
'heart': 3,
'3min': 3,
'close': 3,
'wc1n3xx': 3,
'loan': 3,
'notic': 3,
'ibhltd': 3,
'ldnw15h': 3,
'renew': 3,
'box95qu': 3,
'unsub': 3,
'08715705022': 3,
'542': 3,
'ladi': 3,
'kick': 3,
'euro2004': 3,
'daili': 3,
'postcod': 3,
'20': 3,
'work': 3,
'textoper': 3,
'50': 3,
'listen': 3,
'4t': 3,
'pari': 3,

    'citi': 3,
    'skilgm': 3,
    'ever': 3,
    'life': 3,
    'w45wq': 3,
    '3qxj9': 3,
    '9ae': 3,
    'alfi': 3,
    'moon': 3,
    'children': 3,
    'song': 3,
    '08701417012': 3,
    'profit': 3,
    'cust': 3,
    'five': 3,
    '434': 3,
    'sk3': 3,
    '8wp': 3,
    'address': 3,
    'male': 3,
    'cheap': 3,
    'compani': 3,
    '3g': 3,
    'videophon': 3,
    'videochat': 3,
    'wid': 3,
    'java': 3,
    'dload': 3,
    'nolin': 3,
    'rentl': 3,
    'f': 3,
    'peopl': 3,
    'cant': 3,
    'sipix': 3,
    'receipt': 3,
    'bore': 3,
    'yesterday': 3,
    'mani': 3,
    'let': 3,
    'futur': 3,
    'jordan': 3,
    '84128': 3,
    'set': 3,
    'jamster': 3,
    'origin': 3,
    '3gbp': 3,
    'photo': 3,
    'arsen': 3,
    'pod': 3,
    'alon': 3,
    '09066362231': 3,
    '07xxxxxxxxx': 3,
    'high': 3,
    'shortli': 3,
    '7250i': 3,
    'stay': 3,
    '82242': 3,
    'sorri': 3,
    'connect': 3,
    'ticket': 3,
    '18yr': 3,
    'abta': 3,
    'right': 3,
    'person': 3,
    'well': 3,
    '09050090044': 3,
    'toclaim': 3,
    'pobox334': 3,
    'stockport': 3,
    'max10min': 3,
    'provid': 3,
    'freefon': 3,
    'box97n7qp': 3,
    'men': 3,
    'offici': 3,
    'ip4': 3,
    '5we': 3,
    'cross': 3,
    'also': 3,
    '88066': 3,
    '80082': 3,
    'halloween': 3,
    'sky': 3,

'adult': 3,
'09066612661': 3,
'sept': 3,
'28': 3,
'king': 3,
'super': 3,
'littl': 3,
'50p': 3,
'deliveredtomorrow': 3,
'save': 3,
'callback': 3,
'200': 3,
'7': 3,
'nat': 3,
'goal': 3,
'30': 3,
'fa': 2,
'tkt': 2,
'87121': 2,
'darl': 2,
'chg': 2,
'rcv': 2,
'six': 2,
'team': 2,
'sptv': 2,
'cinema': 2,
'09061209465': 2,
'suprman': 2,
'matrix3': 2,
'starwars3': 2,
'advis': 2,
'simpli': 2,
'password': 2,
'verifi': 2,
'gent': 2,
'45239': 2,
'350': 2,
'q': 2,
'countri': 2,
'ansr': 2,
'tyron': 2,
'69888': 2,
'hardcor': 2,
'randi': 2,
'08700621170150p': 2,
'arrang': 2,
'uniqu': 2,
'enough': 2,
'august': 2,
'dream': 2,
'soon': 2,
'horo': 2,
'sign': 2,
'sam': 2,
'125gift': 2,
'69698': 2,
'qualiti': 2,
'pay': 2,
'sim': 2,
'021': 2,
'blind': 2,
'82468': 2,
'smart': 2,
'fix': 2,
'0844': 2,
'861': 2,
'85': 2,
'prepay': 2,
'lifetim': 2,
'shag': 2,
'avail': 2,
'4mth': 2,
'mobilesdirect': 2,
'08000938767': 2,
'or2stoptxt': 2,
'true': 2,
'box334sk38ch': 2,
'competit': 2,
'80086': 2,
'lover': 2,
'name1': 2,
'name2': 2,
'mobno': 2,

