A picture containing nature

Description automatically generated

***T- Mobile / Sprint Merger***

***Sentiment Analysis***

SODEINDE, AHMED A20187045

ALISETTY, KALYAN A20199542

GANDHI, SAHIL A20207207

*A black and orange sign

Description automatically generated*

**Table of Contents**

**Executive Summary** 1

**Statement of Scope** 2

**Project Schedule** 3

**Data Preparation** 5

*Data Access* 5

*Data Consolidation* 5

*Data Cleaning* 6

*Data Transformation* 7

*Data Reduction* 8

**Data Dictionary** 8

**Descriptive Statistics** 10

**Sprint**  11

*Text Mining*  11

*Sentiment Analysis &* Parts of Speech Recognition  12

**T-Mobile**  17

*Text Mining*  17

*Sentiment Analysis &* Parts of Speech Recognition 18

**T-Mobile Sprint Merger** 23

*Text Mining*  23

*Sentiment Analysis &* Parts of Speech Recognition 24

*Named Entity Recognition and LDA* 30

**Conclusion** 32

**References** 33

**Executive Summary**

In April of 2018, T-Mobile and Sprint announced a 26-billion-dollar merger which would translate to about 126 million customers under one brand bringing it closer to the rivals AT&T (141 million subscribers) and Verizon (150 million). The announcement triggered different reactions from different sectors – the media, government and most importantly the customers.

Would prices go up? What would happen to the existing plans? Is the new network going to be better and more efficient? These are some of the questions that bother customers of both networks. This analysis aims to show how users feel about the merger based on their opinions on twitter. It would benefit both the customers as a collective and the network providers. The customers would get the most popular opinions/sentiments about the merger communicated to the network providers and they can then use these sentiments to provide better considerations. These articles further explore all the [issues](https://www.theatlantic.com/ideas/archive/2019/10/t-mobile-and-sprints-merger-will-hurt-consumers/599245/) that might arise and [questions](https://www.nytimes.com/2019/07/26/technology/personaltech/t-mobile-sprint-merger.html) the customers may have about the merger.

**Statement of Scope**

The team’s objectives and scope of work are to:

* Identify sources of user opinions regarding the merger and the network providers
* Collect data from twitter and the play store using R and Python
* Identify features pertinent to the analysis
* Clean the data by handling missing values and less represented variables

Some of the variables would be:

* Tweet id - which is a unique number that identifies each tweet
* Date and time of tweet

**Project Schedule**

The project is estimated to take about 15 weeks barring unforeseen circumstances in which case expedient decisions will be made. The team will meet physically about 3 times and would mostly communicate through the already created WhatsApp group chat, Zoom and Skype video calls. Kalyan and Sahil will work on gathering and analysis from the data for the deliverables and Ahmed will work on the documentation of both the deliverables. The team will remain fluid and re-assignment of tasks may occur.

**Data Preparation**

***Data Access***

Twitter -Using Twitter’s API, we scraped tweets with the hashtag #TmobileSprintmerger. The choice of twitter was informed by the documented inclination of people expressing opinions about most things on the social media platform.

Play store – The comment sections of the play store apps for T-Mobile and Sprint happen to be rich with opinions about the network providers in general. The code can also be found in the appendix.

While performing the analysis for the 2nd deliverable, we noticed that earlier we included the tweets with hashtag Sprint or T-Mobile along with merger tweets to make sure we had enough data for sentiment analysis. However, the results were a bit skewed so, we used only the tweets with hashtag merger (Refer code for details). Also, since there were not a lot of tweets, so we changed the end date to November 30 to have more tweets text for performing sentiment analysis.

***Data Consolidation***

The extracted data was put into data frames and then stored in a csv file. Since the tweets were generated based on the hashtags, we initially stored the tweets data of each hashtag in a separate csv and consolidated all those csv files into one for performing sentiment analysis. This was done purely using code which could be found in the appendix.

***Data Cleaning***

We have obtained data from Twitter to have the merger tweets and from Google Play store to have the list of review comments posted by the users on apps of both T-Mobile and Sprint for sentiment analysis. The data from twitter consists of tweets from January 2018 to November 2019 to have as much tweets as possible for performing better sentiment analysis and the play store data was derived from April 2017 to October 2019.

We obtained two data files. The First file consists of tweets with hashtag merger and the second file has the data of review comments from Google Play store.

The below workflow is a sequence of three steps aiming at producing high-quality data:

INSPECTION Detect unexpected, incorrect, and inconsistent data



CLEANING Fix or remove the anomalies discovered



VERIFYING After cleaning, the results are inspected to verify correctness



|  |
| --- |
| REPORTING Start with the analysis |

Inconsistent data often leads to false/inconclusive results. Therefore, how well data cleaning is performed has a significant impact on the quality/conclusiveness of results. Thus, data inspection was done, and irrelevant/inconsistent data was excluded from the sentiment analysis.

Since the twitter data was collected using different hashtags related to merger and there were a lot of tweets where multiple hashtags related to merger were there due to which there were duplicates in the combined csv file. Therefore, for data cleaning, we performed the following steps:

* Remove Duplicate Tweets
* Remove special characters from the tweets
* Removed the play store comments containing words like ‘App’, ‘Apps’ etc

***Data Transformation***

After Data cleaning, we performed data transformation by enhancing the usability of the data for our analysis. There were continual updates for this step throughout the semester for improving upon it. Following steps were performed for data transformation:

* Constructing new attributes: For the play store data, the review ratings were in the form like *1 out of 5* and since we only wanted to look at the ratings, therefore, a new attribute was created which only had the rating like *1,* the reason being, out of 5 was something common to every review comment and did not add any value.
* Transforming text into categories: For sentiment analysis and topic labeling while determining whether text is positive or negative, transformation of text was done. Removing stop words, stemming and other steps were performed for deriving meaningful insights.

***Data Reduction***

Although data reduction is not a necessary step, however, in our case, not all variables were important and the variables which did not add value to our analysis were dropped.

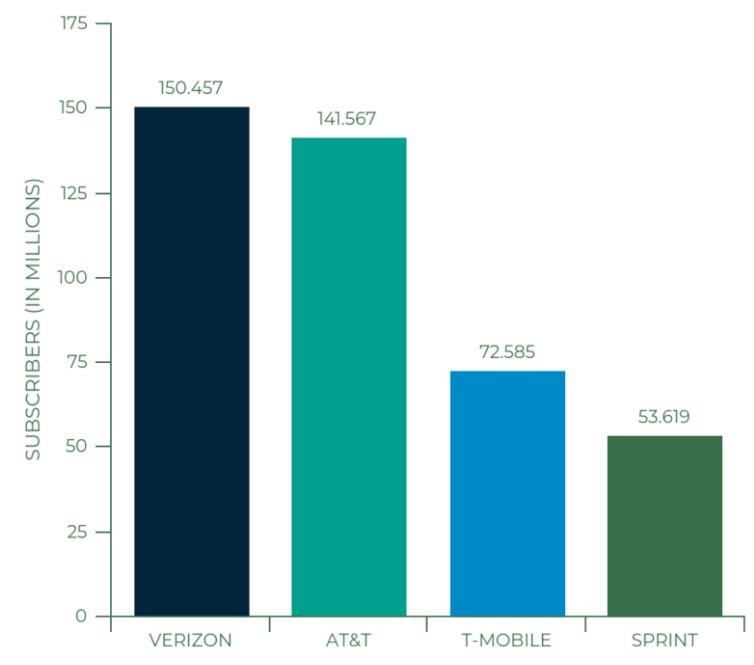
Below steps were performed for data reduction:

* Dropping all variables except date and tweets from the twitter data as those variables are not providing any valuable insights
* Dropping all variables except date, user comments and user ratings from the play store data as those variables are not providing any valuable insights

**Data Dictionary**

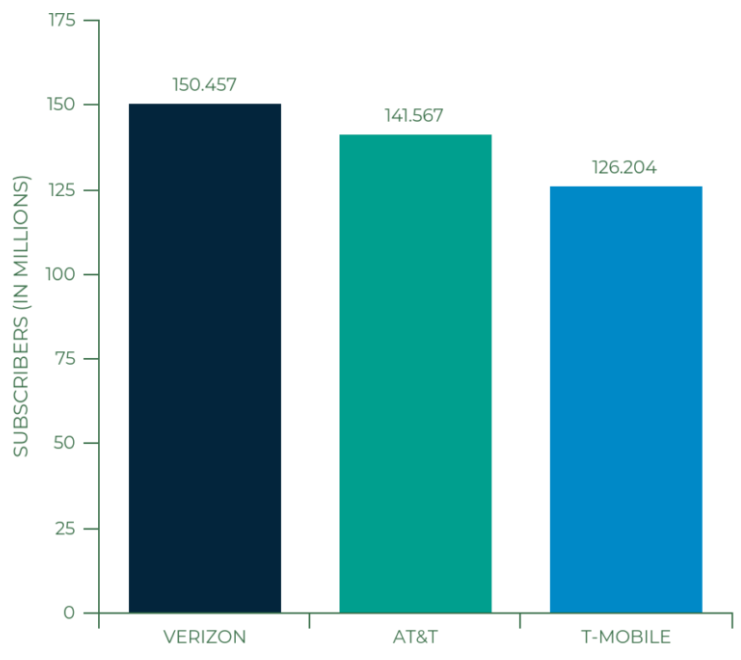
|  |  |  |  |
| --- | --- | --- | --- |
| **Attribute Name** | **Description** | **Data Type** | **Source** |
| ID | Unique float identifier for a tweet | Float | twitter.com or playstore.com |
| Date | Date tweet/comment was created | DateTime | twitter.com or playstore.com |
| Time | Time tweet was created | DateTime | twitter.com or playstore.com |
| Time zone | Geo time zone of tweet | Time Zone | twitter.com or playstore.com |
| User ID | Unique float identifier of person tweeting | Float | twitter.com or playstore.com |
| User Name | Username of person tweeting or commenting | Char(var) | twitter.com or playstore.com |
| Name | Name of commenter or tweeter (May not necessarily be conventional | Char(var) | twitter.com or playstore.com |
| Text | Body of tweet/comment | Char(var) | twitter.com or playstore.com |
| URLs | Link to tweet or comment page | URL | twitter.com or playstore.com |
| Retweet Count | Number of retweets | Int | twitter.com or playstore.com |
| Likes Count | Number of likes | Int | twitter.com or playstore.com |
| Links in text | Other urls found in the body of the tweet/comment | URL | twitter.com or playstore.com |

