HOMEWORK 7
Submitted by: SHARANG BIRADAR N11407631

PART 1

Case:

The CEO has charged your analytics groups with a task: "What are the characteristics of an organization that adapts well to data analytics?" The CEO's intention is to restructure the company to foster adoption and advancement of data analytics capabilities across units.

Steps:

- Find a list of successful data-driven companies (See this Forbes article for inspiration) and select 30 organizations. Create a corpus of their mission statements

	Document	Company.Name	Mission.Statements	Core.Values
1	Docu_1	Oracle	Simplify: Speed information delivery with integrated s...	Integriy. Mutual Respect. Teamwork. Communication....
2	Docu_2	SAP	We don't just create game-changers like the SAP HAN...	Tell it like it is, stay curious, embrace differences, kee...
3	Docu_3	Twitter	To give everyone the power to create and share ideas ...	Grow our business in a way that makes us proud. Rec...
4	Docu_4	PayPal	To build the Web's most convenient, secure, cost-effe...	We operate an open, secure and technology agnostic ...
5	Docu_5	Alibaba	To make it easy to do business anywhere.	Our success and rapid growth is built on the spirit of ...
6	Docu_6	Dell	Dell's mission is to be the most successful computer ...	Customers: We believe our relationships with custom...
7	Docu_7	Facebook	To give people the power to share and make the worl...	Be Bold. Focus on impact. Move Fast. Be open. Build s...
8	Docu_8	Uber	Uber is evolving the way the world moves. By seamles...	Be an owner, not a renter. Take big bold bets. The be...
9	Docu_9	Tesla	Tesla's mission is to accelerate the world's transition t...	Challenege reality and solve problems beyond yourself.
10	Docu_10	Hewlett Packard	Customer Loyaly. Profit. Growth. Market Leadership. ...	Trust and respect for individuals. Achievement and co...
11	Docu_11	Cisco	Shape the future of the Internet by creating unpreced...	Change the world. Focus intensely on customers. Mak...
12	Docu_12	Intel	Utilize the power of Moore<cd>s Law to bring smart, ...	Quality. Risk taking. Customer orientation. Results ori...
13	Docu_13	Adobe	To move the web forward and give web designers and...	genuine, exceptional, innovative, and involved
14	Docu_14	AWS	The AWS Cloud makes a broad set of services, partner...	Customer obsession. Ownership. Invent and simplify. ...
15	Docu_15	Walmart	We save people money so they can live better.	Serve the people.
16	Docu_16	P&G	We will provide branded products and services of sup...	We always try to do the right thing. We are honest an...
17	Docu_17	Southwest	Dedication to the highest quality of Customer Service ...	The development, imporvement, and refinement of th...
18	Docu_18	Trip Advisor	To help people around the world plan and have the p...	A commitment to innovation and excellence.
19	Docu_19	Loreal	Offering all women and men worldwide the best of co...	Passion for adventure, enrichment through diversity, l...
20	Docu_20	Toyota	To attract and attain customers with high-valued pro...	Kaizen. Challenege. Teamwork. Respect.
21	Docu_21	Virgin America	Virgin America is a California-based airline that is on ...	Virgin America is committed to doing business in a w...
22	Docu_22	DU PONT	To create shareholder and societal value while reduci...	Safety & Health, Environmental Stewardship, Highest ...
23	Docu_23	Philips	Improving people's lives through meaningful innovation	We take a two-dimensional approach social and ecolo...
24	Docu_24	Sales Force	Salesforce pioneered a new technology model with ou...	Equality for all, Innovation, Giving back, Customer Su...
25	Docu_25	Tableau	We help people see and understand their data. Eight ...	User Interface Design, Data Access Design
26	Docu_26	SAS Institute	SAS delivers proven solutions that drive innovation an...	Approachable. Customer Driven. Swift & Agile. Innova...
27	Docu_27	Confluence	Confluence Charter Schools provides the highest qual...	Confluence Charter Schools will provide a world-class...
28	Docu_28	Microsoft	our mission is to enable people and businesses throu...	As a company, and as individuals, we value<e6>integ...
29	Docu_29	MongoDB	In the ge ofdigital transformation and disruption, you...	Think Big, Go Far; Make it Matter, Build together, Inte...
30	Docu_30	Apple	Apple designs Macs, the best personal computers in t...	We believe that we're on the face of the Earth to make...

```
> View(companies)
> dim(companies) # dim of file; 107 files with 11 headers
[1] 30  4
> names(companies)   # names of the headers
[1] "Document"           "Company.Name"       "Mission.Statements"
[4] "Core.Values"
> require(quanteda)  #create a corpus with metadata
> help(corpus)
> companies_new<- corpus(companies$Mission.Statements,
+                        docnames=companies$Document)
> names(companies_new)  #to explore the output of the corpus function: "documents" "metad
ata"  "settings"  "tokens"
[1] "documents" "metadata"  "settings"  "tokens"
```

```
> summary(companies_new)  #summary of corpus
Corpus consisting of 30 documents.

    Text Types Tokens Sentences
  Docu_1    31     39         5
  Docu_2    57     74         4
  Docu_3    16     17         1
  Docu_4    14     15         1
  Docu_5     9      9         1
  Docu_6    33     38         2
  Docu_7    14     16         1
  Docu_8    29     37         2
  Docu_9    11     12         1
 Docu_10    13     20         7
 Docu_11    20     24         1
 Docu_12    19     20         1
 Docu_13    15     20         1
 Docu_14    31     34         1
 Docu_15    10     10         1
 Docu_16    18     21         1
 Docu_17    21     24         1
 Docu_18    12     13         1
 Docu_19    17     19         1
 Docu_20    19     21         1
 Docu_21    37     46         1
 Docu_22    18     20         1
 Docu_23     6      6         1
 Docu_24    44     74         1
 Docu_25    22     24         3
 Docu_26    11     11         1
 Docu_27    30     38         2
 Docu_28    22     28         2
 Docu_29    24     25         1
 Docu_30    44     66         3
```

