"Opinion Mining" Project
Szymon Klimowicz, Aleksander Stańczak

Abstract

Businesses today are heavily dependent on data. Sifting through huge volumes of this text data is difficult as well as time-consuming. Also, it requires a great deal of expertise and resources to analyze all of that. Definitely it is not an easy task. Fortunately there is Opinion Mining, field that tries to identify and extract opinions within a given text. Through this tool we can understand clients, find company strengths and weaknesses and then formulate recommendations. It is especially very important for business which sell or repair products for individual clients. Future consumers rely on opinions on the internet and for many of them this is the most important factor.

Keywords: cell phone stores, brand knowledge, semantic analysis, topic modelling, text mining, brand clustering

1. Introduction

In this report, our focus is to analyze opinions about the mobile phones repair and reselling business. In order to do that, we gathered data about seven businesses with most ratings and comments (i.e. Apple, Sprint). It's difficult at times to gauge if the company's services are of high quality by looking and only few of the comments or ratings, so our main objective is to find out which brands and services are the most popular (or unpopular) with customers.

2. Theoretical Background

According to *wikipedia.com* text mining is the process of deriving high-quality information from text. High-quality information is typically derived through the devising of patterns and trends through means such as statistical pattern learning. Text mining usually involves the process of structuring the input text (usually parsing, along with the addition of some derived linguistic features and the removal of others, and subsequent insertion into a database), deriving patterns within the structured data, and finally evaluation and interpretation of the output. Typical text mining tasks include text categorization, text clustering, concept/entity extraction, production of granular taxonomies, sentiment analysis, document summarization, and entity relation modeling (i.e., learning relations between named entities).

3. Research Methodology

3.1. The main Definitions Used in the Project

Definitions used in the project: 'Text preprocessing', 'Topic modelling', 'Comments scraping', 'Comments sourcing', 'Clustering', 'Sentiment Analysis'.

3.2. Research Questions

We're interested in finding out answers to research questions such as:

- What are the main topics in the branch?
- Which brands have a good/bad customer opinion? What is the opinion about the branch as a whole?
- What is the degree of engagement of customers in certain topics in each brand and in the whole branch?
- What topics are common for our business and what opinions do customers have about them?

3.3. Research Plan
1. Comments sourcing and scraping using Python
2. Preprocessing of the texts using Python and R
3. Analyzing the corpus and sentiment in the whole branch
4. Generating topics in the corpus, analyzing topic sentiment for each brand
5. Clustering of the brands based on topics shared by the brands
6. Clustering of the brands based on sentiment and degree of engagement

4. Experiments Results

4.1. Comments Scraping.
We choose 7 cell phone stores from website *trustpilot.com* and using *Python* we conducted comments scrapping.
Firstly, we created links generator to each brand. Here is an example of results for *Apple:*
https://www.trustpilot.com/review/www.apple.com?languages=en&page=1
https://www.trustpilot.com/review/www.apple.com?languages=en&page=2
https://www.trustpilot.com/review/www.apple.com?languages=en&page=3
etc. till page 25

Secondly, using created links and library *BeautifulSoup* we scrapped around 500 comments for each brand. Few exemplary comments for *Smarter Phone:*

```
I sent my phone for refund, and I have no answer since 10 days, the site is down, I hope they still exists.

Not bad ....no very bad ....just very...very ....  very ...bad ..not serious.....

Still waiting for my refund.. I sent back my iPhone on the 12.11 because the general battery was at 78%. The iPhone would turn off in
2 hours. We are the 3rd December and I am still waiting for a refund !!! The customer service is very slow to answer me and I am really really
disappointed. I will continue to write messages since I don't get the refund.

the phone was very good and at a very nice price

A great company to deal with, swift trouble-free transaction with good quality material!

Order a premium grade iphone. I received a phone full a dents and visible scratches on the screen. They sent me a basic grade product
(for me it is poor grade) Try to sent it back but can' t. The phone was delivered very late (~mouth) and the return period expired (it started
from order not after receive) ! They offer a replacement but not with a premium grade only with basic. Very bad experience... Never again

The phone worked fine for about a month, then many things on the phone just stopped working, like the on off button.  And out phone
service, that was always good is horrible now.  They think it is the phone. ☺
```

4.2. Corpus Preprocessing.

Using libraries *nltk, string, re* in *Python* we conducted data cleaning tasks such as:
- Converting document to lower case
- Removing numbers
- Removing punctuation marks
- Removing english stopwords
- Removing extra whitespace
- Removing special characters

Then we performed text stemming (also based on *nltk*). Here is an example of our work for *Smarter Phone:*

```
sent phone refund answer since days site hope still exists bad bad veryvery bad serious still waiting refund sent back iphone general
battery iphone would turn hours rd december still waiting refund customer service slow answer really really disappointed continue
write messages since dont get refund phone good nice price great company deal swift troublefree transaction good quality material
order premium grade iphone received phone full dents visible scratches screen sent basic grade product poor grade try sent back phone
delivered late mouth return period expired started order receive offer replacement premium grade basic bad experience never phone
worked fine month many things phone stopped working like button phone service always good horrible think phone phone didnt genuine
apple battery two months started shutting time fois good satisfied great buys quality merchandise guys needs improve service horrible
phone connect cellular data services phones works well good products good customer service phone never turn everyone attitude pay
money gift embarrassed disappointed great company fast shipping definitly order many thanks first little skeptical ipad pro bought
practically brand new functions exactly like needed detailed shipment details time definitely purchase smarterphone suckssss phone
didnt even work bought two iphone one good one bad start screen issue called told send phone get checked diagnose take approximately
days also told going get call tag mail waited waited n guess call tags arrived take local store get fixed ended coasting replace
screen never service excellent start finish id never used company usual pessimistic phone arrived promised condition said would order
one wife thank faulty phone unable return v disappointed promised month warranty pretty fast able use new one right got general nice
website purchase refurbished phone update prices often phone recieved advertised scratches blemishes nicely packaged side difficult
know carrier type whatwhere phone could used phone came south pacific im usa phone software updated work us choose different network
carrier website claimed bottom line information english website greatly lacking adding guessing phone work ive ordered one iphone gb
shipping fast problem audio loudspeaker phone audio device functioning proper could listen phone calls neither audio video contents
told options go repair center fix smarter would refund maximum went normal speaker phone calls fixed ok cost one werent able fix
substituting would cost around would make would paying repair product hands functionally ive ship back smarter substitution went fast
less weeks another phone also time problems slot sim card missing insert sim card went repair store buy made phone still functioning
already spent extra problems right make another reclaim lose time decided spend money phone worked thing first one battery lasted ok
battery app written indicated battery could achieve around initial capacity second one around charge least twice day fully charge use
one hour facebook app battery completely goes one hour use second third problem didnt made claim mainly mood wait another weeks
another phone mostly sure rd phone would better come without problems ive decided quit pay new battery end ive paid repairs first two
weeks phone expensive price ive paid smarter admit smart wanted try several reasons buy reconditioned phone tc g first phone
functioning took month new one second phone defects great support fast reliable great product great products faster delivery sent
locked phone return peru get replaced shipping cost months aggravation hours dealing dhl customs lima peru live hr bus ride lima sorry
cant recommend sincerely h wooten difficult reach customer service person phone wait around minutes talk someone customer service
hours also unpredictable sometimes would call never get answer even though website said open phone support customer support person
knowledgeable friendly reliable good products issue easily fixed sent empty box charged get credit card company involved take care
charge price hard beat poor quality worse customer service left seeking replacements every months phone looked new works perfect
```

