In [2]:
```
import pandas as pd
import numpy as np
import seaborn as sn
import matplotlib.pyplot as plt
```

In [3]:
```
df=pd.read_csv("C:\\Users\\umar1\\OneDrive\\Documents\\Desktop\\email_spam_classification\\spam2.csv",encoding_
```

Data Cleaning

In [4]:
```
df
```

Out[4]:

	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
...	...	...	...	...	...
5567	spam	This is the 2nd time we have tried 2 contact u...	NaN	NaN	NaN
5568	ham	Will _ b going to esplanade fr home?	NaN	NaN	NaN
5569	ham	Pity, * was in mood for that. So...any other s...	NaN	NaN	NaN
5570	ham	The guy did some bitching but I acted like i'd...	NaN	NaN	NaN
5571	ham	Rofl. Its true to its name	NaN	NaN	NaN

5572 rows × 5 columns

In [5]:
```
df.rename(columns={'v1':'spam','v2':'email'},inplace=True)
```

In [6]:
```
df
```

Out[6]:

	spam	email	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
...	...	...	...	...	...
5567	spam	This is the 2nd time we have tried 2 contact u...	NaN	NaN	NaN
5568	ham	Will _ b going to esplanade fr home?	NaN	NaN	NaN
5569	ham	Pity, * was in mood for that. So...any other s...	NaN	NaN	NaN
5570	ham	The guy did some bitching but I acted like i'd...	NaN	NaN	NaN
5571	ham	Rofl. Its true to its name	NaN	NaN	NaN

5572 rows × 5 columns

In [7]:
```
df=df.drop({'Unnamed: 2','Unnamed: 3','Unnamed: 4'},axis=1)
```

In [8]:
```
df
```

Out[8]:

	spam	email
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...
...	...	...

5567	spam	This is the 2nd time we have tried 2 contact u...
5568	ham	Will _ b going to esplanade fr home?
5569	ham	Pity, * was in mood for that. So...any other s...
5570	ham	The guy did some bitching but I acted like i'd...
5571	ham	Rofl. Its true to its name

5572 rows × 2 columns

In [9]:
```
from sklearn.preprocessing import LabelEncoder
encoder=LabelEncoder()
```

In [10]:
```
df['spam']=encoder.fit_transform(df['spam'])
```

In [11]:
```
df
```

Out[11]:

	spam	email
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 _ 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

5572 rows × 2 columns

In [12]:
```
df.isnull().sum()
#There are no null values
```

Out[12]:
```
spam      0
email     0
dtype: int64
```

In [13]:
```
df.duplicated().sum()
```

Out[13]:
```
403
```

In [14]:
```
df.drop_duplicates(inplace=True)
```

In [15]:
```
df.duplicated().sum()
```

Out[15]:
```
0
```

In [16]:
```
df.shape
```

Out[16]:
```
(5169, 2)
```

Exploratory Data Analysis

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

Out[17]:

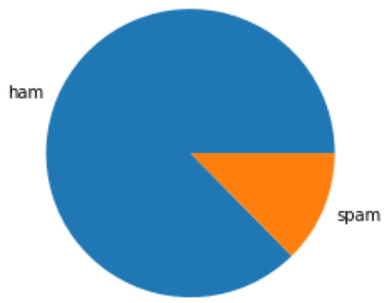
	spam	email
0	0	Go until jurong point, crazy.. Available only ...
1	0	Ok lar... Joking wif u oni...
2	1	Free entry in 2 a wkly comp to win FA Cup fina...
3	0	U dun say so early hor... U c already then say...
4	0	Nah I don't think he goes to usf, he lives aro...

In [18]:
```
df['spam'].value_counts()
```

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

In [19]:
```
# Now here we see that the proportion of ham is much more than that of the spam--- data is imbalanced
plt.pie(df['spam'].value_counts(),labels=['ham','spam'])
```

Out[19]:
```
([<matplotlib.patches.Wedge at 0x186c023aca0>,
  <matplotlib.patches.Wedge at 0x186c08f01c0>],
 [Text(-1.0144997251399075, 0.4251944351600247, 'ham'),
  Text(1.014499764949479, -0.4251943401757036, 'spam')])
```