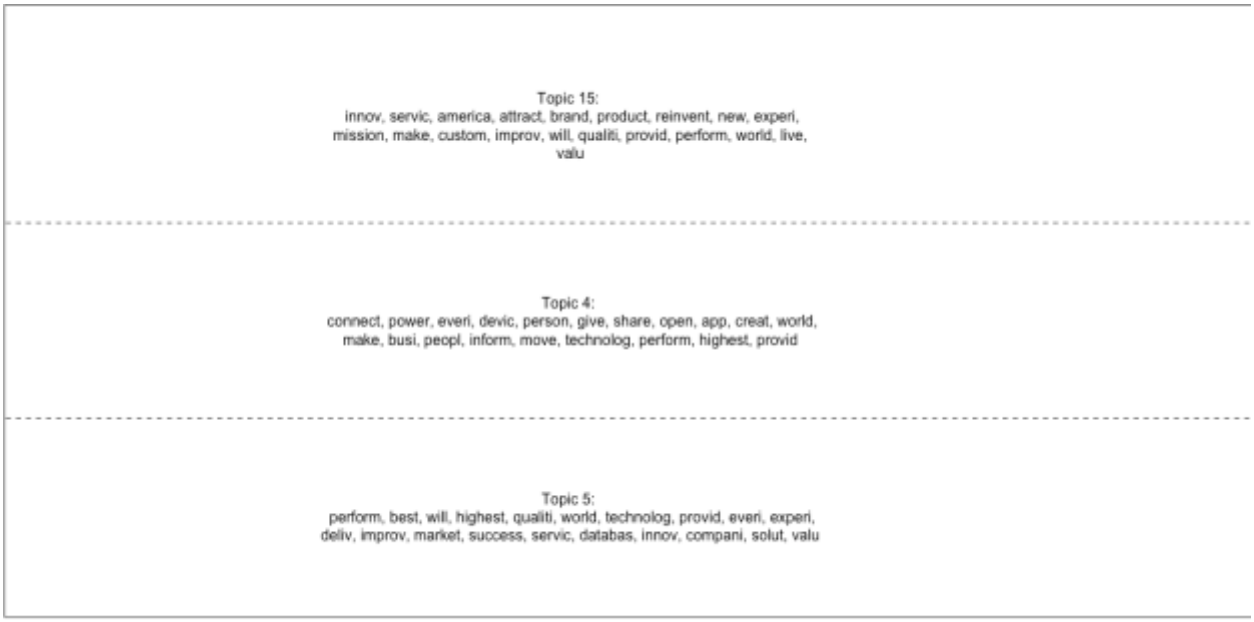
-
```
> companies_new<- corpus(companies$Core.Values,
+                        docnames=companies$Document)
> names(companies_new)  #to explore the output of the corpus function: "documents" "metad
ata"  "settings"  "tokens"
[1] "documents" "metadata"  "settings"  "tokens"
> summary(companies_new)  #summary of corpus
Corpus consisting of 30 documents.

    Text Types Tokens Sentences
  Docu_1    12     19         9
  Docu_2    19     23         1
  Docu_3    46     60        11
  Docu_4    28     30         1
  Docu_5    44     54         3
  Docu_6    63    101         6
  Docu_7    12     17         5
  Docu_8    16     18         3
  Docu_9     8      8         1
 Docu_10    15     20         5
 Docu_11    23     29         6
 Docu_12     8     12         5
 Docu_13     6      8         1
 Docu_14    31     44        11
 Docu_15     4      4         1
 Docu_16    42     61         5
 Docu_17    16     22         1
 Docu_18     7      7         1
 Docu_19    18     21         1
 Docu_20     5      8         4
 Docu_21    30     39         2
 Docu_22    14     16         1
 Docu_23    50     69         3
 Docu_24    14     19         1
 Docu_25     6      7         1
 Docu_26     9     13         5
 Docu_27    24     24         1
 Docu_28    35     54         2
 Docu_29    16     19         1
 Docu_30    79    138         7
```