***Current Scenario:***



*Fig 1.1 “Big Four” subscriber numbers now*

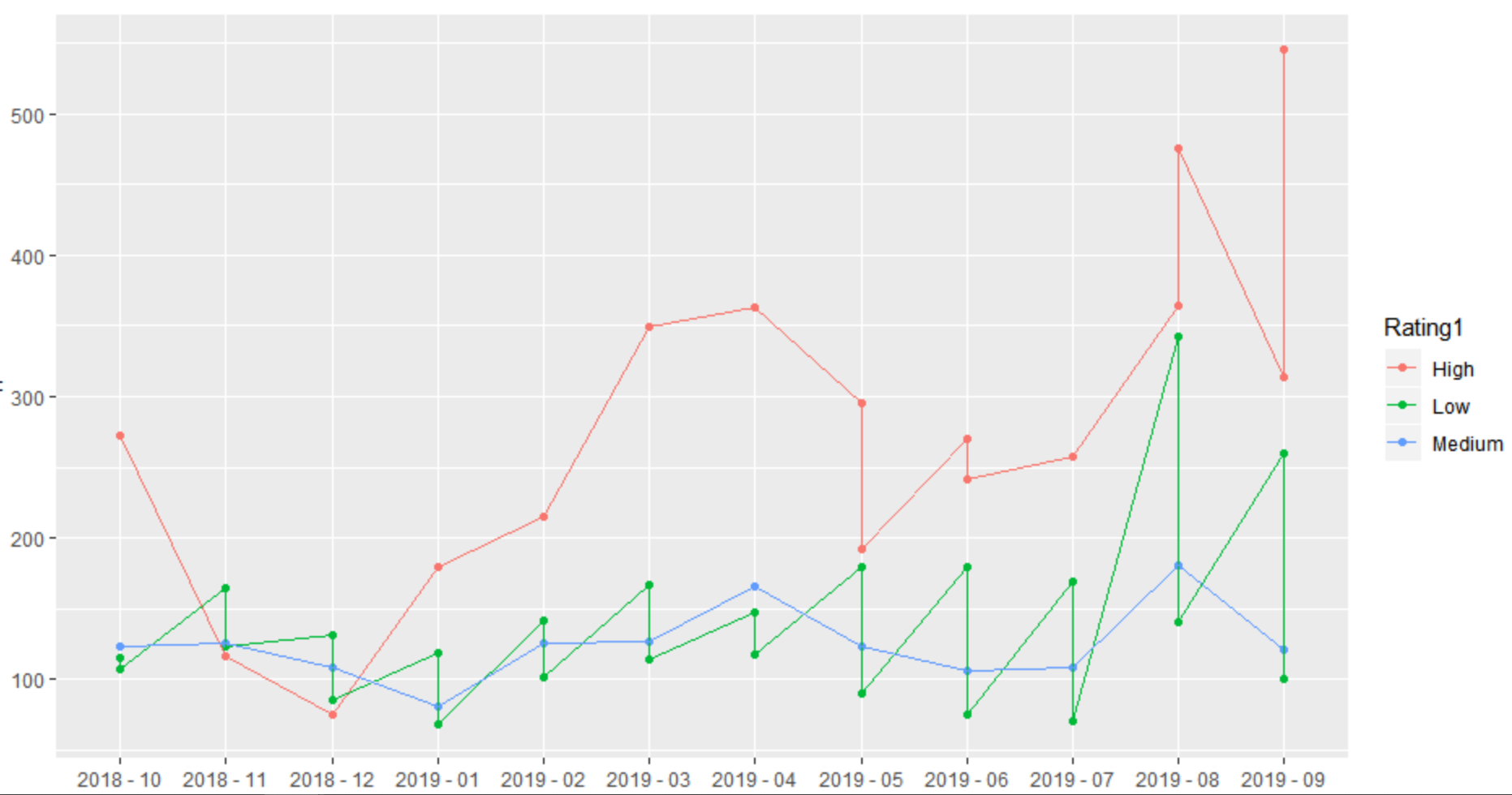
***After Merger:***



*Fig 1.2 “Big Three” subscriber numbers*

**Descriptive Statistics and Analysis**

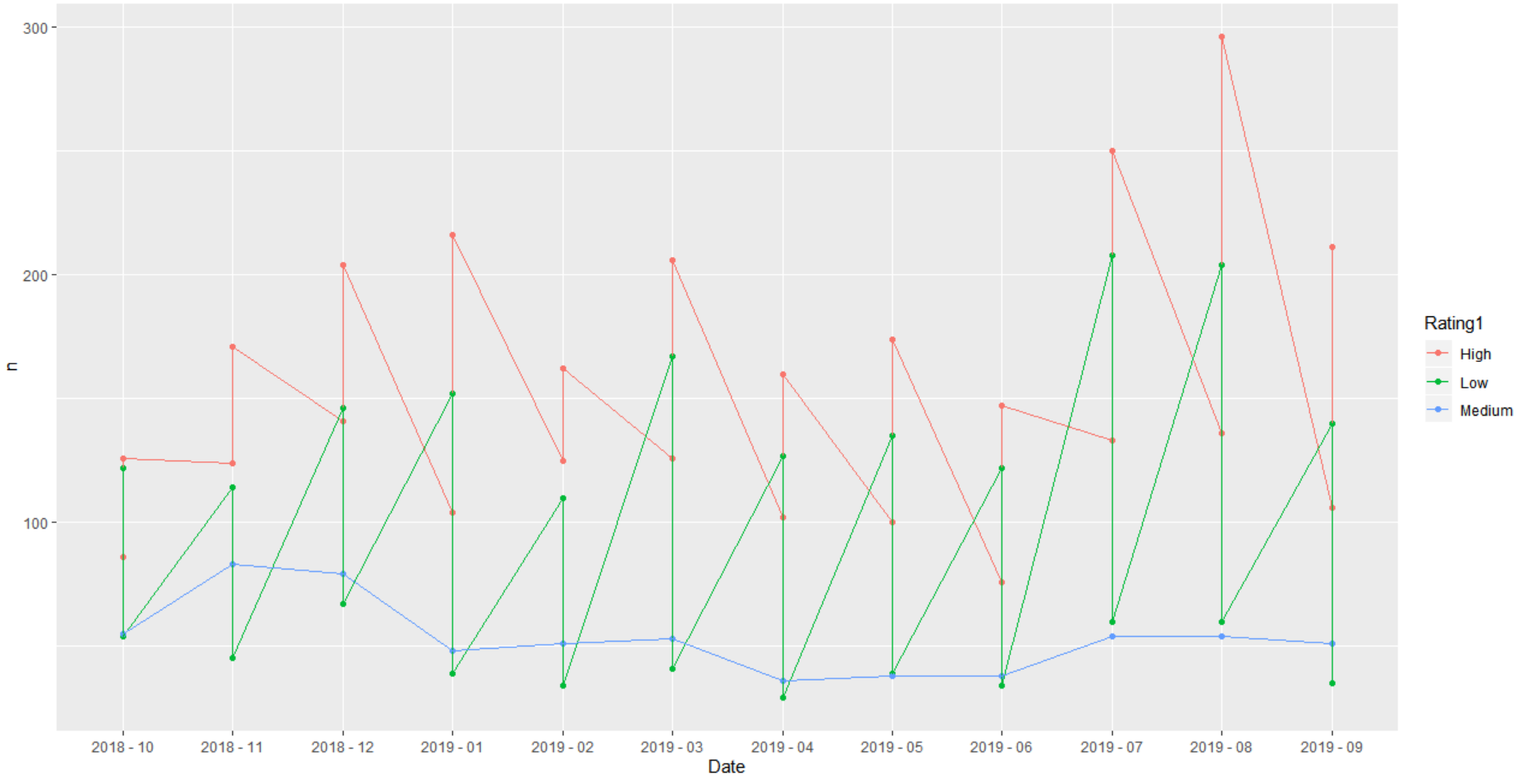
***T-Mobile***



*Fig 2.1 Count of different categories of rating over the last one year*

People have given more *high rating* in the year 2019 and *low rating* from Nov 2018 to Dec 2018

***Sprint***

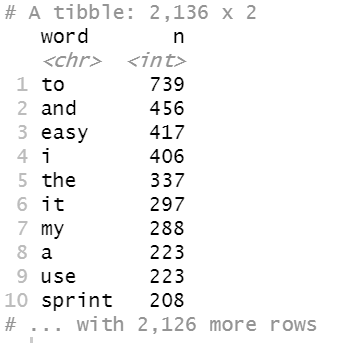
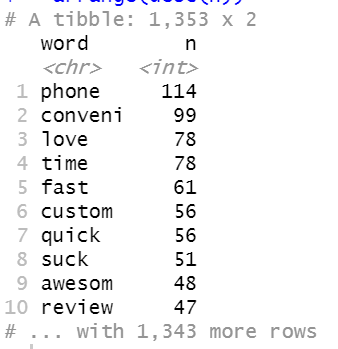
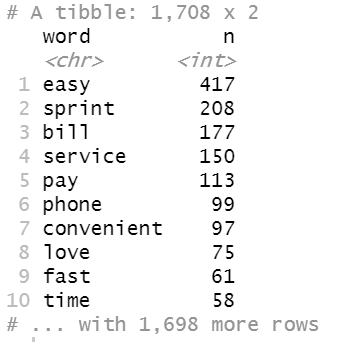


*Fig 2.2 Count of different categories of rating over the last one year*

People have given more *high rating* from the past one year

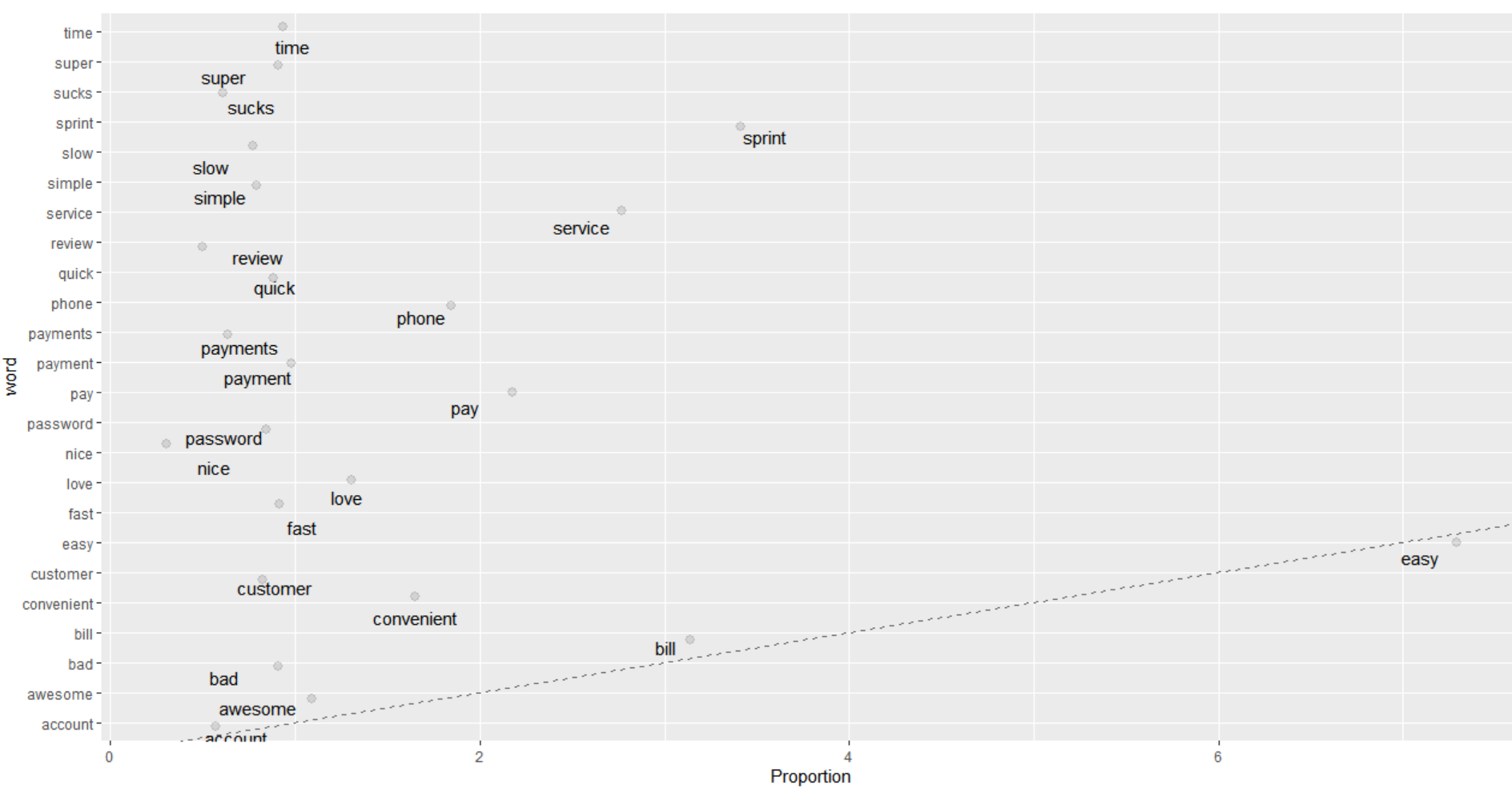
**Text Mining and Sentiment Analysis**

***Analysis of the Sprint comments***



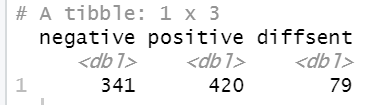
*Fig 3.1 Top 10 words before Fig 3.2 Top 10 words after Fig 3.3 Words after stemming and removal of removing stop words stemming, unnecessary stop words words removed, same*

*frequency was set*



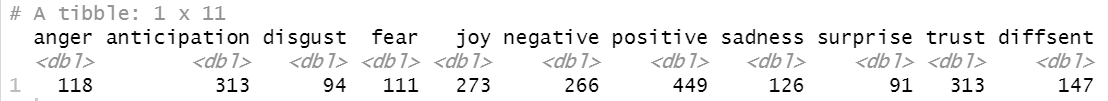
*Fig 3.4 Graphical representation of words from Sprint comments after filtering*

The above graph shows that the word *easy* is mostly used by the user in the comments



*Fig 3.5 Showing number of positive and negative emotions*

The snapshot above shows the number of positive and negative comments from the analysis of the sprint user comments data. The figure suggests us that the people are more positive about the sprint network.

******

*Fig 3.6 Showing the count of other feelings/emotions*

The figure above shows different feeling of the user on the sprint data. As we can see that the people are feeling more joyful than sadness. There is more trust compared to fear.

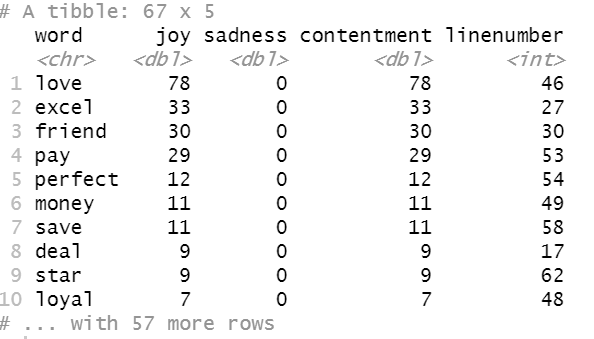
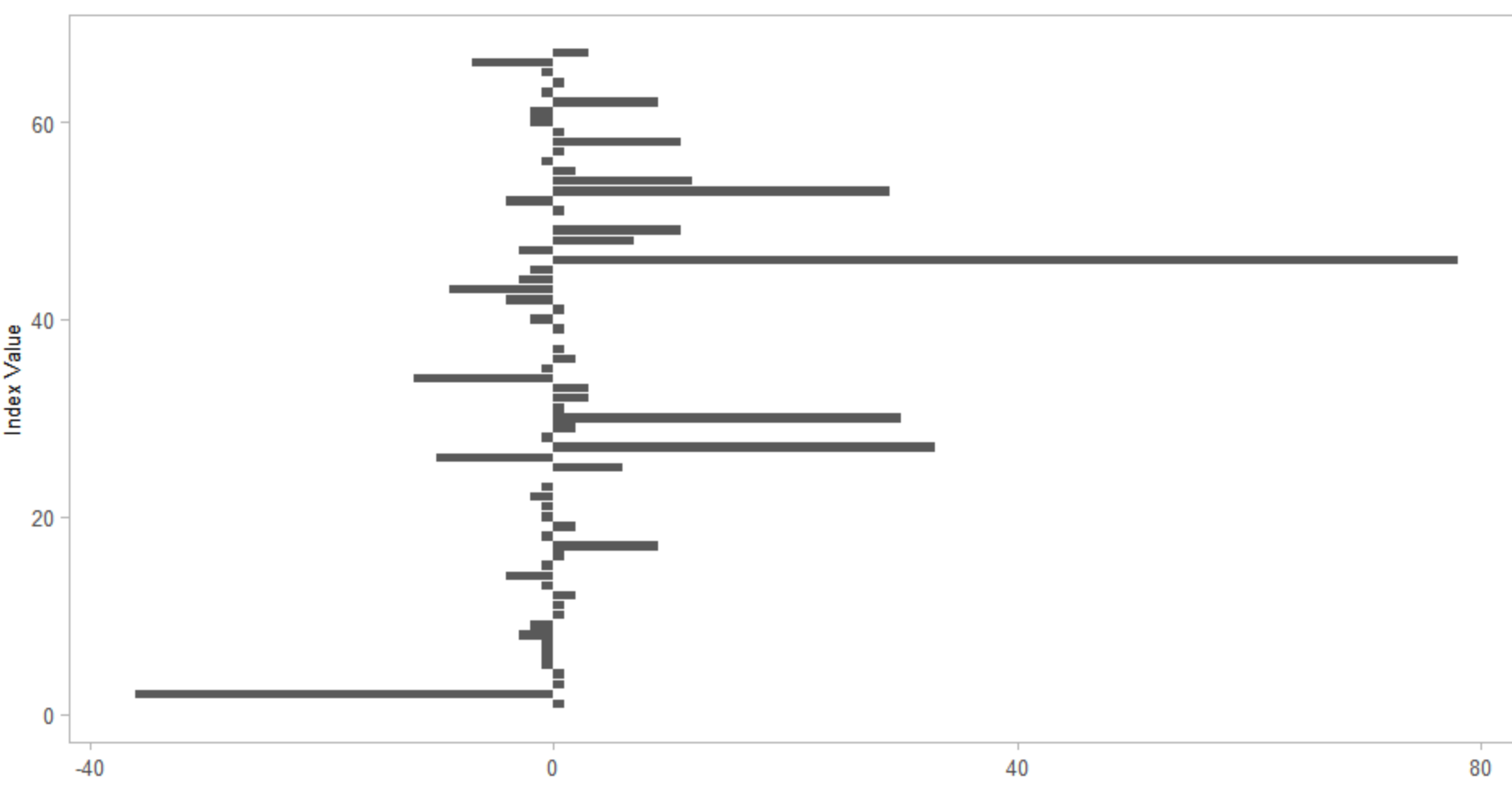
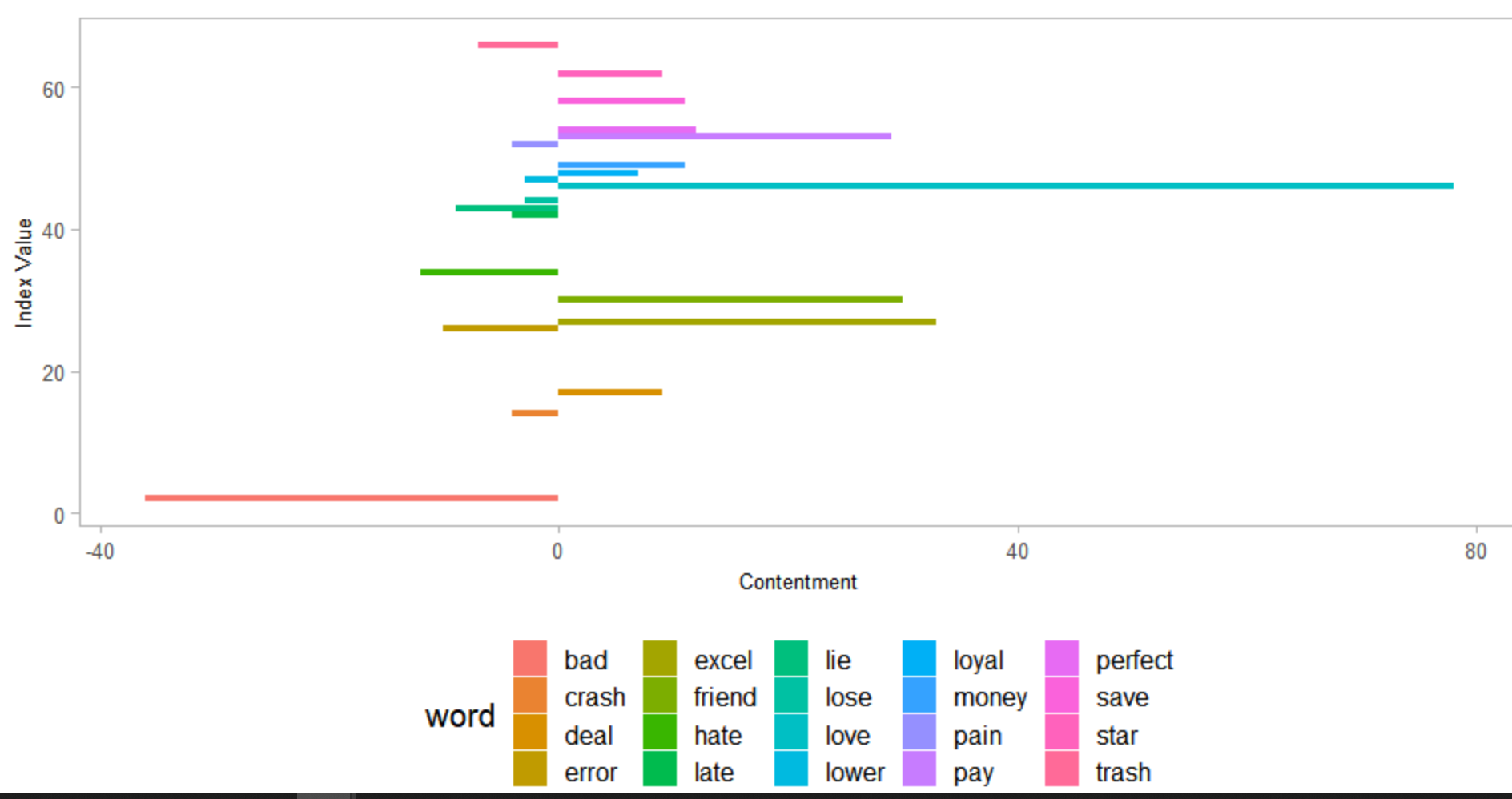


