NAME : MABLE NALUGYA   COHORT 2

PWC AND KMPG ANALYSIS
AS AT 7th Feb, 2019

Introduction

My analysis was focused on two of the big four (Deloitte, PricewaterhouseCoopers (PwC), Ernst & Young (EY) and KPMG.) services networks which offer services such as **audit**, assurance, corporate finance and legal services, among others as we will visually see in the word clouds. Focusing on the 2 firms which **are PWC AND KMPG** these competitive firms offer similar services. But what excited me to analysis both the company reports were to really see what makes them different from each other on the basis of data.

Findings

PWC and KMPG are both international firms based in several countries offering financial services to different entities. PWC is much more developed in terms of its financials and the sizes of the company at large on the bases of its annual revenue and KMPG is relative, it's still catching up with the updated technologies and systems, it still employs developing talents. Currently PWC is sponsoring some social responsibility events like providing Like Provide clean water sources and volunteering at various children homes as a form of teamwork strategies and KMPG is more into developing its technologies.

Analysis:

Coming from the view of the advanced technologies and Cooperate responsibility we can see from the bigrams that PWC engages more in this due it is relatively high cash flows and flexible expenditure schemes due to the relatively high margin. PWC has more advanced engagements such as the Materiality matrix. The Materiality matrix enables a company to decide which corporate social responsibility initiatives to invest in. This process helps to give back to the society ,so from this we can assume that PWC engages in Voluntary items with the society may be as a form of social responsibility, cleaning and building boreholes in rural country in areas its located we can also see from the bigram that the reports of the 2 entities are quite different where can see that one reports profits and KMPG shows its intentional to make profits in the future in different ways as reports in the quality controls measures, risk credits etc.

Another in factor is that the human capital is shown as important in PWC which was not displayed for KMPG. (PWC annual report 2017-2018) Human Capital is an intangible asset or quality not listed on a company's balance sheet. It can be classified as the economic value of a worker's experience and skills. This includes assets like education, training, intelligence, skills, health, and other things employers' value such as loyalty and punctuality. from these we can see that PWC is into taking great talent that can bring up ideas and new innovation this will show as that there is reason for new innovations. Actually, this is true because recently PWC Included health Care Analytics although it's not yet implemented in all its location globally, but it has started.

Similarities, Both these companies have board of Directories and Policies that govern them as we visualized in both the world and the bigrams, customer satisfaction is a big factor for the success of both the entitles and that risk, loss and quality assurances are major factor which shows as that professionalism and specified skills are required in such industries such as CPA and the like to be able to adapt such quality standards as shown in the bigram and also a major require in actual work environment.

Insight Conclusions

From my Analysis based on both the data I have analyzed from the reports of this 2 companies and various reading on these audits. I will have to agree that both the companies have similar goals such as profitability. they are also both people and goal oriented and also promotions are earned in various ways such as being a partner which was also shows as a predominant word in the word cloud this level will be received due to you net worth or net capital. KMPG is a progressing company according to the activities being held its progressing both financially and also as an industry at large. PWC due to its increasing magnitude, its facing more negative sentiments than KMPG as expected because it's under taking new ventures hence being exposed to higher risk leading to higher profitability or losses and high expenditure in the longer run.

References

https://home.kpmg/content/dam/kpmg/nl/pdf/over-ons/integrated-report-2017-2018.pdf
https://www.pwc.nl/nl/assets/documents/pwc-annual-report-2017-2018.pdf
http://markets.ft.com/research/Lexicon/Term?term=materiality-matrix
https://www.investopedia.com/terms/h/humancapital.asp

APPENDIX

```
library(textreadr)
library(dplyr)
library(stringr)
library(tidytext)
library(pdftools)

PWC<-read_pdf(file='/Users/macbookpro/Desktop/TEXT ANALYTICS/PWC REPORT.pdf')
View (PWC)
```

