Assignment 1: Text Mining Social Data

Dr Abubakar Siddique

A GRAFF 2231290

DS6501 Student: 2231290

Table of Contents

Analysis Report

As part of this Text Mining Social Data assignment, the first task was to import tweets into R studio. Data sets were available on Moodle and Petrol Companies was selected for my text mining and analysis.

Fuel Companies folder was created in my one drive, and all three csv data sets were extracted from Moodle (BP, Mobil and Z Energy) and simply pasted into this folder.

R Studio – we were able to perform dir_path<-"C:\\Users\\Student\\OneDrive - Whitireia and WelTec\\BIT 2023\\Documents\\Fuel Companies"

BP, Mobil, and Z Energy tweets csv files were easy to access and read as per our working directory in our R script.

BP Tweets

We start with BP tweets csv file and begin our work. Firstly, we see the first 6 rows of the csv file and obtain text only. We get a first sight indication of what we need to do to clean the text.

```
[1] "<U+0001F1F0><U+0001F1F7> Offshore platform construction <U+0001F3D7><U+FE0F>, automotive lubricants <U+0001F697>, LNG trading <U+0001F6F3> &amp; now #
OffshoreWind! <U+0001F32C><U+FE0F> \n\nWere expand https://t.co/vQYXVvUSUE"
[2] "@TottenhamFlyer Hi there, we're really sorry to hear this but thank you for getting in touch. Can you please send u https://t.co/z88MvaTjAV"
[3] "2022 full year results: We're accelerating investment in our transition and the energy transition while helping pro https://t.co/pzHTLfu8SL"
[4] "300+ new @bppulseuk charge points are live across London - making it even easier for people who want to plug in <U+0001F50C>, https://t.co/Whe6zeAkrd"
[5] "@Susieshoequeen Hi Susie, thanks for confirming this. It does sound like an error at the site so we will make sure https://t.co/Jx5FJfJexb"
[6] "After 9 months at sea, this high-tech Seawatch Wind LiDAR (light detection &amp; ranging) buoy is back on land. Its m https://t.co/bNFOSXRJdP"
```

There are <U+***>, URL's, Twitter handles etc many irrelevant information. We tidy these up using specifically chosen functions to preprocess these tweets. Below is our result.

```
[1] " Offshore platform construction, automotive lubricants, LNG trading  now "
[2] " Hi there, we're really sorry to hear this but thank you for getting in touch. Can you please send u "
[3] "2022 full year results: We're accelerating investment in our transition and the energy transition while helping pro "
[4] "300+ new  charge points are live across London - making it even easier for people who want to plug in, "
[5] " Hi Susie, thanks for confirming this. It does sound like an error at the site so we will make sure "
[6] "After 9 months at sea, this high-tech Seawatch Wind LiDAR (light detection  ranging) buoy is back on land. Its m "
```

Tweets are now ready for sentimental analysis so we begin by getting a score for each tweet based on different emotions

```
                                                                                                             bptweets.df2 anger anticipation disgust fear
1                                               Offshore platform construction, automotive lubricants, LNG trading  now     0            0       0    0
2                     Hi there, we're really sorry to hear this but thank you for getting in touch. Can you please send u     0            0       0    0
3 2022 full year results: We're accelerating investment in our transition and the energy transition while helping pro     0            0       0    0
4                  300+ new  charge points are live across London - making it even easier for people who want to plug in,     0            0       0    0
5                      Hi Susie, thanks for confirming this. It does sound like an error at the site so we will make sure     0            0       0    0
6    After 9 months at sea, this high-tech Seawatch Wind LiDAR (light detection  ranging) buoy is back on land. Its m     0            0       0    0
  joy sadness surprise trust negative positive
1   0       0        0     0        0        0
2   0       0        0     0        0        0
3   0       0        0     0        0        1
4   0       0        0     0        0        0
5   0       1        0     0        1        0
6   0       0        0     0        0        4
```

Highest scoring sentiment tweet with the most emotional content.

```
> most.emotionalcontent #most emotional content tweet
[1] "Todays the day! Our delegates and scholars are buzzing with excitement, and are ready to listen, learn collabo "
```

Words such as "scholars, buzzing, excitement, ready, collaborate" reflect the highest scoring emotional tweet

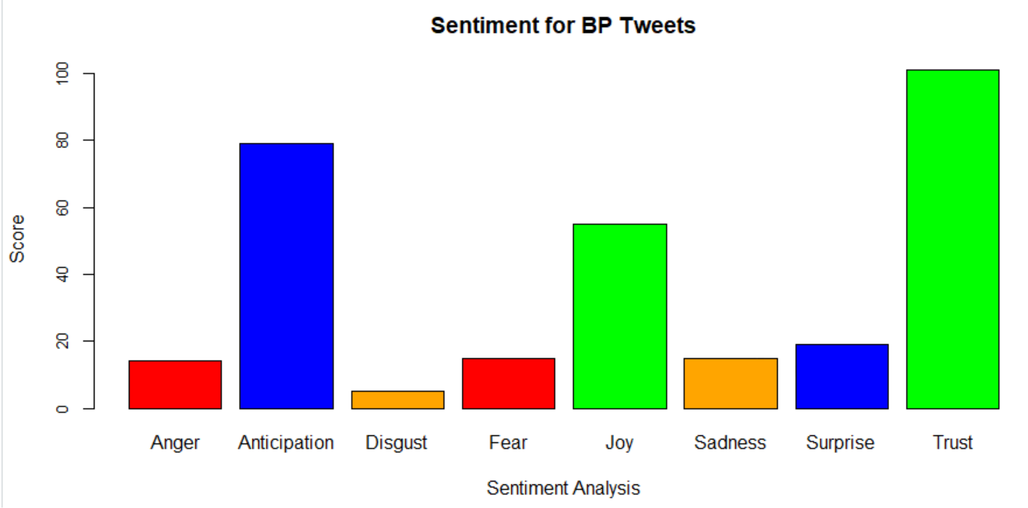
The most positive and negative tweets. Sentimental algorithm is quite mixed here as some rows may not be interpreted as positive. Positive tweets contain "thank you" and "please" indicate its scoring.

```
> head(positive.tweets) #lists first few rows of positive tweets
[1] " Hi there, we're really sorry to hear this but thank you for getting in touch. Can you please send u "
[2] "2022 full year results: We're accelerating investment in our transition and the energy transition while helping pro "
[3] "300+ new  charge points are live across London - making it even easier for people who want to plug in, "
[4] " Hi Susie, thanks for confirming this. It does sound like an error at the site so we will make sure "
[5] "After 9 months at sea, this high-tech Seawatch Wind LiDAR (light detection  ranging) buoy is back on land. Its m "
[6] "Today we're launching this year's Energy Outlook. Join Spencer Dale, our chief economist, at 2pm UK time. There's still time to register"
```

In the negative tweets outcome, we can see "we're really sorry" as the reason for this alignment.

