Task 4

1. Introduction

Sentimental Analysis on our dataset is implemented both in R programming and SAS Enterprise Miner. Sentimental analysis is nothing but text mining with emotional rating given to the given input data. This is a special type of text mining in which we will extract the subjective information from our data input text, and we will gain knowledge about the emotions, opinions, moods. In other words, the overall sentiments that exist input text.

Our goal of implementing sentiment analysis is to analyse the qualitative info of input text data. There are too many tasks sentiment analysis, so we decided on checks of the polarity of the text which could be positive, negative or neutral. We will also mine emotions in our input text like happy, sad, angry and more.

In R programming, we have packages that help us implement sentimental analysis, syuzhet_vector, bing_vector and afinn_vector which will help us obtain the sentimental qualitative information and its frequency count i.e. number of times those qualitative terms occurred in our input text data.

And in SAS EM, we will use text rule builder node to build a model that classifies reviews into positive, negative and neutral quantitative terms. Text Rule builder node is a boolean rule-based categorizer that automatically generates an ordered set of rules for describing and predicting our target variable.

Finally, we take sentimental analysis results from both R and SAS implementations and try to compile the comparison reports generated individually.

2. Background Research

Sentimental Analysis as we know also called opinion mining is used for analysing and determining emotional in the text whether they are having positive, negative, or neutral vibes. In simple words, this helps us categorize pieces of writing in qualitive approach. This saves time and effort because sentimental extraction is fully automated algorithm that analyses sentiment of our dataset. This is becoming a popular topic in artificial intelligence, deep learning, machine learning techniques, and natural language processing technologies. Cambria, E., Das, D., Bandyopadhyay, S. and Feraco, A. eds., 2017. A practical guide to sentiment analysis (pp. 1-196). Cham, Switzerland: Springer International Publishing [1].

In other words, opinion mining or sentimental analysis gives us opportunity to explore mindset of reviewers or audience and study the state of text in from opposite point of view. This makes it a great tool for expanded product analytics and market research with precision targeting and scoring of the reviews. Agarwal, Apoorv, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca J. Passonneau. "Sentiment analysis of twitter data." In Proceedings of the workshop on language in social media (LSM 2011), pp. 30-38. 2011 [2].

With that in mind, sentimental analysis is applied to our dataset in finding and extracting opinionated data for our text data input. We will also determine polarity (positive or negative), subjective terms and identify opinion holder. Gonçalves, P., Araújo, M., Benevenuto, F. and Cha, M., 2013, October. Comparing and combining sentiment analysis methods. In Proceedings of the first ACM conference on Online social networks (pp. 27-38) [3].

Why do sentimental analysis matter, can be answered how it create perception and understanding of our input text data through the lens. Medhat, W., Hassan, A. and Korashy, H., 2014. Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal, 5(4), pp.1093-1113 [4]. It is an influential factor in formulating the segmentation of audience of our review data.

Rule-based approach in sentimental analysis implemented on defined description of an opinion to identify. This involves stemming, tokenization, parsing, and lexicon analysis Hutto, C. and Gilbert, E., 2014, May. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 8, No. 1) [5]. We are not using lexicon analysis in our research.

SAS EM is a powerful tool with integrated data mining techniques like text parsing, text filtering, text profiling and text segmenting. We will use same in our research. Jain, V.K. and Kumar, S., 2017. Improving customer experience using sentiment analysis in e-commerce. In Handbook of Research on Intelligent Techniques and Modeling Applications in Marketing Analytics (pp. 216-224). IGI Global [6].

3. Exploration of Data Set

We are provided with "hotel_accomodation_reviews.csv" dataset to implement Sentimental Analysis or Opinion mining. We are asked to choose 30 hotel or restaurant as a subset of main dataset to work with. In R we slice and partition data by selecting 30 unique hotel and create a new data subset as shown in Table [1].