Fig 3.*7 Top 10 words depicting feelings of joy and sadness*

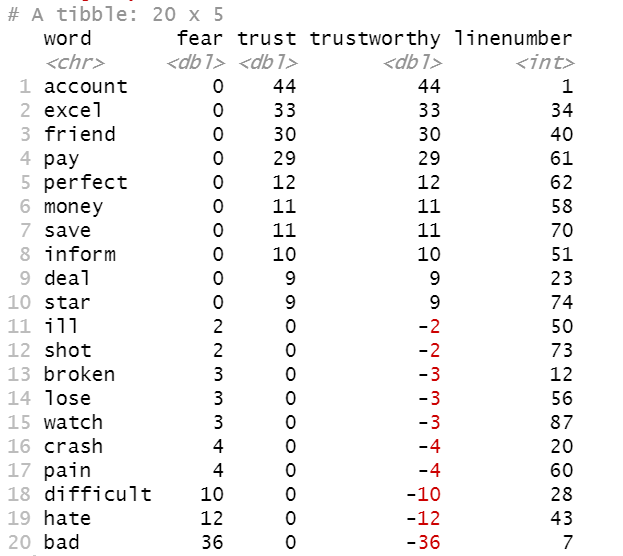


*Fig 3.8 Graphical representation of the feelings ‘Joy’ and ‘Sadness’*

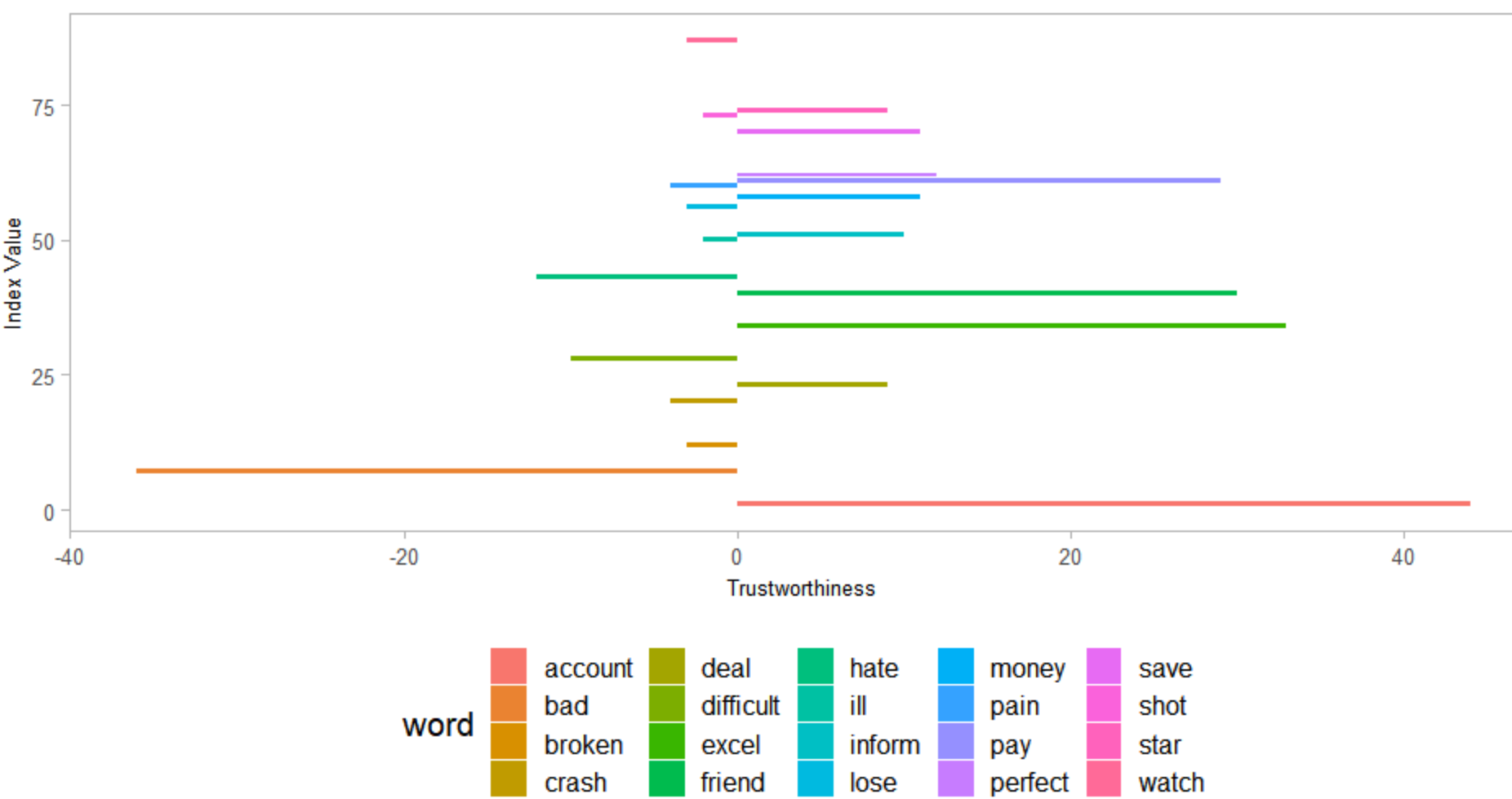


*Fig 3.9 Depicting level of contentment of feelings of joy and sadness*

This suggests us that love stands out as the word with shows the most contentment and bad has the lowest level of contentment. Further, this indicates that people are feeling more joyful compared to sadness while remaining connected with Sprint network.



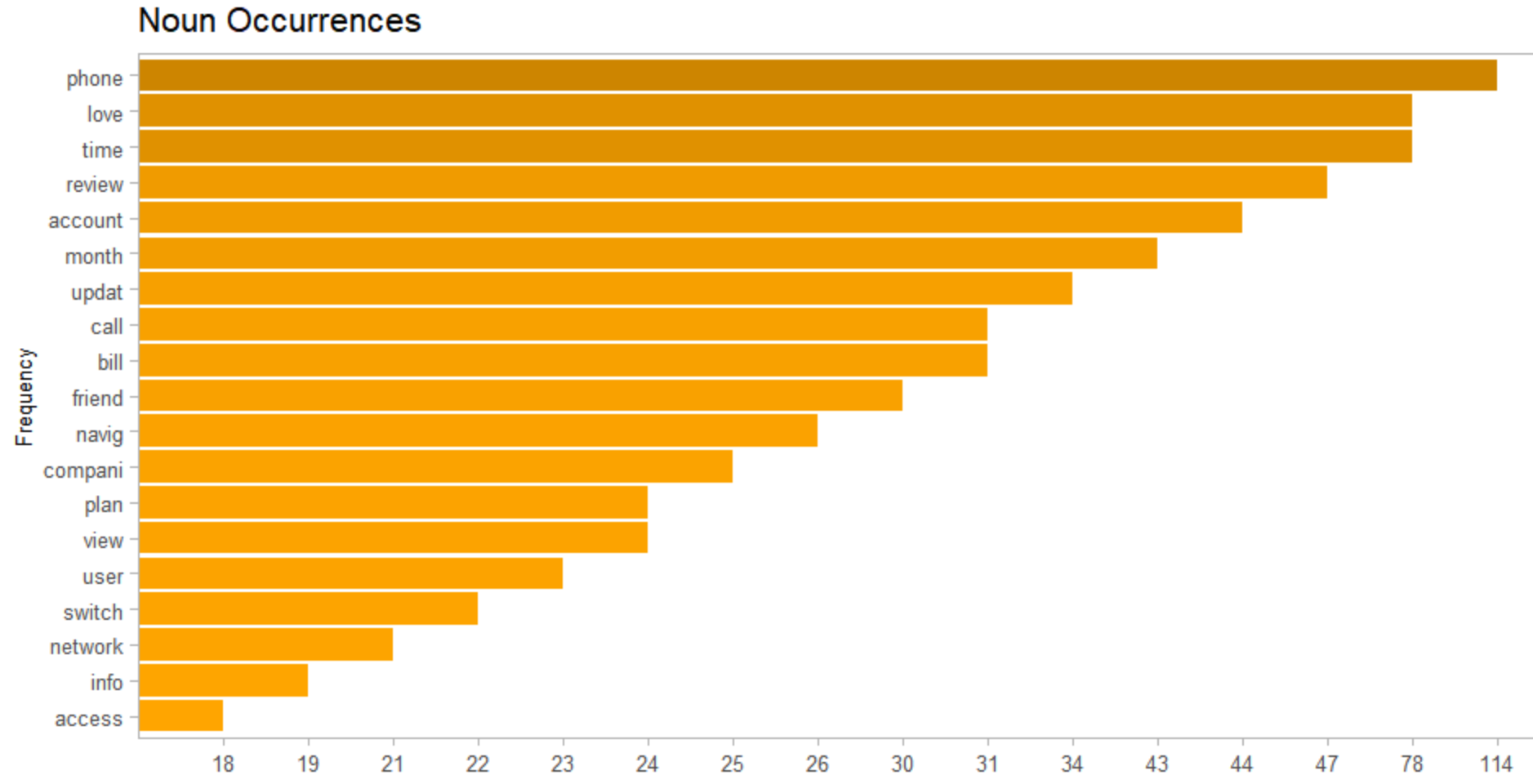
*Fig 3.10 Top 20 words depicting feelings of trust and fear*



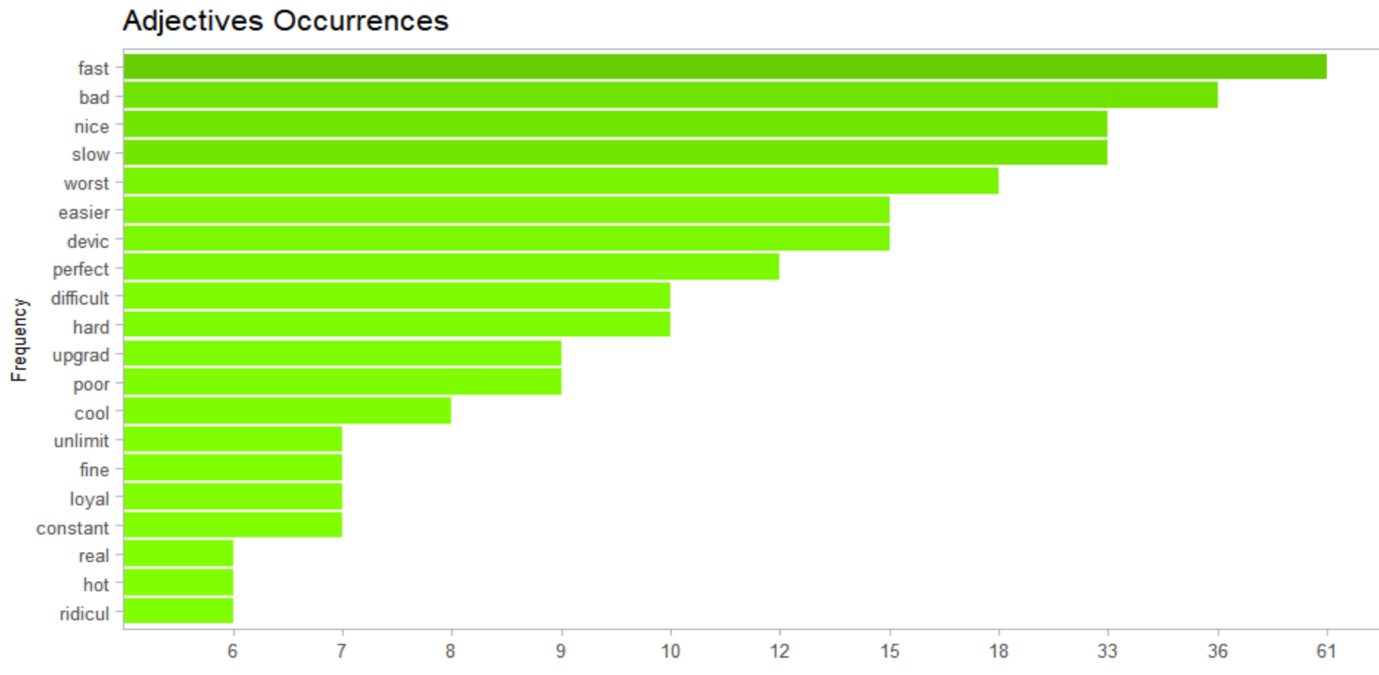
*Fig 3.11 Depicting level of trustworthiness based on feelings of trust and fear*

This ascertains that account stands out as the word with shows the most contentment and bad has the lowest level of contentment. Further, it suggests that people are having more trust compared to fear with Sprint network.