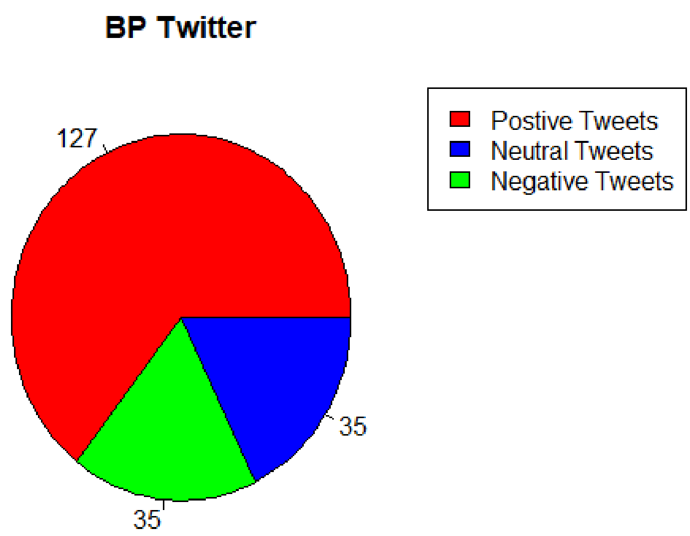
```
> head(negative.tweets) #lists first few rows of negative tweets
[1] " Hi Emily, we're very sorry for any inconvenience this cause you but we're pleased to say this issue h "
[2] " Hi there, we're really sorry to hear this. We have responded to your private message to try and help. "
[3] " Hi Stacey, we're really sorry to hear this. We have no intention of charging customers at our sites, and i "
[4] " Hi Pankaj, thank you for reaching out to us. Unfortunately, this app is a scam and is not owned or "
[5] " Hi there, unfortunately there is a scam app and it is not affiliated with bp. We recommend you do "
[6] " Hi Deepak, unfortunately this is a scam app and is not affiliated with bp. We recommend you do not g "
```

BP Visual Representation



Findings: An interesting sentiment analysis of the BP tweets. This bar graph indicates tweet sentiment scoring was highly favoured in the Trust emotion category, followed by Anticipation and Joy.

This pie chart gives us a clear overall view of BP tweets and how these were emotionally scored. 127 were Positive and 35 for both Neutral and Negative Tweets.



Generating a TDM, and filtering this to find words of 5-15 characters, selected stop words, removing numbers and converting to lower case. We were able to find the most commonly appearing words.

```
word_freqs[head(ordr)] #we can see first six most commonly occuring terms
      sorry         please         thanks          sites         energy unfortunately
         26             19             17             16             12            12
```

We can see “sorry” is the most commonly used term in the BP tweets, as appears a standard mannerisms per a social customer interactive platform

BP TDM and Word Cloud

We would then use three of the most commonly occurring terms and find other correlating words

```
> findAssocs(bptdm,"sorry",0.25) #from above results we use most freq words and find word associations with tweet
$sorry
    really experience      please       sites     receipt       aware
      0.49       0.41        0.28        0.27        0.26        0.26

> findAssocs(bptdm,"thanks",0.3) #correlation figures vary to find good range of associations
$thanks
     interest          owned        current opportunities independently          sound          susie
         0.41           0.37           0.34           0.34           0.34           0.33           0.33

> findAssocs(bptdm,"energy",0.3)
$energy
transition      years    outlook     spencer       chief  economist  launching
      0.60       0.50       0.49        0.49        0.40       0.40       0.40
```

Finally a word cloud is produced,

What we can analyse from the BP twitter word cloud is sorry, please and thanks as commonly used words in the tweets. It may reflect the trust dominance in the sentimental bar graph. Of the 197 tweets, 127 were positive and 35 were both neutral and negative which is a good outcome.

The word association with sorry were experience, receipt, aware so it may be a way of resolve for customer matters. Words associated with thanks and energy were interest and transition. The transition word association with energy. BP possibly moving towards a cleaner greener environment.

Mobil Tweets

As we have analysed the BP tweets we will go through and see what information we can gather from Mobil using different text mining techniques.

```
[1] "How to reduce industrial emissions? #Carboncapture and storage. #CCS could store up to 500 billions of tons of CO2.\x85 https://t.co/yCJCqTTi2G"
[2] "@SPGCIOil Looking forward to hear Dr. Vijay Swarup speaking at #CERAWeek!"
[3] "Across West Texas and Southeast New Mexico, we\x92re helping to build U.S. #energysupply in the Permian Basin: \n<U+0001F6E2><U+FE0F>560K\x85 https://
t.co/kPezLKxrAE"
[4] "Did you know helium isn\x92t just used for balloons? <U+0001F388>\n\nIt\x92s also a critical element in everything from space explorat\x85 https://t.c
o/FNOIlGNa0e"
[5] "@exxonmobil_qa Exciting things happening over at the @exxonmobil_qa booth at Doha Festival City and Villaggio!"
[6] "We\x92re committed to meeting the goals of the Paris Agreement using various approaches. Learn more about how we\x92re ad\x85 https://t.co/vo8AlcvTvm"
```

From the text only data frame from the Mobil tweets csv file, we can see a lot of unwanted characters etc that will not be useful in our text analysis – so we clean this up using few functions. Hashtags, twitter handles, punctuation, control codes will all be removed.

```
[1] "How to reduce industrial emissions Carboncapture and storage CCS could store up to 500 billions of tons of CO2 "
[2] "SPGCIOil Looking forward to hear Dr Vijay Swarup speaking at CERAWeek"
[3] "Across West Texas and Southeast New Mexico were helping to build US energysupply in the Permian Basin560K "
[4] "Did you know helium isnt just used for balloonsIts also a critical element in everything from space explorat "
[5] "exxonmobilqa Exciting things happening over at the exxonmobilqa booth at Doha Festival City and Villaggio"
[6] "Were committed to meeting the goals of the Paris Agreement using various approaches Learn more about how were ad "
```

We can now see how much cleaner the text is, so we can perform some text mining techniques to see what data we can get from sentimental analysis.

```
                                                                                                   mobiltweets.df2 anger anticipation disgust fear
1   How to reduce industrial emissions Carboncapture and storage CCS could store up to 500 billions of tons of CO2     0            1       0    0
2                                                 SPGCIOil Looking forward to hear Dr Vijay Swarup speaking at CERAWeek     0            0       0    0
3        Across West Texas and Southeast New Mexico were helping to build US energysupply in the Permian Basin560K     0            0       0    0
4     Did you know helium isnt just used for balloonsIts also a critical element in everything from space explorat     0            0       0    0
5          exxonmobilqa Exciting things happening over at the exxonmobilqa booth at Doha Festival City and Villaggio     0            2       0    0
6 Were committed to meeting the goals of the Paris Agreement using various approaches Learn more about how were ad     0            0       0    0
  joy sadness surprise trust negative positive
1   0       0        0     0        0        1
2   0       0        0     0        0        1
3   0       0        0     0        0        1
4   0       0        0     0        0        0
5   2       0        2     0        0        2
6   0       0        0     2        0        3
```

Tweets have been scored across different emotional categories. Scoring such as 1 indicates low or high scoring of 3 etc is a strong resemblance to the emotion.

	mobiltweets.df2	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
1	How to reduce industrial emissions Carboncapture and stor...	0	1	0	0	0	0	0	0	0	1
2	SPGCIOil Looking forward to hear Dr Vijay Swarup speaking ...	0	0	0	0	0	0	0	0	0	1
3	Across West Texas and Southeast New Mexico were helping ...	0	0	0	0	0	0	0	0	0	1
4	Did you know helium isnt just used for balloonsIts also a crit...	0	0	0	0	0	0	0	0	0	0
5	exxonmobilqa Exciting things happening over at the exxon...	0	2	0	0	2	0	2	0	0	2
6	Were committed to meeting the goals of the Paris Agreeme...	0	0	0	0	0	0	0	2	0	3

Clicking on the environment pane, and selecting the mobiletweets.df2 data frame we can see a clearer view of the scoring table and how each tweet measures

The highest scoring tweet for each emotional category is listed below

	mobiltweets.df2	anger
639	We are deeply saddened by the loss of life injury and damage caused by the earthquake in the Papua New Guinea High	4

(Highest scoring for anger)

	mobiltweets.df2	anger	anticipation
661	Celebrating one year since the start of our partnership with redbullracing Glad to be fueling DanielRicciardo an	0	3