In [20]:
```
import nltk
```

In [21]:
```
nltk.download('punkt')
```

```
[nltk_data] Error loading punkt: <urlopen error [WinError 10060] A
[nltk_data]     connection attempt failed because the connected party
[nltk_data]     did not properly respond after a period of time, or
[nltk_data]     established connection failed because connected host
[nltk_data]     has failed to respond>
```

Out[21]:
```
False
```

In [22]:
```
#no.of characters present is determined here.
df['num_characters']=df['email'].apply(len)
```

In [23]:
```
df['num_word']=df['email'].apply(lambda x:len(nltk.word_tokenize(x)))
```

In [24]:
```
df
```

Out[24]:

	spam	email	num_characters	num_word
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
...	...	...	...	...
5567	1	This is the 2nd time we have tried 2 contact u...	160	35
5568	0	Will �_ b going to esplanade fr home?	37	9
5569	0	Pity, * was in mood for that. So...any other s...	57	15
5570	0	The guy did some bitching but I acted like i'd...	125	27
5571	0	Rofl. Its true to its name	26	7

5169 rows × 4 columns

In [25]:
```
df['num_sentences']=df['email'].apply(lambda x:len(nltk.sent_tokenize(x)))
```

In [26]:
```
df
```

Out[26]:

	spam	email	num_characters	num_word	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
...	...	...	...	...	...
5567	1	This is the 2nd time we have tried 2 contact u...	160	35	4
5568	0	Will ﬁ_ b going to esplanade fr home?	37	9	1
5569	0	Pity, * was in mood for that. So...any other s...	57	15	2
5570	0	The guy did some bitching but I acted like i'd...	125	27	1
5571	0	Rofl. Its true to its name	26	7	2

5169 rows × 5 columns

In [27]:
```
df[['num_characters','num_word','num_sentences']].describe()
```

Out[27]:

	num_characters	num_word	num_sentences
count	5169.000000	5169.000000	5169.000000
mean	78.923776	18.454246	1.948152
std	58.174846	13.325668	1.363792
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	28.000000

In [28]:
```
df[df['spam']==1][['num_characters','num_word','num_sentences']].describe()
```

Out[28]:

	num_characters	num_word	num_sentences
count	653.000000	653.000000	653.000000
mean	137.479326	27.675345	2.975498
std	30.014336	7.011513	1.487993
min	13.000000	2.000000	1.000000
25%	131.000000	25.000000	2.000000
50%	148.000000	29.000000	3.000000
75%	157.000000	32.000000	4.000000
max	223.000000	46.000000	8.000000

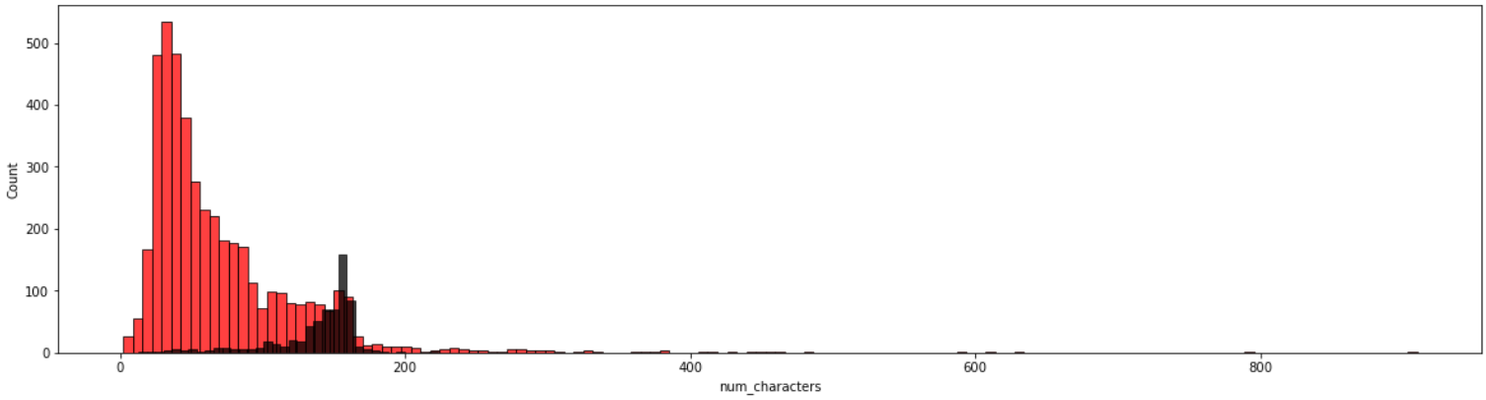
In [29]:
```
df[df['spam']==0][['num_characters','num_word','num_sentences']].describe()
```