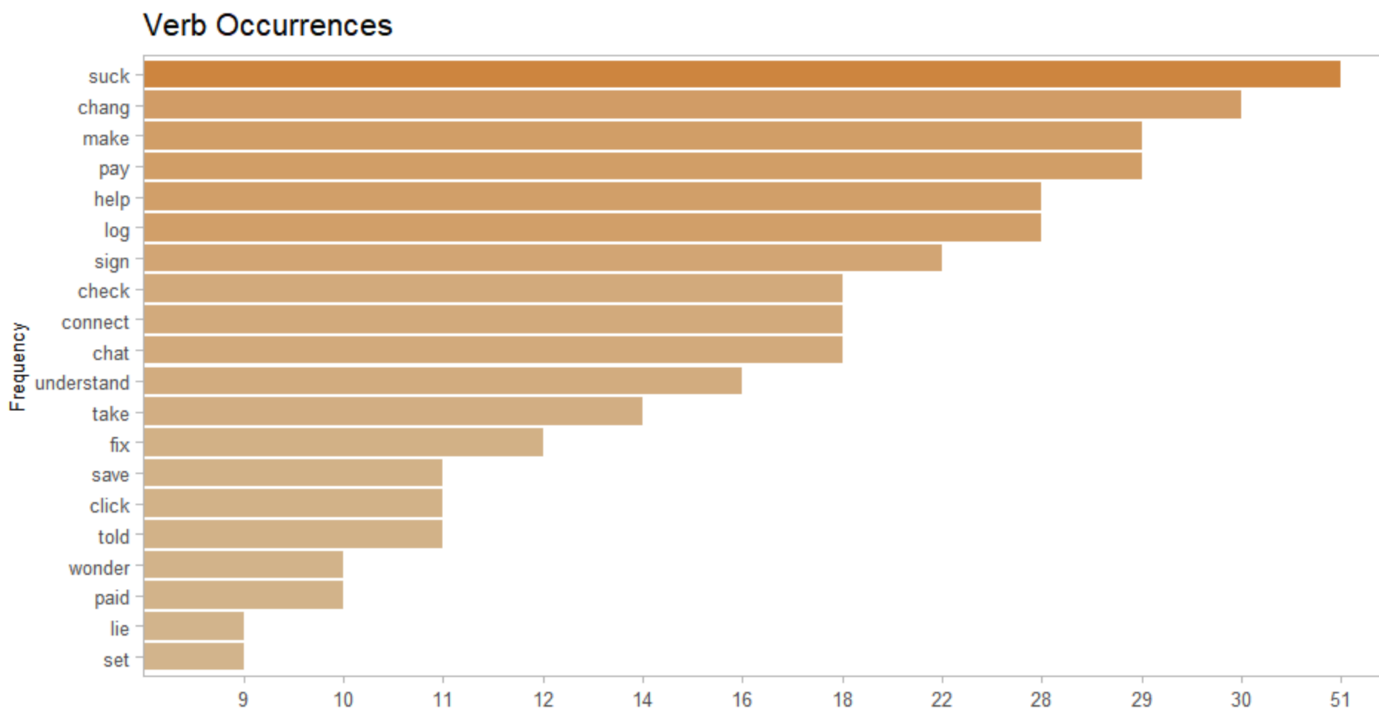
***Parts of Speech Recognition***



*Fig 3.12 Horizontal bar chart shows top nouns and number of their occurrences*

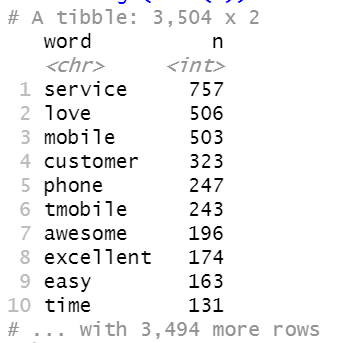


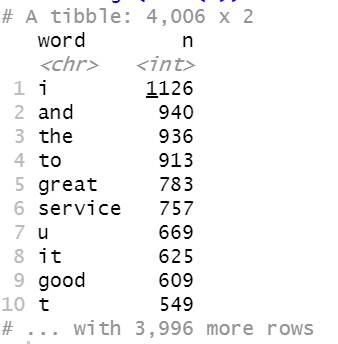
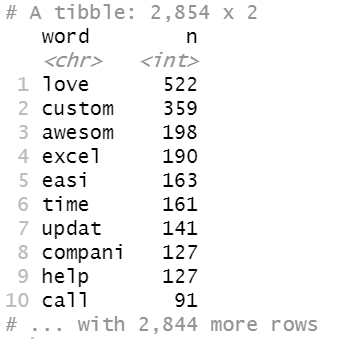
*Fig 3.13 Horizontal bar chart showing top adjectives and number of their occurrences*



*Fig 3.14 Bar chart showing top verbs and number of their occurrences*

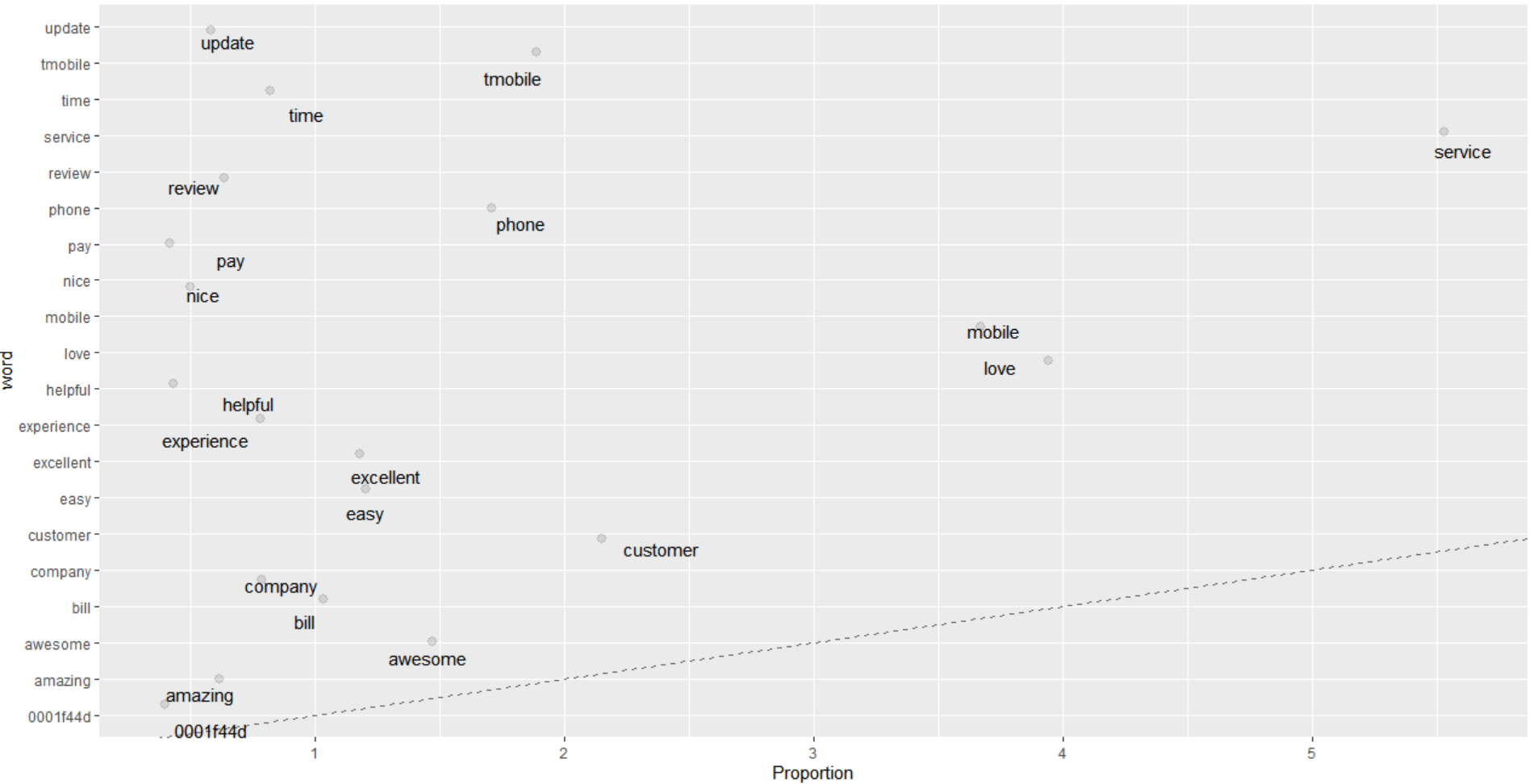
**Analysis of T-Mobile comments**





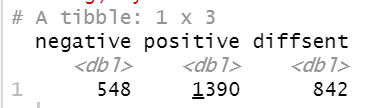
*Fig 4.1 Top 10 words before Fig 4.2 Top 10 words after Fig 4.3 Words after stemming and removal of removing stop words stemming, unnecessary stop words words removed, same*

*frequency was set*



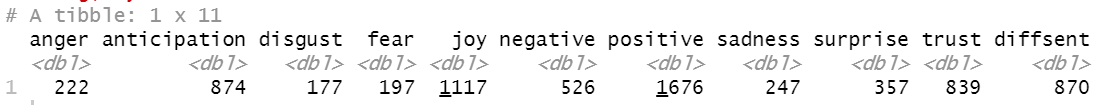
*Fig 4.4 Graphical representation of proportion of words from comments related to T-Mobile*

The above graph shows that the word *Service* is mostly used by the user in the comments



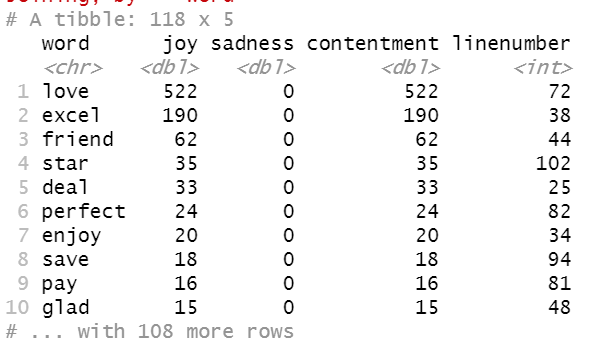
*Fig 4.5 Showing positive and negative emotions*

The snapshot above shows the number of positive and negative comments from the analysis of the T-Mobile user comments data. This suggests us that the people are more positive about the T-Mobile network.

******

*Fig 4.6 Showing other emotions/feeling*

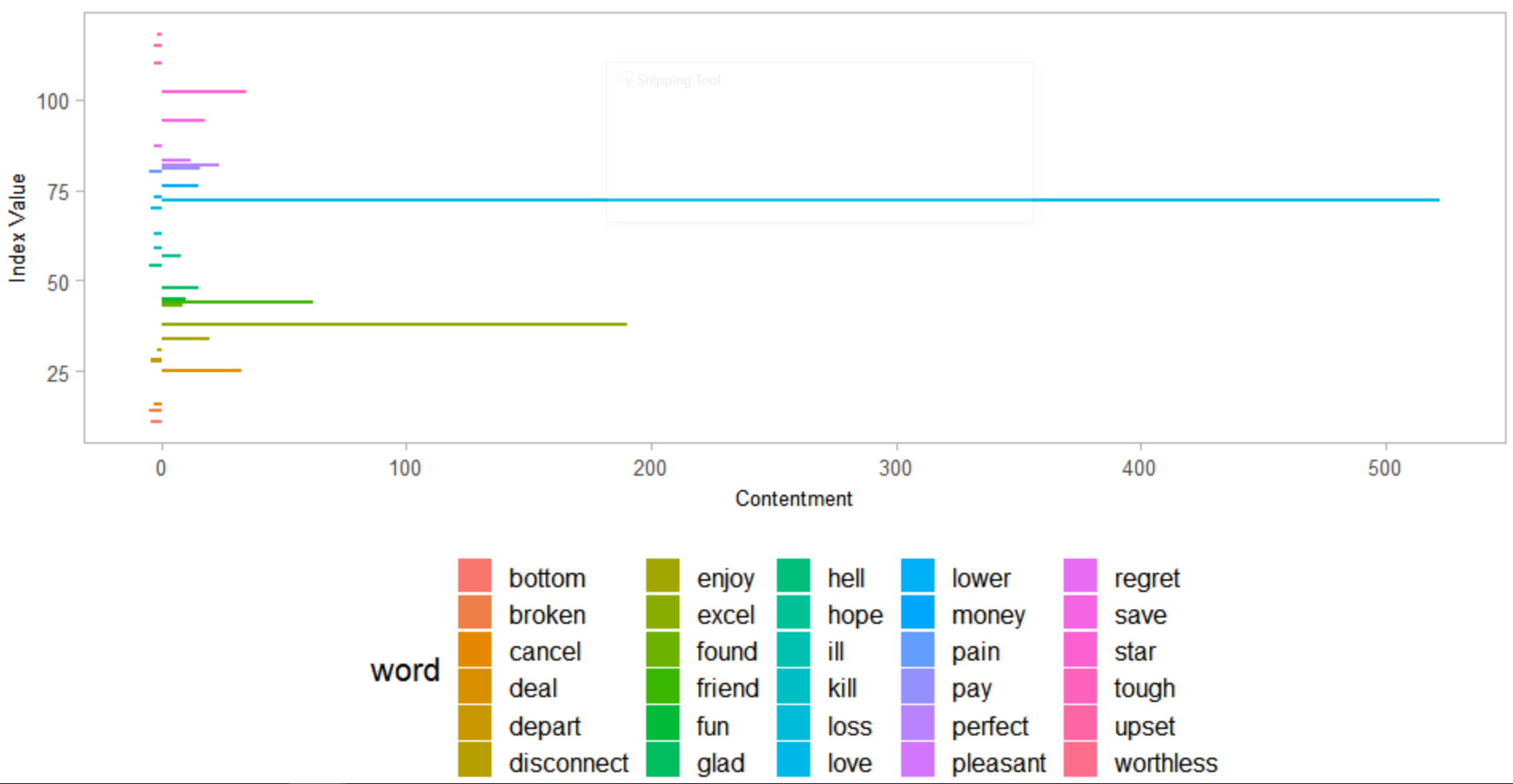
The above figure shows different feeling of the users on the T-Mobile data as we can see that the people are more joyful than sadness. There is more trust compared to fear.