(Highest scoring for anticipation)

(Highest scoring for disgust and fear)

	mobiltweets.df2	anger	anticipation	disgust	fear
274	While death and infection rates have drastically declined in the last two decades malaria continues to impact mill	1	2	2	3

	mobiltweets.df2	anger	anticipation	disgust	fear	joy
493	Were celebrating our Black Employee Success Team BEST for BlackHistoryMonth and their work to advance black l	0	3	0	1	3

(Highest scoring for joy)

	mobiltweets.df2	anger	anticipation	disgust	fear	joy	sadness
639	We are deeply saddened by the loss of life injury and damage caused by the earthquake in the Papua New Guinea High	4	0	1	3	0	4

(Highest scoring for sadness)

	mobiltweets.df2	anger	anticipation	disgust	fear	joy	sadness	surprise
180	Thrilled to have AirLiquideUSA BASFCorporation and Shell join the growing efforts to advance carboncapture and sto	1	1	0	2	1	1	2

(Highest scoring for surprise)

mobiltweets.df2	anger	anticipation	disgust	fear	joy	sadness	surprise	trust
As Mozambique recovers after Cyclone Idai our employees are on the ground providing food shelter and comfort to p	0	2	0	1	3	0	1	5

(Highest scoring for trust)

```
> most.emotionalcontent
[1] "We are making a 1 million aid donation to support the relief effort in PNG plus helping communities recover from "
```

The above tweet was the highest scoring sentiment tweet with the most emotional content. We can see "1 million aid donation", "support", "relief", "effort", "helping communities recover" are all great attributes to the selection of this tweet.

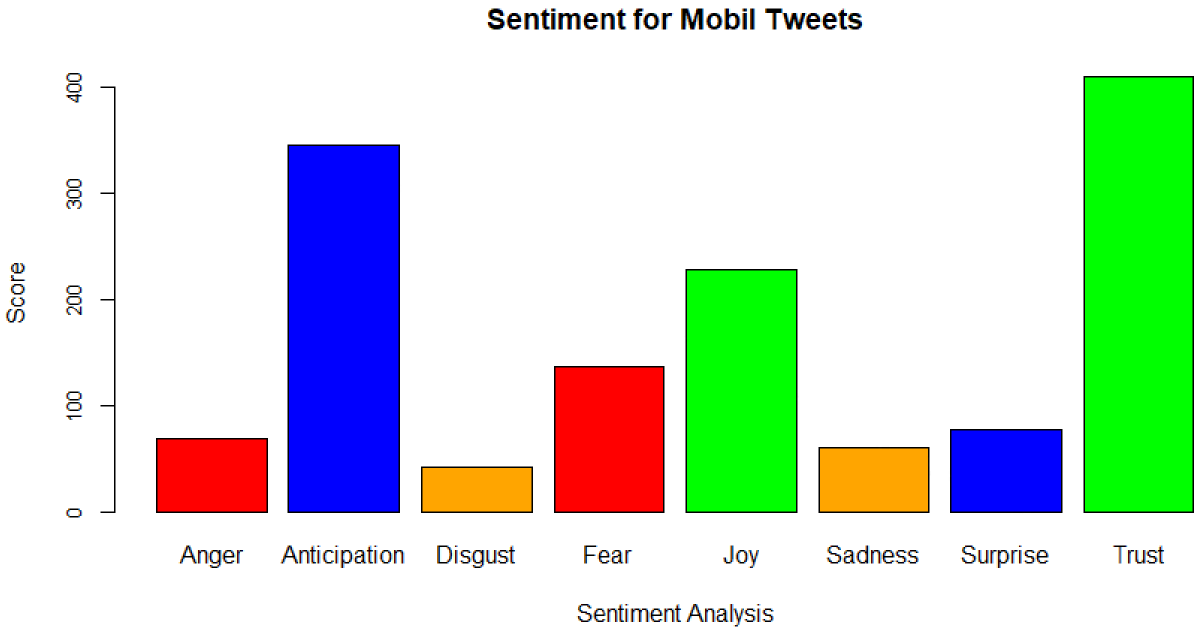
Below, we have the most positive and negative tweets

```
> head(positive.tweets)
[1] "SPGCIOil Looking forward to hear Dr Vijay Swarup speaking at CERAWeek"
[2] "Across West Texas and Southeast New Mexico were helping to build US energysupply in the Permian Basin560K "
[3] "exxonmobilqa Exciting things happening over at the exxonmobilqa booth at Doha Festival City and Villaggio"
[4] "Were committed to meeting the goals of the Paris Agreement using various approaches Learn more about how were ad "
[5] "UAGradSchool Huge congratulations to Dr Ifeanyi Okpala on this notable achievement Were so glad you are a part "
[6] "ExxonMobilNG NBAAfrica PanAfricareNg Great work on the PowerForward Project"
```

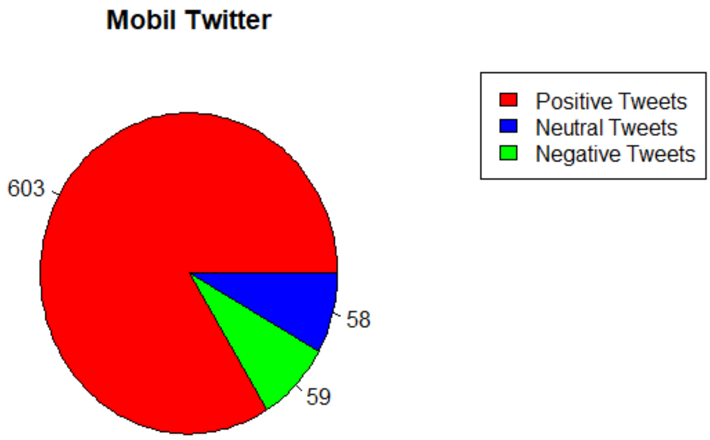
```
> head(negative.tweets)
[1] "Did you know helium isnt just used for balloonsIts also a critical element in everything from space explorat "
[2] "SquawkCNBC WhiteHouse POTUS Here are some of the facts Our employees worked hard to meet demand Upstream p "
[3] " At least 2700 feet to 4000 feet underground Far enough away from fault lines and earthquakeprone areasDr "
[4] "Were using carbon capture and storage in WyomingDid you know weve captured more CO2 here to date than any othe "
[5] "Largescale advancedrecycling unit started can now process gt80M lbs of plastic waste per year "
[6] "Our Strathcona facility will produce renewable diesel with carbon capture and storage technology which could help "
```

The algorithm would have used scoring sentiment and identifying key words to compile scoring of each tweet. Positive tweets we can see good words such as "helping, great work and congratulations" that reflect these. Negative tweets do appear neutral and not necessary negative. It may have picked up words such as "fault, waste, critical" for this outcome.

Mobil Visual Representation



Findings: Interesting to see Trust and Anticipation as both dominant pillars similar to what we saw for the BP tweets. Third is Joy and followed closely by Fear which we could explore more. Given the scoring, it is clear we are dealing with a higher volume of tweets compared to BP.



Another quite evenly balanced neutral and negative tweets, although one edges slighter more than the other. Many positive tweets. The proportion of the pie chart we can easily see this.

Mobil TDM and Word Cloud

After seeing the sentiment bar graph and mobile twitter tweets, we then look at some specific text/tweets. We create a TDM, and filtering this to find words of 5-15 characters, selected stop words, removing numbers and converting to lower case. We were able to find the most commonly appearing words.

```
> word_freqs[head(ordr)]
    energy   emissions       today  exxonmobil      carbon  technology 
       115          63          44          40          37          34 
```

We can see "energy" as the most common word for Mobil and "emissions" both to do with fuel etc as per fuel company.

