



## 1.1 Problem Statement

- Stage-Status phrases for Sales & Support in Slam Activity Remarks
- Using sentiment analysis to find who puts the most positive words and negative words

## 1.2 Data Description

The text data used is the activities entered by the **Sales and the Collection team** in SLAM. 27,5488 activities with 8 variables namely Account Name, Activity Name, Activity Description, Activity Outcome, Event location, Activity Purpose, Users and Activity Direction Name was extracted from SLAM.

## 1.3 Objective

We are looking out for Common terms and relationship between words to suggest Stage wise phrases for Sales & Collection in Slam Activity.

# Table of Contents

- 1 [Introduction](#)
- 2 [Preparation](#)
- 3 [Glimpse of Data](#)
- 4 [Who has entered the most activities?](#)
- 5 [Which User entered long Sentences?](#)
- 6 [Tokenization](#)
  - 6.1 [Removing the Stop words](#)
- 7 [Top Ten most Common Words](#)
  - 7.1 [WordCloud of the Common Words](#)
- 8 [Word Analysis of User](#)
- 9 [Parts of Speech analysis](#)
- 10 TF-IDF
  - 10.1 The Math
  - 10.2 [Twenty Most Important words for the Most Active Characters](#)
- 11 [Similarities](#)
- 12 [Most Common Bigrams](#)
- 13 [Most Common Trigrams](#)
- 14 [Relationship among words](#)
- 15 [Relationship with the word](#)
- 16 [Sentiment Analysis using AFINN sentiment lexicon](#)
- 17 [Positive and Not So Positive Words of Characters](#)
  - 17.1 [Sentiment Analysis Words - Fear for Sagar](#)
  - 17.2 [Sentiment Analysis Words - Joy for Jyotirmay Parida](#)
- 18 [Sentiment Analysis using NRC Sentiment lexicon](#)
  - 18.1 [Sentiment Analysis Words - Trust for Sagar](#)
  - 18.2 [Sentiment Analysis Words - Anger for Sagar](#)
  - 18.3 [Sentiment Analysis Words - Joy for Sagar](#)
  - 18.4 [Sentiment Analysis Words - Disgust for Sagar](#)
- 19 [Stage phrases for Sales & Collection in Slam Activity based on Trigram and Bigrams](#)

## 1 Introduction

The text data used is the activities entered by the Sales and the Collection team in SLAM. The problem observed was, “**Activity**” being an open field, the user use to enter a descriptive summary in the tab (which included copy pasting the whole email conversation). Hence making it difficult to draw insights from the field.

We are looking out for common term and relationship between words to suggest Stage wise phrases for Sales & Collection in Slam Activity.

## 2 Preparation

Following are the data cleaning processes undertaken before processing:

1. Convert all the text to lowercase
2. Removed Punctuation
3. Removed Numbers
4. Remove Whitespace
5. Removed Stop words
6. Removed Custom Stop words

### 2.1 Library used

```
library(tidyverse)
```

```
library(tidytext) # text manipulation
```

```
library(wordcloud) # word cloud
```

```
library(stringr) #string manipulation
```

```
library(igraph)
```

```
library(gggraph)
```

```
library(widyr)
```

```
library(broom)
```

```
library(DT)
```

```
library(irlba)
```

```
library(topicmodels) # for LDA topic modelling
```

```
library(tm) # general text mining functions, making document term matrixes
```

```
library(caret)
```

