Task 1. Print the number of words starting with consonants and the number of words starting with vowels in all messages.

```
The number of words that start with vowels are:
20940

The number of words that start with consonents are:
60223
```

Task 2. Compute the percentage of capitalized words in spam and ham messages.

```
The percenatge of capitalized words in spam messages
7.077805261974365

The percentage of capiatalized words in ham messages
5.044682299436583
```

Task 3. List all email ids and phone/mobile numbers and their respective counts in the dataset.
Compute percentage for spam and ham messages for both email and mobile numbers.
List of email ids and phone numbers. (Samples)

```
0871474032
0906636192
0800093070
customersqueries@netvision.uk.com
0871520527
0800298690
0905009004
0906345813
0845281007
0870141701
0905009004
0800083940
0871520365
0906636431
0906401530
0906121323
0906179012
0906121323
0871240502
0909535030
0871246032
0906122106
0906638061
0871750738
0871872627
0871570502
207-083-6089
0905809458
121 2025050
0800298690
1705572711
0905000033
0906401729
0844835005
```

```
0870002117
0906401100
0800195038
0906636223
0905009004
0871471237
0906506912
0909972648
0906506915
0122358533
0800083940
0871873803
0905624215
0906120946
0087040504
0871872697
tddnewsletter@emc1.co.uk
0906636875
0800083940
0871240290
0870141701
0871212025
0905809456
0905809800
0870284062
0871872020
0871570502
0906661266
0906170144
info@txt82228.co.uk
0870405040
0906345813
0870609179
0905624215
```

Number of emails ids and phone numbers

```
The number of phone Numbers are:
459

The number of email ids are:
6
```

Percentage of phone numbers and email ids in spam and ham messages
Formula (Total number of phone numbers and email ids in spam messages/ total number of phone numbers and email ids)

```
The percentage of phone numbers or email Ids present in spam messages:
99.56989247311829

The percentage of phone numbers or email Ids present in ham messages:
0.43010752688172044
```

Percentage of spam and ham messages with phone numbers or email ids.
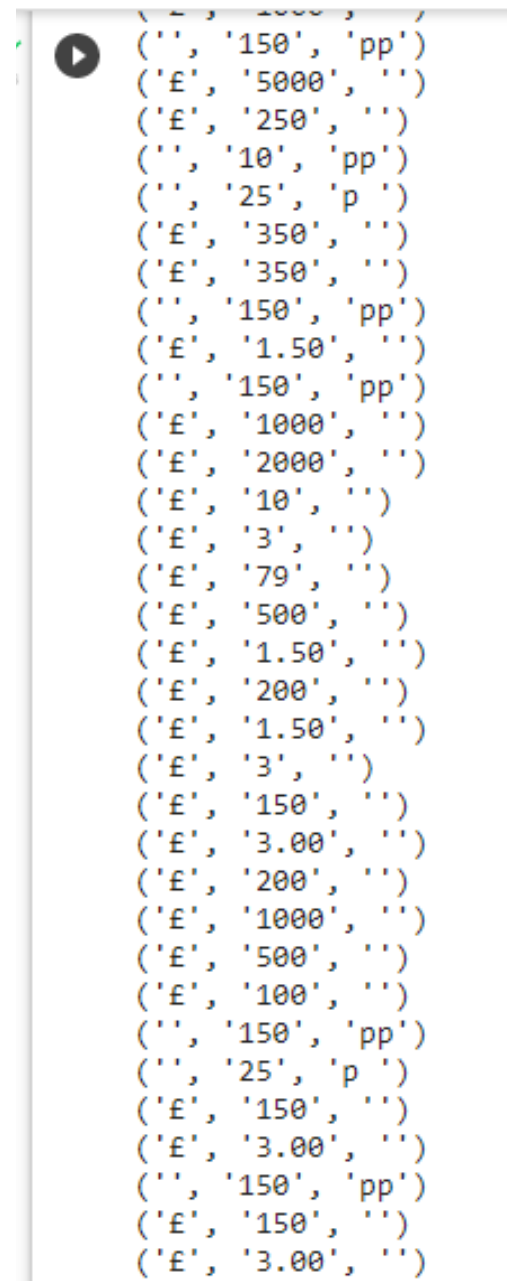Formula = Total number of spam messages with a phone number or email ids/ Total number of spam messages

```
The percentage of spam messages with email Ids or phone numbers
57.42971887550201

The percentage of spam messages with email Ids or phone numbers
0.04145077720207254
```

Task 4. List all monetary quantities, e.g., £1000. Compute the percentage of spam and ham messages with monetary quantities.