Using the most frequent words we find associating/correlating words.

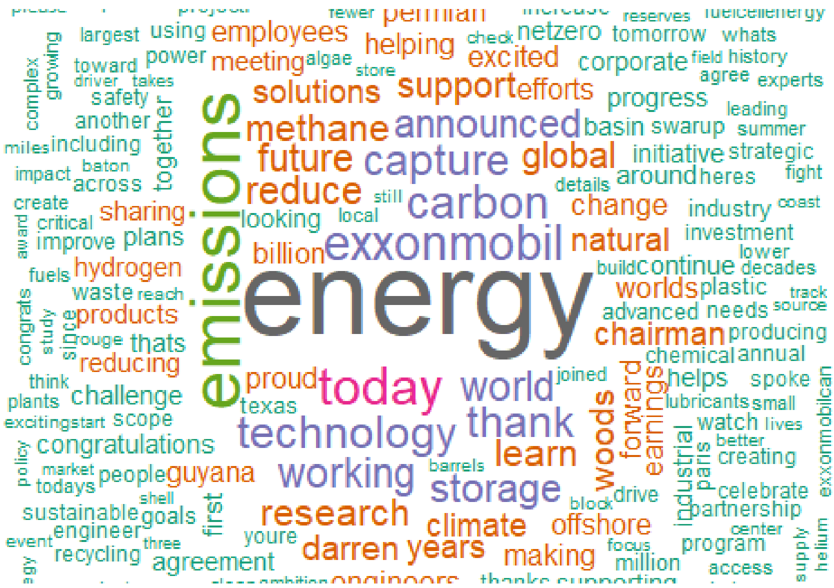
```
> findAssocs(mobiltdm,"energy",0.2)
$energy
reliable security  working 
    0.25     0.25     0.21 

> findAssocs(mobiltdm,"emissions",0.3)
$emissions
  reduce reducing  methane 
    0.45     0.37     0.33 

> findAssocs(mobiltdm,"technology",0.3)
$technology
  formula     racing     factor  radically      shows      truly 
     0.32       0.32       0.32       0.32       0.32       0.32 
```

It seems energy was associated with words like reliable, security and working which may be a marketing term for the Mobil company. Emissions reducing is a global goal that is nice to see the correlation in tweets. Technology associated with formula, racing, factor are quire distinct but are our results.

Finally, our second word cloud is produced,



As per our findings before, energy is the most dominate, we can see carbon, today, emissions, and many several words which at first glance you can see how this would be a text mining from a energy related company.

Z Energy Tweets

We have analysed the previous two fuel companies tweets and grasp some interesting findings. Now it is Z Energy to see what information can be gathered.

```
> head(zenergytweets.df$text)
[1] "@lizmarienes Kia ora Elizabeth-Marie, please go to More &gt; My account &gt; Personal details &gt; Email address &gt; Edit emai\x85 https://t.co/06T5k
HqRu4"
[2] "@nzbirdlaw Kia ora @nzbirdlaw the Marsden Point Refinery is owned by BP, Mobil and Z Energy (since the 1980s) so th\x85 https://t.co/JSVDvKGcBh"
[3] "@YahDxd Kia ora Fatilau, We have no plans to bring back Blokhedz right now. After the Marvel and DC heroes what \"he\x85 https://t.co/cWmCwey4Cc"
[4] "@CarlElphick Kia ora Carl, no sorry, you can't earn Airpoints Dollars using that method. New World Clubcards that a\x85 https://t.co/pgdrLFMrDj"
[5] "@ZureenAli4 Kia ora Zureen, sorry to hear about your experience. Can you please send us a DM letting us know the ti\x85 https://t.co/XBxrjpXTOc"
[6] "@perfectclarity Yep, NZ, New Zealand, Aotearoa - all ways of naming our favourite country^D"
```

A lot of noise in this text, we will clear this up and below is our outcome

```
> head(zenergytweets.df2)
[1] " Kia ora ElizabethMarie please go to More  My account  Personal details  Email address  Edit emai "
[2] " Kia ora  the Marsden Point Refinery is owned by BP Mobil and Z Energy since the 1980s so th "
[3] " Kia ora Fatilau We have no plans to bring back Blokhedz right now After the Marvel and DC heroes what he
[4] " Kia ora Carl no sorry you cant earn Airpoints Dollars using that method New World Clubcards that a "
[5] " Kia ora Zureen sorry to hear about your experience Can you please send us a DM letting us know the ti "
[6] " Yep NZ New Zealand Aotearoa  all ways of naming our favourite countryD"
```

We can now perform sentimental analysis and see how each tweet is scored

```
                                                                                        zenergytweets.df2 anger anticipation disgust fear joy
1          Kia ora ElizabethMarie please go to More  My account  Personal details  Email address  Edit emai     0            0       0    0   0
2                Kia ora  the Marsden Point Refinery is owned by BP Mobil and Z Energy since the 1980s so th     0            0       0    0   0
3 Kia ora Fatilau We have no plans to bring back Blokhedz right now After the Marvel and DC heroes what he     0            0       0    0   0
4         Kia ora Carl no sorry you cant earn Airpoints Dollars using that method New World Clubcards that a     0            0       0    0   0
5   Kia ora Zureen sorry to hear about your experience Can you please send us a DM letting us know the ti     0            0       0    0   0
6                                      Yep NZ New Zealand Aotearoa  all ways of naming our favourite countryD     0            0       0    0   0
  sadness surprise trust negative positive
1       0        0     2        0        0
2       0        0     0        0        0
3       0        1     0        0        1
4       0        0     0        0        1
5       0        0     0        0        0
6       0        0     0        0        0
```

It isn't very clear, so we access the environmental pane and click on zenergytweets.df2 to see a cleaner view

	zenergytweets.df2	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
1	Kia ora ElizabethMarie please go to More My account Personal details Email address Edit emai	0	0	0	0	0	0	0	2	0	0
2	Kia ora the Marsden Point Refinery is owned by BP Mobil and Z Energy since the 1980s so th	0	0	0	0	0	0	0	0	0	0
3	Kia ora Fatilau We have no plans to bring back Blokhedz right now After the Marvel and DC heroes what he	0	0	0	0	0	0	1	0	0	1
4	Kia ora Carl no sorry you cant earn Airpoints Dollars using that method New World Clubcards that a	0	0	0	0	0	0	0	0	0	1
5	Kia ora Zureen sorry to hear about your experience Can you please send us a DM letting us know the ti	0	0	0	0	0	0	0	0	0	0
6	Yep NZ New Zealand Aotearoa all ways of naming our favourite countryD	0	0	0	0	0	0	0	0	0	0

Below will list all the highest emotion category for Z Energy tweets with words relating to emotion

Anger ("mean")

	zenergytweets.df2	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
117	We dont mean to taunt During Level 4 we cooked less hot food than we usually would given there were fe	2	0	0	1	1	1	0	2	1	2

Anticipation ("reckon")

	zenergytweets.df2	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
73	Were proud to support Mike Kings I Am Hope charity and we reckon their plan for free counselling to any young perso	0	4	0	0	4	0	2	2	0	4

Disgust ("sorry")

	zenergytweets.df2	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
43	Hey Mark sorry for the delay in getting our new Good in the Hood content up Check it out here for 2019	2	1	2	2	2	1	1	2	2	2

Fear ("sorry/delay")

	zenergytweets.df2	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
343	Hey Mark sorry for the delay in getting our new Good in the Hood content up Check it out here for 2019	2	1	2	2	2	1	1	2	2	2

