**Text Message Predictive Analysis**

This is the report of the model “Text Message Predictive Analysis”. I did the backup of the text messages of my cell phone using the software “SMS Backup & Restore”. The backup was in xml format and I am using this as my data set for analysis.

I made a word cloud of the text messages and made a model which predicts the contact who might have sent the message. The prediction works by matching the input message with the body of the text message.

**Program:**

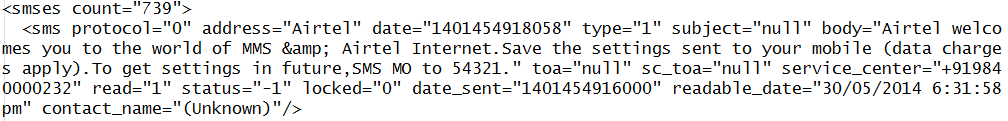
We first find the working directory and set it to our workspace.



We now take the backup which we made as our input. The name of the backup file is **sms.xml**. We save the input in a file called **file.** We then find the **xmlroot** and save that in a variable called **root.**



The **root** file is:



We now get all the received messages and store it in **s1.**





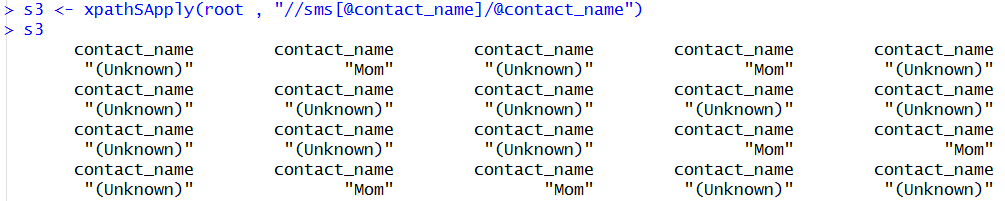
We now get all the messages from the contact named **Mom** and store it in **s2.**





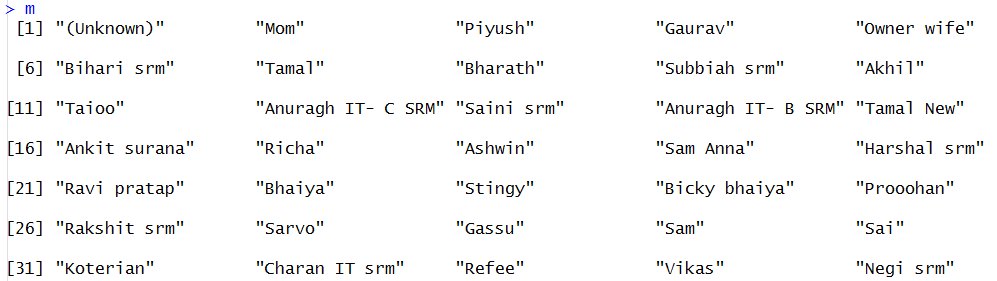
We now get all the contacts from the **file** and store it in **s3**.





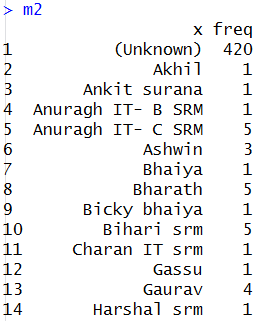
We now get the unique contacts from **s3** and store it in **m.**





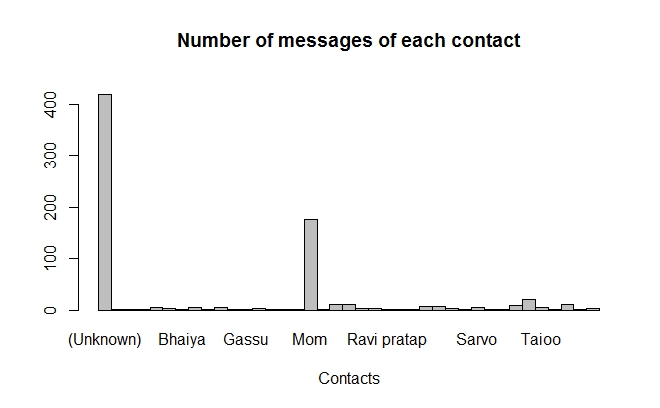
Now, we get the count of messages of each contact i.e. the number of messages of each contact and store it in **m2.**





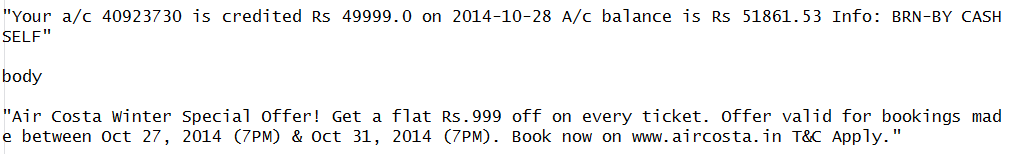
We now plot a bar graph for the contacts and their messages.