R-OUTPUT

	page_id	element_id	text
1	1	1	2017-2018 PwC Annual Report pwc.nl
2	2	1	Contents Foreword Key statistics Report of the ...
3	3	1	Contents Foreword Foreword Key s...
4	4	1	Contents Foreword We can become...
5	5	1	Contents Foreword Key statistics Report of the Superv...
6	6	1	Contents Foreword Key statistic...
7	7	1	Contents Foreword Key statistics Looking...
8	8	1	Contents Foreword Key statistics Report of the Superv...
9	9	1	
10	10	1	Contents Foreword The SB is curre...
11	11	1	
12	12	1	Contents Foreword • A ...
13	13	1	Contents Foreword • R ...
14	14	1	Contents Foreword Key statistics Report of the Superv...

```
a <- 136
b <- 1
my_df<- as.data.frame(matrix(nrow=a, ncol=b))

for(z in 1:b){
  for(i in 1:a){
    my_df[i,z]<- PWC$text[i*b+z-b]
  }
}

## This is the de

PWC_reloaded <- my_df %>%
unnest_tokens(word, V1) %>%
  anti_join(stop_words) %>%
  count(word, sort=TRUE)
PWC_reloaded
```

R-OUTPUT

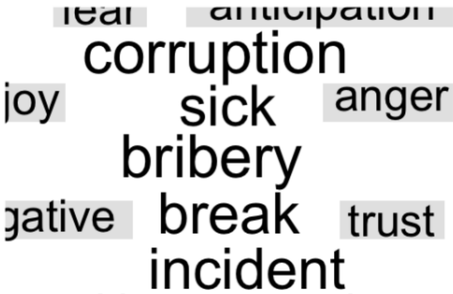
```
word                  n
  <chr>              <int>
 1 pwc                 740
 2 board                669
 3 report               622
 4 2018                610
 5 2017                496
 6 management           479
 7 financial            467
 8 pricewaterhousecoopers   316
 9 statements           310
10 supervisory          285
# … with 5,438 more rows
>
```

```
library(wordcloud)
PWC_reloaded %>%
  top_n(100) %>%
  wordcloud()
```

R-OUTPUT


```
library(reshape2)
PWC_reloaded %>%
  inner_join(get_sentiments("nrc")) %>%
  count(word, sentiment, sort=TRUE) %>%
  acast(word ~sentiment, value.var="n", fill=0) %>%
  comparison.cloud(colors = c("grey20", "gray80"),
                max.words=50)
```

R-OUTPUT



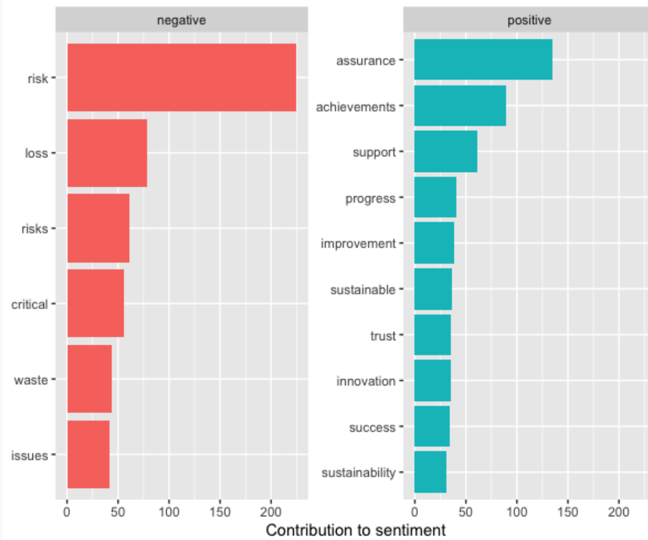
CONCLUSIONS

The most apparent word that appear for PWC is the corruption and bribery .. this may mean that they are the insightful words but also they could just be the ones that appear often even with out any insights attached to them

```
PWC_sentiment <- PWC_reloaded %>%
  inner_join(get_sentiments("bing")) %>%
  ungroup()

PWC_sentiment %>%
  group_by(sentiment) %>%
  filter(n>30) %>%
  ungroup() %>%
  mutate(word=reorder(word, n)) %>%
  ggplot(aes(word, n, fill=sentiment)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~sentiment, scales = "free_y")+
  labs(y="Contribution to sentiment", x=NULL)+
  coord_flip()
```