Joy (“proud”)

	zenergytweets.df2	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
273	Were proud to support Mike Kings I Am Hope charity and we reckon their plan for free counselling to any young perso	0	4	0	0	4	0	2	2	0	4

Sadness (“problem”)

	zenergytweets.df2	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
269	Kia Ora KiwiAnon20 right now we have the opposite problem a surplus not a shortage Whereas we w	1	0	0	2	0	2	0	0	2	0

Surprise (“good/vote”)

	zenergytweets.df2	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
165	Come into Z and vote for good today	1	2	0	0	2	1	2	2	1	2

Trust (“official/conformation”)

	zenergytweets.df2	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
277	Weve had official confirmation that while were able to sell our premade hot food during Level 4 coffee and other	1	0	0	0	1	0	0	4	0	2

Now we find the tweet with the highest scoring sentiment having the most emotional content

```
> most.sentimental
[1] "Happy Diwali to everyone here in New Zealand celebrating We wish you a joyful year "
```

Its clear to say this would score as the highest scoring sentiment, it has very positive words such as “Happy, celebrating, wish and joyful”. Many expressions in this tweet.

```
> head(positive.tweets)
[1] " Kia ora ElizabethMarie please go to More  My account  Personal details  Email address  Edit emai "
[2] " Kia ora Fatilau We have no plans to bring back Blokhedz right now After the Marvel and DC heroes what he "
[3] " Kia ora Carl no sorry you cant earn Airpoints Dollars using that method New World Clubcards that a "
[4] " Kia ora Zureen sorry to hear about your experience Can you please send us a DM letting us know the ti "
[5] " Yep NZ New Zealand Aotearoa  all ways of naming our favourite countryD"
[6] " Hi there While the ownership has changed who we are as a company has not We remain firmly committed "
```

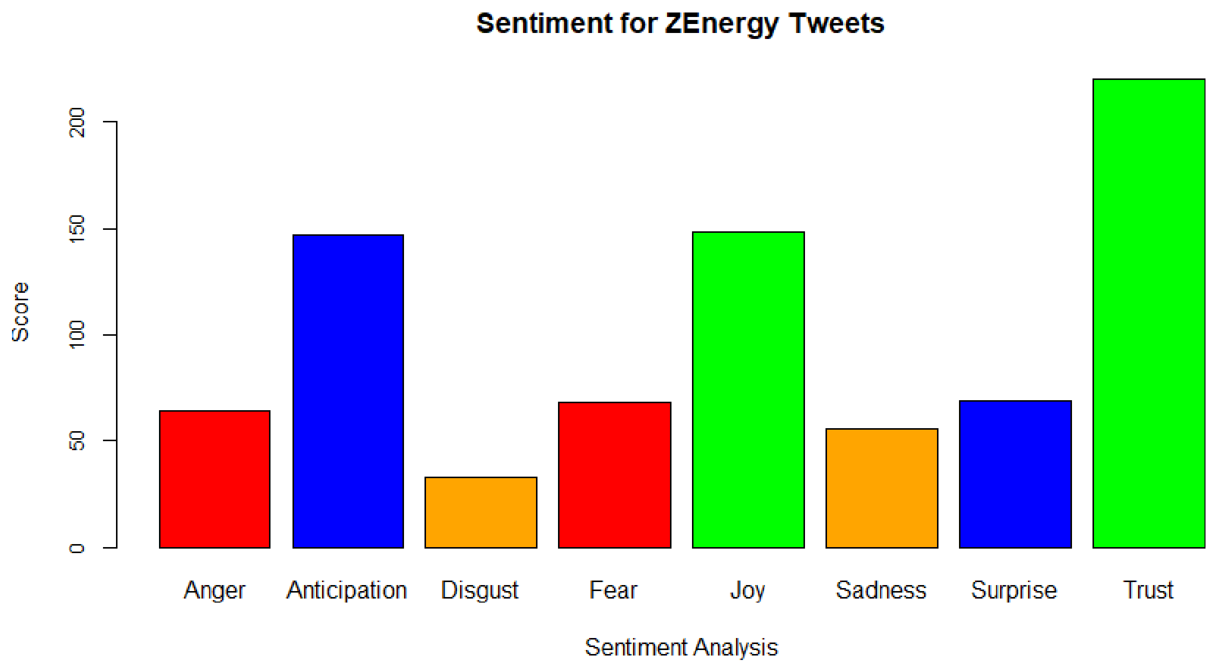
Z Energy positive tweets, although it appears the algorithm appears to have few a misunderstandings in the interpretation of the tweets as a few tweets indicate not so positive situation.

For example, rows 2, 3, 4; Rows 2 has the word “heroes” which may overwhelm the “no plans”. Rows 3 “No sorry you can’t” may have been overlooked by “Airpoints dollars” and “new world club cards”. Finally Row 4, “experience, please” may have been seen as positive – overlooking the word “sorry”.

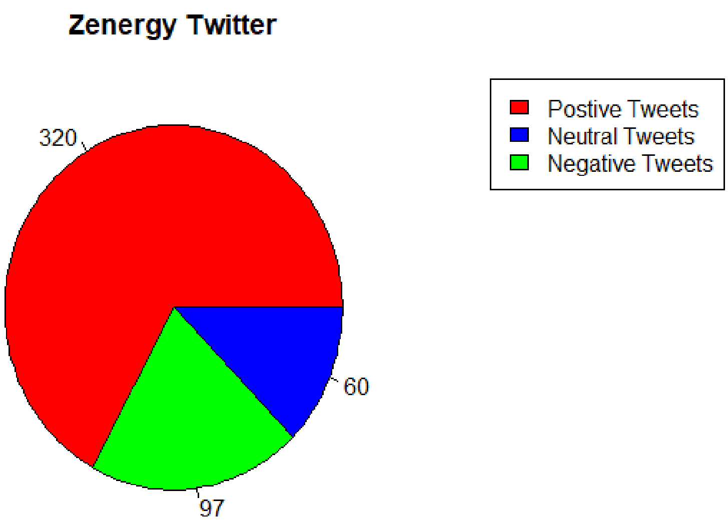
```
> head(negative.tweets)
[1] " Kia ora Z Dunedin truck stop has temporarily closed due to maintenance until midDecember We apologi "
[2] " Weve put measures in place such as warning signage contactless payment methods to reduce levels of c "
[3] " Were really sorry to hear that we dont want anyone feeling pressured to come in store Thats wh "
[4] " Unfortunately only a small number of our customers use the feature However offsettings only one par "
[5] " Kia Ora We dont deny that we have an impact on the environment its something we are working har "
[6] " Hi Claire Sorry to hear the app was crashing for you Ive asked our app team to look into this Are "
```

Most negative tweets outcome with some tweets could be neutral. Such as row 2, “putting measures in place to reduce levels”.

Z Energy Visual Representation



Findings: Compared to BP and Mobile Sentiment Bar Graphs, here we have seen Joy equal to anticipation as the second highest scoring tweet emotion.



Z Energy Twitter Pie Chart

A slightly higher margin of negative tweets in comparison to previous two fuel companies. Positive tweets are dominate here.

Z Energy TDM and Word Cloud

We now create a TDM for word cloud processing and first see a list for most frequently commonly appearing words in the Z Energy Tweets

```
> word_freqs[head(ordr)]
 thanks   sorry  please   great   right message 
     58      43      28      24      19      18 
```

This output is similar to BP the first twitter we perform text mining. We can see the most frequent word "thanks, "sorry", "please" which were all the top three in the BP most frequent words.