List of monetary quantities. (Samples)

```
( 2 ,  1000 ,    )
('', '150', 'pp')
('£', '5000', '')
('£', '250', '')
('', '10', 'pp')
('', '25', 'p ')
('£', '350', '')
('£', '350', '')
('', '150', 'pp')
('£', '1.50', '')
('', '150', 'pp')
('£', '1000', '')
('£', '2000', '')
('£', '10', '')
('£', '3', '')
('£', '79', '')
('£', '500', '')
('£', '1.50', '')
('£', '200', '')
('£', '1.50', '')
('£', '3', '')
('£', '150', '')
('£', '3.00', '')
('£', '200', '')
('£', '1000', '')
('£', '500', '')
('£', '100', '')
('', '150', 'pp')
('', '25', 'p ')
('£', '150', '')
('£', '3.00', '')
('', '150', 'pp')
('£', '150', '')
('£', '3.00', '')
```

Number of monetary quantities

Percentage of monetary quantities in spam and ham messages
Formula (Total number of phone numbers and email ids in spam messages/ total number of
phone numbers and email ids)

Percentage of spam and ham messages with monetary quantities.
Formula = Total number of spam messages with a phone number or email ids/ Total number of
spam messages

```
The number of monetary quantities are:
447

The percentage of monetray quantities in spams:
95.9731543624161

The percentage of monetary quantities in hams:
4.026845637583892

The percentage of spam messages with monetary quantities
42.03480589022758

The percentage of ham messages with monetary quantities
0.2694300518134715
```

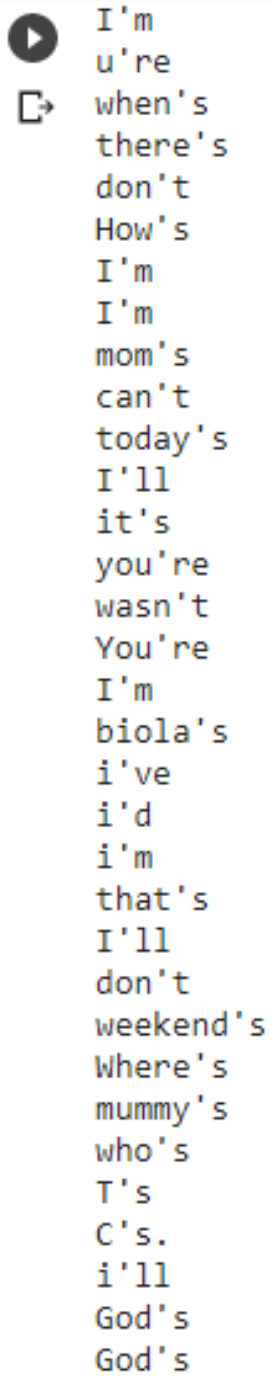5. Count and print all emoticons (use NLTK only) in the dataset.

```
:-P
:-)
:(
:)
d:
:)
d:
:8
:)
:-)
;-(
:)
:)
:)
:)
```

The number of emoticons.

```
The number of emoticons in the dataset is:
556
```

6. Print a list of words with clitics

```
I'm
u're
when's
there's
don't
How's
I'm
I'm
mom's
can't
today's
I'll
it's
you're
wasn't
You're
I'm
biola's
i've
i'd
i'm
that's
I'll
don't
weekend's
Where's
mummy's
who's
T's
C's.
i'll
God's
God's
```

The number of words with clitics:

```
The number of clitics in the dataset:
1672
```

Task 7. Print the messages and number of messages starting with a given word as a custom input.

Sample.