	ID	Review.Date	Location	Hotel.Restaurant.name	Review
1	rn286430164	Reviewed July 7, 2015	Patong	Khon Kaen Buffet BBQ	Hubby wasn't feeling well so I consulted Trip Advisor to find...
2	rn531106916	Reviewed October 8, 2017	Patong	Laimai Courtyard Restaurant and Bar	Best restaurant at Patong Beach. They provide good service...
3	rn510366857	Reviewed August 8, 2017	Thalang District	Tatonka	We stay in the Laguna area quite often. The problem with Ta...
4	rn552352570	Reviewed January 7, 2018	Mai Khao	JJ's Bar & Food	We went to JJÕs bar last night. The food and service are exc...
5	rn399759399	Reviewed July 31, 2016	Karon	Enjoy's Italian Bar & Restaurant	A good restaurant, the food is delicious, convenient...will def...
6	rn405159638	Reviewed August 14, 2016	Kata Beach	Pomodoro Pizza Restaurant	The service is very good the food was colorful and have the ...
7	rn388976990	Reviewed July 4, 2016	Kata Beach	En Vogue Restaurant	We visited this restaurant on several occasions during our t...
8	rn464189800	Reviewed March 2, 2017	Choeng Thale	Dee Plee - Anantara Layan Phuket Resort	We chose the restaurant of the Anantara for our last night i...
9	rn538928787	Reviewed November 6, 2017	Chalong	Mamma Mia Grill & Restaurant Bangtao	We have been patronising Mamma Mia on and off since 20...
10	rn573473481	Reviewed 5 weeks ago	Patong	Sam's Steaks and Grill	Sam's Steakhouse and grill is located within the Holiday Inn ...
11	rn288044127	Reviewed July 11, 2015	Karon	Coral Restaurant	This restaurant is literally a 5 minute walk from the Centara ...
12	rn533812447	Reviewed October 17, 2017	Patong	Jeffer	My partner and I visited Jeffer for dinner on Sunday and he ...
13	rn411633256	Reviewed August 27, 2016	Mai Khao	Tropica Restaurant & Beer Garden	Stop by for a beer. Didn't had food here but I quite like the ...
14	rn518002642	Reviewed August 27, 2017	Kata Beach	Pomodoro Pizza Restaurant	If you don't know to go dinner. I'd like to come to here goo...
15	rn540123464	Reviewed November 10, 2017	Mai Khao	JJ's Bar & Food	You HAVE to try the Green Curry with chicken! Simply the be...
16	rn549268394	Reviewed December 26, 2017	Kathu	Thong Dee The Kathu Brasserie	Came here on a whim and couldnÕt be more happy with ou...
17	rn365837756	Reviewed April 20, 2016	Karon	Bob's Restaurant & Bar	A few drinks in Bobs bar, after a meal in a Kamala beach res...
18	rn477604241	Reviewed April 20, 2017	Kata Beach	Restaurant Mama Kata (Seafood)	We dinned here for a family meal, I have got to say the food...
19	rn479631463	Reviewed April 27, 2017	Cape Panwa	Uncle Nan's Italian Restaurant	This was a nice experience especially as we had free voucher...
20	rn488932925	Reviewed May 29, 2017	Kata Beach	Tiger Bar	First time walk along Bangla street, many pubs and finally c...
21	rn175881295	Reviewed September 5, 2013	Kamala	H.C. Andersen - Family Restaurant	I dropped by this place with 3 of my friend because the rest...
22	rn547596134	Reviewed December 16, 2017	Patong	The Green Gourmet	I walked past the place for about 3 days before I came in I al...
23	rn429047858	Reviewed October 17, 2016	Rawai	Cafe Java	If you want to indulge with a good french dinner, this is the ...
24	rn527927436	Reviewed September 27, 2017	Kathu	Enjoy Restaurant	Excellent value for money and great location, highly recom...
25	rn487588574	Reviewed May 24, 2017	Kamala	Baan Rim Pa Kalim	Do you want to eat very good Thai food, but in classy surro...
26	rn360669190	Reviewed April 1, 2016	Rawai	The Islander	Another fantastic visit to the islander great food and compa...
27	rn413386334	Reviewed August 31, 2016	Kamala	Kaab Gluay	Enroute our drive around phuket, we asked our driver to sto...
28	rn560372256	Reviewed February 13, 2018	Karon	Two Chefs - Karon Beach	Live music sets the atmosphere, excellent service at the tabl...
29	rn471285969	Reviewed March 29, 2017	Koh Kaew	Laem Hin Seafood	Best seafood by far in Phuket, worth the drive. Try the steam...
30	rn557740540	Reviewed February 1, 2018	Patong	Salute Italian Restaurant	Very welcoming and friendly staff and some nice little extras...