'adam': 2,
'07123456789': 2,
'yahoo': 2,
'txtno': 2,
'ave': 2,
'purpos': 2,
'tenant': 2,
'refus': 2,
'0207': 2,
'153': 2,
'juli': 2,
'dave': 2,
'09061743806': 2,
'box326': 2,
'cha': 2,
'24hr': 2,
'08718738001': 2,
'web': 2,
'call09050000327': 2,
'2000': 2,
'08712402050': 2,
'10ppm': 2,
'ag': 2,
'promo': 2,
'kept': 2,
'result': 2,
'83222': 2,
'textbuddi': 2,
'search': 2,
'89693': 2,
'40533': 2,
'rstm': 2,
'sw7': 2,
'3ss': 2,
'premium': 2,
'sue': 2,
'old': 2,
'09071512433': 2,
'050703': 2,
'csbcm4235wc1n3xx': 2,
'callcost': 2,
'mobilesvari': 2,
'email': 2,
'nokia6650': 2,
'81151': 2,
'ctxt': 2,
'helplin': 2,
'08706091795': 2,
'40': 2,
'around': 2,
'premier': 2,
'88039': 2,
'tscs087147403231winawk': 2,
'0578': 2,
'commun': 2,
'subpoli': 2,
'08718727870': 2,
'm263uz': 2,
'cashto': 2,
'08000407165': 2,
'getstop': 2,
'88222': 2,
'php': 2,
'08002888812': 2,
'wed': 2,
'09066350750': 2,
'world': 2,
'discreet': 2,
'suppli': 2,
'virgin': 2,
'41685': 2,
'till': 2,
'10k': 2,
'liverpool': 2,
'09058094565': 2,
'remind': 2,
'cheaper': 2,
'peak': 2,
'88600': 2,
'lux': 2,
'filthi': 2,
'valid12hr': 2,
'09063458130': 2,

'polyph': 2,
'somebodi': 2,
'secretli': 2,
'datebox1282essexcm61xn': 2,
'ls15hb': 2,
'euro': 2,
'flower': 2,
'84025': 2,
'lot': 2,
'09050003091': 2,
'c52': 2,
'p': 2,
'09061790121': 2,
'speedchat': 2,
'08000776320': 2,
'wish': 2,
'forget': 2,
'request': 2,
'4fil': 2,
'ralli': 2,
'energi': 2,
'celeb': 2,
'voicemail': 2,
'choic': 2,
'87021': 2,
'txtin': 2,
'4info': 2,
'complet': 2,
'hungri': 2,
'feel': 2,
'lion': 2,
'lionm': 2,
'lionp': 2,
'potter': 2,
'phoenix': 2,
'harri': 2,
'among': 2,
'reader': 2,
'balanc': 2,
'sauci': 2,
'bear': 2,
'upload': 2,
'08718730666': 2,
'textpod': 2,
'clair': 2,
'luci': 2,
'hubbi': 2,
'fri': 2,
'leav': 2,
'80488': 2,
'jsco': 2,
'sing': 2,
'7250': 2,
'british': 2,
'sw73ss': 2,
'dracula': 2,
'ghost': 2,
'addamsfa': 2,
'munster': 2,
'exorcist': 2,
'twilight': 2,
'respons': 2,
'cam': 2,
'msg150p': 2,
'2rcv': 2,
'hlp': 2,
'08712317606': 2,
'08712405022': 2,
'wiv': 2,
'marri': 2,
'81303': 2,
'hard': 2,
'121': 2,
'biggest': 2,
'wow': 2,
'boy': 2,
'2007': 2,
'tour': 2,
'chosen': 2,
'bahama': 2,
'buy': 2,
'press': 2,
'0': 2,