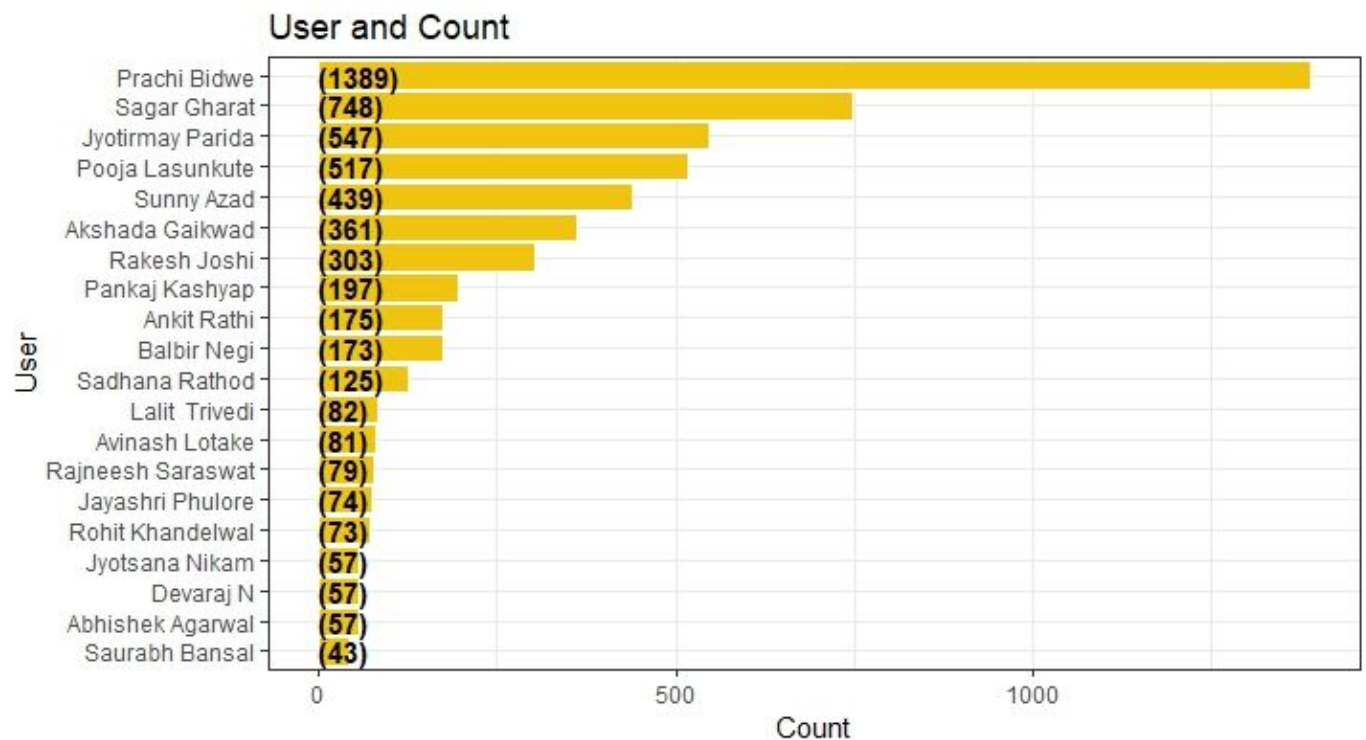
```
library(glmnet)
```

## 3 Glimpse of Data

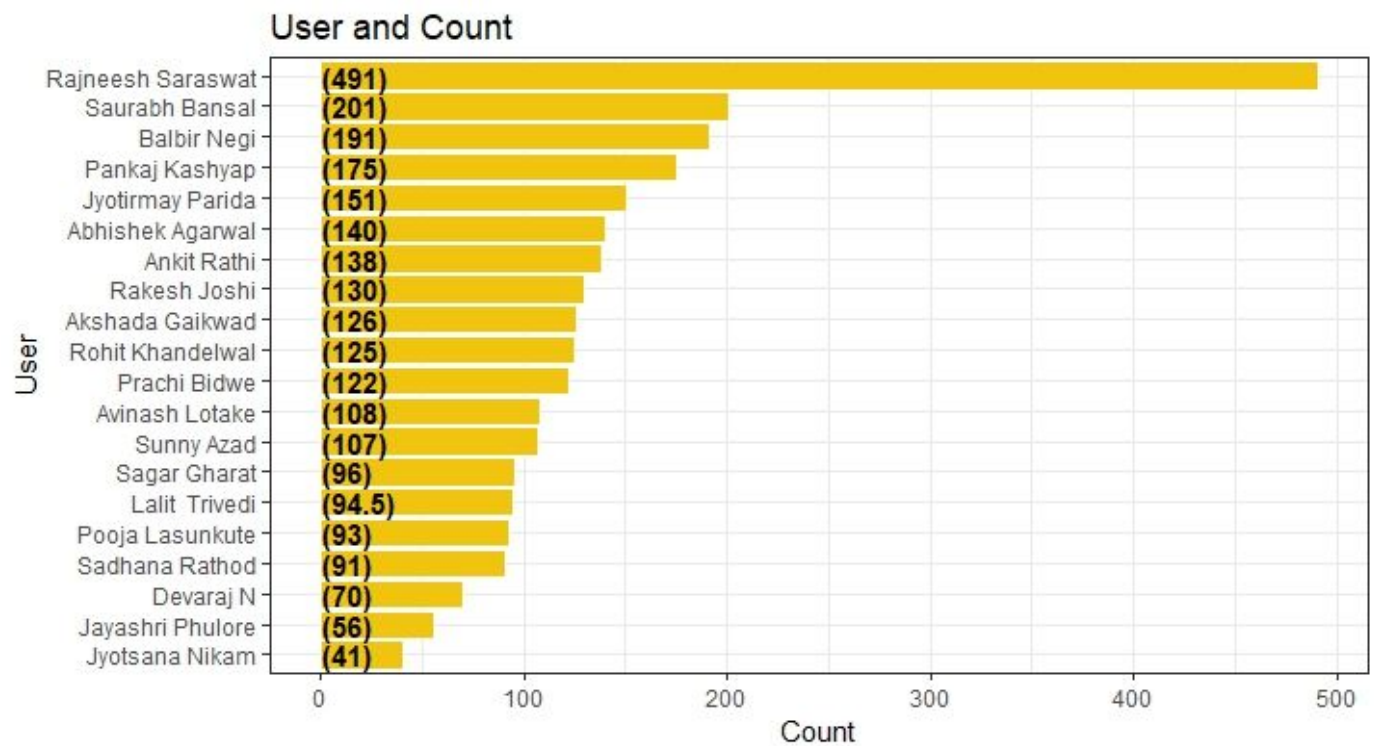
```
> glimpse(data)
observations: 6,803
variables: 8
$ AccountName      <chr> " Hindustan Motor Finance Corporation...
$ ActivityName     <chr> "sms-pme", "100 co founder", "FS Call...
$ ActivityOutcome  <chr> "Successful", "Successful", "Successf...
$ ActivityDescription <chr> "Client says management has to take d...
$ EventLocation    <chr> "TamilNadu", "North", "Pune", "Pune",...
$ ActivityPurposeName <chr> "Sales Assist", "Sales Assist", "Fiel...
$ User             <chr> "Devaraj N", "Rakesh Joshi", "Jyotsan...
$ ActivityDirectionName <chr> "Inbound", "Outbound", "Outbound", "O..."
```

#### 4 Who has entered the most activities?

The plot shows the users who has entered the most number of activities.