R-OUTPUT

```
A tibble: 411 x 3
 word              n sentiment
 <chr>         <int> <chr>
1 risk           225 negative
2 assurance      135 positive
3 achievements    89 positive
4 loss            78 negative
5 risks           61 negative
6 support         61 positive
7 critical        56 negative
8 waste           44 negative
9 issues          42 negative
0 progress        41 positive
… with 401 more rows
```



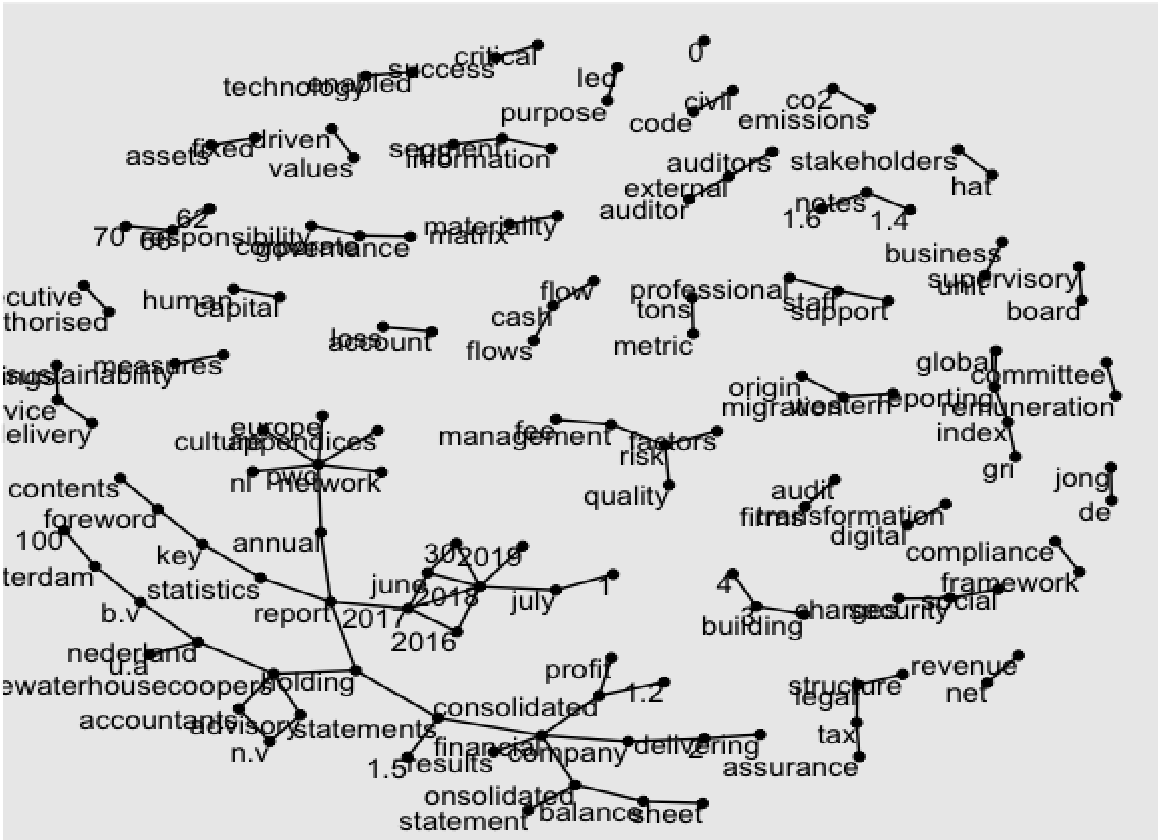
CONCLUSIONS

from the sentiment, we can use that the positive sentiments are greater than the negatives but what we cannot see from the graph is that, First risk and risks are really different risk can be a positive thing especially from the business and transaction point of view because where is risk there is no return yet from the sentiment before the risk and risks are negative sentiments. this should inform as to make further analysis even after the sentimentations.

```
library(dplyr)
library(tidytext)
library(janeaustenr)
library(tidyr)

PWC$text <- as.character(PWC$text)


my_bigrams <- PWC %>%
  unnest_tokens(bigram, text, token = "ngrams", n=2)

my_bigrams #We want to see the bigrams (words that appear together, "pairs")

PWC_bigrams <- my_bigrams %>%
  count(bigram, sort = TRUE) #this has many stop words, need to remove them

#to remove stop words from the bigram data, we need to use the separate function:
library(tidyr)
bigrams_separated <- my_bigrams %>%
  separate(bigram, c("word1", "word2"), sep = " ")
```

```
bigrams_filtered <- bigrams_separated %>%
  filter(!word1 %in% stop_words$word) %>%
  filter(!word2 %in% stop_words$word)

#creating the new bigram, "no-stop-words":
bigram_counts <- bigrams_filtered %>%
  count(word1, word2, sort = TRUE)
#want to see the new bigrams
bigram_counts



library(igraph)
bigram_graph <- bigram_counts %>%
  filter(n>13) %>%
  graph_from_data_frame()

bigram_graph

# install.packages("ggraph")
library(ggraph)
ggraph(bigram_graph, layout = "fr") +
  geom_edge_link()+
  geom_node_point()+
  geom_node_text(aes(label= name), vjust =1, hjust=1)
```

R-OUTPUT



KMPG###
###########

KMPG <-read_pdf(file='/Users/macbookpro/Desktop/TEXT ANALYTICS/integrated-report-2017-2018.pdf')

View(KMPG)

R-OUTPUT

	page_id	element_id	text
1	1	1	People driven progress Integrated Report 2017-2018 ...
2	2	1	People-driven progress We ...
3	3	1	People-driven progress KPMG at ...
4	4	1	Contents 1 KPMG at a glance 7 ...
5	5	1	Contents Technology accelerates success ...