'purchas': 2,
'big': 2,
'excit': 2,
'3100': 2,
'straight': 2,
'wet': 2,
'89555': 2,
'g696ga': 2,
'pm': 2,
'upto': 2,
'12mth': 2,
'0776xxxxxxx': 2,
'326': 2,
'total': 2,
'80182': 2,
'08452810073': 2,
'09066358152': 2,
'prompt': 2,
'mind': 2,
'done': 2,
'61610': 2,
'yo': 2,
'tsc': 2,
'aom': 2,
'gender': 2,
'yer': 2,
'eng': 2,
'box39822': 2,
'w111wx': 2,
'hol': 2,
'med': 2,
'satisfi': 2,
'home': 2,
'pix': 2,
'ref': 2,
'respond': 2,
'300p': 2,
'heard': 2,
'u4': 2,
'rude': 2,
'8552': 2,
'88877': 2,
'lost': 2,
'upd8': 2,
'21870000': 2,
'mailbox': 2,
'09056242159': 2,
'retriev': 2,
'pub': 2,
'street': 2,
'duchess': 2,
'cornwal': 2,
'008704050406': 2,
'08701417012150p': 2,
'score': 2,
'town': 2,
'babi': 2,
'thing': 2,
'slo': 2,
'24': 2,
'acl03530150pm': 2,
'brought': 2,
'bx420': 2,
'amaz': 2,
'picsfree1': 2,
'instruct': 2,
'knock': 2,
'norm': 2,
'69696': 2,
'nyt': 2,
'3lp': 2,
'includ': 2,
'fastest': 2,
'grow': 2,
'2stoptxt': 2,
'luxuri': 2,
'canari': 2,
'island': 2,
'burger': 2,
'interflora': 2,
'everyon': 2,
'sinc': 2,
'warm': 2,

```
                'bring': 2,
                'sound': 2,
                '88888': 2,
                'realiti': 2,
                'yet': 2,
                'drive': 2,
                'improv': 2,
                'lotr': 2,
                '08708034412': 2,
                'lookatm': 2,
                '09061221066': 2,
                'fromm': 2,
                'pobox45w2tg150p': 2,
                'truli': 2,
                'wo': 2,
                'movi': 2,
                'bank': 2,
                'granit': 2,
                'issu': 2,
                'explos': 2,
                'nasdaq': 2,
                'symbol': 2,
                'cdgt': 2,
                'thur': 2,
                'sat': 2,
                'unabl': 2,
                'class': 2,
                'iscom': 2,
                'blu': 2,
                'wml': 2,
                '09061743386': 2,
                'spanish': 2,
                'indic': 2,
                'accid': 2,
                '2find': 2,
                'anoth': 2,
                'polo': 2,
                'suit': 2,
                '373': 2,
                'w1j': 2,
                '6hl': 2,
                'busi': 2,
                'explicit': 2,
                'sec': 2,
                '02073162414': 2,
                '89070': 2,
                'mnth': 2,
                '86888': 2,
                '3hr': 2,
                'txtstop': 2,
                ...})
```
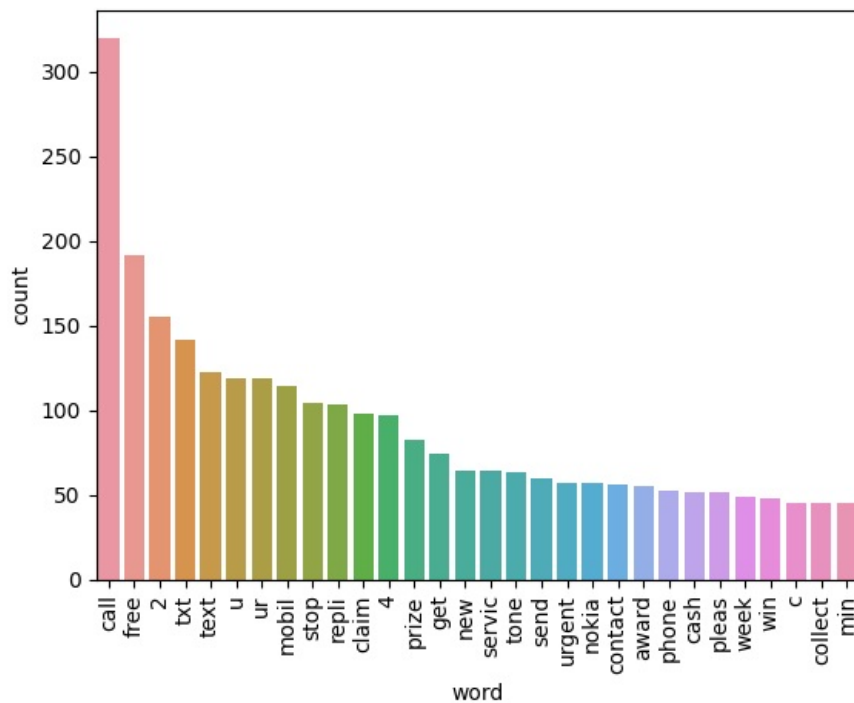
In [61]:
```python
spam_counter = Counter(spam_corpus)
```

In [62]:
```python
df_counts = pd.DataFrame(spam_counter.most_common(30), columns=['word', 'count'])
```

In [63]:
```python
sns.barplot(x='word', y='count', data=df_counts)
plt.xticks(rotation='vertical')
plt.show()
```

```
In [64]: df[df['Target']==0]
```

Out[64]:

	Target	Text	num_characters	num_words	num_sentences	transformed_text
0	0	Go until jurong point, crazy.. Available only ...	111	24	2	go jurong point crazi avail bugi n great world...
1	0	Ok lar... Joking wif u oni...	29	8	2	ok lar joke wif u oni
3	0	U dun say so early hor... U c already then say...	49	13	1	u dun say earli hor u c alreadi say
4	0	Nah I don't think he goes to usf, he lives aro...	61	15	1	nah think goe usf live around though
6	0	Even my brother is not like to speak with me. ...	77	18	2	even brother like speak treat like aid patent
...	...	...	...	...	...	...
5565	0	Huh y lei...	12	4	1	huh lei
5568	0	Will Ì_ b going to esplanade fr home?	37	9	1	b go esplanad fr home
5569	0	Pity, * was in mood for that. So...any other s...	57	15	2	piti mood suggest
5570	0	The guy did some bitching but I acted like i'd...	125	27	1	guy bitch act like interest buy someth els nex...
5571	0	Rofl. Its true to its name	26	7	2	rofl true name

4516 rows × 6 columns

```
In [65]: ham_corpus = []
         for msg in df[df['Target']==0]['transformed_text'].tolist():
             for word in msg.split():
                 ham_corpus.append(word)
```
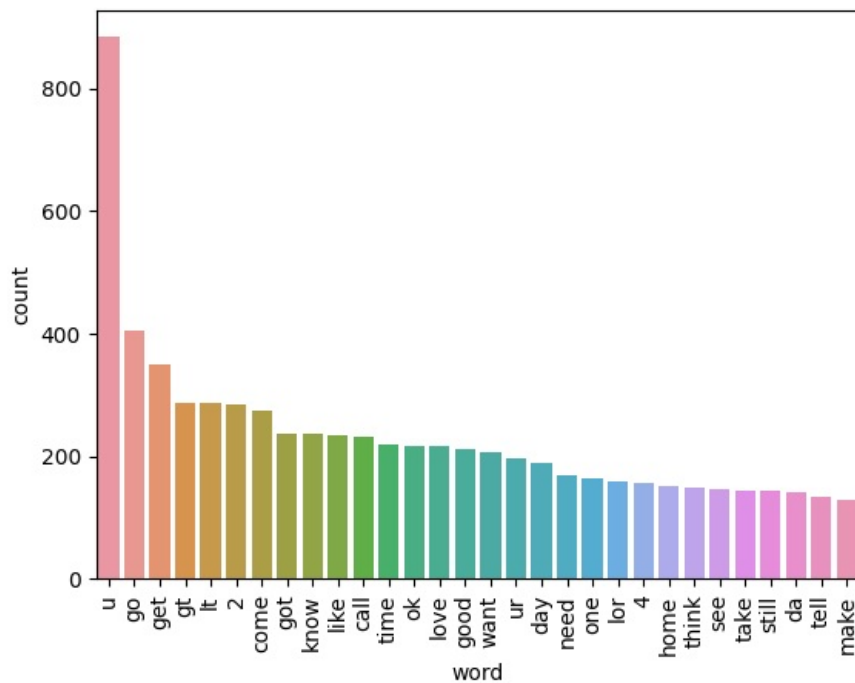
```
In [66]: len(ham_corpus)
```

Out[66]: 35404

```
In [67]: ham_counter = Counter(ham_corpus)
```

```
In [68]: df_counts = pd.DataFrame(ham_counter.most_common(30), columns=['word', 'count'])
```

```
In [69]: sns.barplot(x='word', y='count', data=df_counts)
         plt.xticks(rotation='vertical')
         plt.show()
```

```
In [70]:  # text vectorization using Bag of words
          df.head()
```

Out[70]:

	Target	Text	num_characters	num_words	num_sentences	transformed_text
0	0	Go until jurong point, crazy.. Available only ...	111	24	2	go jurong point crazi avail bugi n great world...
1	0	Ok lar... Joking wif u oni...	29	8	2	ok lar joke wif u oni
2	1	Free entry in 2 a wkly comp to win FA Cup fina...	155	37	2	free entri 2 wkli comp win fa cup final tkt 21...
3	0	U dun say so early hor... U c already then say...	49	13	1	u dun say earli hor u c alreadi say
4	0	Nah I don't think he goes to usf, he lives aro...	61	15	1	nah think goe usf live around though

```
In [71]:  from sklearn.feature_extraction.text import CountVectorizer,TfidfVectorizer
          cv = CountVectorizer()
          tfidf = TfidfVectorizer(max_features=3000)
```

```
In [72]:  X = tfidf.fit_transform(df['transformed_text']).toarray()
```

```
In [73]:  X.shape
```