Given word: hi.

```
Hi. I'm sorry i missed your call. Can you pls call back.
Hi, Mobile no.  &lt;#&gt;  has added you in their contact list on www.fullonsms.com It s a great place to send free sms to people For more visit fullonsms.com
Hi! This is Roger from CL. How are you?
Hi happy birthday. Hi hi hi hi hi hi hi
Hi! You just spoke to MANEESHA V. We'd like to know if you were satisfied with the experience. Reply Toll Free with Yes or No.
Hi, can i please get a  &lt;#&gt;  dollar loan from you. I.ll pay you back by mid february. Pls.
Hi Chachi tried calling u now unable to reach u .. Pl give me a missed cal once u c tiz msg  Kanagu
Hi Harish's rent has been transfred to ur Acnt.
HI DARLIN I HOPE YOU HAD A NICE NIGHT I WISH I HAD COME CANT WAIT TO SEE YOU LOVE FRAN PS I WANT DIRTY ANAL SEX AND I WANT A 10 MAN GANG BANG
Hi' Test on  &lt;#&gt; rd ....
Hi babe its Jordan, how r u? Im home from abroad and lonely, text me back if u wanna chat xxSP visionsms.com Text stop to stopCost 150p 08712400603
Hi. Happy New Year. I dont mean to intrude but can you pls let me know how much tuition you paid last semester and how much this semester is. Thanks
Hi neva worry bout da truth coz the truth will lead me 2 ur heart. It®s the least a unique person like u deserve. Sleep tight or morning
Hi, my love! How goes that day? Fuck, this morning I woke and dropped my cell on the way down the stairs but it seems alright ... *phews* I miss you !
HI ITS KATE CAN U GIVE ME A RING ASAP XXX
Hi there. We have now moved in2 our pub . Would be great 2 c u if u cud come up.
HI DARLIN I FINISH AT 3 DO U 1 2 PICK ME UP OR MEET ME? TEXT BACK ON THIS NUMBER LUV KATE XXX
HI DARLIN HOW WAS WORK DID U GET INTO TROUBLE? IJUST TALKED TO YOUR MUM ALL MORNING! I HAD A REALLY GOOD TIME LAST NIGHT IM GOIN OUT SOON BUT CALL ME IF U CAN
Hi darlin did youPhone me? Im atHome if youwanna chat.
Hi darlin i cantdo anythingtomorrow as myparents aretaking me outfor a meal. when are u free? Katexxx
Hi Petey!noi®m ok just wanted 2 chat coz avent spoken 2 u 4 a long time-hope ur doin alrite.have good nit at js love ya am.x
Hi 07734396839 IBH Customer Loyalty Offer: The NEW NOKIA6600 Mobile from ONLY £10 at TXTAUCTION!Txt word:START to No:81151 & get Yours Now!4T&
Hi there, 2nights ur lucky night! Uve been invited 2 XCHAT, the Uks wildest chat! Txt CHAT to 86688 now! 150p/MsgrcvdHG/Suite342/2Lands/Row/W1J6HL LDN 18yrs
Hi, wlcome back, did wonder if you got eaten by a lion or something, nothing much
Hi! You just spoke to MANEESHA V. We'd like to know if you were satisfied with the experience. Reply Toll Free with Yes or No.
Hi. I'm always online on yahoo and would like to chat with you someday
Hi juan. Im coming home on fri hey. Of course i expect a welcome party and lots of presents. Ill phone u when i get back. Loads of love nicky x x x x x x x x x x
Hi dude hw r u da realy mising u today
Hi di is yijue we're meeting at 7 pm at esaplanade tonight.
Hi dear we saw dear. We both are happy. Where you my battery is low
Hi baby ive just got back from work and i was wanting to see u allday! I hope i didnt piss u off on the phone today. If u are up give me a call xxx
Hi:)did you asked to waheeda fathima about leave?
Hi good mornin.. Thanku wish u d same..
```

The number of messages with input word in beginning
99

8. Print the sentences and number of sentences ending with a given word as a custom input.
Sample.
Given word = No

```
Thanks for your subscription to Ringtone UK your mobile will be charged £5/month Please confirm by replying YES or NO.
As a valued customer, I am pleased to advise you that following recent review of your Mob No.
Your Mobile No.
No no.
Reply Toll Free with Yes or No.
Lol no.
MY NO.
: Your Mobile No.
I'm gonna say no.
No.
Oh yes I can speak txt 2 u no!
No.
No.
Its ok., i just askd did u knw tht no?
No.
If you're thinking of lifting me one then no.
No.
Nothing, i got msg frm tht unknown no..
No.
Si.como no?
Hi, Mobile no.
Do u knw dis no.
Reply Toll Free with Yes or No.
There r many model..sony ericson also der.. &lt;#&gt; ..it luks good bt i forgot modl no
No.
As a valued customer, I am pleased to advise you that following recent review of your Mob No.
As a valued customer, I am pleased to advise you that following recent review of your Mob No.
No.
Disconnect the call,coz it iz an attempt of 'terrorist' to make use of the sim card no.
Reply Toll Free with Yes or No.
No, its true..k,Do u knw dis no.
No!
No.
No
```
✓ 4s   completed at 10:29 PM

The number of sentences with no in the end.

```
The number of messages with input word in the end of sentence
47
```

9. Spam/Ham classification
– Utilizing the features in 2, 3, and 4, identify the category of the message.
Note: For a given message, identify whether it contains features as per points 2, 3, and 4.
Propose heuristics ([only 2], [only 3], [only 4], [2,3], [2,4], [3,4], or [2,3,4]) to obtain the
maximum accuracy for the spam/ham classification. You should report accuracy for all
cases.