## 5 Which User spoke long Sentences?



## 6 Tokenization

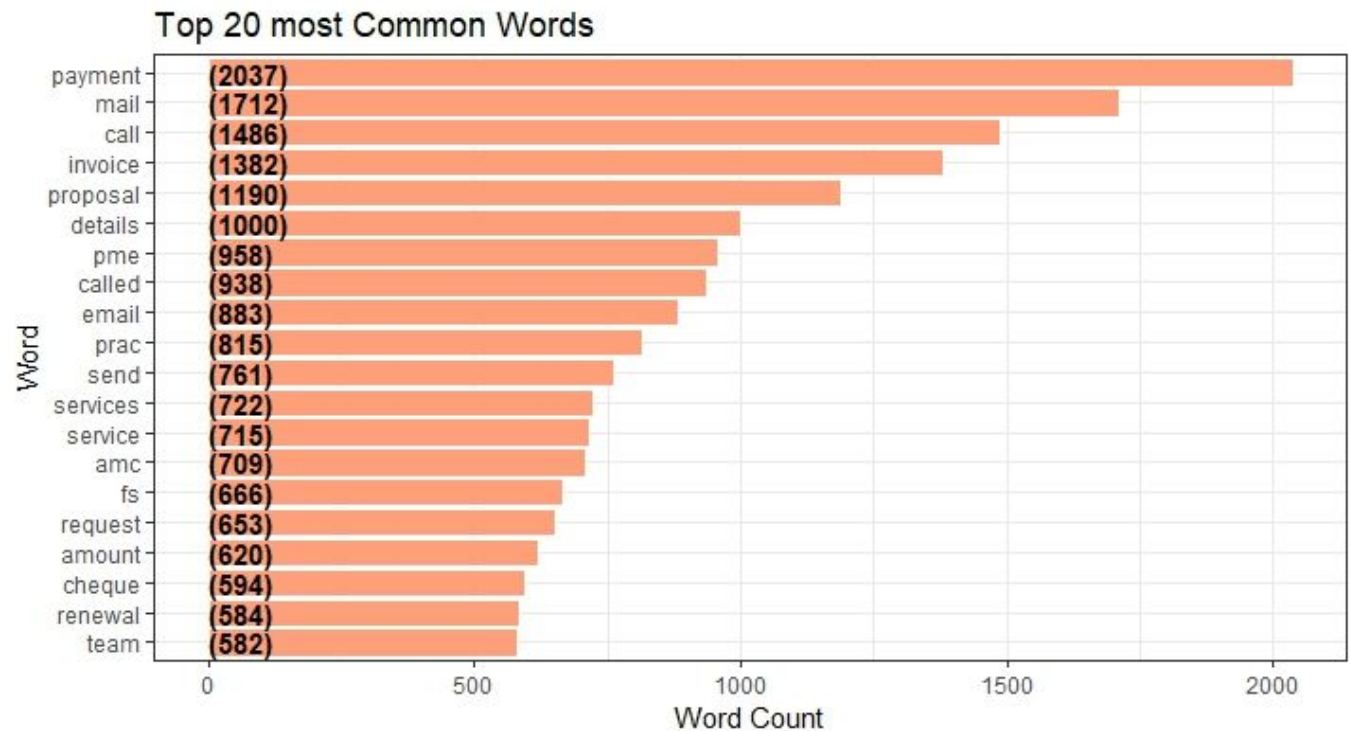
We break the text into individual tokens which are simply individual words. This process is called tokenization.

### 6.1 Removing the Stop words

Removed the most Commonly occurring words in English language.

```
> stop_words %>% head(20)
# A tibble: 20 x 2
  word      lexicon
  <chr>    <chr>
1 a        SMART
2 a's      SMART
3 able     SMART
4 about    SMART
5 above    SMART
6 according SMART
7 accordingly SMART
8 across   SMART
9 actually SMART
10 after    SMART
11 afterwards SMART
12 again    SMART
13 against  SMART
14 ain't    SMART
15 all      SMART
16 allow    SMART
17 allows   SMART
18 almost   SMART
19 alone    SMART
20 along    SMART
```

## 7 Top Ten most Common Words



## 7.1 WordCloud of the Common Words



## 8 Word Analysis of User(s)

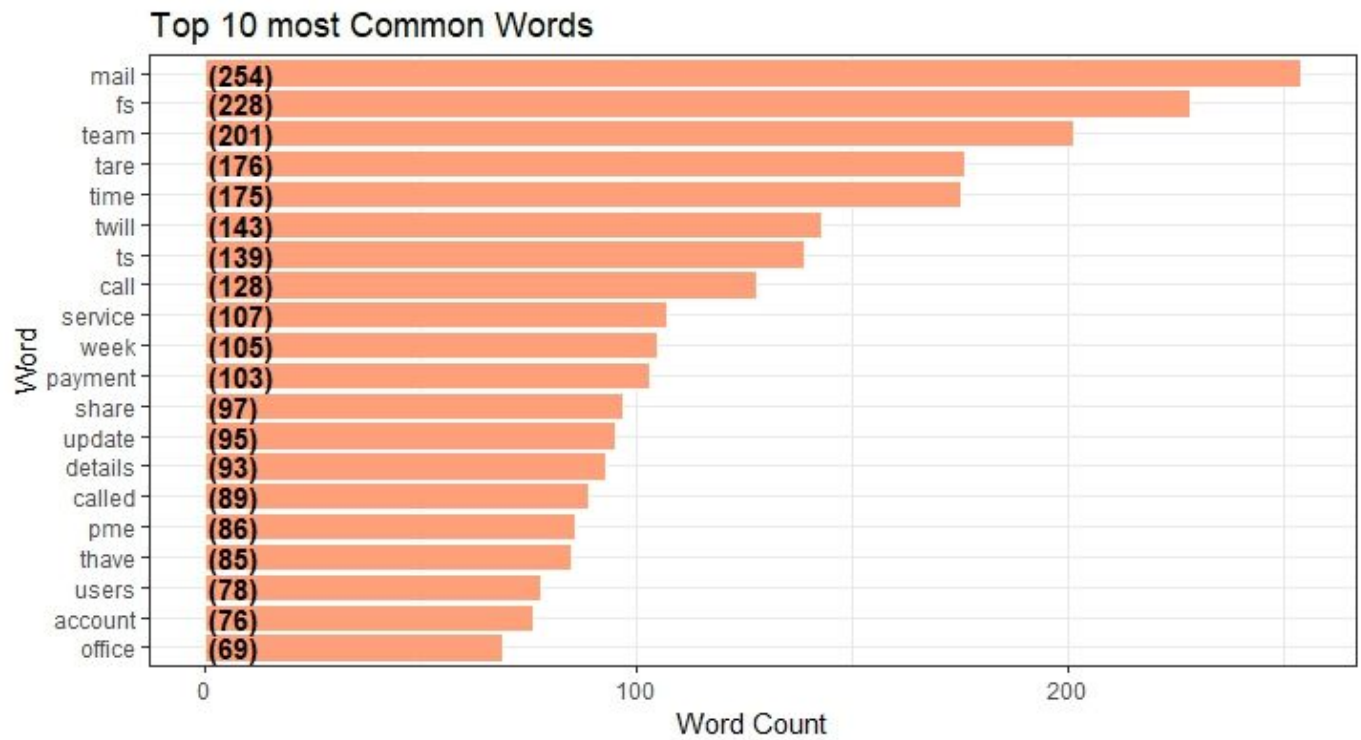
We explore the various words spoken by the specific User

- Sagar's favourite words are payment,release,called, invoice and Status.





- Jyotirmay Parida's favourite words are mail, fs, team, service and Week.