Out[29]:

	num_characters	num_word	num_sentences
count	4516.000000	4516.000000	4516.000000
mean	70.456820	17.120903	1.799601
std	56.356802	13.493725	1.278465
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	28.000000

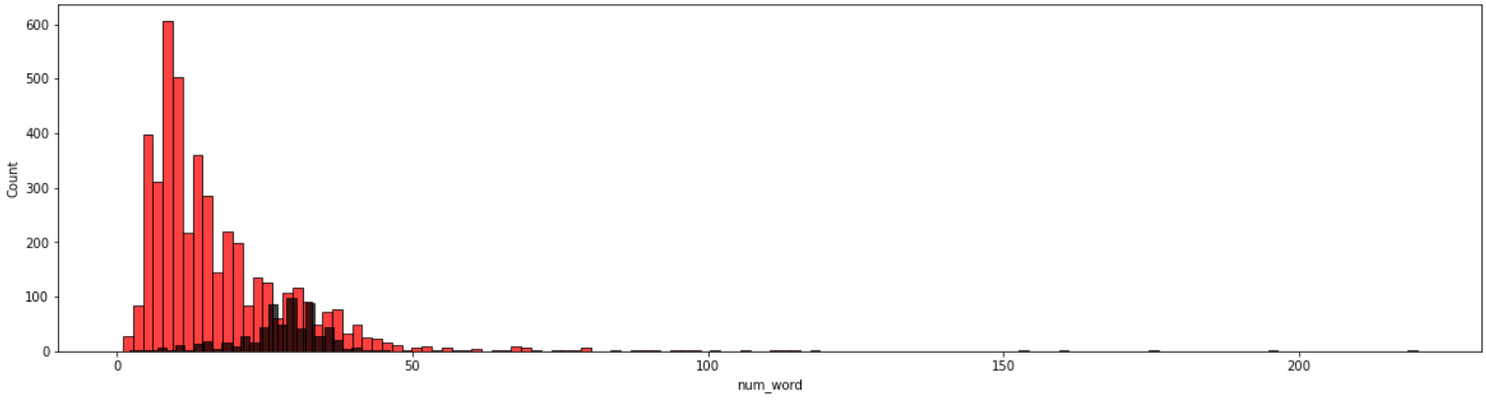
In [30]:
```
# Now let us visualize the cases of number of words and characters for the two cases of spam and ham
fig, ax = plt.subplots(figsize=(20, 5))
sn.histplot(df[df['spam']==0]['num_characters'],color='red')
sn.histplot(df[df['spam']==1]['num_characters'],color='black')
```