4.3. General Corpus Analysis.

Firstly we were interested in corpus created from all seven brands. We began from creating Document Term Matrix. Next, there there was a time for mining the corpus. We plotted all words with 500 as a minimum frequency:
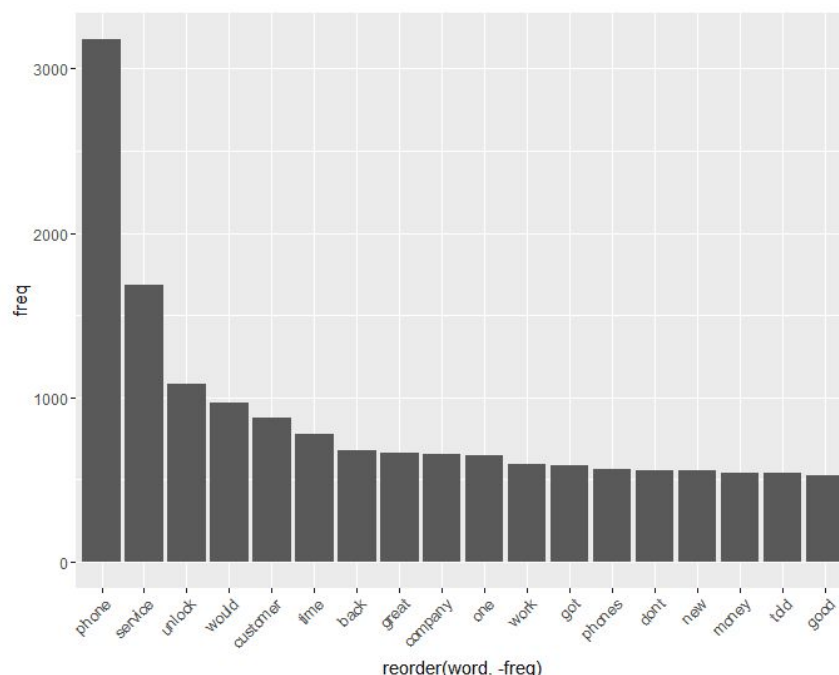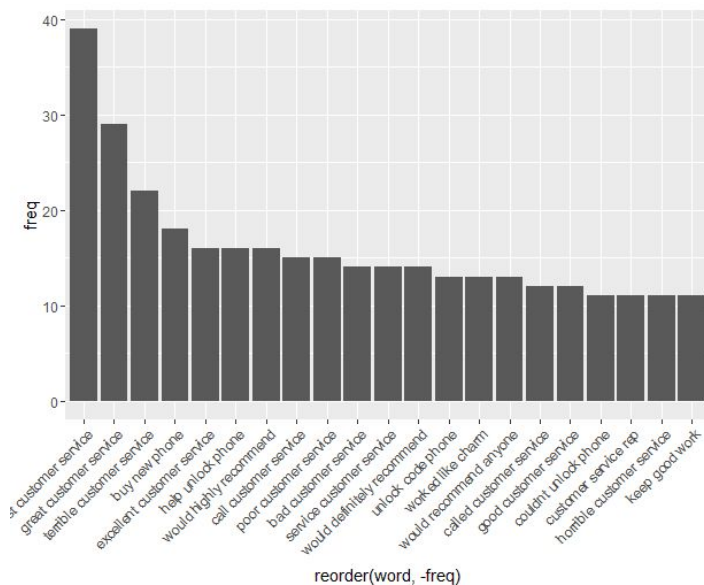
Moreover we created a wordcloud with 150 as a minimum frequency:
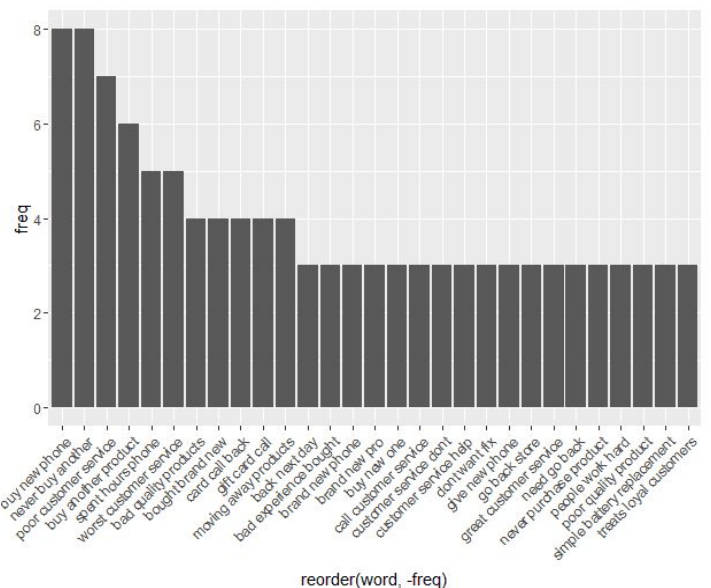

This steps were repeated for all 7 brands and charts for each brand are presented in Appendix (7.2.1 Single terms graphs).

After analysis of all 8 corpuses we decided to create trigrams. That decision was accurate, trigrams help us to see more relevant results:
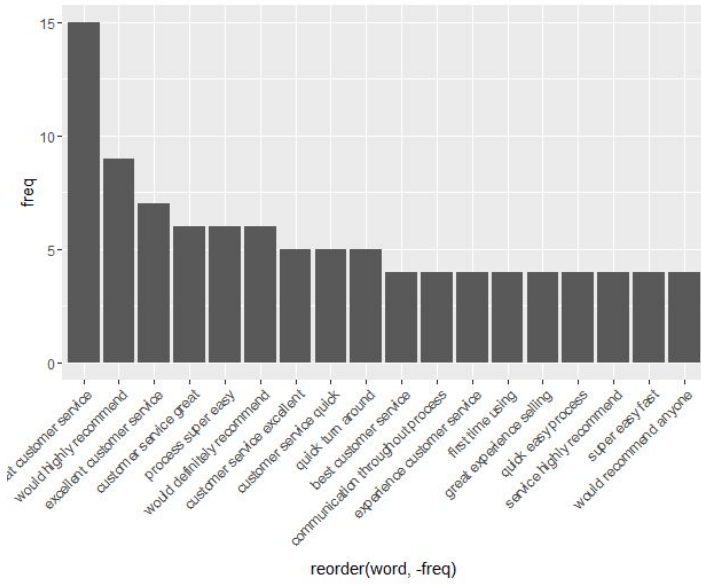
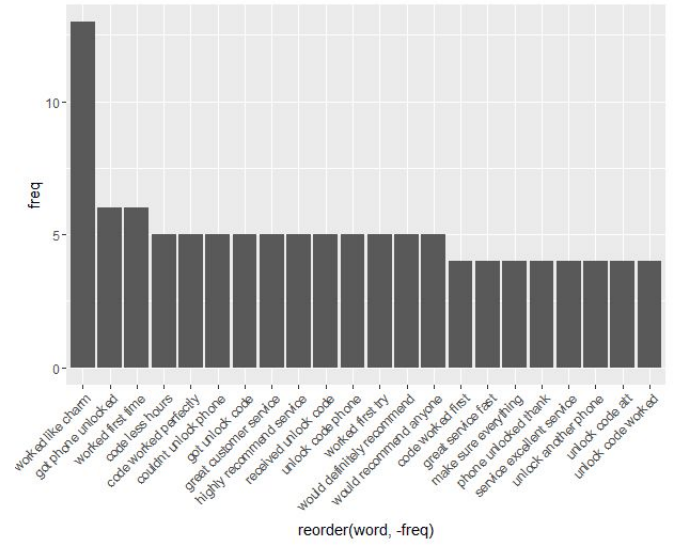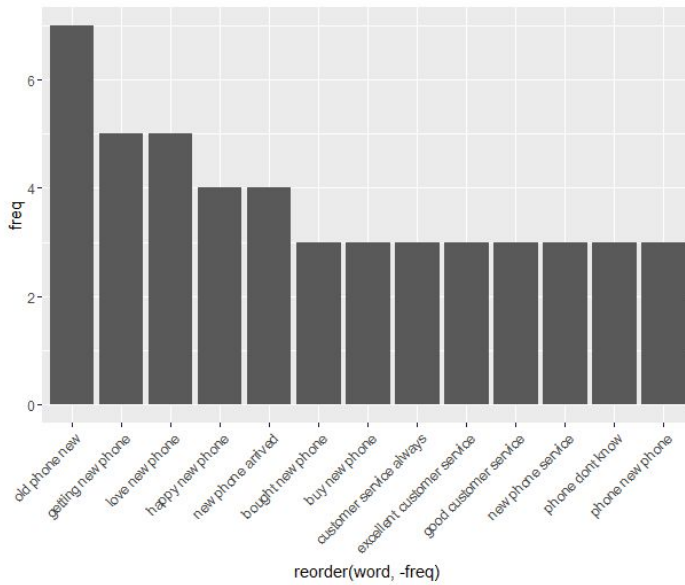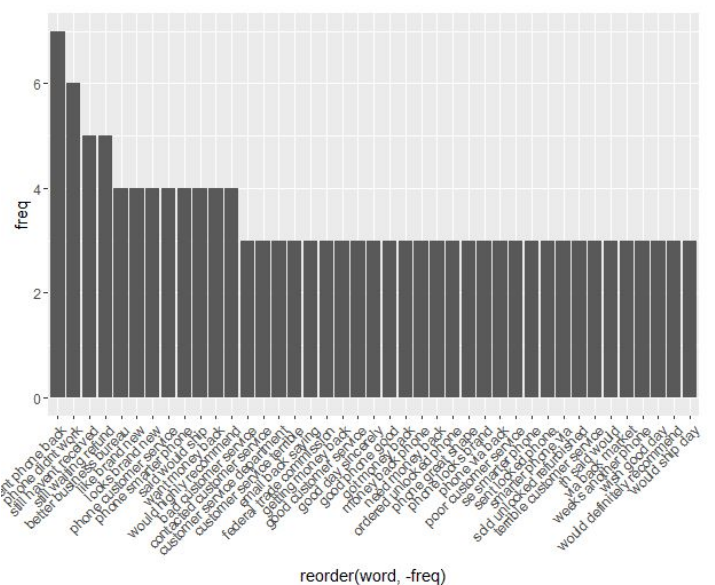● **Corpus created from all brands**   ● **Apple**