6	6	1	Contents 5 Financial Statements 1...
7	7	1	'The human side of'... Ali Alam KPMG at a glance...
8	8	1	KPMG at a glance The publ Theicpublic trust...
9	9	1	The public trusts us KPMG at a glance Our peopl...
10	10	1	Our people are extraordinary KPMG at a glance ...
11	11	1	Our clients see a difference in us KPMG at a glance ...
12	12	1	Technology accelerates success KPMG at a glance ...
13	13	1	'The human side of'... Willem Bonekamp Overview ...
14	14	1	Overview and strategy Trusted...
15	15	1	Overview and strategy This is exactly the reason th

```
a <- 233
b <- 1
my_df<- as.data.frame(matrix(nrow=a, ncol=b))

for(z in 1:b){
  for(i in 1:a){
    my_df[i,z]<- KMPG$text[i*b+z-b]
  }
}
```

```
#######################SECOND DATAFRAME####################
 KMPG_reloaded <- my_df %>% ##   names the KMPG with a new variable that we call after
tokenzing which is KMPG_reloaded for this case
  unnest_tokens(word, V1) %>%        #tokenizing  seperates word per word that appears in the
dataset
  anti_join(stop_words) %>% #here's where we remove tokens
  count(word, sort=TRUE)
KMPG_reloaded
```

R-OUTPUT

```
   word          n
   <chr>     <int>
 1 kpmg        825
 2 2018        652
 3 2017        520
 4 financial   484
 5 n.v         416
 6 1           383
 7 audit       383
 8 board       328
 9 report      315
10 2           300
# … with 5,688 more rows
```

CONCLUSIONS

The original file before it becomes a data frame or even tokenized is called KMPG. to make it callable we make it a data frame using a loop. where we name the fie my_df and after tokenized and gave it a new name KMPG_reloaded.

After tokenizing the data frame we get the number of times each word appears in the data frame as we can see above

```
KMPG_reloaded <- my_df %>%
  unnest_tokens(word, V1) %>%
  anti_join(stop_words) %>%
  count(word, sort=TRUE)
KMPG_reloaded

library(wordcloud2)
KMPG_reloaded %>%
  top_n(100) %>%
  wordcloud2()
```

R- OUTPUT

CONCLUSIONS

First, before the out put we first tokenized the data frame by unnesting that tokens and removing the stop words and gave it a new name KMPG_reloaded different from its original name which as KMPG

```
library(wordcloud)
library(reshape2)
KMPG_reloaded %>%
  inner_join(get_sentiments("nrc")) %>%
  count(word, sentiment, sort=TRUE) %>%
  acast(word ~sentiment, value.var="n", fill=0) %>%
  comparison.cloud(colors = c("grey20", "gray80"),
                   max.words=50)
```