Table [1]

Once our data is ready for implementation, we take one last final look into our data. We'll needing only text field which "review" in our dataset, so we further clean the data by choosing only review column as our data as shown in Table [2]. We import our data to respective platform now to implement opinion analysis in R and SAS EM in our next step.

	Review
1	Good food. Friendly staff. Good customer service and value ...
2	A group of 11 of us enjoyed lunch here sitting on the beach...
3	The Navrang Mahal Indian Restaurant is easy to find at the ...
4	This place was great value for money. The food was pretty g...
5	Stunning views here but the pizza's were very bland. The Ste...
6	Amazing beach front dinning experience. If you are lucky yo...
7	We had food from O-Oh a few times as it was right next do...
8	No 24 is one of the more modern places on the Naithon b
9	They have very good Thai food here at a really good price
10	Just a heads up Mr. Wanchai has another restaurant dowr
11	Had a very very good evening here . The Food , service an
12	the staff were so friendly and professional, the food was go...
13	Nothing much to say than OMG what a beautiful pizza i had...
14	Nice ambiance and really tasty food. The seafood is nice an...
15	been there today for a quick lunch. It was my second visit to...
16	We were not staying guest of the resort but decided to visit ...
17	Great food!great service!we loved it!everything was fresh,ho...
18	We had a table outside with a beautiful view of the ocean. T...
19	Went with the family for dinner and ate lots for 500bt. Food ...
20	Great pizzas lots toppings, lots to choose from, good size a...
21	People there are simply beautiful, generous and sooo nice! ...
22	On The Rock was our favourite dinner restaurant during our...
23	I truly really enjoyed my visit at Eleven Two & Co. The restau...
24	Horrible service! I was physically attact by the manager for a...
25	Located slightly off the beaten track - and a bit of a walk fro...
26	Food was very good. We asked not spicy curry and it was pe...
27	The low season is hard for the people of Phuket. They have ...
28	Simply Taste like home .. a must visit for all..keep up the goo...
29	Had a great lunch here. Good variety of simple food - pizzas...
30	Absolutely loved our meal ... staff were fantastic. Our meals ...

Table [2]

4. Sentimental Analysis implementation in R and SAS EM

A. Sentimental Analysis implementation in R

After importing the final dataset acquired after data formatting, we proceed to import it to R studio. We Can also execute data formatting in R as show in output [1]. Here we used sample slicing method to distinguish 30 hotels form the rest and selecting on review column for analysis and save the file to localhost for SAS EM to have same dataset which could be useful in comparison of results purposes as shown in Output [1].

```
> reviewed <- Reviews %>% slice_sample(n=30)
> View(reviewed)
> reviewed <- reviewed["Review"]
> View(reviewed)
> #Saving the random sample file for SAS EM usage
> write.csv(reviewed,"C:/Users/Phani/Downloads/sentisample.csv", row.names = FALSE)
```

Output [1]