● Buyback Boss

freq

reorder(word, -freq)

● Cellunlocker

freq

reorder(word, -freq)

● Republic Wireless

freq

reorder(word, -freq)

● Smarter Phone

freq

reorder(word, -freq)

● Sprint

freq

reorder(word, -freq)

● UnlockBoot

freq

reorder(word, -freq)
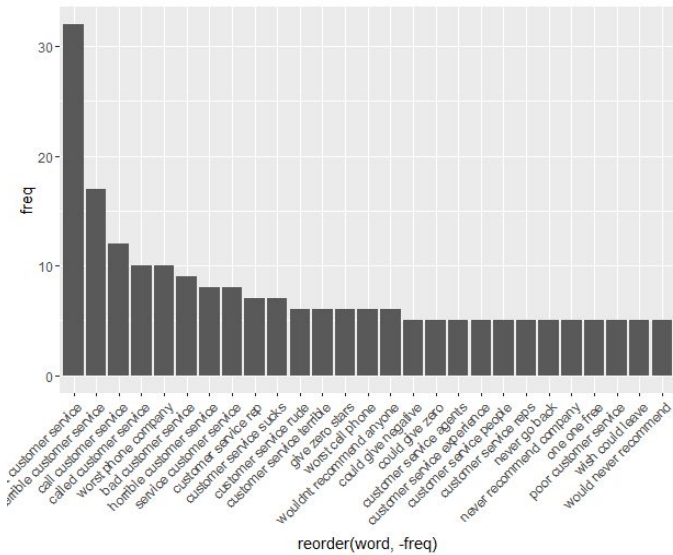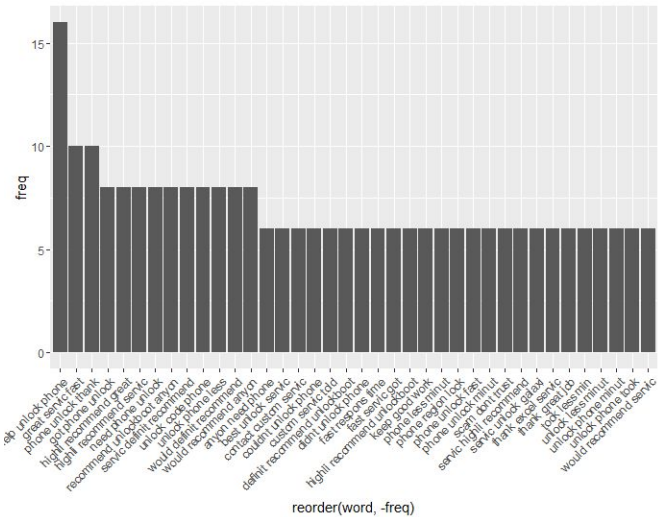
4.4. Topic Modelling

4.4.1. Whole corpus
We chose 3 as a number of topics within the whole Corpus. Here are contextual names to each received topic and names of the documents, which were assigned to particular Topics.

Topics	Product and delivery	Customer service and product	Unlocking phone
Terms	"phone" "service" "great" "received" "good" "easy" "would" "company" "days" "sent"	phone" "service" "customer" "would" "told" "dont" "never" "back" "one" "new"	unlock" "phone" "work" "thank" "recommend" "fast" "help" "great" "code" "time"
Documents	buybackboss_formatted.txt cellunlocker_formatted.txt republicwireless_formatted.txt smarter-phone_formatted.txt	sprint_formatted.txt apple_formatted.txt	unlockboot_formatted.txt

We selected 2 topics discussed by customers in each brand and we gave them contextual names.

4.4.2. Apple

Topics	Products	Service and support
Terms	"phone" "store" "products" "customer" "would" "one" "told"	"service" "new" "back" "time" "said" "support" "could"

	"dont"	"product"

4.4.3. BuyBack Boss

Topics	Delivery	Service
Terms	"quick" "fast" "experience" "customer" "received" "check"	"phone" "easy" "great" "service" "process" "would"

4.4.4. CellUnlocker

Topics	Service	Phone unlocking
Terms	"service" "code" "got" "good" "thank" "first"	"unlock" "phone" "unlocked" "cellunlocker" "worked" "great"

4.4.5. Republic Wireless

Topics	Product and service	Facilities
Terms	"phone" "service" "new" "good" "phones" "customer"	"easy" "wireless" "time" "would" "old" "price"

4.4.6. Smarter Phone

Topics	Service	Refund
	"phone" "back"	"refund" "smarter"

Terms	"order" "service" "days" "got"	"would" "ordered" "company" "received"

4.4.7. Sprint

Topics	Customer service	Products
Terms	"service" "customer" "would" "bill" "company" "pay"	"phone" "phones" "told" "one" "back" "call"

4.4.8. Unlockboot

Topics	Unlocking phone	Service
Terms	"unlock" "work" "thank" "recommend" "help" "unlockboot"	"phone" "service" "great" "fast" "day" "took"

4.5. Topic-Oriented Sentiment and degree of engagement analysis of each Brand comments

4.5.1. First of all, we performed sentiment analysis of each brand in general. Plots are in Appendix (7.2.2 General sentiment analysis of brands).