## 9 Parts of Speech analysis

### 9.1 Word Cloud - Most common Adjectives used





## 9.2 Word Cloud - Transitive Verb



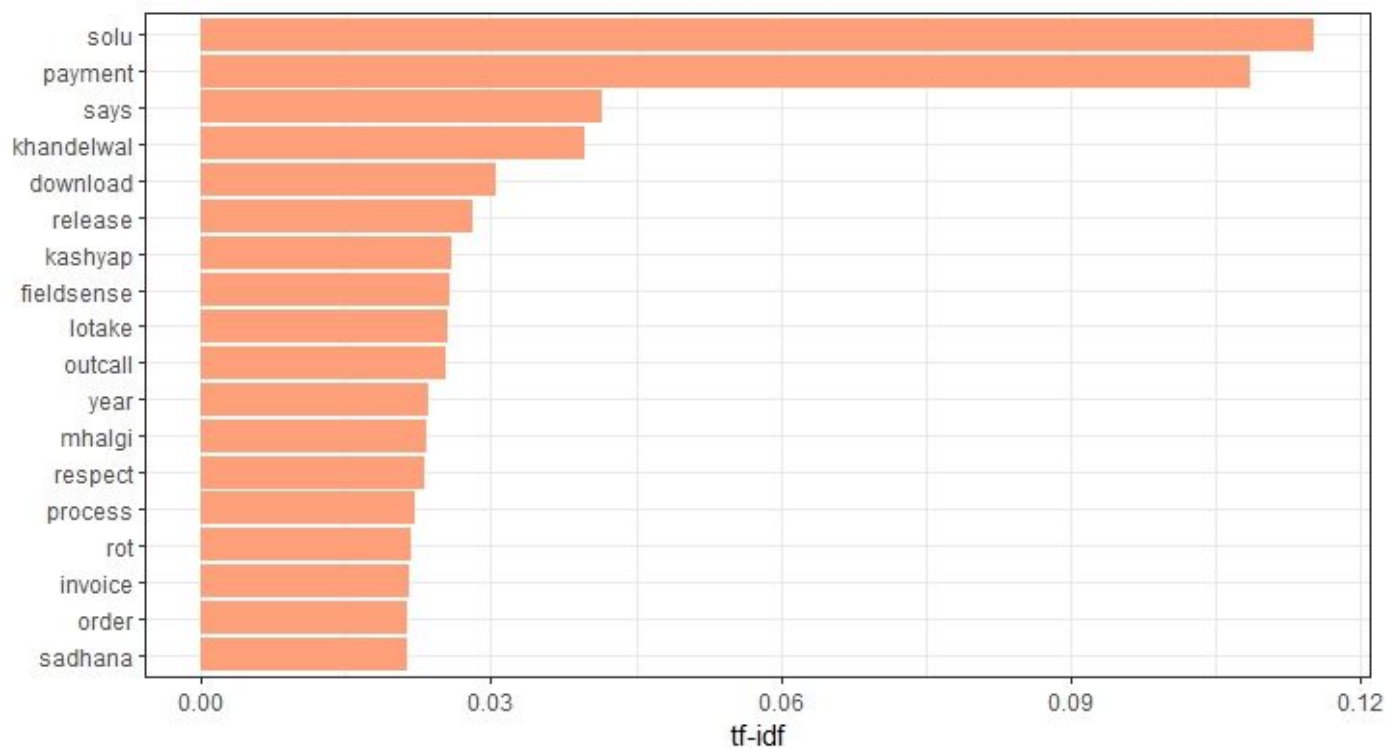
## 9.2 Word Cloud of In-Transitive Verb



## 10.1 TF-IDF

TF-IDF is a weighted model commonly used for information retrieval problems. It aims to convert the text documents into vector models on the basis of occurrence of words in the documents without taking considering the exact ordering.

## 10.2 Twenty Most Important words for the Most Active User



## 11 Similarities

We evaluate the similarities between the word spoken between Users..

1. Similarity between words spoken by Sagar and Jyotirmay is 29%

```
> getCosine(plot_trainwords_dtm["Sagar Gharat",],plot_trainwords_dtm["Jyotirmay Parida",])
[1] 0.2996546
> |
```

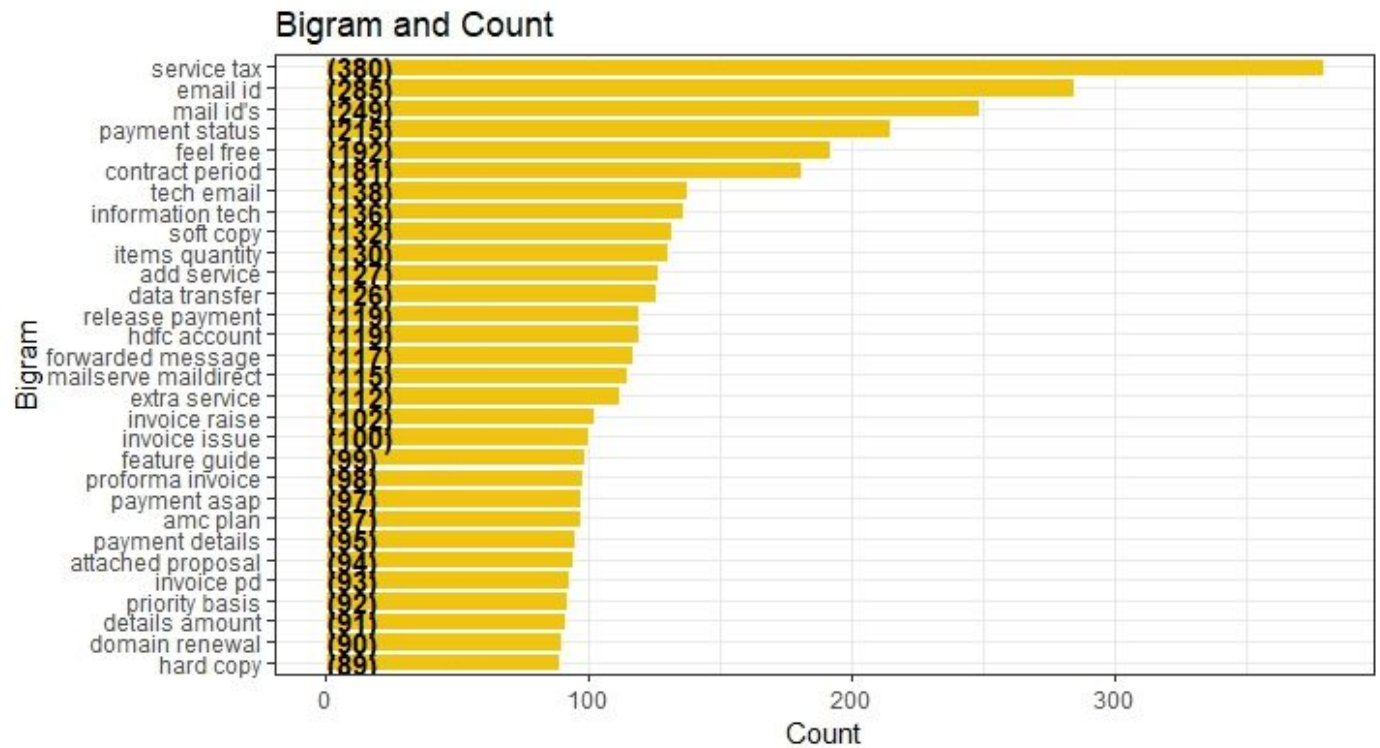
2. Similarity between words spoken by Pooja and Jyotirmay is 27%

```
> getCosine(plot_trainwords_dtm["Pooja Lasunkute",],plot_trainwords_dtm["Jyotirmay Parida",])
[1] 0.277289
```