Using the most frequent words, we will find the associating words correlating to our findings.

```
> findAssocs(zenergytdm,"thanks",0.25)
$thanks
 letting  sharing feedback bringing    looks 
    0.36     0.30     0.28     0.27     0.26 

> findAssocs(zenergytdm,"sorry",0.25)
$sorry
    direct experience     issues      error 
      0.30       0.29       0.25       0.25 

> findAssocs(zenergytdm,"message",0.25)
$message
private  direct  please  detail   error 
   0.52    0.49    0.32    0.26    0.26 
```

The three outputs here, are all to what seem like generic replies to tweets on a customer interaction basis. We have thanks associated with "feedback, sharing, letting", sorry associated with "experience, issues, error" and message "private, direct, error". These all coincide into a direct communicative response.

Lastly, the word cloud

A relatively easy, conversational style word cloud. We can see words like "thanks, sorry, please" which are all of generic terms.

Some interesting findings such as emissions, prices and company.

Opposing competition companies are seen "caltext"

Conclusion

Key insights gained from analysing the BP, Mobile and Z Energy Twitter accounts these were my major findings.

BP and Z Energy had similar results especially with the word cloud as the three commonly frequent words between these twitter accounts were “please, sorry and thanks”. I was able to grasp a conceptual understanding these were local NZ businesses, using the twitter social media tool to communicate with its customers as main audience.

The correlation of frequently used words was sentence based than topic based, given perhaps generic replies by a social media team using these two companies. These were along the lines of “thank you for your reply etc”.

Mobil tweets had a lot more depth in the way of what the frequently used words were. It was not the generic replies you’d get in an interactive basis, but more so broadcasting information to wider audience of all.

Mobil had a higher positive tweet ratio in the pie chart compared to BP and Z Energy but it also had a higher fear sentimental score indicating the news style could be received differently.

We can see this in the initial output of tweets where the tweets of Mobil were news based, and BP and Z Energy were customer-based tweets.