*Fig 4.7 Top 10 words representing feelings of Joy and Sadness*

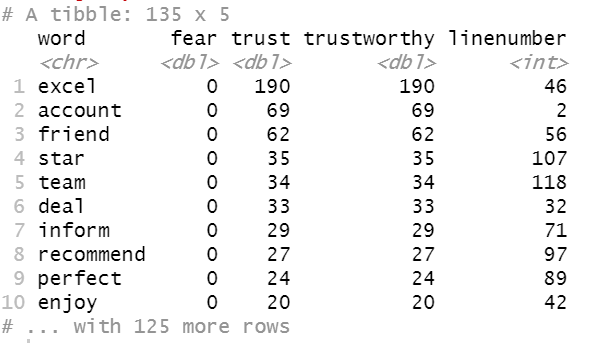


*Fig 4.8 Graphical representation of Fig 4.7, showing level of contentment of each word*

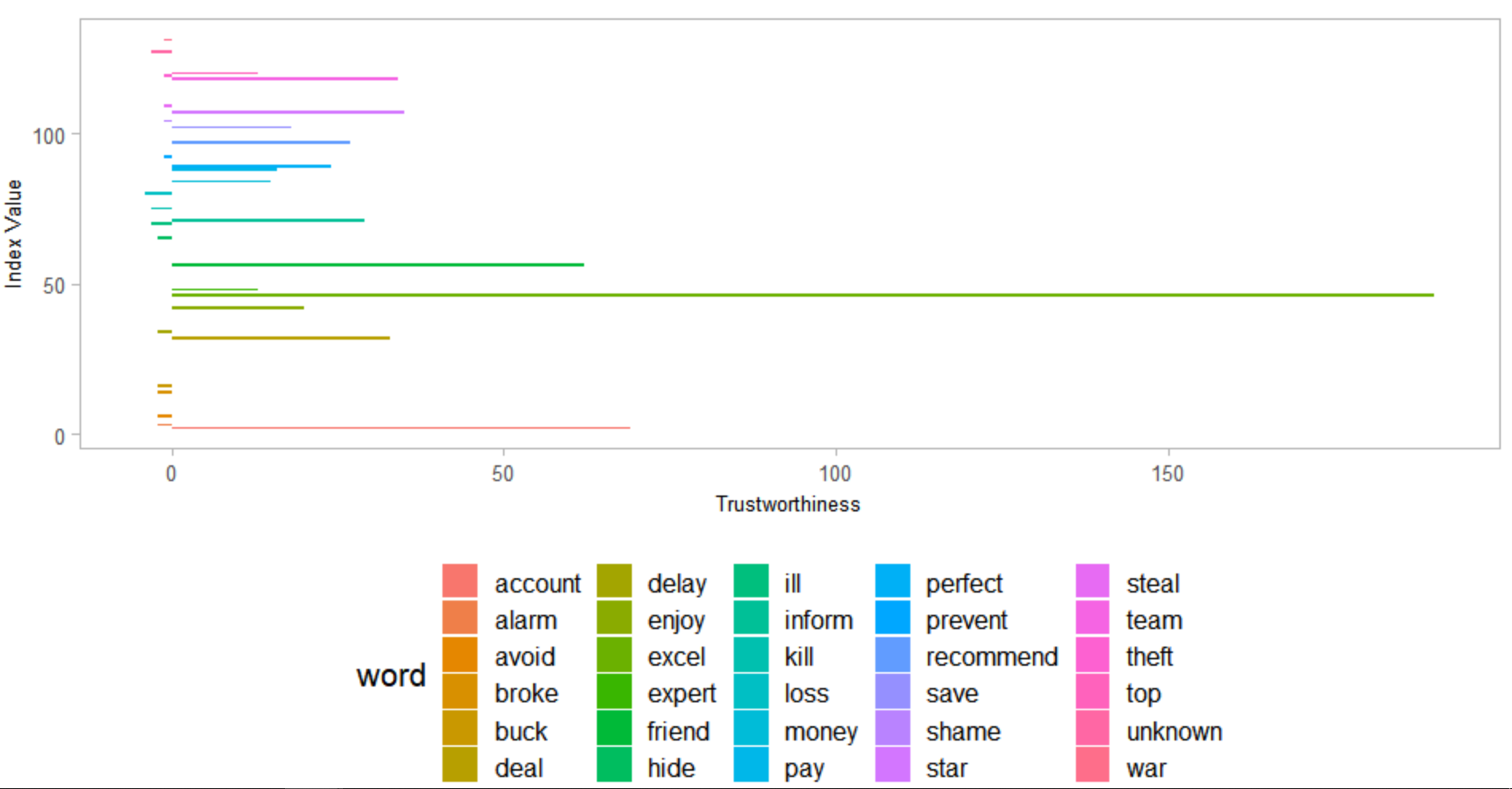


*Fig 4.9 Showing words representing feelings of joy and sadness*

This suggests us that love stands out as the word with shows the most contentment and bottom has the lowest level of contentment. Further, this indicates that people are feeling more joyful while remaining connected with T-Mobile network.



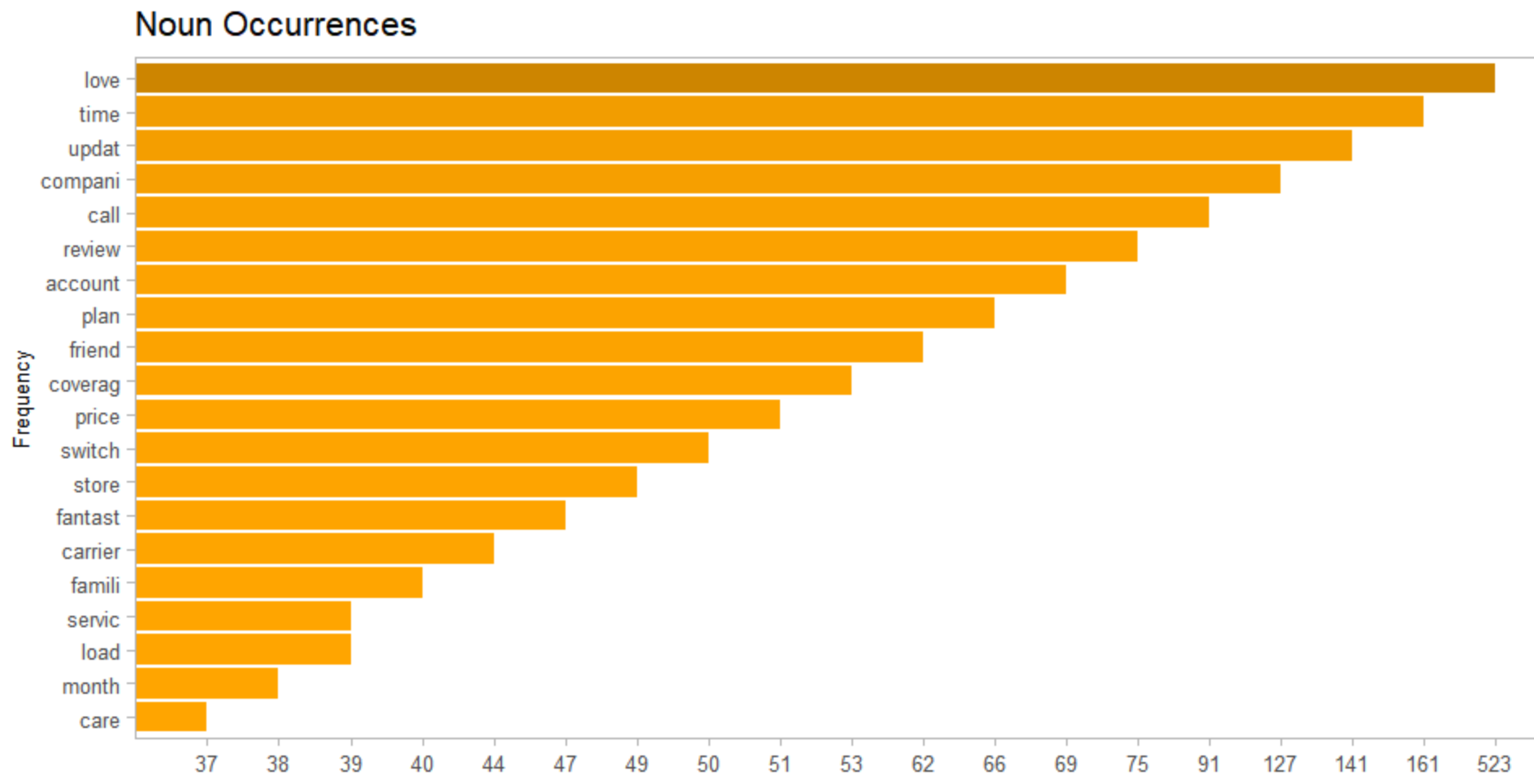
*Fig 4.10 Table showing top 10 words representing feelings of Trust and fear*



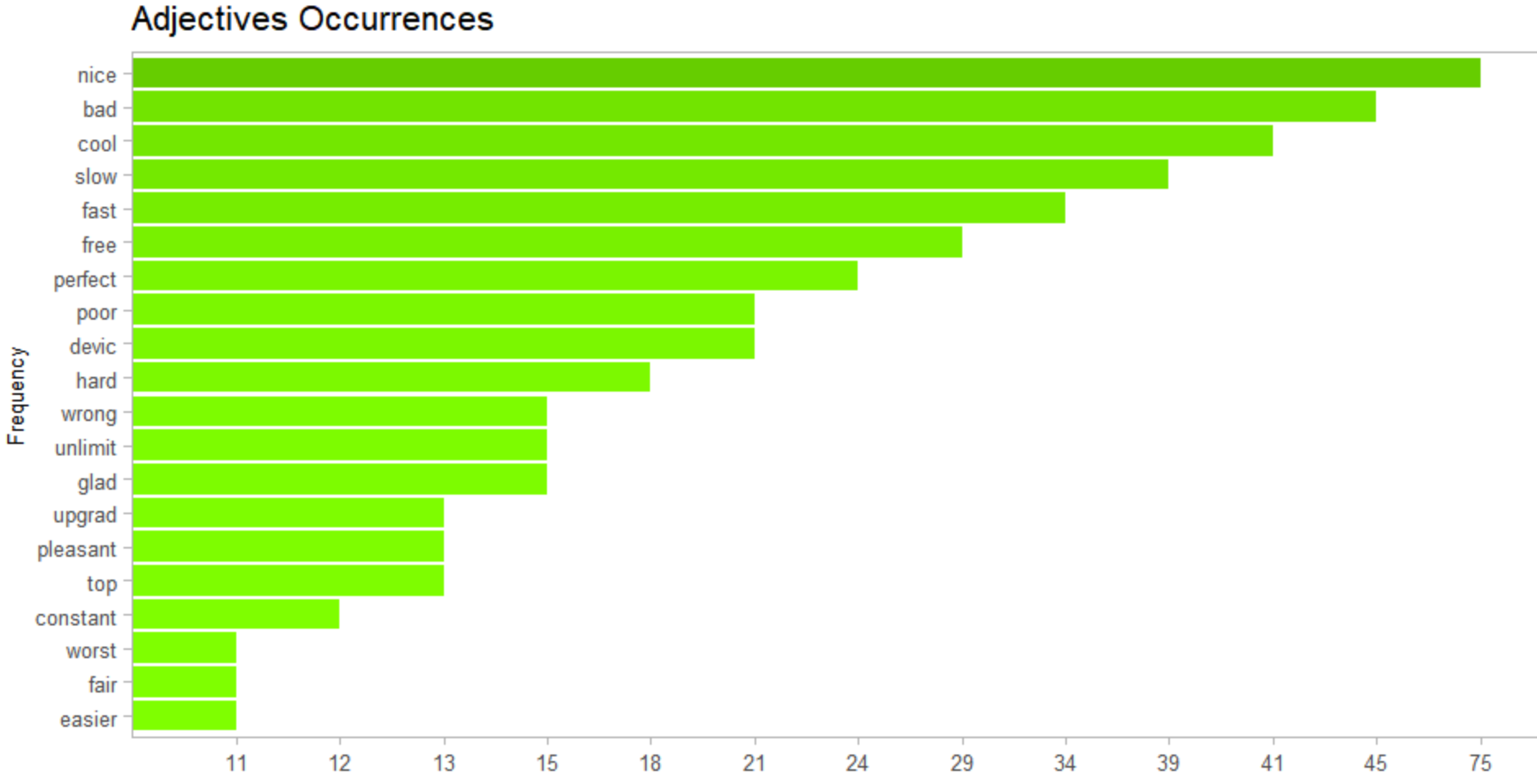
*Fig 4.11 Words depicting feelings of trust and fear*

This tells us that love stands out as the word with shows the most contentment and bad has the lowest level of contentment. This tell us that people are feeling more joyful compared to sadness with T-Mobile network.

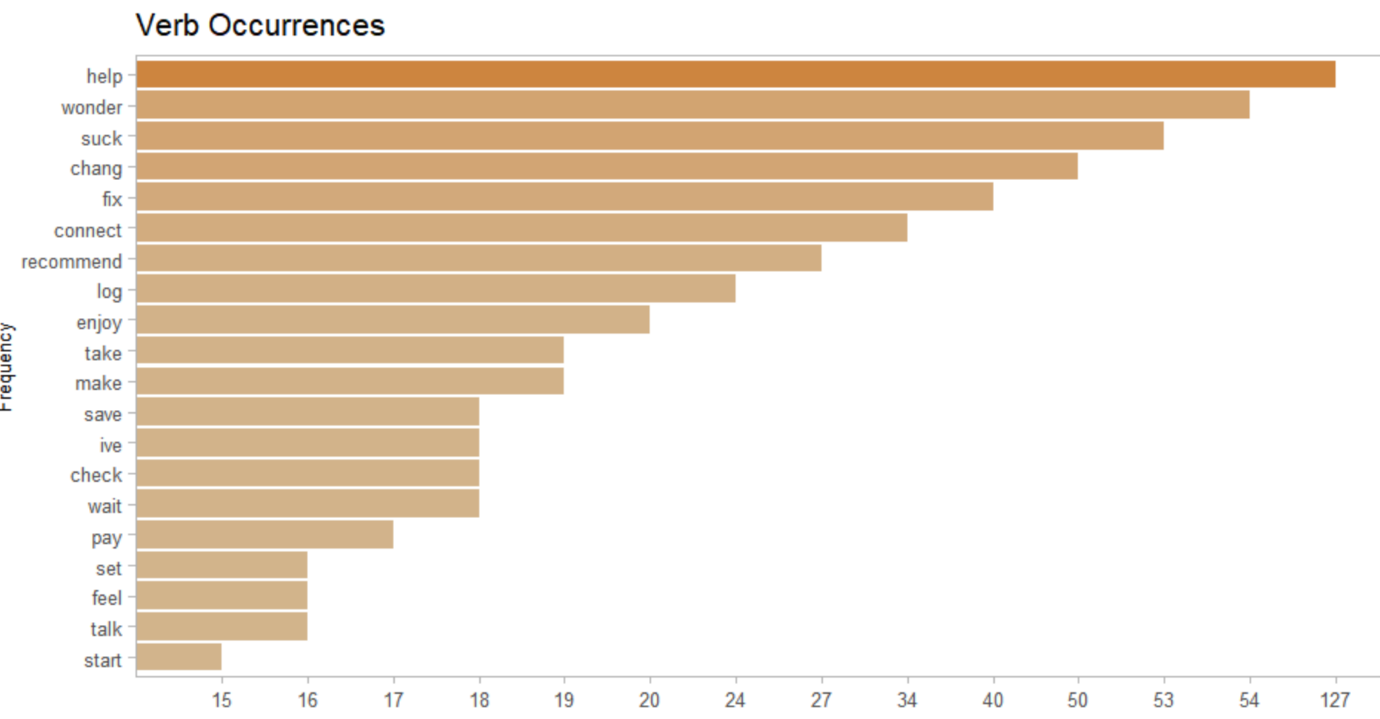
***Parts of Speech recognition***



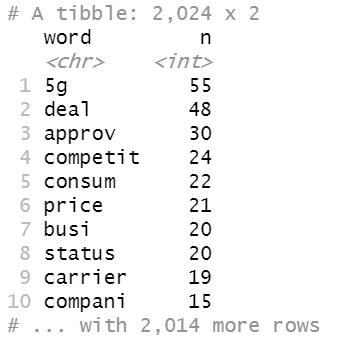
*Fig 4.12 Horizontal bar chart showing top nouns*

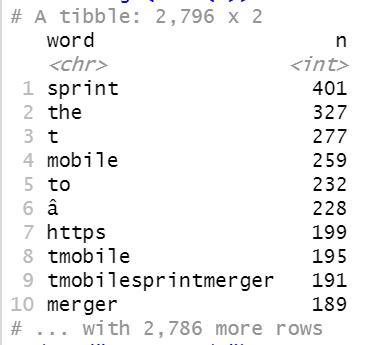
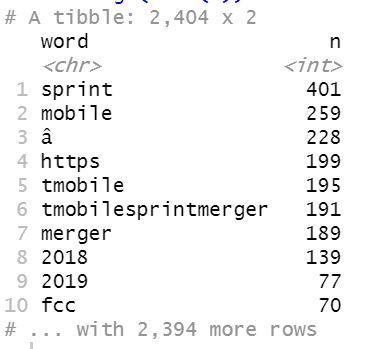