## 12 Most Common Bigrams

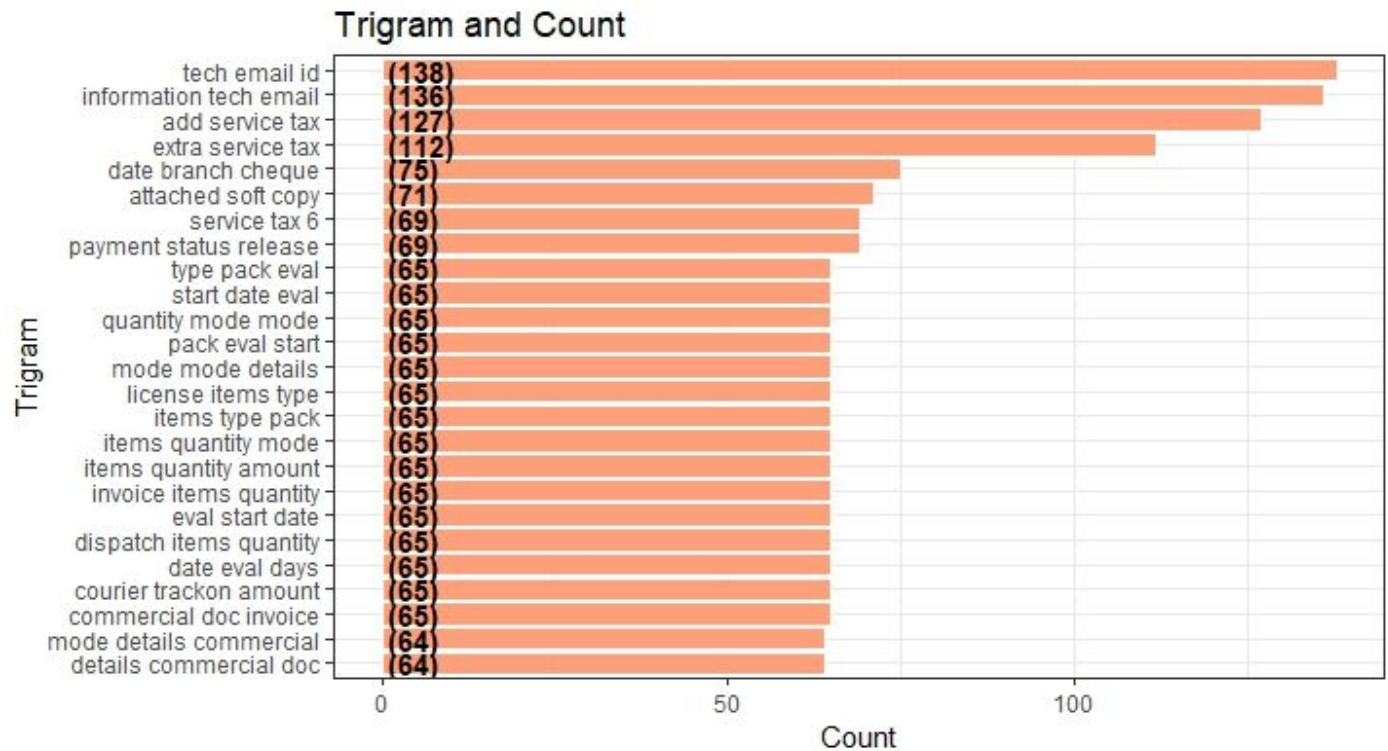
A **Bigram** is a collection of Two words. We examine the most common Bigrams and plot them in a bar plot.

The most common bigrams are **service tax, contract period, data transfer, release payment, Invoice raise, invoice issue, proforma invoice, domain renewal.**



## 13 Most Common Trigrams

The most common trigrams are **add service tax, attach soft copy, type pack eval, payment status release, courier trackon amount**