R-OUTPUT

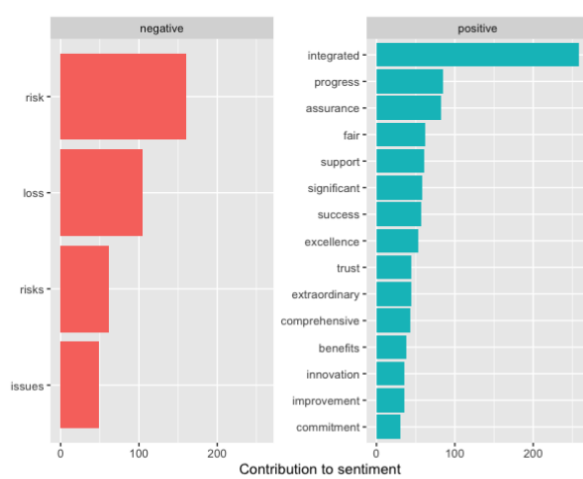
CONCLUSIONS

The cloud displays the words that appear most often and the more the word appears the big the size of the word as we can use above but does this mean the more insightful the word?, may be, may be not its just gives the number of times the word appears in the text and for this case bribery is most apparent.

```
KMPG_sentiment <- KMPG_reloaded %>%
  inner_join(get_sentiments("bing")) %>%
  ungroup()

KMPG_sentiment %>%
  group_by(sentiment) %>%
  filter(n>30) %>%
  ungroup() %>%
  mutate(word=reorder(word, n)) %>%
  ggplot(aes(word, n, fill=sentiment)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~sentiment, scales = "free_y")+
  labs(y="Contribution to sentiment", x=NULL)+
  coord_flip()
KMPG_sentiment
```

R-OUTPUT

```
   word             n sentiment
   <chr>        <int> <chr>
 1 integrated     259 positive
 2 risk           160 negative
 3 loss           105 negative
 4 progress        85 positive
 5 assurance       82 positive
 6 fair            62 positive
 7 risks           62 negative
 8 support         61 positive
 9 significant     59 positive
10 success         57 positive
# … with 403 more rows
>
```