4.5.2. The Polarity of customer opinions on each topic about each Brand:

Company	Topic	Neutral %	Negative %	Positive %	Overall, Compound
Apple	Products	78.17	11,15	10.29	Negative -0.16

	Service and support	78.84	9.78	10.99	Negative -0.08
Buyback Boss	Delivery	60.52	1.25	37.93	Positive 0.68
	Service	73.47	3.40	22.52	Positive 0.61
Cellunlocker	Service	64.02	2.59	33.07	Positive 0.65
	Phone unlocking	79.02	4.87	15.59	Positive 0.42
Republic Wireless	Product and service	67.57	3.74	28.33	Positive 0.49
	Facilities	73.51	4.91	21.14	Positive 0.42
Smarter Phone	Refund	69.53	6.94	23.15	Positive 0.35
	Service	79.94	8.99	10.64	Negative -0.05
Sprint	Customer service	83.38	10.26	5.91	Negative -0.36
	Products	76.84	15.96	6.85	Negative -0.43
Unlockboot	Unlocking phone	77.05	6.00	16.39	Positive 0.23
	Service	64.20	1.83	33.66	Positive 0.63

4.5.3. The Polarity of customer opinions throughout the Branch as a whole: results in Appendix (7.3.1 Sentiment analysis of branch as a whole)

4.5.4. The degree of Customer Engagement in the discussion of certain Topics for each Brand

Company	Topic	Degree of engagement
Apple	Products	50.00%
	Service and support	50.00%
Buyback Boss	Delivery	67.33%
	Service	32.67%
Cellunlocker	Service	61.75%
	Phone unlocking	38.25%
Republic Wireless	Product and service	55.38%
	Facilities	44.62%
Smarter Phone	Refund	53.19%
	Service	46.81%
Sprint	Customer service	44.02%
	Products	55.98%
Unlockboot	Unlocking phone	35.88%
	Service	64.12%

4.5.5. The degree of Customer Engagement in the discussion of certain Topics throughout the Branch as a whole: results in Appendix (7.3.2 Degree of engagement throughout the Branch)

4.6. Clustering of brands based on shared topics

4.6.1 Clustering was performed on two topics: Customer Service and Products. The former of the topics was shared by all of the brands, the latter by three of them : Apple, RepublicWireless, and Sprint.

4.6.2 Topic - Customer Service

CLUSPLOT(as.matrix(d))

Component 1
These two components explain 97.39 % of the point variability.
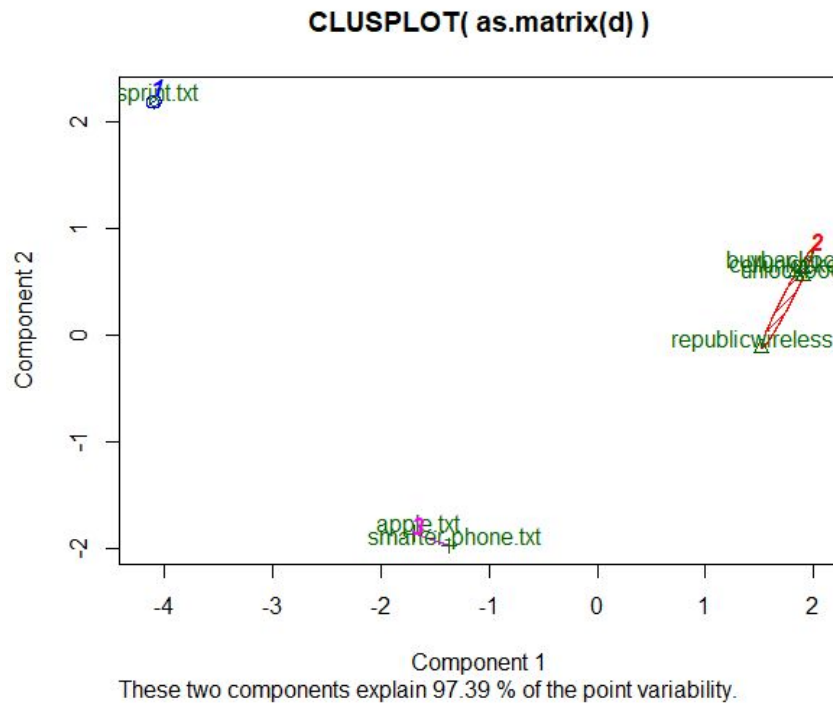
The result is representative of what the sentiment analysis told us. The customer service in Sprint was the worst of the brands. Apple and SmarterPhone had mixed reviews, but their customer service was close to on par/neutral. The rest of the brands were received well, so they're represented together.

4.6.2 Topic - Products

CLUSPLOT(as.matrix(d))

Component 1
These two components explain 100 % of the point variability.
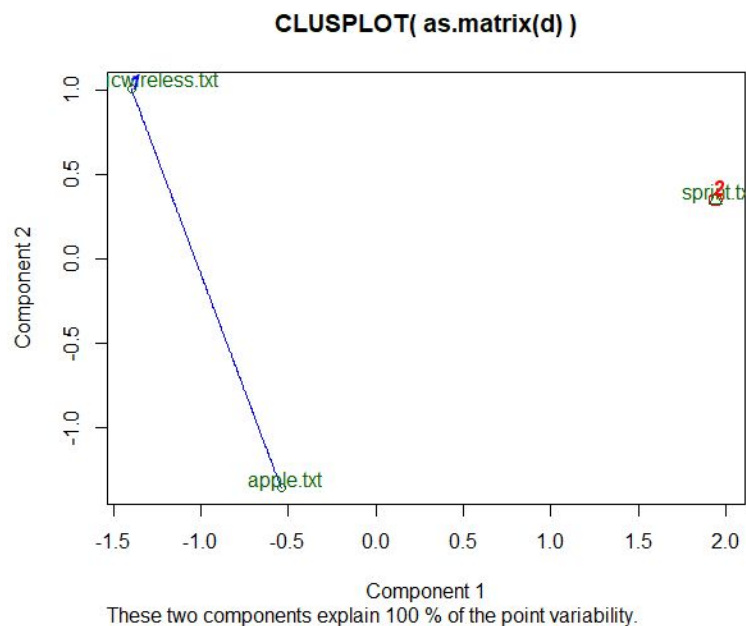
4.7. General clustering of brands based on degree of engagement and sentiment

CLUSPLOT(as.matrix(d))
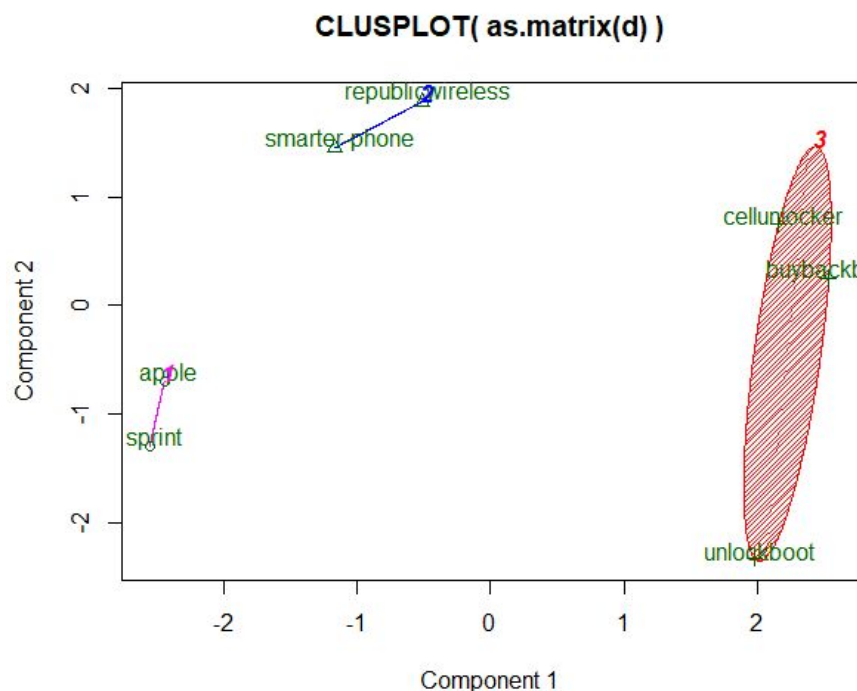
These two components explain 87.77 % of the point variability.

Here, the clustering was performed with regard to sentiment of each brand and also degree of engagement of main topics (from point 4.4.1) in each brand. The clusters formed closely refer to topics formulated with topic modelling. Cluster 1 - Customer service and product, cluster 2 and 3 - Product and delivery, Unlocking phone. Sentiment of the brands obviously also had an impact in clustering as, Apple and Sprint - cluster 1, were the most negative brands, while brands from cluster 3 were the most positive.

5. Conclusions
5.1. Branch
For customers in our branch the most important topics are product or service (unlocking phone or repairing phone), customer service and delivery. The most common problems are connected with customer service or quality of the product. Consumers often feel lack of support from company, loneliness in their problems. In the case of product customers complain about poor quality or damaged device.