*Fig 4.13 Horizontal bar chart showing top adjectives*



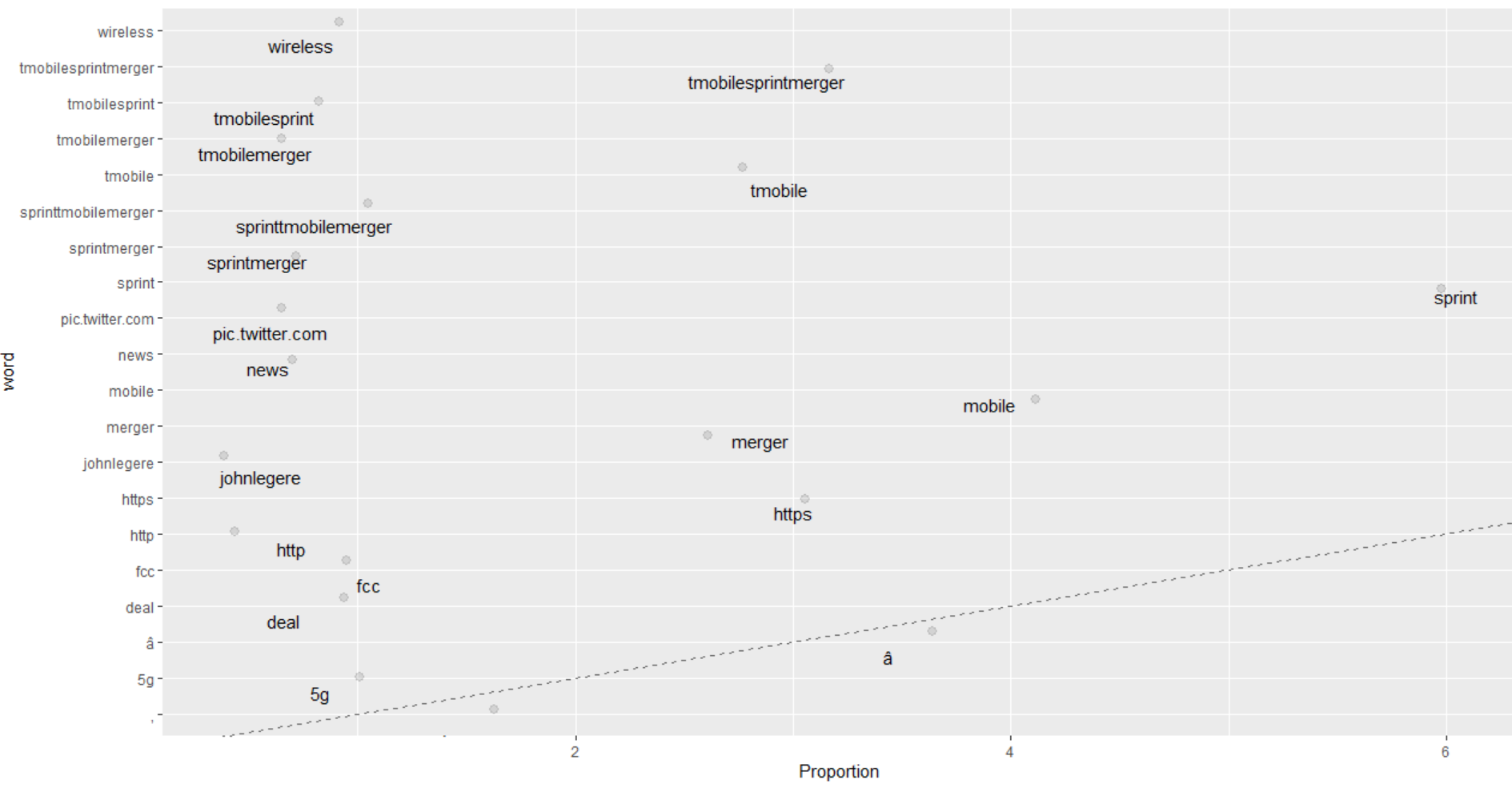
*Fig 4.14 Chart showing top adjectives*

**Analysis of twitter data**



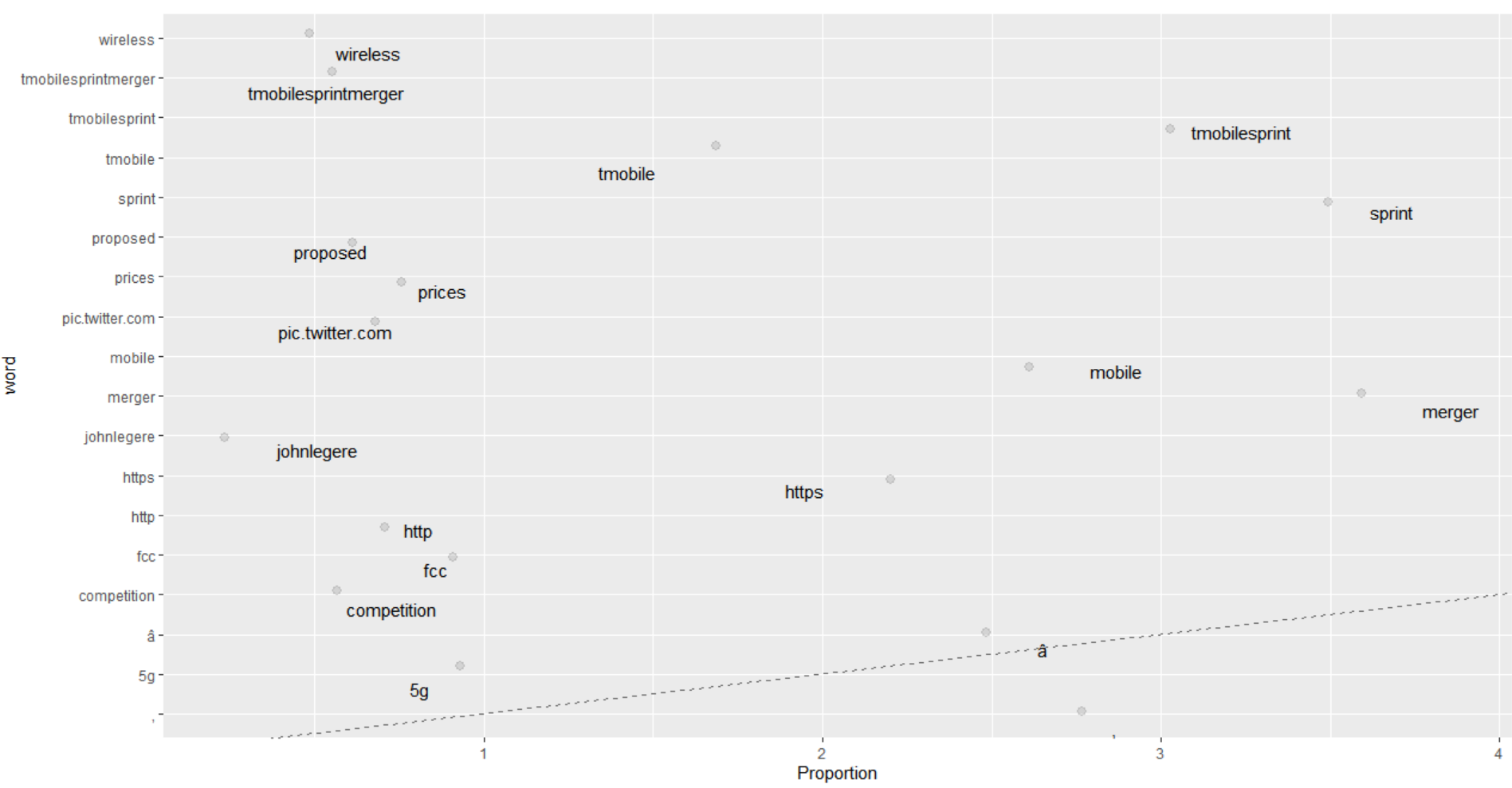
*Fig 5.1 Top 10 words before Fig 5.2 Top 10 words after Fig 5.3 Words after stemming and removal of removing stop words stemming, unnecessary stop words words removed, same*

*frequency was set*



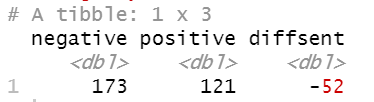
*Fig 5.4 Graph showing word by proportion*

The above figure shows the graphical representation of the proportion of words after removing digits, spaces and unnecessary words. This shows that the word sprint occurred the most times in all the tweets.

******

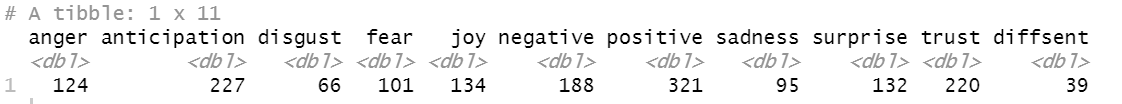
*Fig 5.5 Words set to the same frequency*

Upon setting the words to the same frequency, this figure shows the new proportion of words.

******

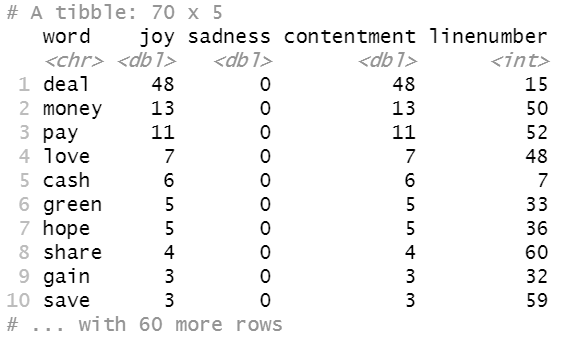
*Fig 5.6 Number of positive and negative sentiments*

The above snapshot shows the number of positive and negative sentiments. This gives us a sense that people are more negative about the merger happening.

******

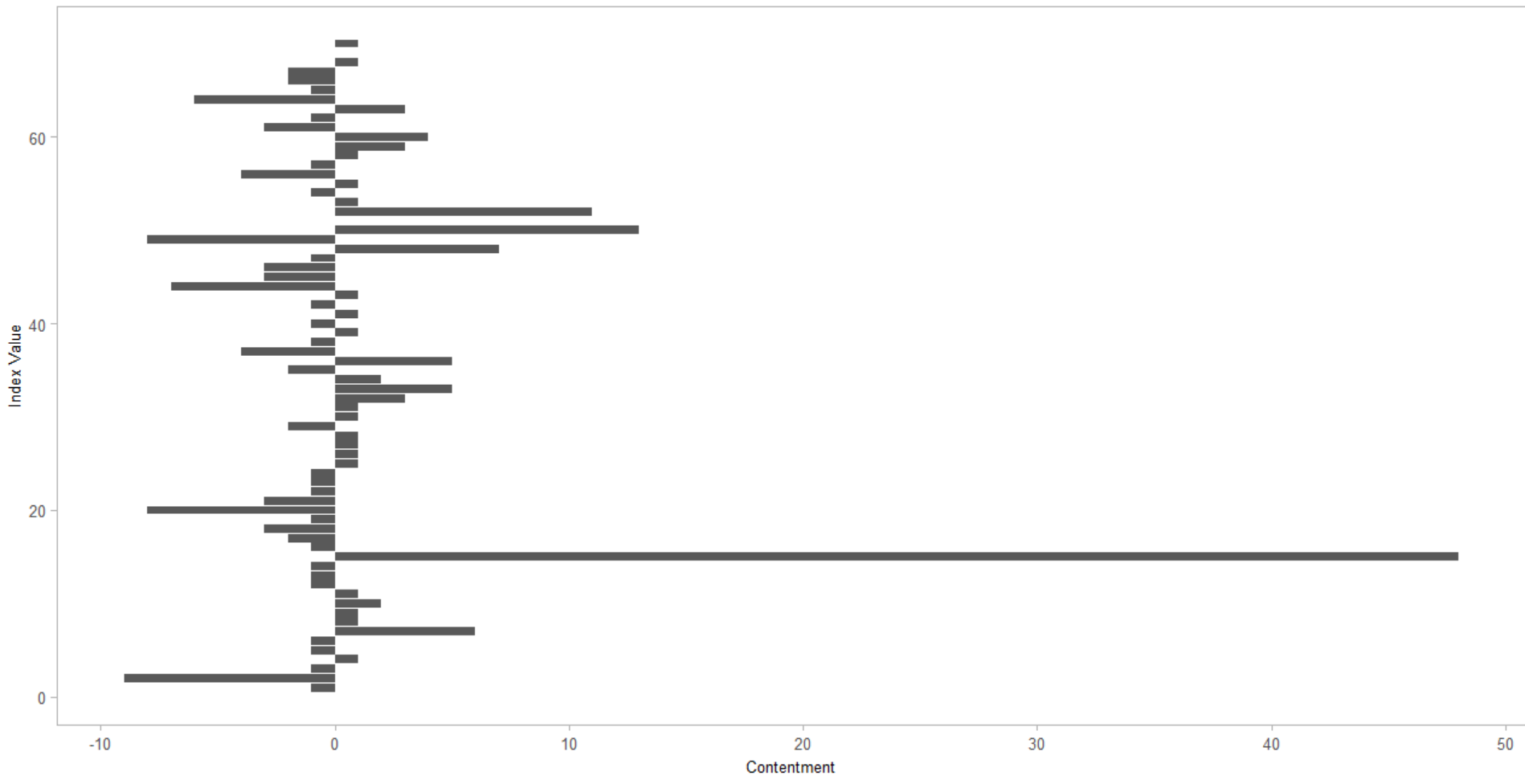
*Fig 5.7 Number of other sentiments*

The figure above shows the number of all other sentiments

******

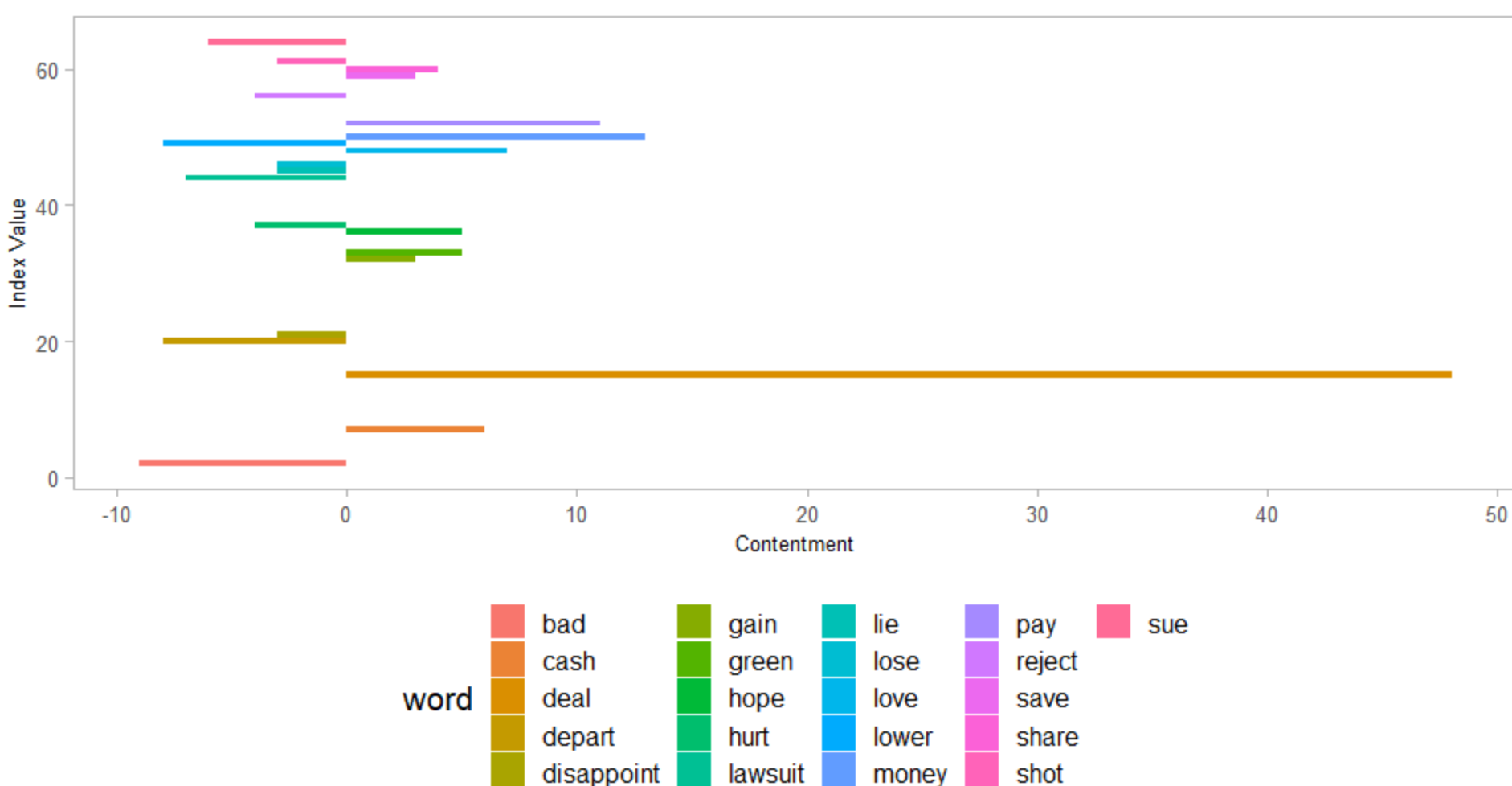
*Fig 5.8 Joy/Sadness and overall contentment*

The figure above shows contentment by each word.