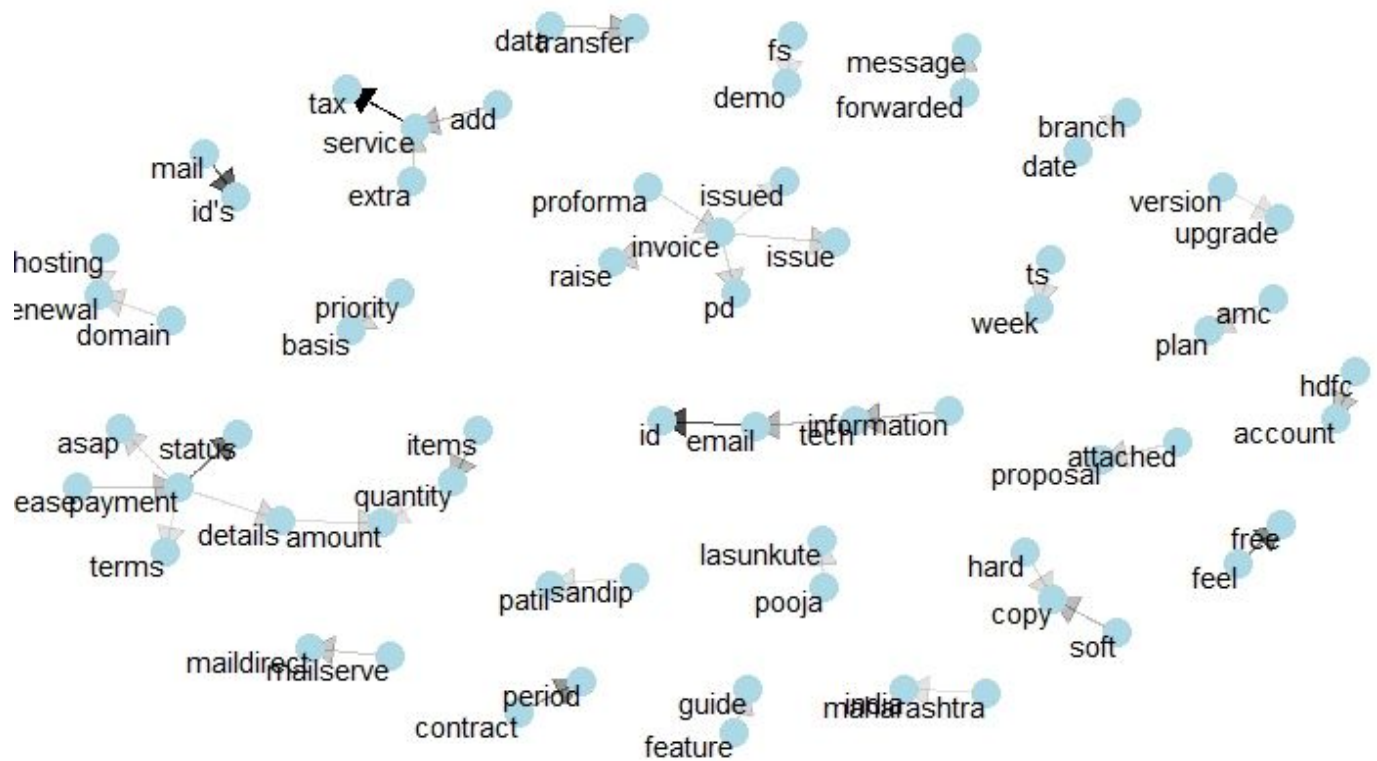


## 14 Relationship among words

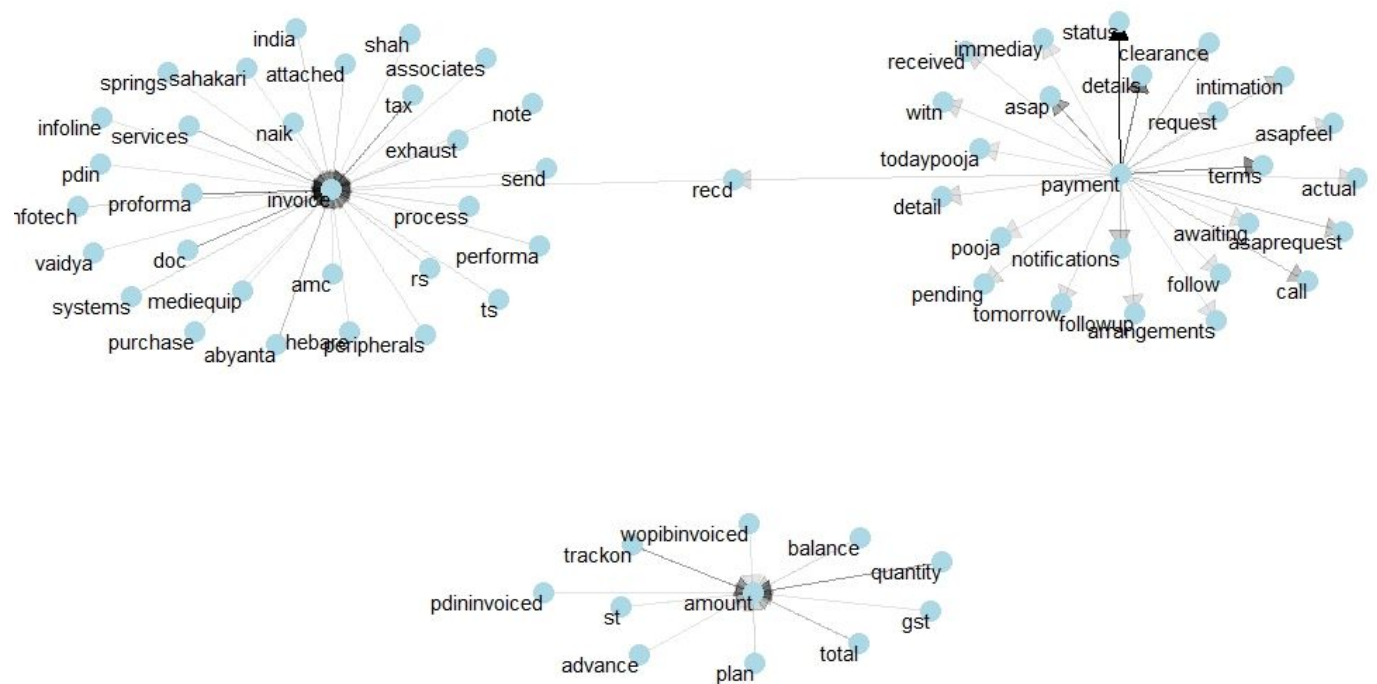
Til now, we have explored the most important words.

Now, we will explore the relationship between words which is repeated over 75 times .The infographic shows the words which follow another word.