Out[30]:
```
<AxesSubplot:xlabel='num_characters', ylabel='Count'>
```



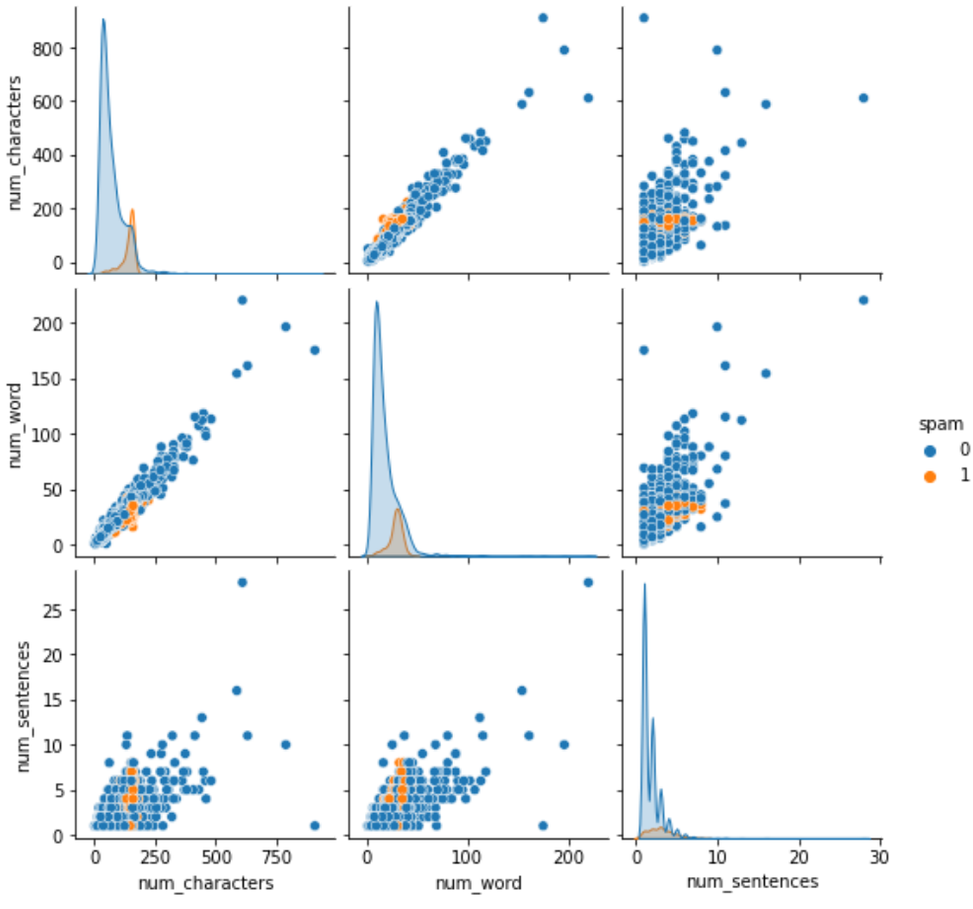
In [31]:
```
fig, ax = plt.subplots(figsize=(20, 5))
sn.histplot(df[df['spam']==0]['num_word'],color='red')
sn.histplot(df[df['spam']==1]['num_word'],color='black')
```

Out[31]:
```
<AxesSubplot:xlabel='num_word', ylabel='Count'>
```



In [32]:
```
sn.pairplot(df,hue='spam')    ## Tells us the relation between each variables.
```

Out[32]:
```
<seaborn.axisgrid.PairGrid at 0x186c2b586d0>
```



In [33]:
```
sn.heatmap(df.corr(),annot=True)
```

Out[33]:
```
<AxesSubplot:>
```

In [34]:
```
#There we can see that the correlation between the coefficient therefore multilinearity is present , Therefore
```

Text Preprocessing

In [35]:
```
nltk.download('stopwords')
from nltk.corpus import stopwords
import string
from string import punctuation
from nltk.stem import PorterStemmer
ps=PorterStemmer()
```

```
[nltk_data] Error loading stopwords: <urlopen error [WinError 10060] A
[nltk_data]     connection attempt failed because the connected party
[nltk_data]     did not properly respond after a period of time, or
[nltk_data]     established connection failed because connected host
[nltk_data]     has failed to respond>
```

In [36]:
```
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 [37]:
```
transform_text("Hi How are You %&&&& i love dancing")
```

Out[37]:
```
'hi love danc'
```

In [38]:
```
df['transformed_text']=df['email'].apply(transform_text)
```

In [39]:
```
df
```

Out[39]:

	spam	email	num_characters	num_word	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
...	...	...	...	...	...	...

5567	1	This is the 2nd time we have tried 2 contact u...	160	35	4	2nd time tri 2 contact u pound prize 2 claim e...
5568	0	Will _ 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

5169 rows × 6 columns

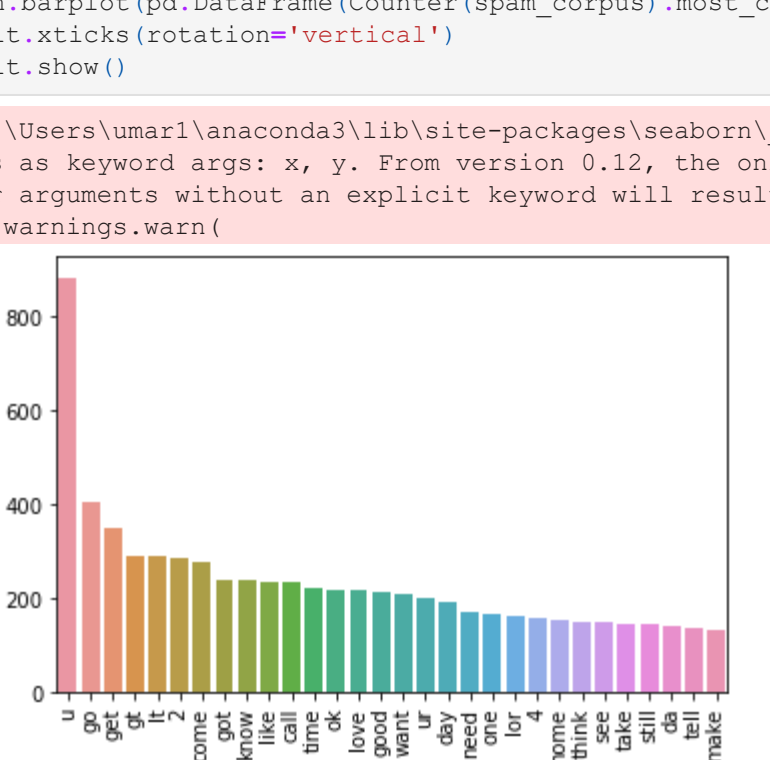
In [40]:
```
#Shows the top 30 most frequent values in case of ham messages