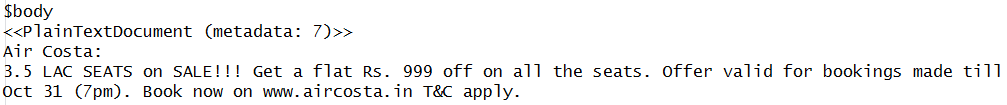
Now, we get the body of the text messages and store it in **s4**.



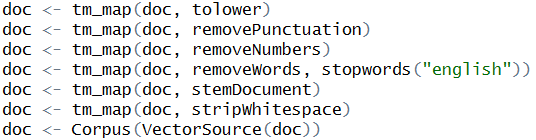


We now create a VCorpus of the body of messages and store it in **doc.**



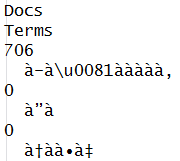


Now we trim the document by making it lower, removing the punctuations, removing the numbers etc.



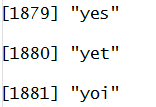
We, now make a TermDocumentMatrix of the corpus and store it in **tdm**.





We then get the unique terms in the tdm and store them in **tem.**

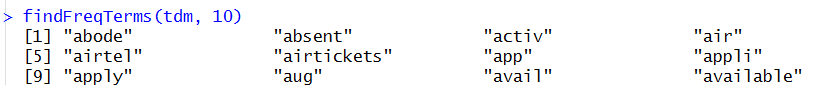




We can then calculate the frequent terms. Here I calculate the terms which occur

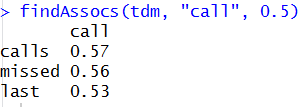
more than 10 times.





We can find the terms which are associated to the term **call.**

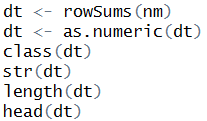


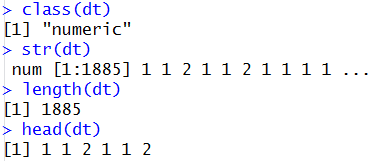


We now convert the TermDocumentMatrix **tdm** to matrix **nm.**

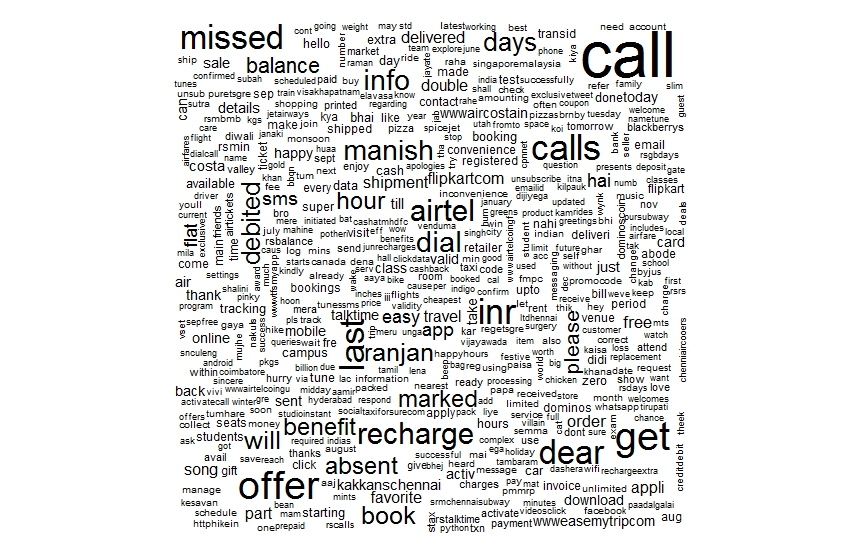


We now calculate the number of occurrences of each term in the matrix **nm** and save them in **dt**. We convert them into numeric.



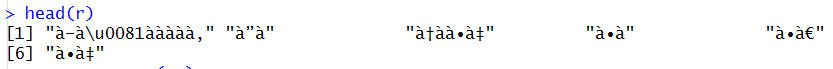


We create a WordCloud of the terms in the VCorpus.



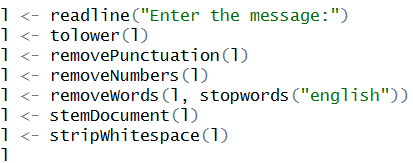
We store the row names of the matrix **nm** in **r.** The row names are the unique terms. So now **r** contains the unique terms of the VCorpus.

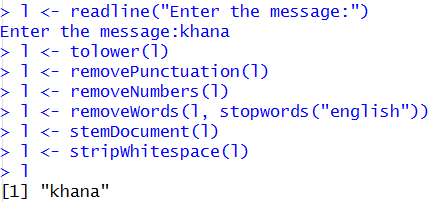




We now take the input message from the user. The system gives the names of the contacts who could have sent the message.

We take the input message and trim the message as we trimmed the VCorpus.





The system now runs the following code and finds out the contacts who could have sent that message.