Out[73]: (5169, 3000)

```
In [74]:  y = df['Target'].values
          y
```

Out[74]: array([0, 0, 1, ..., 0, 0, 0])

```
In [75]:  from sklearn.model_selection import train_test_split
```

```
In [76]:  X_train,X_test,y_train,y_test = train_test_split(X,y,random_state=2,test_size=0.2)
```

```
In [77]:  from sklearn.naive_bayes import GaussianNB,MultinomialNB,BernoulliNB
```

```
In [78]:  gnb = GaussianNB()
          mnb = MultinomialNB()
          bnb = BernoulliNB()
```

```
In [79]:  gnb.fit(X_train,y_train)
```

Out[79]: ▾ GaussianNB

          GaussianNB()

```
In [80]:  y_pred1 = gnb.predict(X_test)
```

```
In [81]:  from sklearn.metrics import accuracy_score,precision_score,confusion_matrix
```

```
In [82]:  print(accuracy_score(y_test,y_pred1))
          print(precision_score(y_test,y_pred1))
          print(confusion_matrix(y_test,y_pred1))

          0.8694390715667312
          0.5068493150684932
          [[788 108]
           [ 27 111]]

In [83]:  mnb.fit(X_train,y_train)

Out[83]:  ▾ MultinomialNB

          MultinomialNB()


In [84]:  y_pred2 = mnb.predict(X_test)

In [87]:  print(accuracy_score(y_test,y_pred2))
          print(precision_score(y_test,y_pred2))
          print(confusion_matrix(y_test,y_pred2))

          0.9709864603481625
          1.0
          [[896    0]
           [ 30 108]]

In [88]:  bnb.fit(X_train,y_train)

Out[88]:  ▾ BernoulliNB

          BernoulliNB()


In [89]:  y_pred3 =bnb.predict(X_test)

In [90]:  print(accuracy_score(y_test,y_pred3))
          print(precision_score(y_test,y_pred3))
          print(confusion_matrix(y_test,y_pred3))

          0.9835589941972921
          0.991869918699187
          [[895    1]
           [ 16 122]]

In [91]:  from sklearn.linear_model import LogisticRegression
          from sklearn.svm import SVC
          from sklearn.naive_bayes import MultinomialNB
          from sklearn.tree import DecisionTreeClassifier
          from sklearn.neighbors import KNeighborsClassifier
          from sklearn.ensemble import RandomForestClassifier
          from sklearn.ensemble import AdaBoostClassifier
          from sklearn.ensemble import BaggingClassifier
          from sklearn.ensemble import ExtraTreesClassifier
          from sklearn.ensemble import GradientBoostingClassifier

In [92]:  svc = SVC(kernel='sigmoid', gamma=1.0)
          knc = KNeighborsClassifier()
          mnb = MultinomialNB()
          dtc = DecisionTreeClassifier(max_depth=5)
          lrc = LogisticRegression(solver='liblinear', penalty='l1')
          rfc = RandomForestClassifier(n_estimators=50, random_state=2)
          abc = AdaBoostClassifier(n_estimators=50, random_state=2)
          bc = BaggingClassifier(n_estimators=50, random_state=2)
          etc = ExtraTreesClassifier(n_estimators=50, random_state=2)
          gbdt = GradientBoostingClassifier(n_estimators=50,random_state=2)

In [94]:  clfs = {
              'SVC' : svc,
              'KN' : knc,
              'NB': mnb,
              'DT': dtc,
              'LR': lrc,
              'RF': rfc,
              'AdaBoost': abc,
              'BgC': bc,
              'ETC': etc,
              'GBDT':gbdt
          }

In [95]:  def train_classifier(clf,X_train,y_train,X_test,y_test):
              clf.fit(X_train,y_train)
              y_pred = clf.predict(X_test)
```

```
        accuracy = accuracy_score(y_test,y_pred)
        precision = precision_score(y_test,y_pred)

        return accuracy,precision
```

In [96]:
```
train_classifier(svc,X_train,y_train,X_test,y_test)
```

Out[96]: (0.9758220502901354, 0.9747899159663865)