spam_corpus=[]
for i in df[df['spam']==0]['transformed_text'].to_list():
    for word in i.split():
      spam_corpus.append(word)
spam_corpus
```

```
 'activ',
 'got',
 'c',
 'lazi',
 'type',
 'forgot',
 'lect',
 'saw',
 'pouch',
 'like',
 'v',
 'nice',
 'k',
 'text',
 'way',
 'sir',
 'wait',
 'mail',
 'swt',
 'thought',
 'get',
 'tire',
 'littl',
 'thing',
 '4',
 'lovabl',
 'person',
 'coz',
 'somtim',
 'littl',
 'thing',
 'occupi',
 'biggest',
 'part',
 'heart',
 'gud',
 'ni8',
 'know',
 'pl',
 'open',
 'back',
 'ye',
 'see',
 'ya',
 'dot',
 'what',
 'staff',
 'name',
 'take',
 'class',
 'us',
 'call',
 'check',
 'life',
 'begin',
 'qatar',
 'pl',
 'pray',
 'hard',
 'k',
 'delet',
 'contact',
 'sindu',
 'got',
 'job',
 'birla',
 'soft',
 'wine',
 'flow',
 'never',
 'yup',
 'thk',
 'cine',
 'better',
 'co',
 'need',
 '2',
 'go',
 '2',
 'plaza',
 'mah',
 'ok',
 'ur',
 'typic',
 'repli',
 'everywher',
 'dirt',
 'floor',
 'window',
 'even',
 'shirt',
 'sometim',
 'open',
 'mouth',
 'come',
 'flow',
 'dream',
 'world',
 'without',
 'half',
 ...]
```

In [41]:

```
from collections import Counter
sn.barplot(pd.DataFrame(Counter(spam_corpus).most_common(30))[0],pd.DataFrame(Counter(spam_corpus).most_common(
plt.xticks(rotation='vertical')
plt.show()
```

```
C:\Users\umar1\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variabl
es as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing oth
er arguments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```