5.2. Sprint
Sprint was the most negative brand of the whole branch. From the corpus analysis and topic clustering is seems that the company should work more on their customer service and on product servicing, which are both very lacking.

5.3. Apple and Smarter Phone

Some people were very enthusiastic about the products and services of these brands, however most of the negative comments disclosed that the customer service should be improved.

5.4. Others

Other brands had very high scores, which was clearly visible as they collected the highest compound sentiment score. There were some areas eligible for improvement, although the overall sentiment was highly positive.

6. References

https://medium.com/analytics-vidhya/simplifying-social-media-sentiment-analysis-using-vader-in-python-f9e6ec6fc52f

https://en.wikipedia.org/wiki/Text_mining

https://enauczanie.pg.edu.pl/moodle/course/view.php?id=2407

7. Appendix

7.1 Program codes

All program codes are supplied in the .zip file.

7.2 Additional graphs

7.2.1. Single terms graphs

● **Apple**
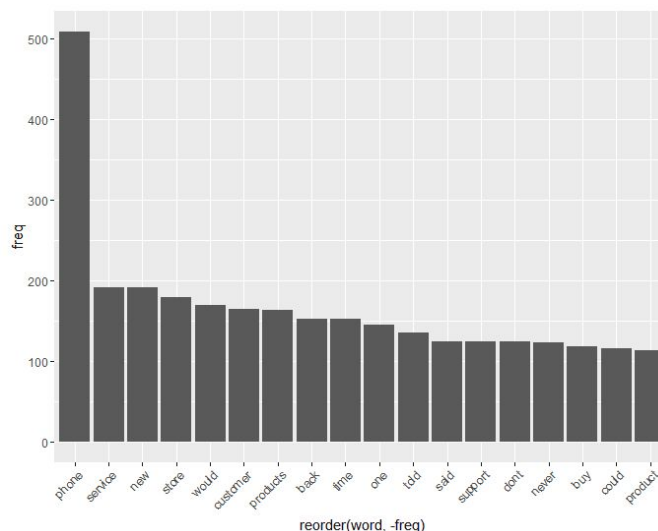
● **Buyback Boss**
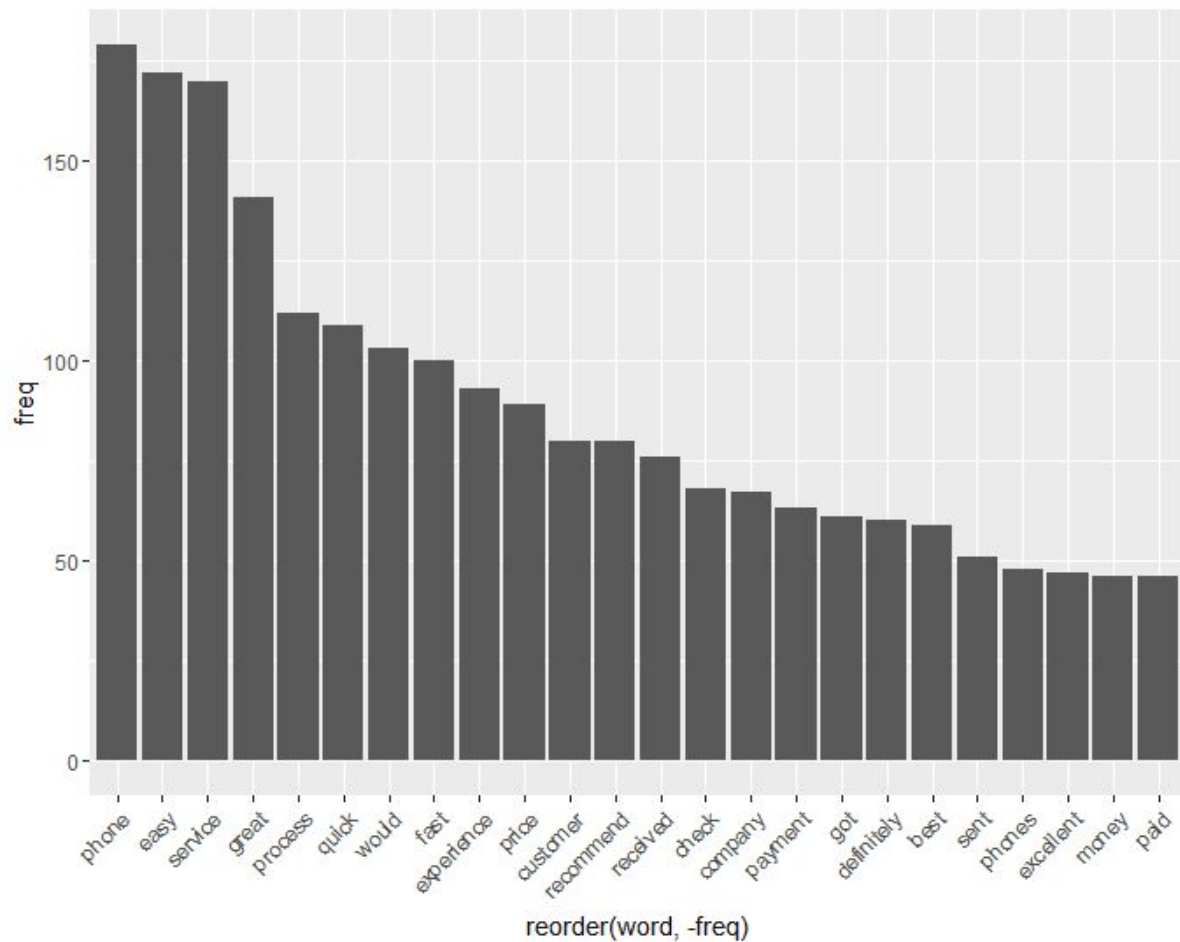
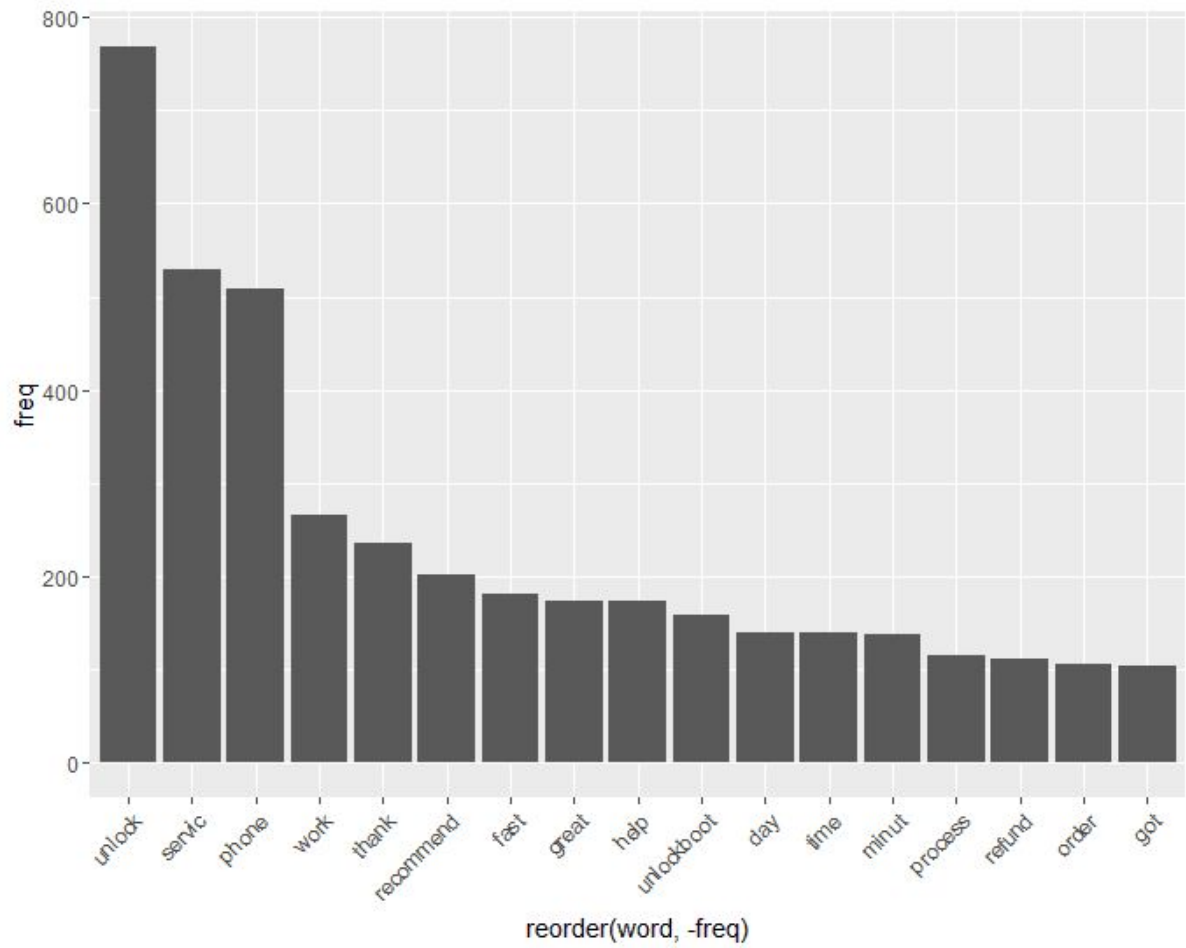● **Cellunlocker**
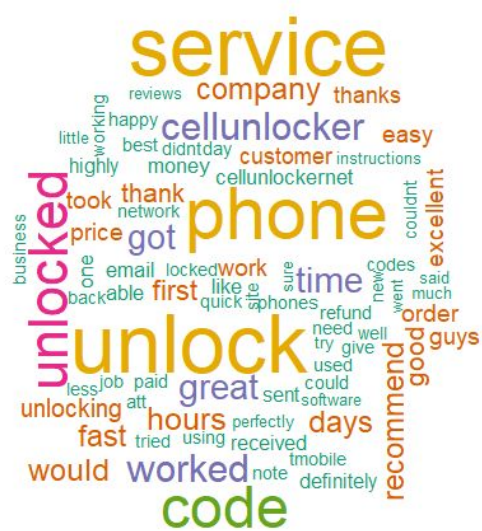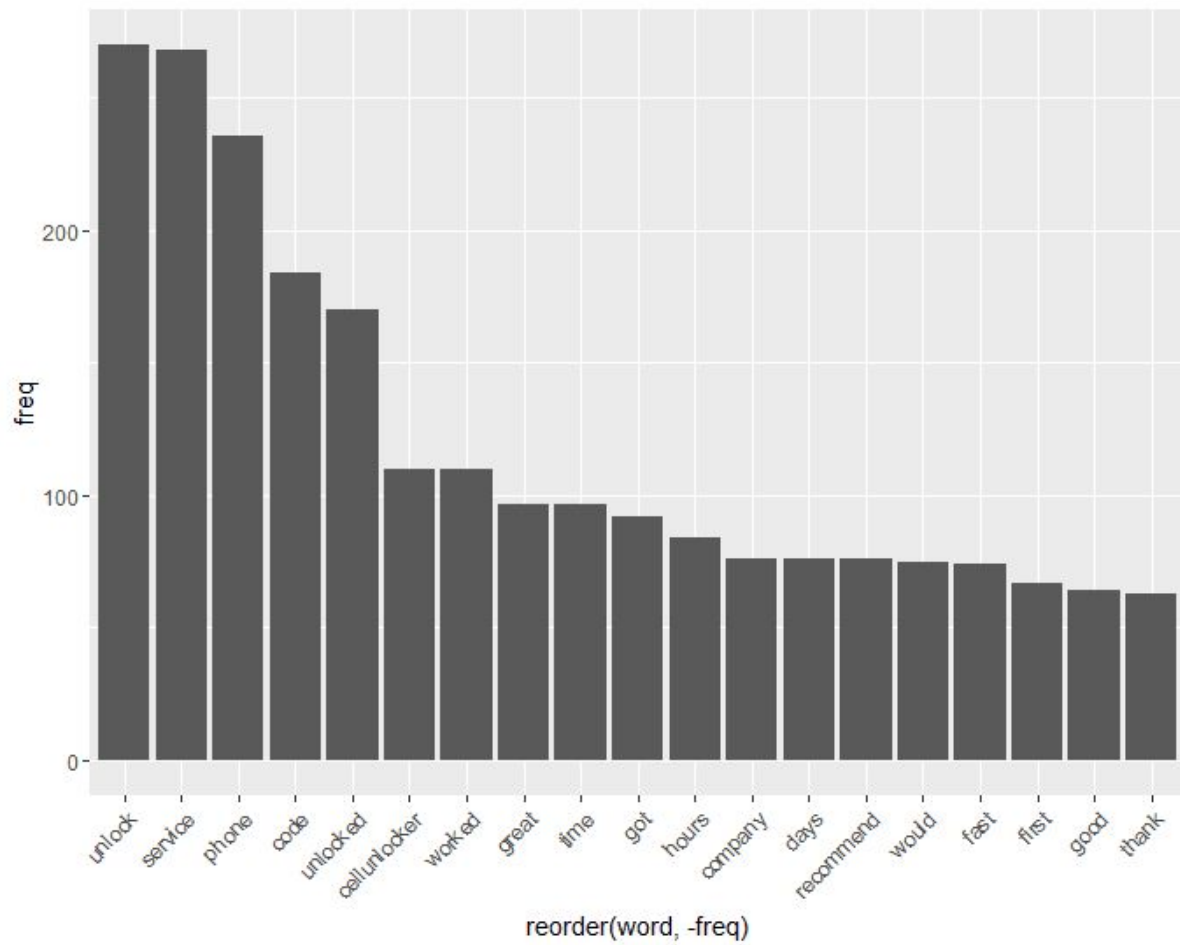
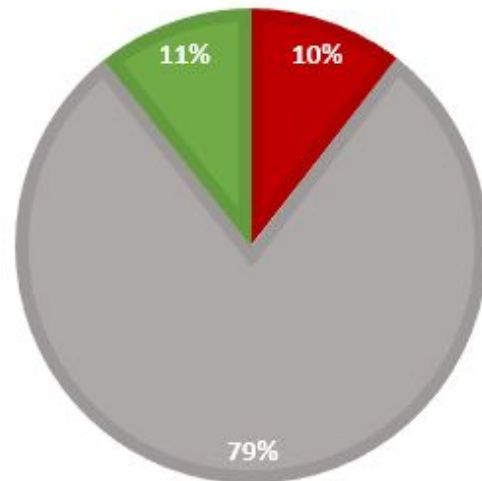## ● Republic Wireless
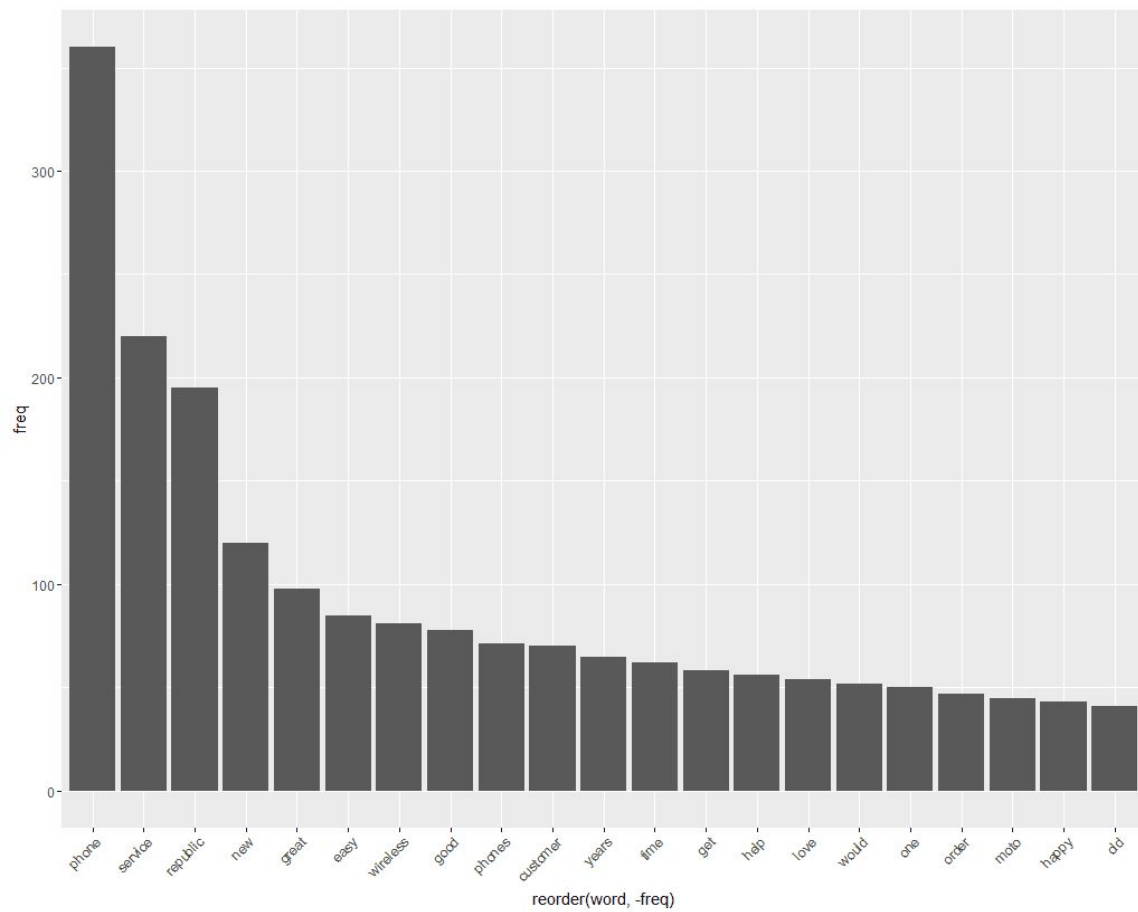
● **Smarter Phone**
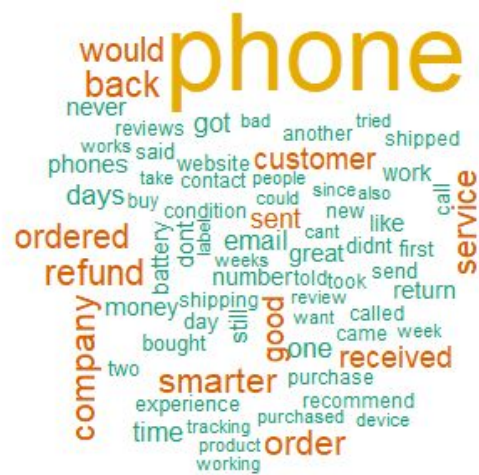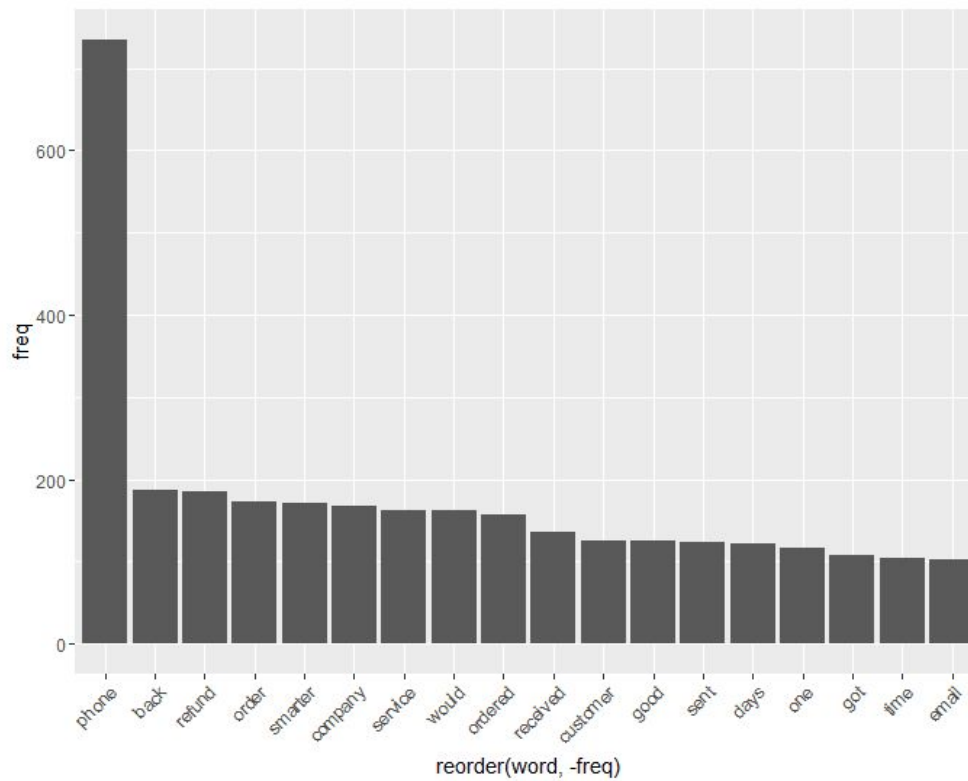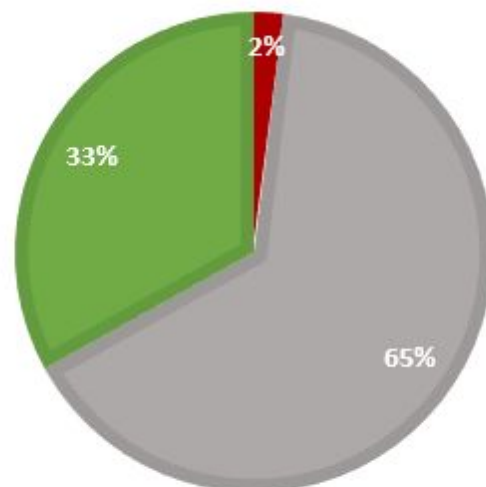
● **Sprint**
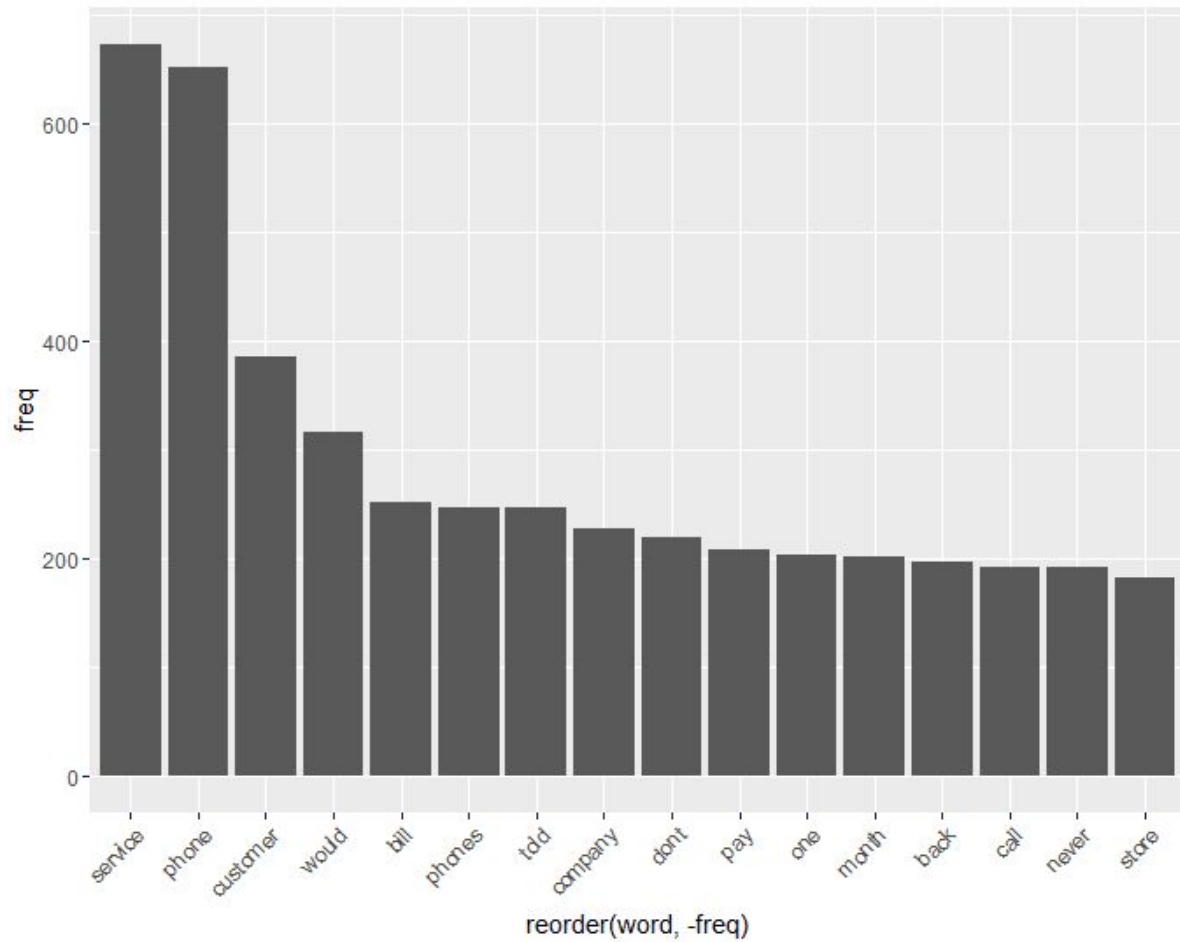
● **UnlockBoot**