R Script

```
#ASSIGNMENT

#loading required libraries
library(SnowballC) #helps with text stemming getting words to word stem
library(tm) #text mining package
library(syuzhet) #generates a score for each of the tweets for 10 diff emotions (sentiment
processing)
library(wordcloud) #creates a world cloud of the commonly occuring words
dir_path<-"C:\\Users\\Student\\OneDrive - Whitireia and WelTec\\BIT 2023\\Documents\\Fuel
Companies"
setwd(dir_path) #setting the above as our working directory - where all csv files are located
getwd() #get working directory, confirming the base of which we operate

#BP
TWEETS=================================================================
=

bptweets.df <- read.csv("bptweets.csv") #reading the bptweets file saved in Fuel Companies
folder and converting into a new data frame/variable
head(bptweets.df) #head function listing the first few rows of our new data frame
head(bptweets.df$text) #we display the text field only

bptweets.df2 <- bptweets.df$text #converting the text only field to a new data frame

head(bptweets.df2) #we see the new first few rows of text only data frame

#Converting text to readable string text to perform gsub operations
rmv_ivalid_char <- function(x){ iconv(x, "latin1", "ASCII", sub="")} #new functions to convert
latin1 to acsii
rmv_ivalid_char2 <- function(x){ iconv(x, "utf-8", "ASCII", sub="")} #new functions to convert utf-8
to acsii

bptweets.df2 <- rmv_ivalid_char(bptweets.df2) #new remove invalid char function performed on
tweets
bptweets.df2 <- rmv_ivalid_char2(bptweets.df2) #repeated cycle of remove invalid char2
function on tweets

head(bptweets.df2) #displays new converted text in order to now perform gsub operation

#We remove URLs, hashtags, control codes, punctuation
#gsub meaning global substitution to find and replace so we can clean up tweets for text mining
and sentiment analysis
bptweets.df2 <- gsub("\\s*<U\\+\\w+>","",bptweets.df2) #removing unwanted patterns <U+...>
bptweets.df2 <- gsub("http.*","",bptweets.df2) #removing http, URLs
bptweets.df2 <- gsub("https.*","",bptweets.df2) #removing https
bptweets.df2 <- gsub("[\\]+","",bptweets.df2) #removing \\ char
bptweets.df2 <- gsub("#.*","",bptweets.df2) #removing hashtags
bptweets.df2 <- gsub("*","",bptweets.df2) #removing asterisks
bptweets.df2 <- gsub("&amp","",bptweets.df2) #removing &amp
```

```
bptweets.df2 <- gsub(";","",bptweets.df2) #removing semi colon
bptweets.df2 <- gsub("@\\w+","",bptweets.df2) #removing twitter handles

head(bptweets.df2) #shows newly cleaned up bptweets.df2 for sentimental analysis

#BP SENTIMENTAL
ANALYSIS=======================================================

bpword.df <- as.vector(bptweets.df2) #to get sentiment we must convert dataframe to vector
bpemotion.df <- get_nrc_sentiment(bpword.df) #now we're able to get scoring using nrc
function
bpemotion.df2 <- cbind(bptweets.df2, bpemotion.df) #joining tweets (bptweets.df and
bpemotion.df) to scoring using cbind function
head(bpemotion.df2) #listing the first few scoring of tweets of the new bpemotion.df2

sent.value <- get_sentiment(bpword.df) #extract sentiment scoring for each tweet

most.emotionalcontent <- bpword.df[sent.value == max(sent.value)] #most positive sentiment
(max = highest scoring)
most.emotionalcontent #most emotional content tweet

least.emotionalcontent <- bpword.df[sent.value <= min(sent.value)] #most negative sentiment
(min = lowest scoring)
least.emotionalcontent #least emotional content tweet

positive.tweets <- bpword.df[sent.value > 0] #positive tweets given a value of more than 0
negative.tweets <- bpword.df[sent.value < 0] #negative tweets given a value of less than 0
neutral.tweets <- bpword.df[sent.value == 0] #neutral tweets given a value of equal than 0

head(positive.tweets) #lists first few rows of positive tweets
head(negative.tweets) #lists first few rows of negative tweets
head(neutral.tweets) #list of neutral tweets

#BP SENTIMENT BAR
CHART=====================================================================
=======================================================

#sentiment analysis so we can see scoring with the highest emotional content in different
categories
bpsentiment<-get_nrc_sentiment(bpword.df)

#collect sentiment score for each emotion
bpsentiment.anger =sum(bpsentiment$anger)
bpsentiment.anticipation =sum(bpsentiment$anticipation)
bpsentiment.disgust =sum(bpsentiment$disgust)
bpsentiment.fear =sum(bpsentiment$fear)
bpsentiment.joy =sum(bpsentiment$joy)
bpsentiment.sadness =sum(bpsentiment$sadness)
bpsentiment.surprise =sum(bpsentiment$surprise)
bpsentiment.trust =sum(bpsentiment$trust)
#creating the yaxis variables to plot bar graph
```

```
yAxis <- c(bpsentiment.anger,
        + bpsentiment.anticipation,
        + bpsentiment.disgust,
        + bpsentiment.fear,
        + bpsentiment.joy,
        + bpsentiment.sadness,
        + bpsentiment.surprise,
        + bpsentiment.trust)
#creating xaxis to plot labels for bar graph
xAxis <- c("Anger","Anticipation","Disgust","Fear","Joy", "Sadness","Surprise","Trust")
colors <- c("red","blue","orange","red","green","orange", "blue", "green")
yRange <- range(0,yAxis) #in accordance with
barplot(yAxis, names.arg = xAxis, #barplot function creating our bar graph
      xlab = "Sentiment Analysis", ylab = "Score", main = "Sentiment for BP Tweets", col = colors,
border = "black", ylim = yRange, xpd = F, axisnames = T, cex.axis = 0.8, cex.sub = 0.8, col.sub =
"blue")

#BP PIE
CHART==============================================================

Positive <- length(positive.tweets) #counts number of positive tweets in new data frame
Neutral <- length(neutral.tweets) #counts number of neutral tweets in new data frame
Negative <- length(negative.tweets) #counts number of negative tweets in new data frame

count <-c(Positive, Neutral, Negative) #adding above values into new data frame count to plot
pie chart
labels <-c("Positive Tweets", "Neutral Tweets", "Negative Tweets") #naming the labels
head(count) #can see the total numbers of each scored tweet category

pie(count, labels = count, main = "BP Twitter", col = rainbow(length(count))) #plotting pie chart.
including count, labels, title and colour
legend("topright", legend = c ("Postive Tweets", "Neutral Tweets", "Negative Tweets"), fill =
c("red", "blue", "green")) #adding a legend to indicate numbers on pie chart for detail

#BP GENERATE
TDM==============================================================

bptweet_corpus <- Corpus(VectorSource(bpword.df)) #using the newly clean tweets in
bpword.df we create a corpus

bptdm <- TermDocumentMatrix(bptweet_corpus, #applying some transformations for our term
document matrix using corpus
                     control = list(removePunctuation = TRUE, wordLengths=c(5, 15), #removes
punctuation and list words charcters 5-15 in length
                            stopwords = c("thank", "bp",stopwords("english")), #stopwords that add no
value to analysis, we find what to put here after finding freq commonly terms
                            removeNumbers = TRUE, tolower = TRUE)) #remove numbers, and set to
lower case
bptdm.matrix <- as.matrix(bptdm) #defining bptdm into a matrix to calculate word frequencies
word_freqs <- sort(rowSums(bptdm.matrix), decreasing=FALSE) #creating word_freqs variable
that counts the words in bptdm.matrix in decreasing order
```

```
ordr <- order(word_freqs, decreasing=TRUE) #creating a sort order for the above variable, notice
decreasing true so most common terms first
word_freqs[head(ordr)] #we can see first six most commonly occuring terms

findAssocs(bptdm,"sorry",0.25) #from above results we use most freq words and find word
associations with tweet
findAssocs(bptdm,"thanks",0.3) #correlation figures vary to find good range of associations
findAssocs(bptdm,"energy",0.3)

bpdm <- data.frame(word=names(word_freqs), freq=word_freqs) #create data frame with words
and their frequencies

#generate wordcloud of bpdm word and bpdm freq of the words, with adjustments of colour
and sort
wordcloud(bpdm$word, bpdm$freq, max.freq = 50, #minimum frequency 50 (only words
showing at least 50 times)
        random.order=FALSE, colors=brewer.pal(8, "Dark2")) #mixes the order of words and
colours


#MOBIL
TWEETS==================================================================

#reading csv from our current working directory
mobiltweets.df <- read.csv("mobiltweets.csv")
head(mobiltweets.df)
head(mobiltweets.df$text)

#new mobiletweets dataframe of text only
mobiltweets.df2 <- mobiltweets.df$text
head(mobiltweets.df2)

#Converting text to readable string text to perform operations
rmv_ivalid_char <- function(x){ iconv(x, "latin1", "ASCII", sub="")}
rmv_ivalid_char2 <- function(x){ iconv(x, "utf-8", "ASCII", sub="")}

#removing invalid characters using the above function
mobiltweets.df2 <- rmv_ivalid_char(mobiltweets.df2)
mobiltweets.df2 <- rmv_ivalid_char2(mobiltweets.df2)

#preprocessing stage cleaning up text to perform text mining analysis

mobiltweets.df2 <- gsub("\\s*<U\\+\\w+>","",mobiltweets.df2)
mobiltweets.df2 <- gsub("http.*","",mobiltweets.df2)
mobiltweets.df2 <- gsub("[\\]+","",mobiltweets.df2)
mobiltweets.df2 <- gsub("#","",mobiltweets.df2)
mobiltweets.df2 <- gsub("*","",mobiltweets.df2)
mobiltweets.df2 <- gsub("https.*","",mobiltweets.df2)
mobiltweets.df2 <- gsub("&amp","and",mobiltweets.df2)