In [42]:

```
#Shows the top 30 most frequent values in case of spam messages
spam_corpus=[]
for i in df[df['spam']==1]['transformed_text'].to_list():
    for word in i.split():
        spam_corpus.append(word)
spam_corpus
```

```
Out[42]: ['free',
 'entri',
 '2',
 'wkli',
 'comp',
 'win',
 'fa',
 'cup',
 'final',
 'tkt',
 '21st',
 'may',
 'text',
 'fa',
 '87121',
 'receiv',
 'entri',
 'question',
 'std',
 'txt',
 'rate',
 'c',
 'appli',
 '08452810075over18',
 'freemsg',
 'hey',
 'darl',
 '3',
 'week',
 'word',
 'back',
 'like',
 'fun',
 'still',
 'tb',
 'ok',
 'xxx',
 'std',
 'chg',
 'send',
 'rcv',
 'winner',
 'valu',
 'network',
 'custom',
 'select',
 'receivea',
 'prize',
 'reward',
 'claim',
 'call',
 'claim',
 'code',
 'kl341',
 'valid',
 '12',
 'hour',
 'mobil',
 '11',
 'month',
 'u',
 'r',
 'entitl',
 'updat',
 'latest',
 'colour',
 'mobil',
 'camera',
 'free',
 'call',
 'mobil',
 'updat',
 'co',
 'free',
 '08002986030',
 'six',
 'chanc',
 'win',
 'cash',
 '100',
 'pound',
 'txt',
 'csh11',
 'send',
 'cost',
 '6day',
 'tsandc',
 'appli',
 'repli',
 'hl',
 '4',
 'info',
 'urgent',
 '1',
 'week',
 'free',
 'membership',
 'prize',
 'jackpot',
 'txt',
 'word',
 'claim',
 '81010',
 'c',
 'lccltd',
 'pobox',
 '4403ldnw1a7rw18',
 'xxxmobilemovieclub',
 'use',
 'credit',
 'click',
 'wap',
 'link',
 'next',
 'txt',
 'messag',
 'click',
 'http',
 'england',
 'v',
 'macedonia',
 'dont',
 'miss',
 'news',
 'txt',
 'ur',
 'nation',
 'team',
 '87077',
 'eg',
 'england',
 '87077',
 'tri',
 'wale',
 'scotland',
 'poboxox36504w45wq',
 'thank',
 'subscript',
 'rington',
 'uk',
 'mobil',
 'charg',
 'pleas',
 'confirm',
 'repli',
 'ye',
 'repli',
 'charg',
 '07732584351',
 'rodger',
 'burn',
 'msg',
 'tri',
 'call',
 'repli',
 'sm',
 'free',
 'nokia',
 'mobil',
 'free',
 'camcord',
 'pleas',
 'call',
 '08000930705',
 'deliveri',
 'tomorrow',
 'sm',
 'ac',
 'sptv',
 'new',
 'jersey',
 'devil',
 'detroit',
 'red',
 'wing',
 'play',
 'ice',
 'hockey',
 'correct',
 'incorrect',
 'end',
 'repli',
 'end',
 'sptv',
 'congrat',
 '1',
 'year',
 'special',
 'cinema',
 'pass',
 '2',
 'call',
 '09061209465',
 'c',
 'suprman',
 'v',
 'matrix3',
 'starwars3',
 'etc',
 '4',
 'free',
 '150pm',
 'dont',
 'miss',
 'valu',
 'custom',
 'pleas',
 'advis',
 'follow',
 'recent',
 'review',
 'mob',
 'award',
 'bonu',
 'prize',
 'call',
 '09066364589',
 'urgent',
 'ur',
 'award',
 'complimentari',
 'trip',
 'eurodisinc',
 'trav',
 'aco',
 'entry41',
 'claim',
 'txt',
 'di',
 '87121',
 'morefrmmob',
 'shracomorsglsuplt',
 '10',
 'ls1',
 '3aj',
 'hear',
 'new',
 'come',
 'ken',
 'stuff',
 'pleas',
 'call',
 'custom',
 'servic',
 'repres',
 '0800',
 '169',
 '6031',
 'guarante',
 'cash',
 'prize',
 'free',
 'rington',
 'wait',
 'collect',
 'simpli',
 'text',
 'password',
 '85069',
 'verifi',
 'get',
 'usher',
 'britney',
 'fml',
 'gent',
 'tri',
 'contact',
 'last',
 'weekend',
 'draw',
 'show',
 'prize',
 'guarante',
 'call',
 'claim',
 'code',
 'k52',
 'valid',
 '12hr',
 '150ppm',