7.2.2. General sentiment analysis of brands

Brand: **Apple**
Text Overall Rated As **Negative**
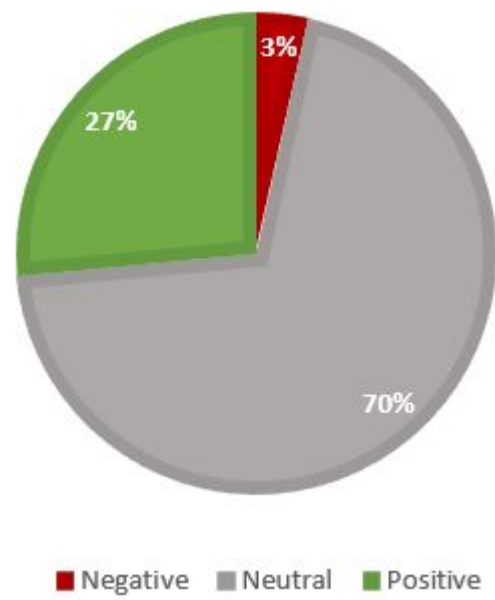Compound: -0,12

Negative 10% Neutral 79% Positive 11%

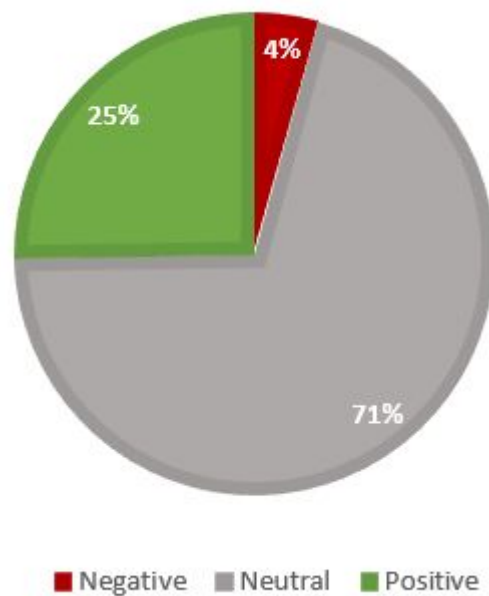Brand: **Buyback Boss**
Text Overall Rated As **Positive**
Compound: 0,66

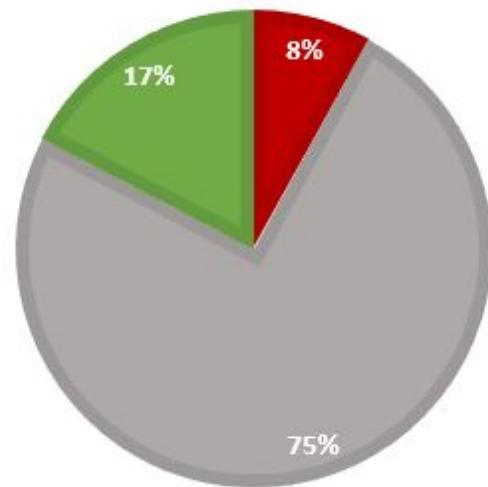Negative 2% Neutral 65% Positive 33%

Brand: **Cellunlocker**
Text Overall Rated As **Positive**
Compound: 0,56

3%
27%
70%

■ Negative   ■ Neutral   ■ Positive

Brand: **Republic Wireless**
Text Overall Rated As **Positive**
Compound: 0,46

4%
25%
71%

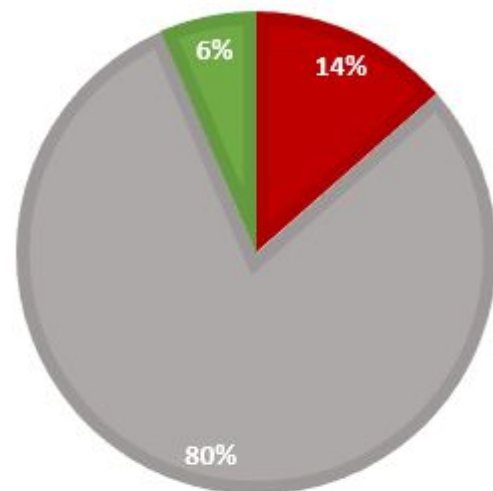■ Negative   ■ Neutral   ■ Positive

Brand: Smarter Phone
Text Overall Rated As **Positive**
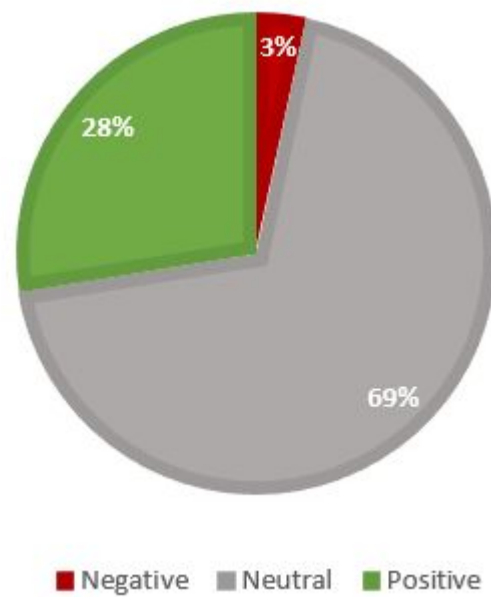Compound: 0,16

8%

17%

75%

■ Negative ■ Neutral ■ Positive

Brand: Sprint
Text Overall Rated As **Negative**
Compound: -0,40

6%

14%

80%

■ Negative ■ Neutral ■ Positive

Brand: **UnlockBoot**
Text Overall Rated As **Positive**
Compound: 0,49

■ Negative ■ Neutral ■ Positive

7.3 Additional tables

7.3.1 Sentiment analysis of branch as a whole

Topic	Neutral %	Negative %	Positive %	Overall, Compound
Product and delivery	70.3	11.8	17.8	0.37
Customer service and product	69.7	18.4	11.9	-0.27
Unlocking phone	66.5	5.1	28.4	0.49

7.3.2 Degree of engagement throughout the Branch

Topic	Degree of engagement
Product and delivery	54.44%
Customer service and product	22.78%
Unlocking phone	22.78%