******

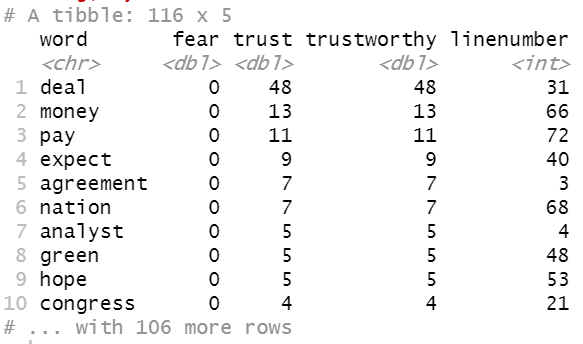
*Fig 5.9 Figure showing degree of contentment*

The above figure shows that more people are happy about the merger.

******

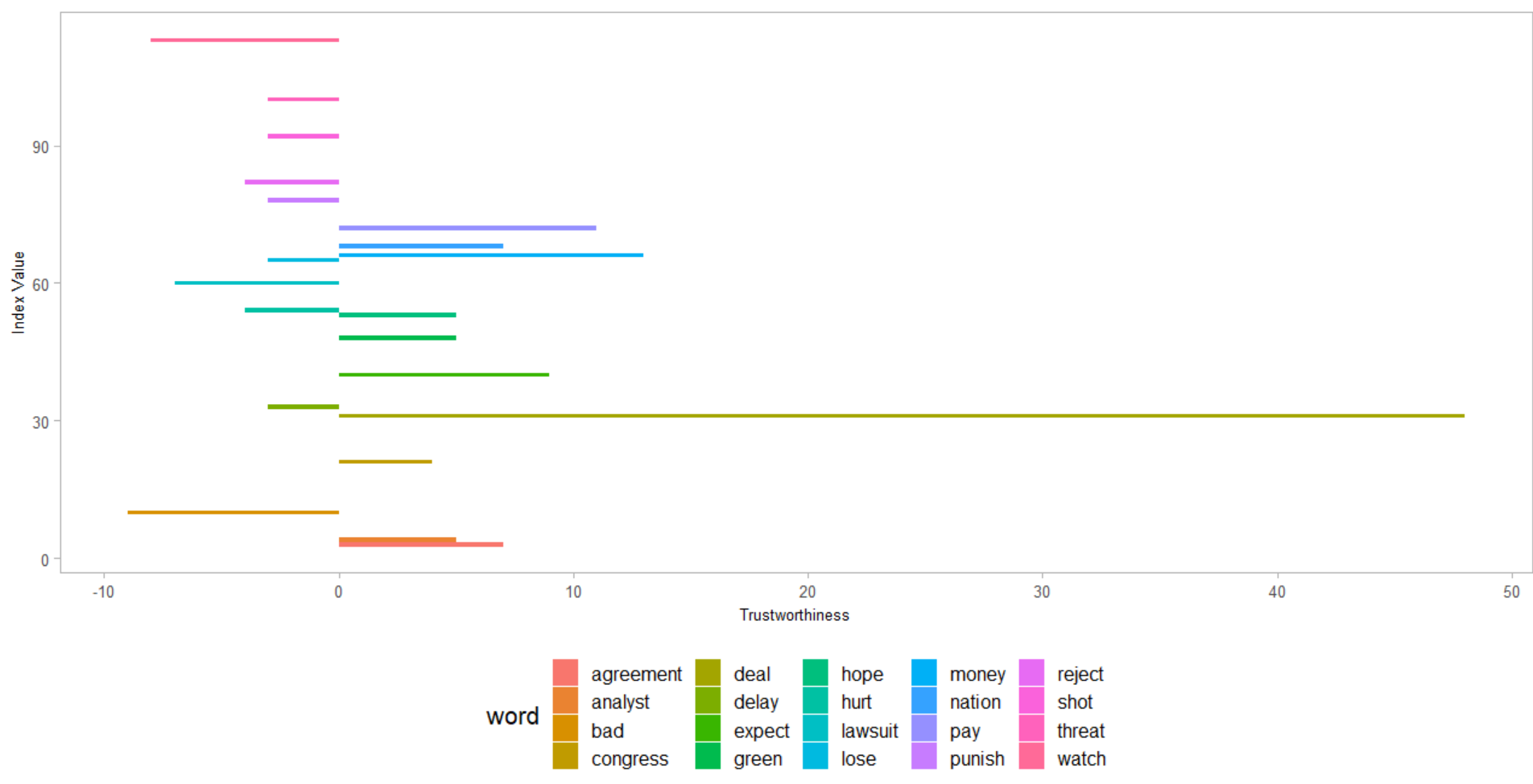
*Fig 5.10 Words with level of contentment*

The figure above shows words and their feelings for joy and sadness. Deal is the top word with highest contentment.

******

*Fig 5.11 Top 10 Words with Trust and Fear Feelings*

Trust and fear words and their contentment.

******

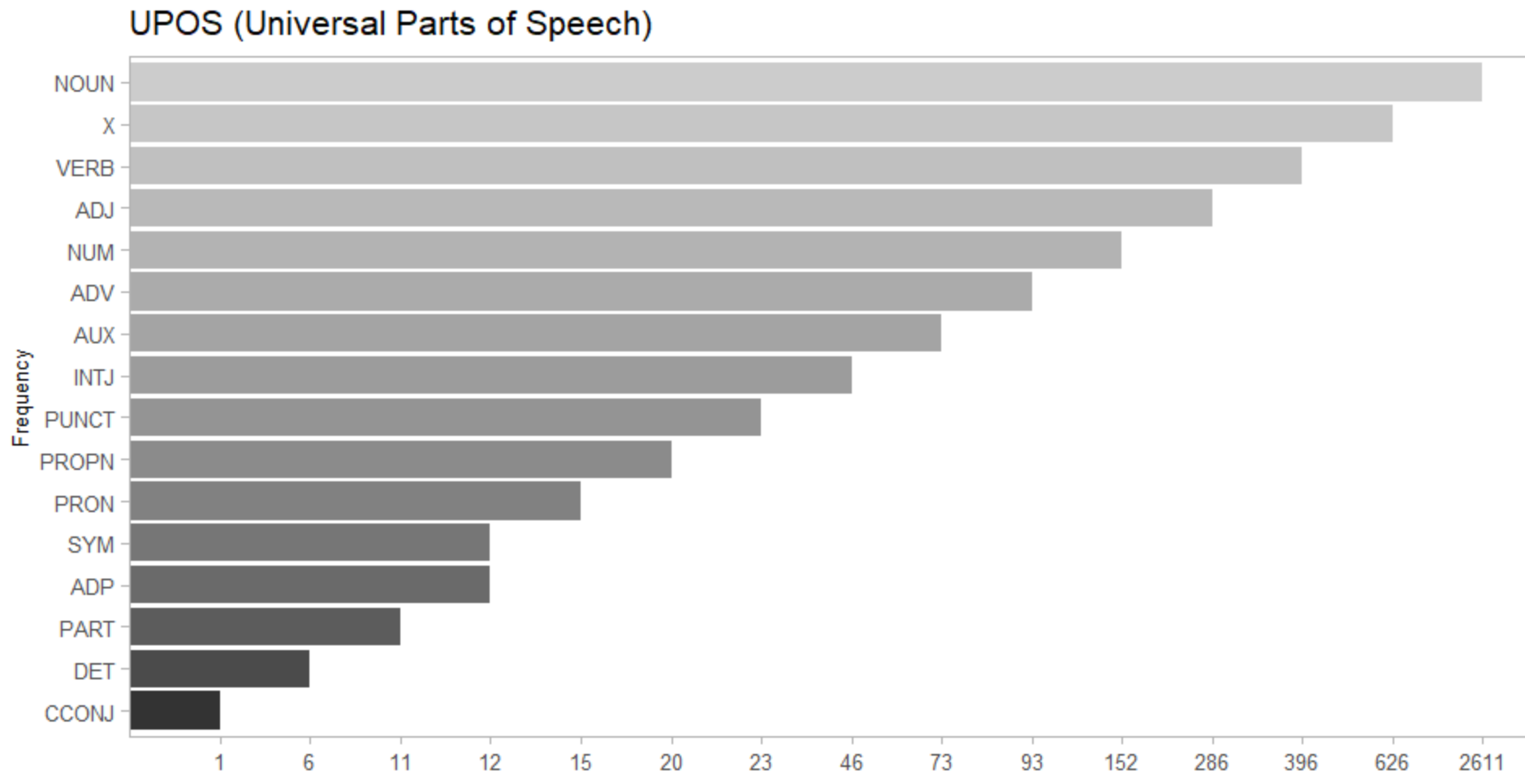
*Fig 5.12 Top words for feelings of Trust and Fear*

The figure above shows words and their feelings for joy and sadness. Deal is the top word with highest contentment.

******

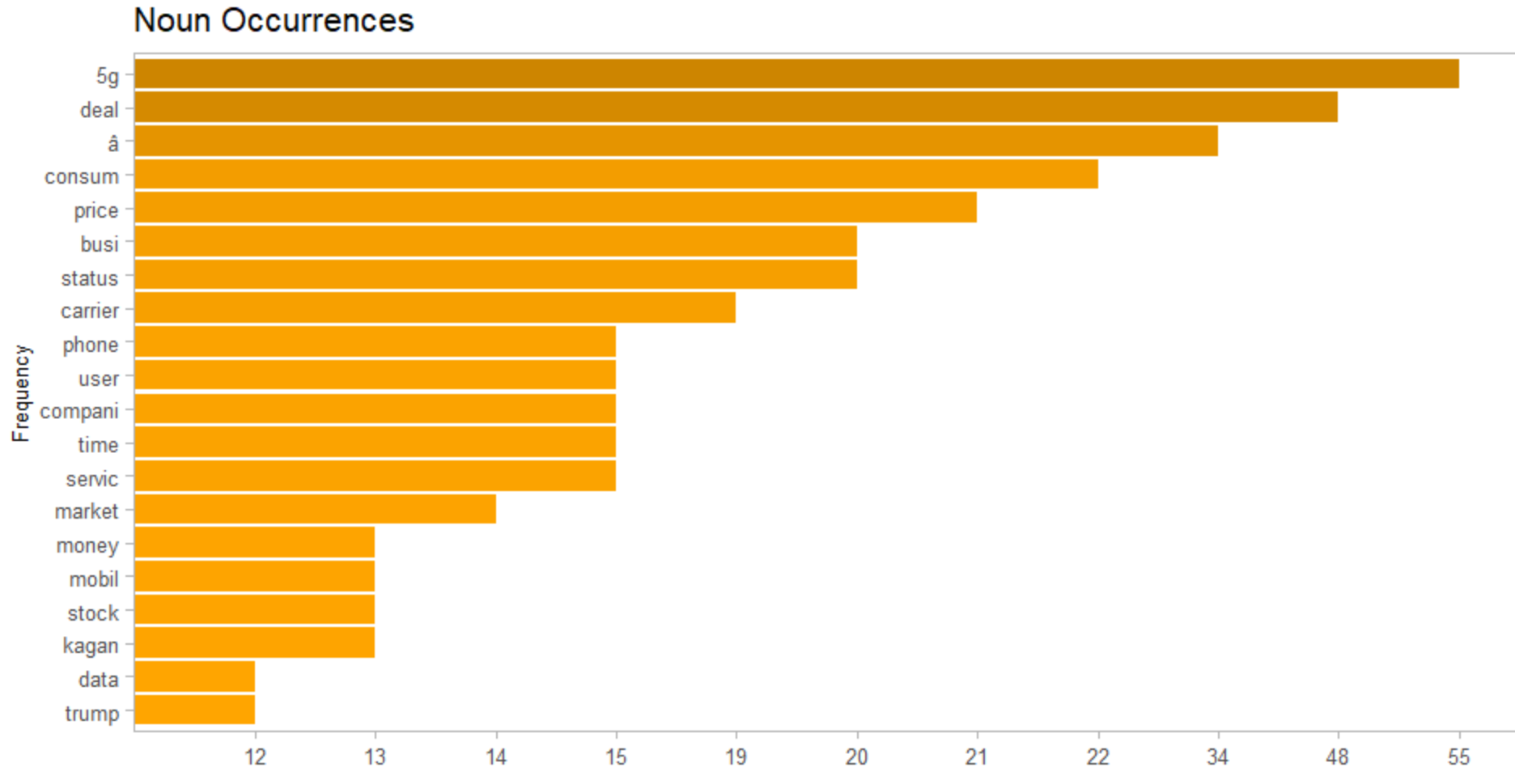
*Fig 5.13 Word cloud showing feelings of trust and fear*