 'winner',
 'u',
 'special',
 'select',
 '2',
 'receiv',
 '4',
 'holiday',
 'flight',
 'inc',
 'speak',
 'live',
 'oper',
 '2',
 'claim',
 'privat',
 '2004',
 'account',
 'statement',
 '07742676969',
 'show',
 '786',
 'unredeem',
 'bonu',
 'point',
 'claim',
 'call',
 '08719180248',
 'identifi',
 'code',
 '45239',
 'expir',
 'urgent',
 'mobil',
 'award',
 'bonu',
 'caller',
 'prize',
 'final',
 'tri',
 'contact',
 'u',
 'call',
 'landlin',
 '09064019788',
 'box42wr29c',
 '150ppm',
 'today',
 'voda',
 'number',
 'end',
 '7548',
 'select',
 'receiv',
 '350',
 'award',
 'match',
 'pleas',
 'call',
 '08712300220',
 'quot',
 'claim',
 'code',
 '4041',
 'standard',
 'rate',
 'app',
 'sunshin',
 'quiz',
 'wkli',
 'q',
 'win',
 'top',
 'soni',
 'dvd',
 'player',
 'u',
 'know',
 'countri',
 'algarv',
 'txt',
 'ansr',
 '82277',
 'sp',
 'tyron',
 'want',
 '2',
 'get',
 'laid',
 'tonight',
 'want',
 'real',
 'dog',
 'locat',
 'sent',
 'direct',
 '2',
 'ur',
 'mob',
 'join',
 'uk',
 'largest',
 'dog',
 'network',
 'bt',
 'txting',
 'gravel',
 '69888',
 'nt',
 'ec2a',
 '150p',
 'rcv',
 'msg',
 'chat',
 'svc',
 'free',
 'hardcor',
 'servic',
 'text',
 'go',
 '69988',
 'u',
 'get',
 'noth',
 'u',
 'must',
 'age',
 'verifi',
 'yr',
 'network',
 'tri',
 'freemsg',
 'repli',
 'text',
 'randi',
 'sexi',
 'femal',
 'live',
 'local',
 'luv',
 'hear',
 'netcollex',
 'ltd',
 '08700621170150p',
 'per',
 'msg',
 'repli',
 'stop',
 'end',
 'custom',
 'servic',
 'annonc',
 'new',
 'year',
 'deliveri',
 'wait',
 'pleas',
 'call',
 '07046744435',
 'arrang',
 'deliveri',
 'winner',
 'u',
 'special',
 'select',
 '2',
 'receiv',
 'cash',
 '4',
 'holiday',
 'flight',
 'inc',
 'speak',
 'live',
 'oper',
 '2',
 'claim',
 '0871277810810',
 'stop',
 'bootydeli',
 'invit',
 'friend',
 'repli',
 'see',
 'stop',
 'send',
 'stop',
 'frnd',
 '62468',
 'bangbab',
 'ur',
 'order',
 'way',
 'u',
 'receiv',
 'servic',
 'msg',
 '2',
 'download',
 'ur',
 'content',
 'u',
 'goto',
 'wap',
 'bangb',
 'tv',
 'ur',
 'mobil',
 'menu',
 'urgent',
 'tri',
 'contact',
 'last',
 'weekend',
 'draw',
 'show',
 'prize',
 'guarante',
 'call',
 'claim',
 'code',
 's89',
 'valid',
 '12hr',
 'pleas',
 'call',
 'custom',
 'servic',
 'repres',
 'freephon',
 '0808',
 '145',
 '4742',
 'guarante',
 'cash',
 'prize',
 'uniqu',
 'enough',
 'find',
 '30th',
 'august',
 '500',
 'new',
 'mobil',
 '2004',
 'must',
 'go',
 'txt',
 'nokia',
 '89545',
 'collect',
 'today',
 '2optout',
 'u',
 'meet',
 'ur',
 'dream',
 'partner',
 'soon',
 'ur',
 'career',
 '2',
 'flyng',
 'start',
 '2',
 'find',
 'free',
 'txt',
 'horo',
 'follow',
 'ur',
 'star',
 'sign',
 'horo',
 'ari',
 'text',
 'meet',
 'someon',
 'sexi',
 'today',
 'u',
 'find',
 'date',
 'even',
 'flirt',
 'join',
 '4',
 '10p',
 'repli',
 'name',
 'age',
 'eg',
 'sam',
 '25',
 '18',
 'recd',
 'thirtyeight',
 'penc',
 'u',
 '447801259231',
 'secret',
 'admir',
 'look',
 '2',
 'make',
 'contact',
 'r',
 'reveal',
 'think',
 'ur',
 '09058094597',
 'congratul',
 'ur',