Accuracies.

```
The accuracy by considering only 2.
72.34386216798278

The accuracy by considering only 3.
94.25699928212491

The accuracy by considering only 4.
91.99569274946158

The accuracy by considering 2 and 3.
89.64465183058148

The accuracy by considering 2 and 4.
89.321608040201

The accuracy by considering 3 and 4.
90.50610193826274

The accuracy by considering 2, 3 and 4
88.35247666905958

The winning heuristic is 3 with accuracy 94.25699928212491
```
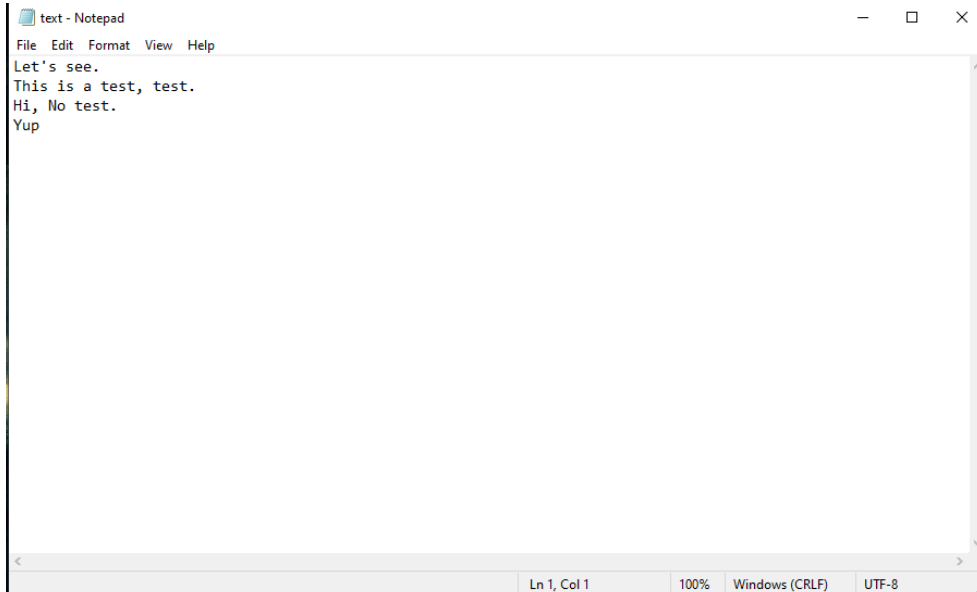
Task 10. Given a word (string) and a .txt file as input, print the count of that word and sentences containing that word in the input file.

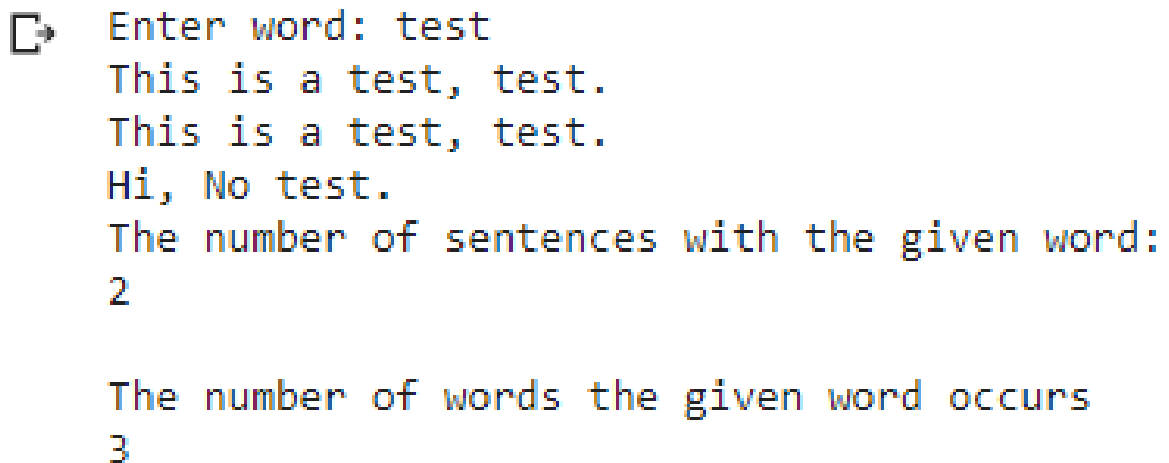Sample.
Input File


```
text - Notepad                                          −   □   ×
File  Edit  Format  View  Help
Let's see.
This is a test, test.
Hi, No test.
Yup
                                                          
          Ln 1, Col 1        100%   Windows (CRLF)    UTF-8
```

Output.


```
Enter word: test
This is a test, test.
This is a test, test.
Hi, No test.
The number of sentences with the given word:
2

The number of words the given word occurs
3
```