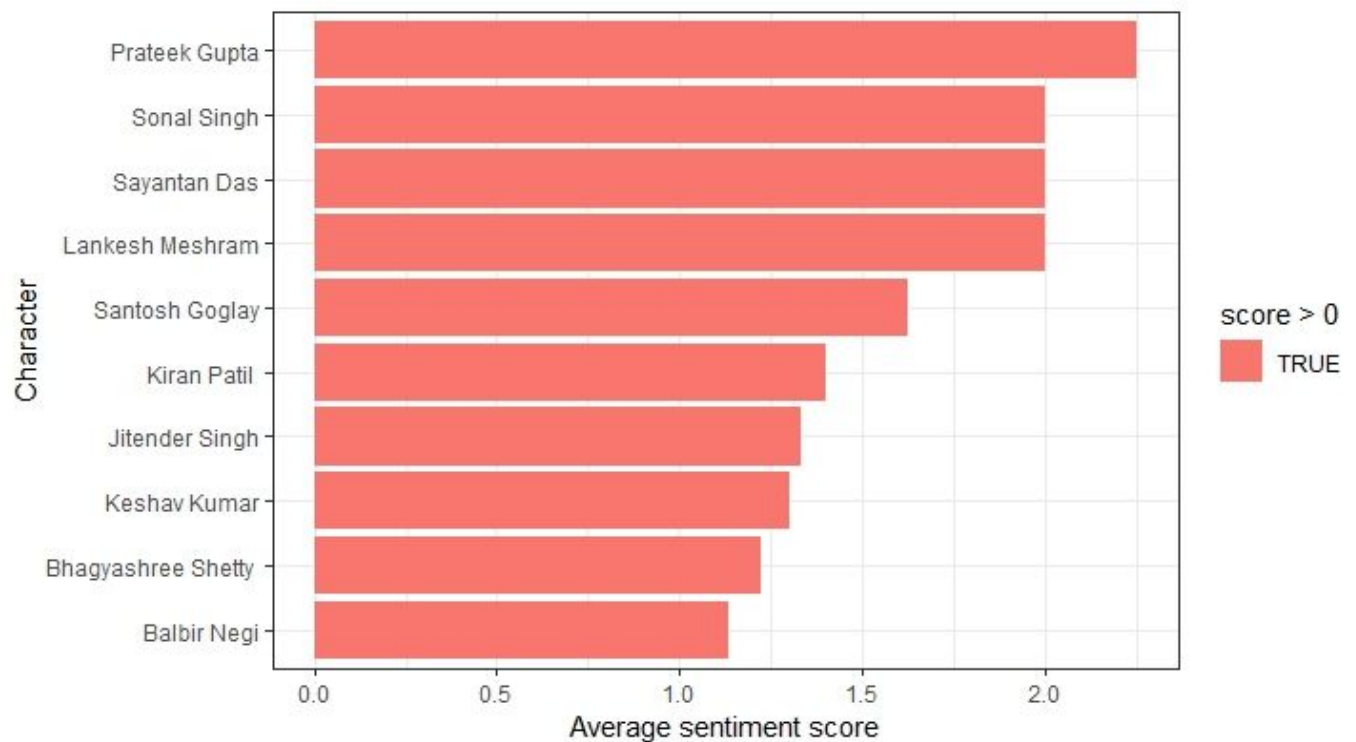
## 15 Relationship with the word **Invoice**, **Payment** and **Amount**



## 16 Sentiment Analysis using **AFINN sentiment lexicon**

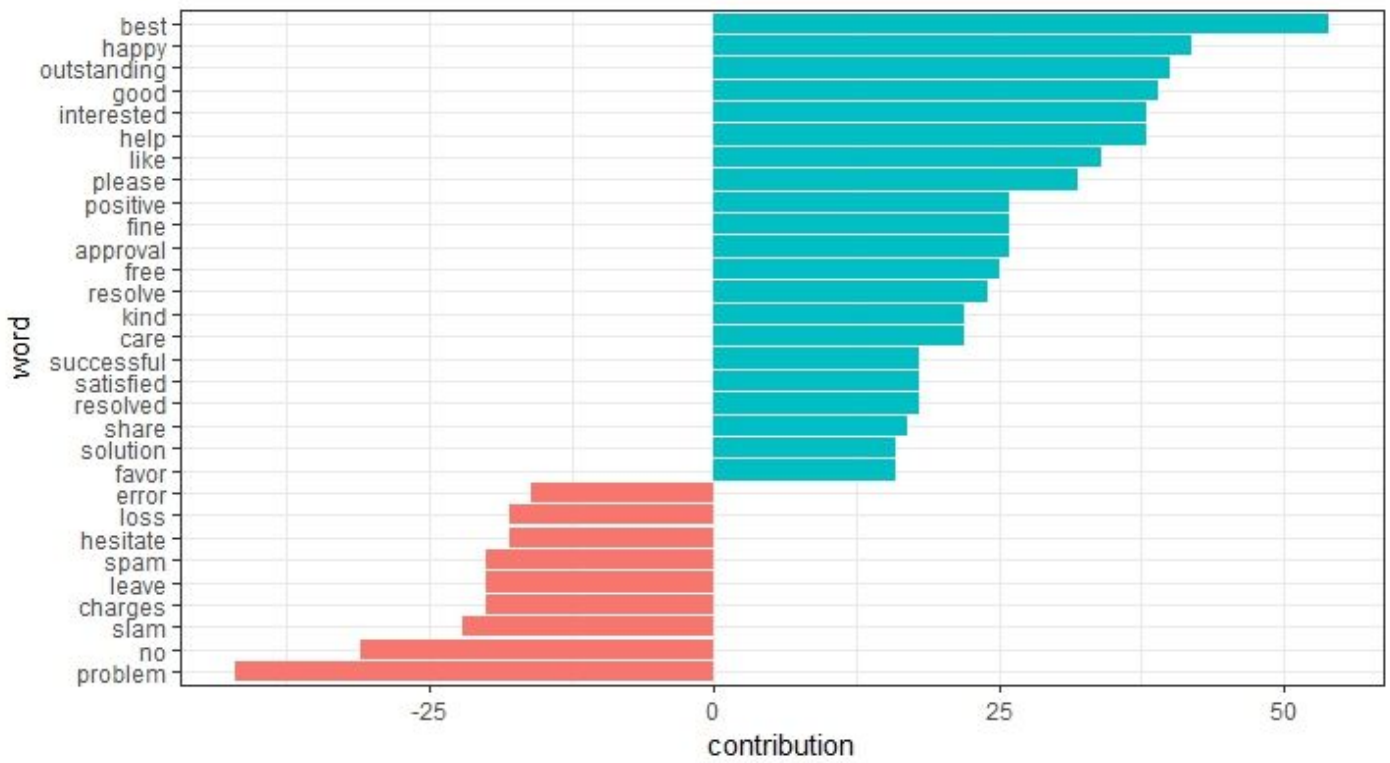
We investigate how often positive and negative words occurred in the text spoken by the Characters. Which Character was the most positive or negative overall?