In [97]:
```
accuracy_scores = []
precision_scores = []

for name,clf in clfs.items():

    current_accuracy,current_precision = train_classifier(clf, X_train,y_train,X_test,y_test)

    print("For ",name)
    print("Accuracy - ",current_accuracy)
    print("Precision - ",current_precision)

    accuracy_scores.append(current_accuracy)
    precision_scores.append(current_precision)
```

```
For  SVC
Accuracy -  0.9758220502901354
Precision -  0.9747899159663865
For  KN
Accuracy -  0.9052224371373307
Precision -  1.0
For  NB
Accuracy -  0.9709864603481625
Precision -  1.0
For  DT
Accuracy -  0.9294003868471954
Precision -  0.8282828282828283
For  LR
Accuracy -  0.9584139264990329
Precision -  0.9702970297029703
For  RF
Accuracy -  0.9758220502901354
Precision -  0.9829059829059829
For  AdaBoost
Accuracy -  0.960348162475822
Precision -  0.9292035398230089
For  BgC
Accuracy -  0.9584139264990329
Precision -  0.8682170542635659
For  ETC
Accuracy -  0.9748549323017408
Precision -  0.9745762711864406
For  GBDT
Accuracy -  0.9468085106382979
Precision -  0.9191919191919192
```

In [98]:
```
performance_df = pd.DataFrame({'Algorithm':clfs.keys(),'Accuracy':accuracy_scores,'Precision':precision_scores}
```
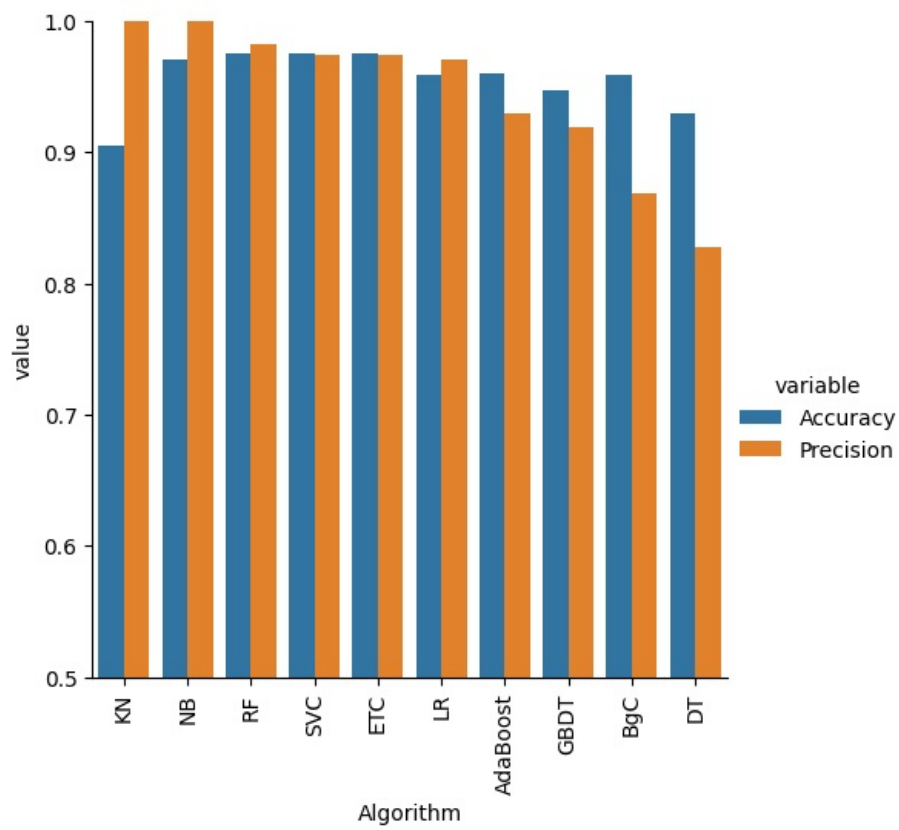
In [99]:
```
performance_df
```

Out[99]:

	Algorithm	Accuracy	Precision
1	KN	0.905222	1.000000
2	NB	0.970986	1.000000
5	RF	0.975822	0.982906
0	SVC	0.975822	0.974790
8	ETC	0.974855	0.974576
4	LR	0.958414	0.970297
6	AdaBoost	0.960348	0.929204
9	GBDT	0.946809	0.919192
7	BgC	0.958414	0.868217
3	DT	0.929400	0.828283

In [100...
```
performance_df1 = pd.melt(performance_df,id_vars='Algorithm')
```

In [101...
```
sns.catplot(x='Algorithm',y='value',hue='variable',data=performance_df1,kind='bar',height=5)
plt.ylim(0.5,1.0)
plt.xticks(rotation='vertical')
```

```
plt.show()
```