- Analyze the corpus and provide insight on how to structure a firm for data-analysis readiness
  For Core Values

  The company should structure itself by embody the above mentioned core values like (Believe, Innovation, Customer, Success, Excellence etc) when considering the data analysis

readiness.

```
> companies_new<- toLower(companies_new, keepAcronyms = FALSE)
> companies_clean <- tokenize(companies_new,
+                         removeNumbers=TRUE,
+                         removePunct = TRUE,
+                         removeSeparators=TRUE,
+                         removeTwitter=FALSE,
+                         verbose=FALSE)
> dfm.simple<- dfm(companies_clean,
+                  toLower = TRUE,
+                  ignoredFeatures =stopwords("english"),
+                  verbose=FALSE,
+                  stem=FALSE)
> topfeatures<-topfeatures(dfm.simple, n=50)
> topfeatures
     believe   innovation      respect    customers     customer      success
          11            8            7            6            5            5
  excellence        build           us        right        focus       people
           5            4            4            4            4            4
     results    integrity contribution         make     teamwork          way
           4            4            4            4            3            3
     passion       matter        trust        every     innovate       values
           3            3            3            3            3            3
     company     together        value  individuals       always       social
           3            3            3            3            3            3
         big        great     products       mutual         grow     business
           3            3            3            2            2            2
 personality         user     simplify      operate         open   technology
           2            2            2            2            2            2
      growth       spirit          fun  exceptional          win         bold
           2            2            2            2            2            2
        take         best
           2            2
```



The company should structure by personifying itself to the above mentioned core values like (Believe, Innovation, Customer, Success, Excellence etc) when considering the data analysis readiness.

- Are there any other data-driven approaches you would recommend the CEO to implement?
  The CEO should implement the core values of believing in the product/service and be innovative at the same time. The CEO should also focus on improving customer experience with respect to its product/service.

PART 2

Case:

This video discusses Donald Trump's linguistic style. Sometimes when he speaks he seems erratic and unfocused, yet many people like him and connect with him. Assume you are providing "intelligence" to Trump's campaign manager on what is making this candidate effective with people

Steps:

- Search for "Donald Trump speech transcript" and select 3 speeches of your choice
- Create a corpus for the speeches

```
> dim(speech) # dim of file; 107 files with 11 headers
[1] 3 2
> names(speech)   # names of the headers
[1] "Document"  "Full.text"
> require(quanteda)  #create a corpus with metadata
> speech_new<- corpus(speech$Full.text,docnames=speech$Document)
> names(speech_new)   #to explore the output of the corpus function: "documents" "metadat
a"  "settings"  "tokens"
[1] "documents" "metadata"  "settings"  "tokens"

> summary(speech_new)  #summary of corpus
Corpus consisting of 3 documents.

       Text Types Tokens Sentences
 Document 1  1273   4633       206
 Document 2   862   2943       156
 Document_3    44     55         5
```

- Complete a frequency analysis of word usage

  Create a corpus with metadata and

  Clean corpus: removes punctuation, digits, converts to lower case

```
> speech_new<- toLower(speech_new, keepAcronyms = FALSE)
> speech_clean <- tokenize(speech_new,
+                          removeNumbers=TRUE,
+                          removePunct = TRUE,
+                          removeSeparators=TRUE,
+                          removeTwitter=FALSE,
+                          verbose=TRUE)
Starting tokenization...
  ...preserving Twitter characters (#, @)...total elapsed: 0 seconds.
  ...tokenizing texts...total elapsed:  0.00800000000162981 seconds.
  ...replacing Twitter characters (#, @)...total elapsed: 0.00100000000384171 seconds.
  ...replacing names...total elapsed:  0.000999999996565748 seconds.
Finished tokenizing and cleaning 3 texts.
>
```