We will use the **AFINN sentiment lexicon**, which provides numeric positivity scores for each word, and visualize it with a bar plot.

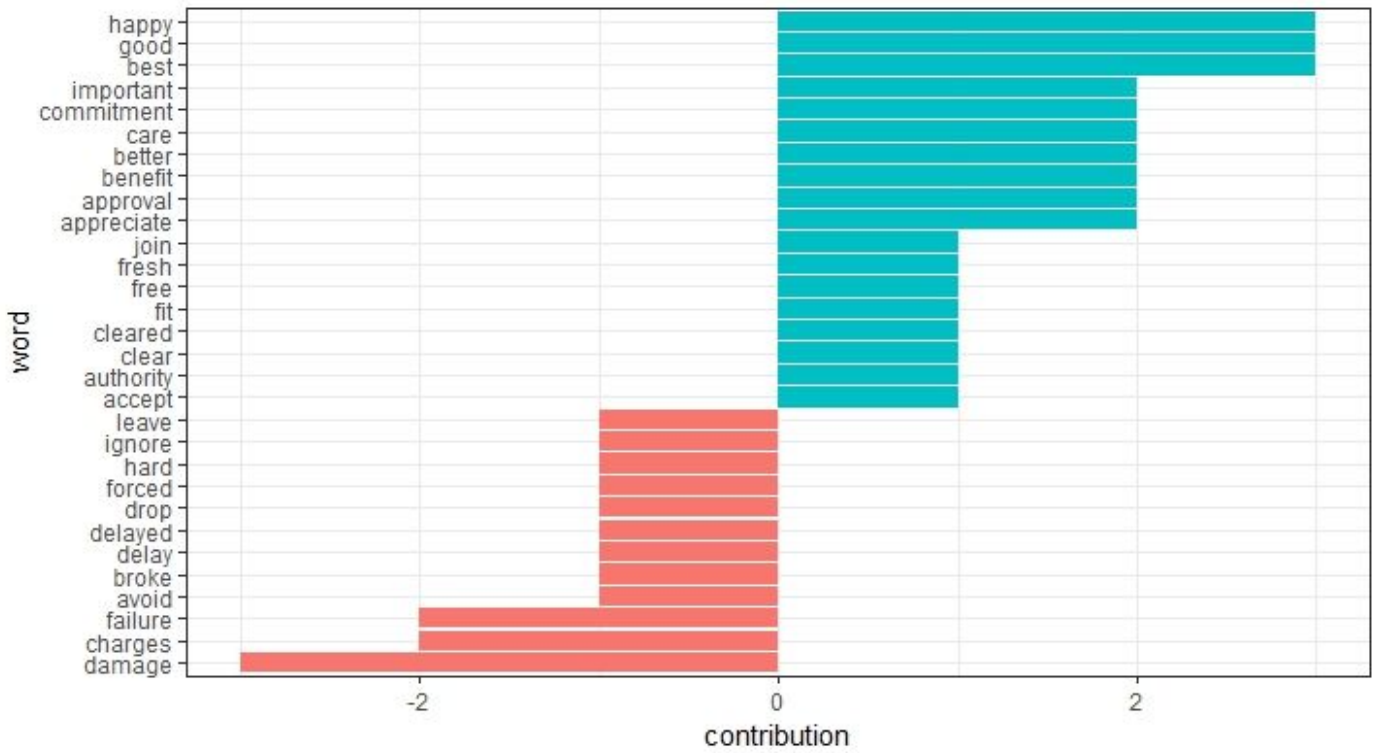


## 17 Positive and Not So Positive Words of Characters





## 17.1 Positive and Not So Positive Words of Sagar



## 17.2 Positive and Not So Positive Words of Jyotirmay Parida



## 18.2 Sentiment Analysis Words - Anger for Sagar



## 18.3 Sentiment Analysis Words - Joy for Sagar



## 18.4 Sentiment Analysis Words - Disgust for Sagar



## 19 Stage phrases for Sales & Collection in Slam Activity

REASON/CONTACT PERSON	PURPOSE/RESPONSE	OUTCOME
1 Called ABC company / Spoke with XYZ	Busy in a meeting	Call back
2 Called for Appointment	Agreed	Scheduled Appointment
3 Emailed XYZ	Send Quote	Will renew the same Plan
4 Spoke to XYZ	Send Quote for AMC	Awaiting approval
5 Order Confirmation	Attached proforma invoice and bank details	Awaiting payment
6 Emailed XYZ	Process WO #	
7 Called XYZ	Payment follow up	Payment status released

8	Emailed XYZ	Find Proposal for MD upgrade   Extra Service Tax/GST applicable	Awaiting response
9	Invoice Raised	Date,Amount, Cheque No. xxx-xxxx collected	Close WO
10	Emailed XYZ	To schedule Demo & start evaluation	Awaiting response
11	Emailed XYZ	FieldSense presentation with video link	Awaiting response
12	Called XYZ	Awaiting CEO'S Approval	Follow back on xx-xx-xxxx
13	Called XYZ	Contact Busy   No response	Will follow back on xx-xx-xxxx

### Guidelines

1. Check if account exists before entering new account in SLAM
2. Use Full Company Name
3. Do not Abbreviate (e.g AE - Account Executive, COB — Close of Business)
4. If abbreviations are acceptable, eg WO or MD, then do not put spaces between the letters.
5. Including country / division at end is acceptable
6. Beware of spaces, hyphens, etc.
7. Beware of case (e.g. QGate not QGATE)
8. Beware of misspelling / typing errors
9. Where possible, take the precise company name from reliable sources.