Conclusion

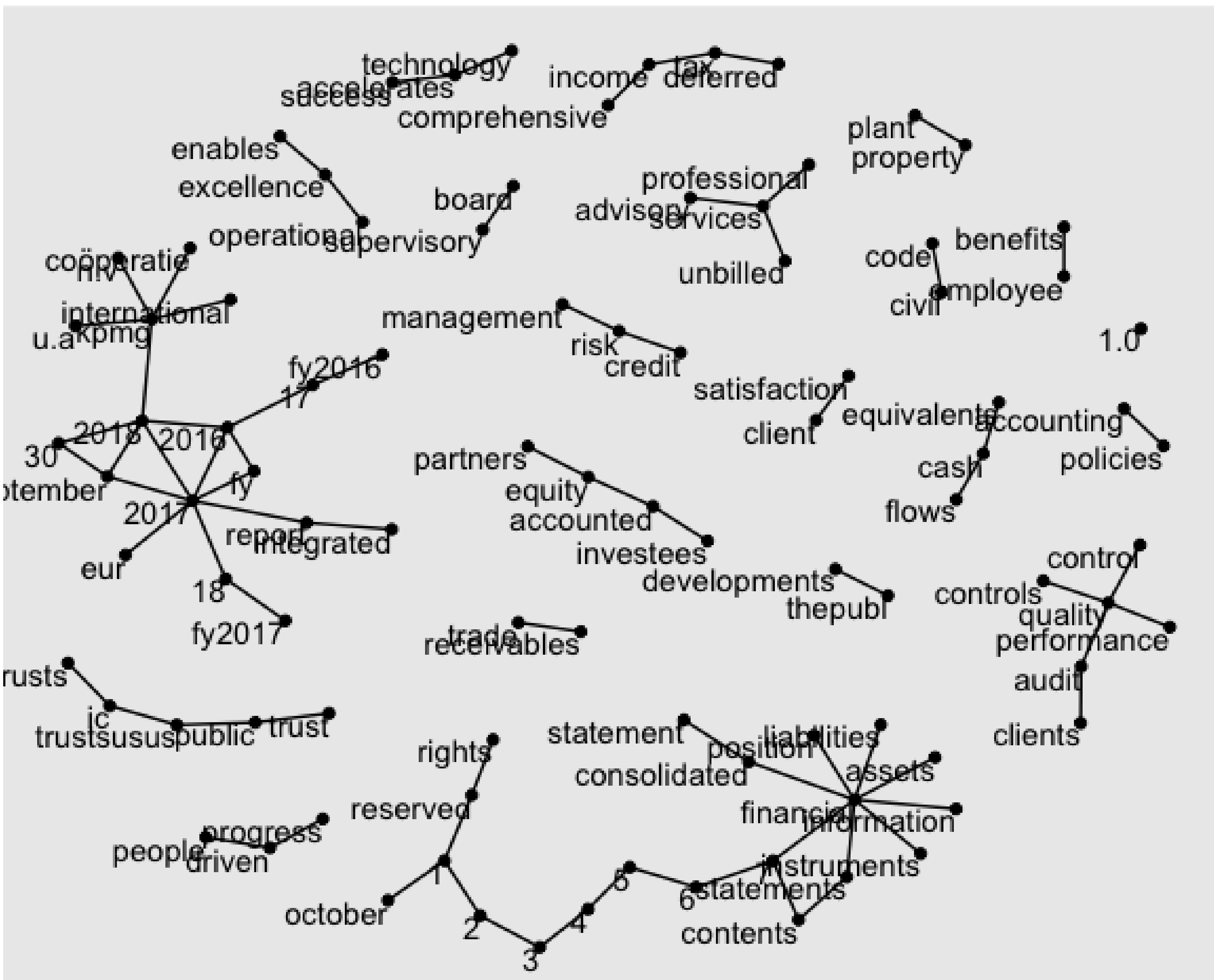
As we can visualize from the graph and the text displayed that we have more positive sentiments than the negative sentiments. this is because it’s what we can see but sometimes the sentiments are wrongly positioned for example risk is being viewed as a negative sentiment but actually risk is not always negative given the fact that this is a financial company we can have risk management, risky profitable ventures and the risk departments and the organization, not necessary in a negative context like the graph above displays.

3##

```
KMPG$text <- as.character(KMPG$text)


my_bigrams <- KMPG %>%
  unnest_tokens(bigram, text, token = "ngrams", n=2)

my_bigrams #We want to see the bigrams (words that appear together, "pairs")

KPMG_bigrams <- my_bigrams %>%
  count(bigram, sort = TRUE) #this has many stop words, need to remove them

#to remove stop words from the bigram data, we need to use the separate function:
library(tidyr)
bigrams_separated <- my_bigrams %>%
  separate(bigram, c("word1", "word2"), sep = " ")

bigrams_filtered <- bigrams_separated %>%
  filter(!word1 %in% stop_words$word) %>%
  filter(!word2 %in% stop_words$word)

#creating the new bigram, "no-stop-words":
bigram_counts <- bigrams_filtered %>%
  count(word1, word2, sort = TRUE)
#want to see the new bigrams
bigram_counts



library(igraph)
bigram_graph <- bigram_counts %>%
  filter(n>13) %>%
```

```
  graph_from_data_frame()

bigram_graph

# install.packages("ggraph")
library(ggraph)
ggraph(bigram_graph, layout = "fr") +
  geom_edge_link()+
  geom_node_point()+
  geom_node_text(aes(label= name), vjust =1, hjust=1)
```

R-OUTPUT



CONCLUSIONS

From this bigram we can assume a lot of meaning by just analysis the 2 words such as, we can see this organization has accounting policies that govern it, the company is trusted by the public, the focus on the quality control of their audit clients. In general we can see that its people driven oriented to satisfy client needs. These words can lead as to deeper analysis and foundation for the start.