Create document feature matrix from clean corpus + stem

```
> dfm.simple<- dfm(speech_clean,
+                  toLower = TRUE,
+                  ignoredFeatures =stopwords("english"),
+                  verbose=TRUE,
+                  stem=FALSE)
Creating a dfm from a tokenizedTexts object ...
   ... lowercasing
   ... indexing documents: 3 documents
   ... indexing features: 1,539 feature types
   ... removed 106 features, from 174 supplied (glob) feature types
   ... created a 3 x 1433 sparse dfm
   ... complete.
Elapsed time: 0.019 seconds.
>
```

To display most Frequent terms in dfm

```
> topfeatures<-topfeatures(dfm.simple, n=50)
> topfeatures
          will    immigration         people        clinton        hillary
            97             47             46             46             36
             s        illegal       american        country              t
            35             33             33             28             26
           one         border         system     immigrants            now
            22             21             20             20             19
      criminal            law          going            new           want
            18             18             17             17             17
         state           make           jobs        america       citizens
            16             16             16             16             15
      security           time           just           laws       policies
            15             15             14             14             14
        number           back administration          years          crime
            14             13             13             13             13
            ve          great           work       americans        federal
            13             12             12             12             12
       million            can        special          today            put
            12             12             11             11             11
          many            let    enforcement           like         united
            11             11             11             10             10
```

- Complete a sentiment analysis

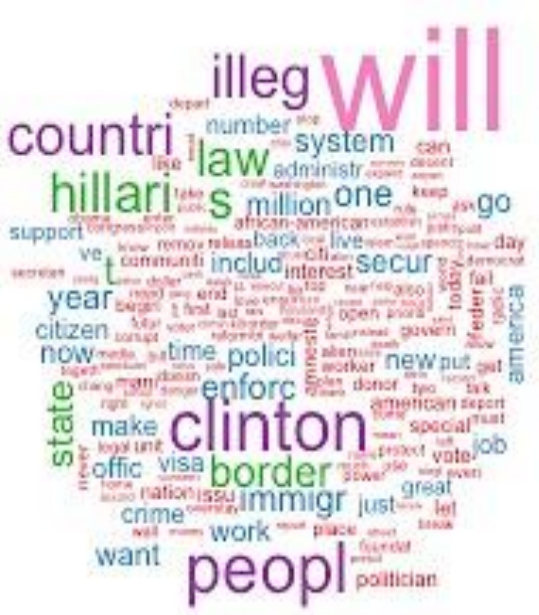
```
> mydict <- dictionary(list(negative = c("illeg*", "border*", "enforc*", "crime*"),
+                           postive = c("american", "vote", "administr*", "system")))
> dfm.sentiment <- dfm(speech_clean, dictionary = mydict)
Creating a dfm from a tokenizedTexts object ...
   ... indexing documents: 3 documents
   ... indexing features: 1,539 feature types
   ... applying a dictionary consisting of 2 keys
   ... created a 3 x 2 sparse dfm
   ... complete.
Elapsed time: 0.028 seconds.
> topfeatures(dfm.sentiment)
negative  postive
     105       76
```

View(dfm.sentiment)

	negative	postive
Document 1	86	45
Document 2	19	31
Document_3	0	0

- What are the common topics in the corpus

```
> View(dfm.sentiment)
> library(wordcloud)
> set.seed(142)   #keeps cloud' shape fixed
> dark2 <- brewer.pal(8, "Set1")
> freq<-topfeatures(dfm.stem, n=500)
> wordcloud(names(freq),
+           freq, max.words=200,
+           scale=c(3, .1),
+           colors=brewer.pal(8, "Set1"))
```



- Write a memo style report summarizing Trump's linguistic effectiveness

  People understand Trump on emotional level. The American people look for authenticity and spontaneity in a president, which trump seems to have mastered, they also know to value discipline in their leaders. He also calls his supporters "Folks" to show he is one of them. He keeps calling Hillary "crooked", or keeps referring to terrorists as "radical Muslims", he's strengthening the association through repetition. He makes use of terms like "Believe me", "Many people are saying…", etc.. which tends to sound more trustworthy to the listeners since he tries to imply that he is having a direct experience with what he is talking about.