***Parts of Speech Recognition***

******

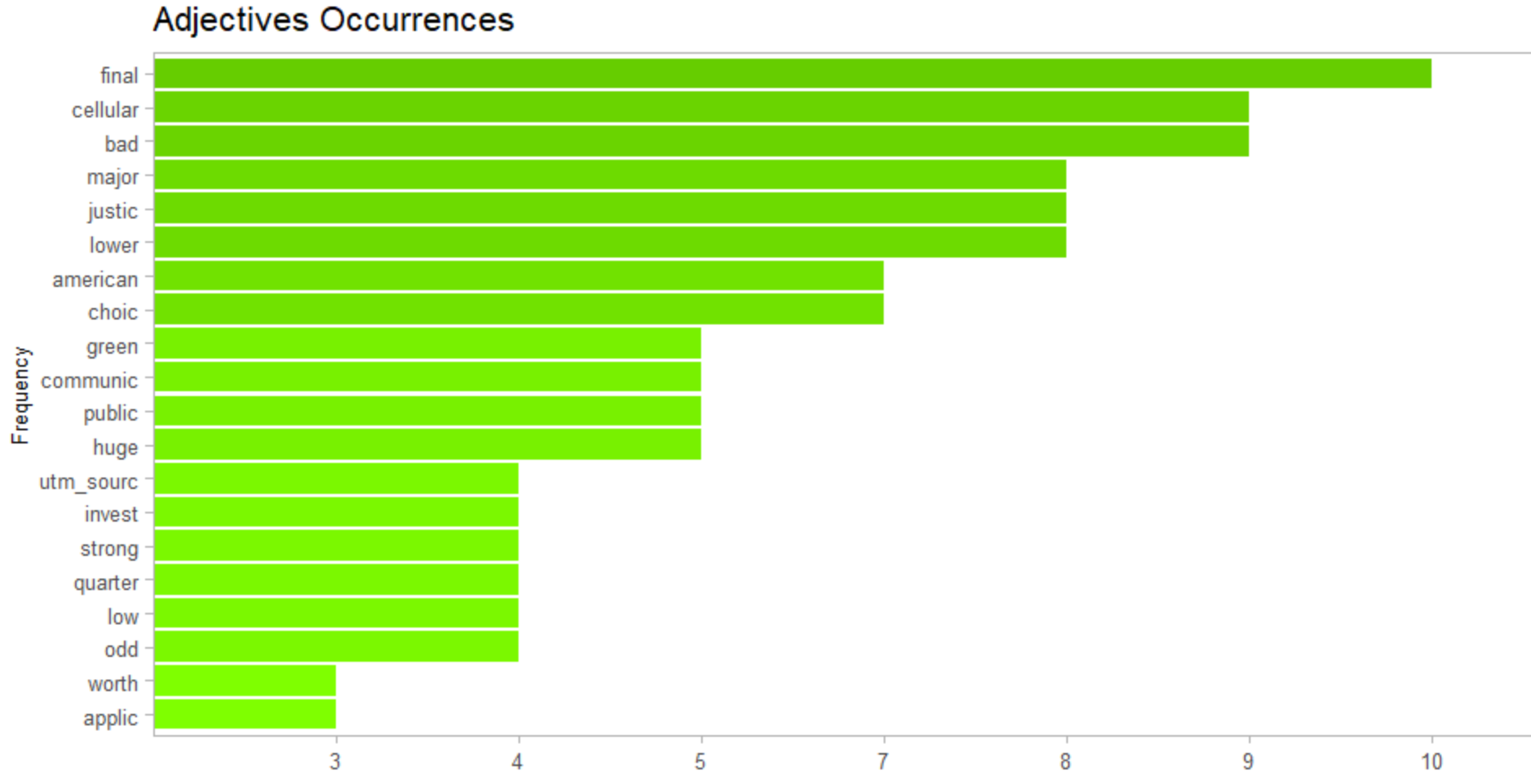
*Fig 5.14 Horizontal bar chart showing universal parts of speech*

The numbers of words of nouns, verbs, adjectives in the text. Diving deeper into each of the top part of speeches like noun, verb, adjectives and adverbs.

******

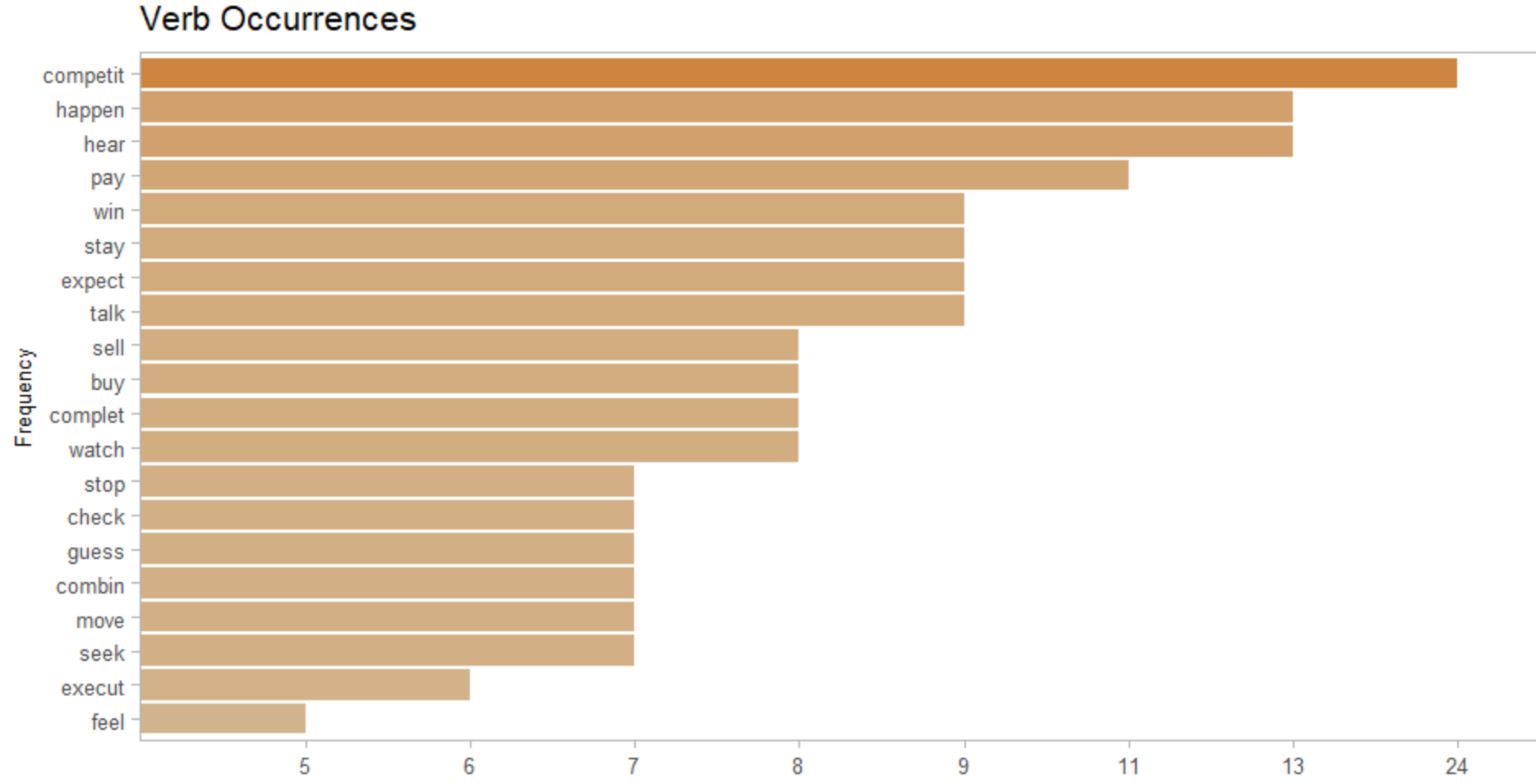
*Fig 5.15 Chart showing top 20 nouns*

According to the bar chart, the noun 5g occurred the most times.

******

*Fig 5.16 Chart showing top 20 adjectives*

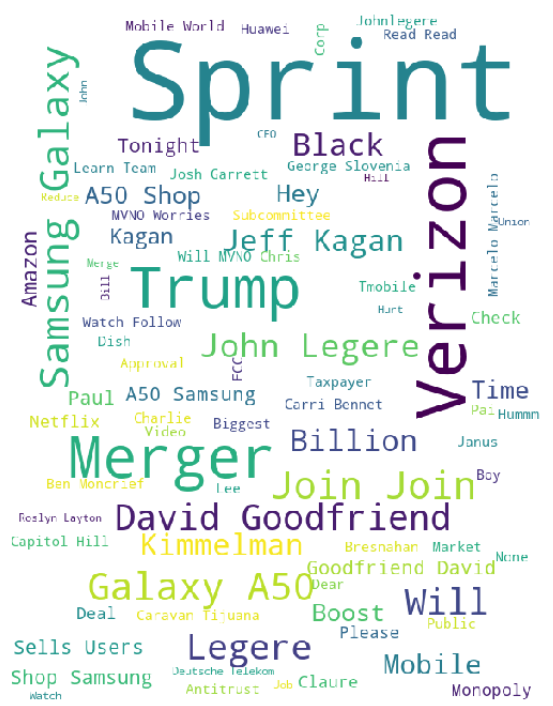
According to the bar chart, the adjective final was the most common.

******

*Fig 5.17 Chart showing top 20 verbs*

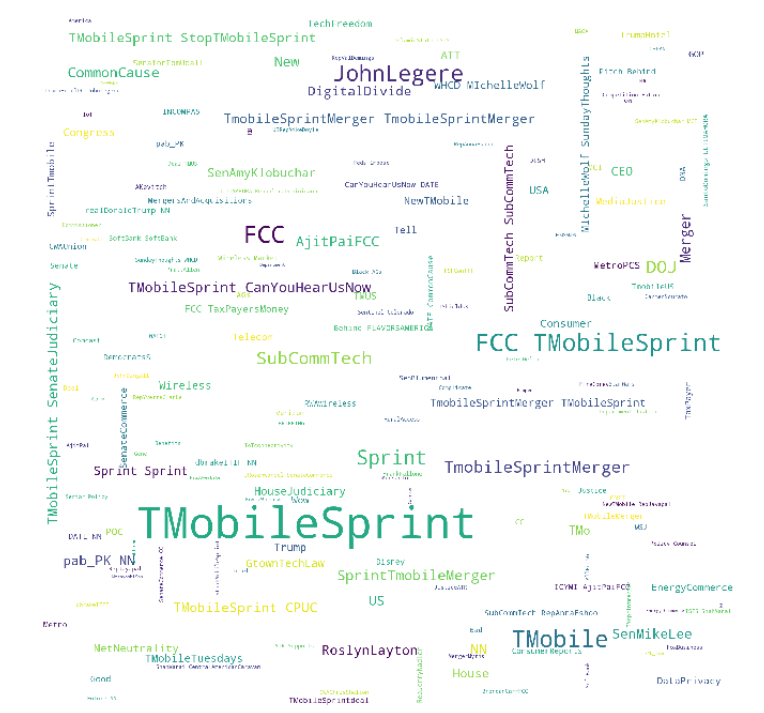
According to the bar chart, the stemmed version of the verb compete was the most common.

***Named entity recognition***

******

*Fig 5.18 Word cloud showing the named entity - person*

The persons such as Trump, David standout from our analysis on the merger comments data



*Fig 5.19 Word cloud showing the named entity - Organization*

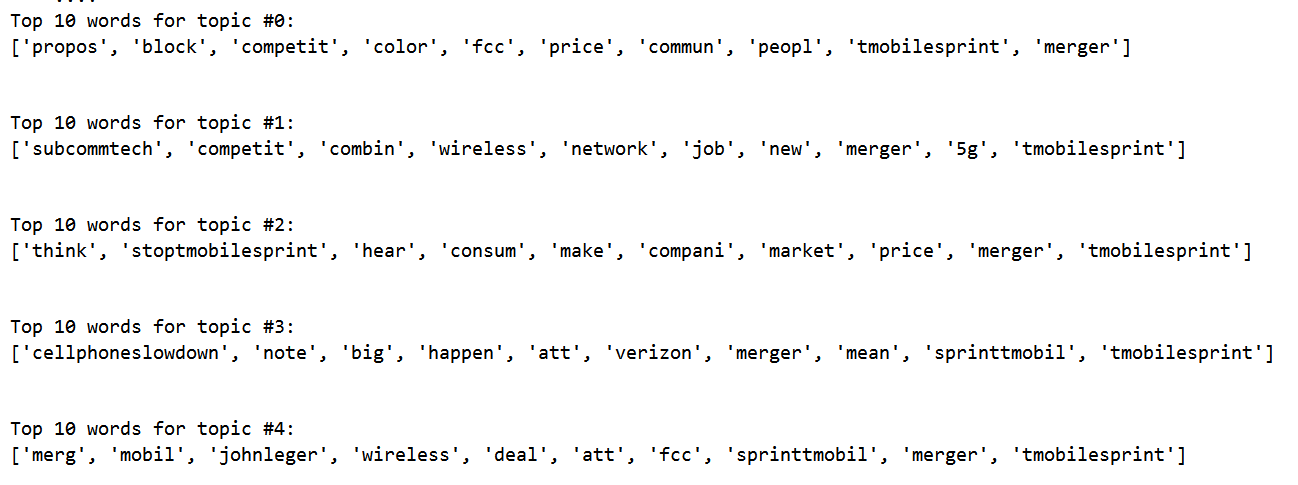
The organizations such as Sprint, T-Mobile standout majorly from our analysis on the merger comments data



*Fig 5.20 Word cloud showing the named entity – Geographical Locations*

The locations such as California, Washington standout majorly from our analysis on the merger comments data.

***Topic analysis on the merger data using Latent Dirichlet Allocation (LDA)***

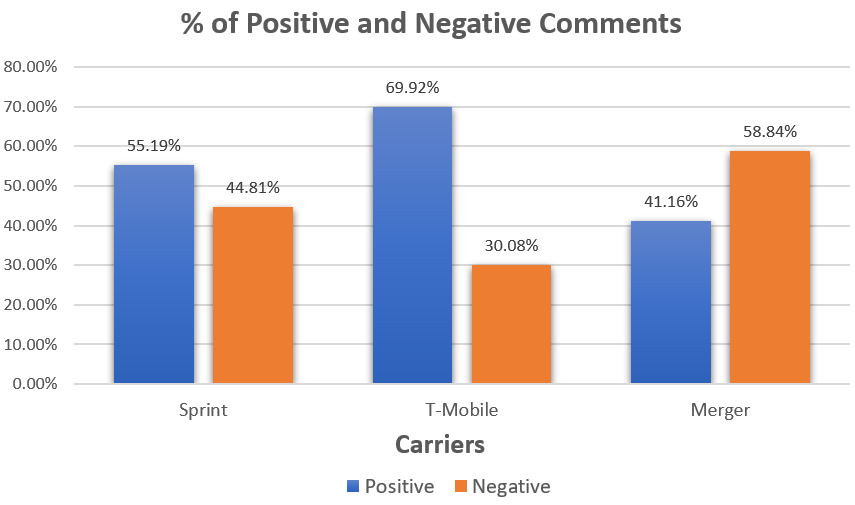
******

*Fig 5.21 Top 10 words from each of the top 5 topics*

Users are mostly worried about the reduction in the jobs and because of the merger the whole market would be dependent only on three major carriers. Some of users are happier about the merger happening.

**Conclusion**

* Customers still seem to provide more positive feedbacks for Sprint and T-Mobile individually, however, the merger has an overall response inclination towards negative
* Overall negative response towards the Merger could be partly due to the reasons that customers might think the prices would go higher, innovation might suffer, loss of jobs or combinations of any or all of these
* Since the merger benefits the T-Mobile and especially Sprint shareholders, there are ought to be some positive response combined to small section of customers who probably like the merger
* The locations such as California, Washington and persons such as Trump, David stood out on the merger comments data
* The possible advantages of this merger include deployment of 5G services. T-Mobile and Sprint have committed to deploy 5G services for covering 97% of Americans within the next 3 years, and inside 6 years to arrive at 99% all things considered

****

**References**

Omar Elgabry (2019/03/02). The Ultimate Guide to Data Cleaning. Retrieved December 10, 2019, from <https://towardsdatascience.com/the-ultimate-guide-to-data-cleaning-3969843991d4>.

[C. Scott Brown](https://www.androidauthority.com/author/cscottbrown/) (2019/10/16). T-Mobile Sprint Merger: Everything you need to know

Retrieved December 10, 2019,<https://www.androidauthority.com/t-mobile-sprint-merger-plans-921398/>