 'award',
 '500',
 'cd',
 'voucher',
 '125gift',
 'guarante',
 'free',
 'entri',
 '2',
 '100',
 'wkli',
 'draw',
 'txt',
 'music',
 '87066',
 'tnc',
 'tri',
 'contact',
 'repli',
 'offer',
 'video',
 'handset',
 '750',
 'anytim',
 'network',
 'min',
 'unlimit',
 'text',
 'camcord',
 'repli',
 'call',
 '08000930705',
 'hey',
 'realli',
 'horni',
 'want',
 'chat',
 'see',
 'nake',
 'text',
 'hot',
 '69698',
 'text',
 'charg',
 '150pm',
 'unsubscrib',
 'text',
 'stop',
 '69698',
 'ur',
 'rington',
 'servic',
 'chang',
 '25',
 'free',
 'credit',
 'go',
 'choos',
 'content',
 'stop',
 'txt',
 'club',
 'stop',
 '87070',
 'club4',
 'po',
 'box1146',
 'mk45',
 '2wt',
 'rington',
 'club',
 'get',
 'uk',
 'singl',
 'chart',
 'mobil',
 'week',
 'choos',
 'top',
 'qualiti',
 'rington',
 'messag',
 'free',
 'charg',
 'hmv',
 'bonu',
 'special',
 '500',
 'pound',
 'genuin',
 'hmv',
 'voucher',
 'answer',
 '4',
 'easi',
 'question',
 'play',
 'send',
 'hmv',
 '86688',
 'info',
 'custom',
 'may',
 'claim',
 'free',
 'camera',
 'phone',
 'upgrad',
 'pay',
 'go',
 'sim',
 'card',
 'loyalti',
 'call',
 '0845',
 '021',
 'end',
 'c',
 'appli',
 'sm',
 'ac',
 'blind',
 'date',
 '4u',
 'rodds1',
 'aberdeen',
 'unit',
 'kingdom',
 'check',
 'http',
 'sm',
 'blind',
 'date',
 'send',
 'hide',
 'themob',
 'check',
 'newest',
 'select',
 'content',
 'game',
 'tone',
 'gossip',
 'babe',
 'sport',
 'keep',
 'mobil',
 'fit',
 'funki',
 'text',
 'wap',
 '82468',
 'think',
 'ur',
 'smart',
 'win',
 'week',
 'weekli',
 'quiz',
 'text',
 'play',
 '85222',
 'cs',
 'winnersclub',
 'po',
 'box',
 '84',
 'm26',
 '3uz',
 'decemb',
 'mobil',
 'entitl',
 'updat',
 'latest',
 'colour',
 'camera',
 'mobil',
 'free',
 'call',
 'mobil',
 'updat',
 'co',
 'free',
 '08002986906',
 'call',
 'germani',
 '1',
 'penc',
 'per',
 'minut',
 'call',
 'fix',
 'line',
 'via',
 'access',
 'number',
 '0844',
 '861',
 '85',
 'prepay',
 'direct',
 'access',
 'valentin',
 'day',
 'special',
 'win',
 'quiz',
 'take',
 'partner',
 'trip',
 'lifetim',
 'send',
 'go',
 '83600',
 'rcvd',
 'fanci',
 'shag',
 'txt',
 'xxuk',
 'suzi',
 'txt',
 'cost',
 'per',
 'msg',
 'tnc',
 'websit',
 'x',
 'ur',
 'current',
 '500',
 'pound',
 'maxim',
 'ur',
 'send',
 'cash',
 '86688',
 'cc',
 '08708800282',
 'xma',
 'offer',
 'latest',
 'motorola',
 'sonyericsson',
 'nokia',
 'free',
 'bluetooth',
 'doubl',
 'min',
 '1000',
 'txt',
 'orang',
 'call',
 'mobileupd8',
 '08000839402',
 'discount',
 'code',
 'rp176781',
 'stop',
 'messag',
 'repli',
 'stop',
 'custom',
 'servic',
 '08717205546',
 'thank',
 'rington',
 'order',
 'refer',
 't91',
 'charg',
 'gbp',
 '4',
 'per',
 'week',
 'unsubscrib',
 'anytim',
 'call',
 'custom',
 'servic',
 '09057039994',
 'doubl',
 'min',
 'txt',
 '4',
 '6month',
 'free',
 'bluetooth',
 'orang',
 'avail',
 'soni',
 'nokia',
 'motorola',
 'phone',
 'call',
 'mobileupd8',
 '08000839402',
 '4mth',
 'half',
 'price',
 'orang',
 'line',
 'rental',
 'latest',
 'camera',
 'phone',
 '4',
 'free',
```