mobiltweets.df2 <- gsub(";","",mobiltweets.df2)
mobiltweets.df2 <- gsub("w/!","",mobiltweets.df2)
```

```
mobiltweets.df2 <- gsub("[\n\n]", "", mobiltweets.df2)
mobiltweets.df2 <- gsub("[[:punct:]]", "", mobiltweets.df2) #removing punctuation
mobiltweets.df2 <- gsub("(RT|via)((?:\\b\\W*@\\w+)+)", "", mobiltweets.df2) #removing RT|via
other control codes

#we can see the first few rows of the newly cleaned text for sentimental analysis
head(mobiltweets.df2)

#MOBIL SENTIMENTAL
ANALYSIS====================================================

mobilword.df <- as.vector(mobiltweets.df2)
mobilemotion.df <- get_nrc_sentiment(mobilword.df)
mobilemotion.df2 <- cbind(mobiltweets.df2, mobilemotion.df)
head(mobilemotion.df2)

sent.value <- get_sentiment(mobilword.df)
#highest scoring terms
most.emotionalcontent <- mobilword.df[sent.value == max(sent.value)]
most.emotionalcontent

least.emotionalcontent <- mobilword.df[sent.value <= min(sent.value)]
least.emotionalcontent

#getting scoring values and inputting into new data frames
positive.tweets <- mobilword.df[sent.value > 0]
negative.tweets <- mobilword.df[sent.value < 0]
neutral.tweets <- mobilword.df[sent.value == 0]

#see all the most commonly scored tweets in each category
head(positive.tweets)
head(negative.tweets)
head(neutral.tweets)

#MOBIL SENTIMENT BAR
CHART=====================================================

# Perform sentiment analysis so we can see scoring with the highest emotional content in
different categories
mobilsentiment<-get_nrc_sentiment(mobilword.df)

# Get the sentiment score for each emotion
mobilsentiment.anger =sum(mobilsentiment$anger)
mobilsentiment.anticipation =sum(mobilsentiment$anticipation)
mobilsentiment.disgust =sum(mobilsentiment$disgust)
mobilsentiment.fear =sum(mobilsentiment$fear)
mobilsentiment.joy =sum(mobilsentiment$joy)
mobilsentiment.sadness =sum(mobilsentiment$sadness)
mobilsentiment.surprise =sum(mobilsentiment$surprise)
mobilsentiment.trust =sum(mobilsentiment$trust)
```

```
# Create the bar chart
yAxis <- c(mobilsentiment.anger,
        + mobilsentiment.anticipation,
        + mobilsentiment.disgust,
        + mobilsentiment.fear,
        + mobilsentiment.joy,
        + mobilsentiment.sadness,
        + mobilsentiment.surprise,
        + mobilsentiment.trust)

xAxis <- c("Anger","Anticipation","Disgust","Fear","Joy", "Sadness","Surprise","Trust")

colors <- c("red","blue","orange","red","green","orange", "blue", "green")

yRange <- range(0,yAxis)

barplot(yAxis, names.arg = xAxis,
      xlab = "Sentiment Analysis", ylab = "Score", main = "Sentiment for Mobil Tweets", col =
colors, border = "black", ylim = yRange, xpd = F, axisnames = T, cex.axis = 0.8, cex.sub = 0.8,
col.sub = "blue")


#MOBIL PIE
CHART=================================================================

#creating new dataframes to plot data for pie chart
Positive <- length(positive.tweets)
Neutral <- length(neutral.tweets)
Negative <- length(negative.tweets)

#couting how many so we can how many tweets per category
count <- c(Positive, Neutral, Negative)
labels <- c("Positive Tweets", "Neutral Tweets", "Negative Tweets")
head(count) #can inspect the numbers of tweets

#creating pie chart with given dataframes
pie(count, labels = count, main = "Mobil Twitter", col = rainbow(length(count)))
legend("topright", legend = c ("Positive Tweets", "Neutral Tweets", "Negative Tweets"), fill =
c("red", "blue", "green"))

#MOBILE TDM AND WORD
CLOUD=======================================================

#creating a corpus to create tdm thus a word cloud
mobiltweet_corpus <- Corpus(VectorSource(mobilword.df))
mobiltdm <- TermDocumentMatrix(mobiltweet_corpus,
                    control = list(removePunctuation = TRUE, wordLengths=c(5, 15),
                              stopwords = c("mobil",stopwords("english")),
                              removeNumbers = TRUE, tolower = TRUE))
mobiltdm.matrix <- as.matrix(mobiltdm)
word_freqs <- sort(rowSums(mobiltdm.matrix), decreasing=FALSE)
```

```
ordr <- order(word_freqs, decreasing=TRUE)
word_freqs[head(ordr)]

findAssocs(mobiltdm,"energy",0.2)
findAssocs(mobiltdm,"emissions",0.3)
findAssocs(mobiltdm,"technology",0.3)

mobildm <- data.frame(word=names(word_freqs), freq=word_freqs)
wordcloud(mobildm$word, mobildm$freq, max.freq = 50, #create word cloud of words
appearing at least 50 times
        random.order=FALSE, colors=brewer.pal(8, "Dark2"))

#Z ENERGY
TWEETS===============================================================

zenergytweets.df <- read.csv("zenergytweets.csv")
head(zenergytweets.df)
head(zenergytweets.df$text)
zenergytweets.df2 <- zenergytweets.df$text

head(zenergytweets.df2)

rmv_ivalid_char <- function(x){ iconv(x, "latin1", "ASCII", sub="")}
rmv_ivalid_char2 <- function(x){ iconv(x, "utf-8", "ASCII", sub="")}
zenergytweets.df2 <- rmv_ivalid_char(zenergytweets.df2)
zenergytweets.df2 <- rmv_ivalid_char2(zenergytweets.df2)

zenergytweets.df2 <- gsub("\\s*<U\\+\\w+>","",zenergytweets.df2)
zenergytweets.df2 <- gsub("http.*","",zenergytweets.df2)
zenergytweets.df2 <- gsub("[\\]+","",zenergytweets.df2)
zenergytweets.df2 <- gsub("#.*","",zenergytweets.df2)
zenergytweets.df2 <- gsub("*","",zenergytweets.df2)
zenergytweets.df2 <- gsub("https.*","",zenergytweets.df2)
zenergytweets.df2 <- gsub("&amp","and",zenergytweets.df2)
zenergytweets.df2 <- gsub(";","",zenergytweets.df2)
zenergytweets.df2 <- gsub("@\\w+","",zenergytweets.df2)
zenergytweets.df2 <- gsub("&gt","",zenergytweets.df2)
zenergytweets.df2 <- gsub("[[:punct:]]", "", zenergytweets.df2)
zenergytweets.df2 <- gsub("(RT|via)((?:\\b\\W*@\\w+)+)", "", zenergytweets.df2)
zenergytweets.df2 <- gsub("http\\w+", "", zenergytweets.df2)

#cleaned preprocessed tweets

head(zenergytweets.df2)

#Z ENERGY SENTIMENTAL
ANALYSIS=====================================================

zenergyword.df <- as.vector(zenergytweets.df2)
zenergyemotion.df <- get_nrc_sentiment(zenergyword.df)
zenergyemotion.df2 <- cbind(zenergytweets.df2, zenergyemotion.df)
```

```
head(zenergyemotion.df2)

sent.value <- get_sentiment(zenergyword.df)

most.sentimental <- zenergyword.df[sent.value == max(sent.value)]
most.sentimental
least.sentimental <- zenergyword.df[sent.value <= min(sent.value)]
least.sentimental

positive.tweets <- zenergyword.df[sent.value > 0]
negative.tweets <- zenergyword.df[sent.value < 0]
neutral.tweets <- zenergyword.df[sent.value == 0]

head(positive.tweets)
head(negative.tweets)
head(neutral.tweets)

# Perform sentiment analysis so we can see scoring with the highest emotional content in
different categories
zenergysentiment<-get_nrc_sentiment(zenergyword.df)

# Get the sentiment score for each emotion
zenergysentiment.anger =sum(zenergysentiment$anger)
zenergysentiment.anticipation =sum(zenergysentiment$anticipation)
zenergysentiment.disgust =sum(zenergysentiment$disgust)
zenergysentiment.fear =sum(zenergysentiment$fear)
zenergysentiment.joy =sum(zenergysentiment$joy)
zenergysentiment.sadness =sum(zenergysentiment$sadness)
zenergysentiment.surprise =sum(zenergysentiment$surprise)
zenergysentiment.trust =sum(zenergysentiment$trust)

# Create the bar chart
yAxis <- c(zenergysentiment.anger,
        + zenergysentiment.anticipation,
        + zenergysentiment.disgust,
        + zenergysentiment.fear,
        + zenergysentiment.joy,
        + zenergysentiment.sadness,
        + zenergysentiment.surprise,
        + zenergysentiment.trust)

xAxis <- c("Anger","Anticipation","Disgust","Fear","Joy", "Sadness","Surprise","Trust")

colors <- c("red","blue","orange","red","green","orange", "blue", "green")

yRange <- range(0,yAxis)

barplot(yAxis, names.arg = xAxis,
      xlab = "Sentiment Analysis", ylab = "Score", main = "Sentiment for ZEnergy Tweets", col =
colors, border = "black", ylim = yRange, xpd = F, axisnames = T, cex.axis = 0.8, cex.sub = 0.8,
col.sub = "blue")
```

```
#Z ENERGY
PIECHART=============================================================

Positive <- length(positive.tweets)
Neutral <- length(neutral.tweets)
Negative <- length(negative.tweets)
count <- c(Positive, Neutral, Negative)
labels <- c("Positive Tweets", "Neutral Tweets", "Negative Tweets")
head(count)
pie(count, labels = count, main = "Zenergy Twitter", col = rainbow(length(count)))
legend("topright", legend = c ("Postive Tweets", "Neutral Tweets", "Negative Tweets"), fill =
c("red", "blue", "green"))

#Z ENERGY TDM AND WORD
CLOUD===================================================================

zenergytweet_corpus <- Corpus(VectorSource(zenergyword.df))

zenergytdm <- TermDocumentMatrix(zenergytweet_corpus,
                    control = list(removePunctuation = TRUE, wordLengths=c(5, 15),
                            stopwords = c("Zenergy","youre",stopwords("english")),
                            removeNumbers = TRUE, tolower = TRUE))
zenergytdm.matrix <- as.matrix(zenergytdm)
word_freqs <- sort(rowSums(zenergytdm.matrix), decreasing=FALSE)
ordr <- order(word_freqs, decreasing=TRUE)
word_freqs[head(ordr)]

findAssocs(zenergytdm,"thanks",0.25)
findAssocs(zenergytdm,"sorry",0.25)
findAssocs(zenergytdm,"message",0.25)

zenergydm <- data.frame(word=names(word_freqs), freq=word_freqs)
wordcloud(zenergydm$word, zenergydm$freq, max.freq = 50,
     random.order=FALSE, colors=brewer.pal(8, "Dark2"))

#===========================================================================
======================
```