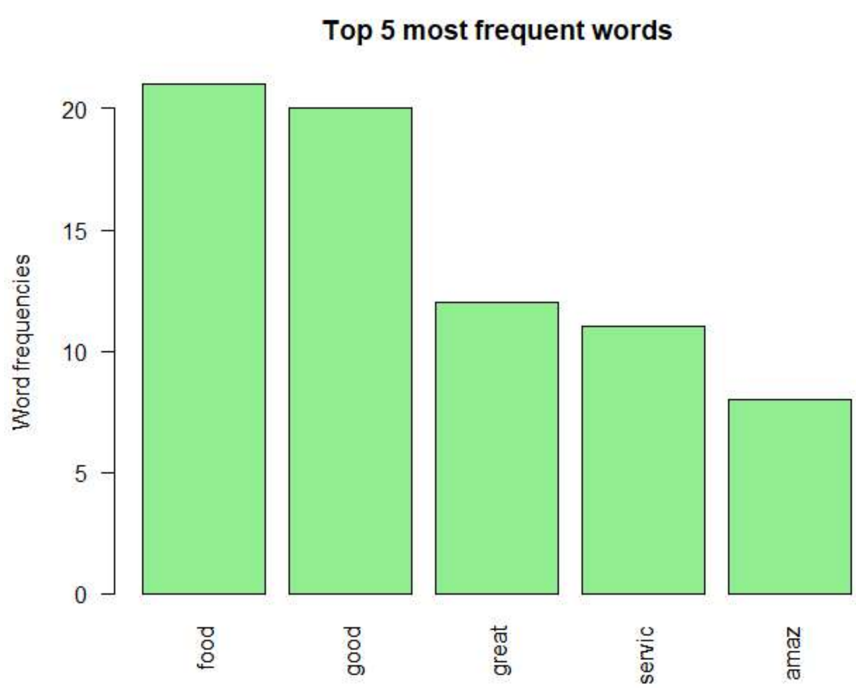
We start cleansing our data in by transforming our data using corpus function. Removing English stopwords, converting higher cases to lower cases alphabet, removing specific repeated word like “restaurant” and “review”, removing punctuations, numbers, spaces, and stemmed words by creating a stem document.

The data then converting into matrix table to identify top frequent word with count as show in Output[2].

```
> head(dtm_d, 5)
          word freq
food      food   21
good      good   20
great    great   12
servic  servic   11
amaz      amaz    8
```

Output[2]

Plot view of the top 5 most frequent terms is shown in Plot [1].



Plot [1]

With the use of wordcloud package, our top frequent terms can be visualized with size factor as shown in Figure [1].

Figure [1]

In our next step we will try find association of top 3 words as show in Output [3] and Output [4].

```
> findAssocs(reviewed_c_dtm, terms = c("good","food","great"), corlimit = 0.25)
$good
numeric(0)

$food
numeric(0)

$great
numeric(0)
```

Output [3]

```
> findAssocs(reviewed_c_dtm, terms = findFreqTerms(reviewed_c_dtm, lowfreq = 10), corlimit = 0.25)
$food
numeric(0)

$good
numeric(0)
```

Output [4]

Unfortunately, we dint find any association for terms “good”, “food” and “great”. Out frequency output shows the same found no evidence of association.

For finding, sentimental qualitative information of our text data, syuzhet_vector, bing_vector and afinn_vector methods are introduced as shown in output [5]. A matrix comparison vector is created as well.

```
> #regular sentiment score using get_sentiment() function
> # please note that different methods may have different scales
> syuzhet_vector <- get_sentiment(Reviews, method="syuzhet")
> # see the first row of the vector
> head(syuzhet_vector)
[1]  0.00  0.00  0.00 18.35 20.60
> # see summary statistics of the vector
> summary(syuzhet_vector)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00    0.00    0.00    7.79   18.35   20.60
> # bing
> bing_vector <- get_sentiment(Reviews, method="bing")
> head(bing_vector)
[1]    0    0    0    9 -319
> summary(bing_vector)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   -319       0       0     -62       0       9
> #affin
> afinn_vector <- get_sentiment(Reviews, method="afinn")
> head(afinn_vector)
[1]   0   0   0  29 -94
> summary(afinn_vector)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    -94       0       0     -13       0      29
> #compare the first row of each vector using sign function
> rbind(
+   sign(head(syuzhet_vector)),
+   sign(head(bing_vector)),
+   sign(head(afinn_vector))
+ )
     [,1] [,2] [,3] [,4] [,5]
[1,]    0    0    0    1    1
[2,]    0    0    0    1   -1
[3,]    0    0    0    1   -1
```

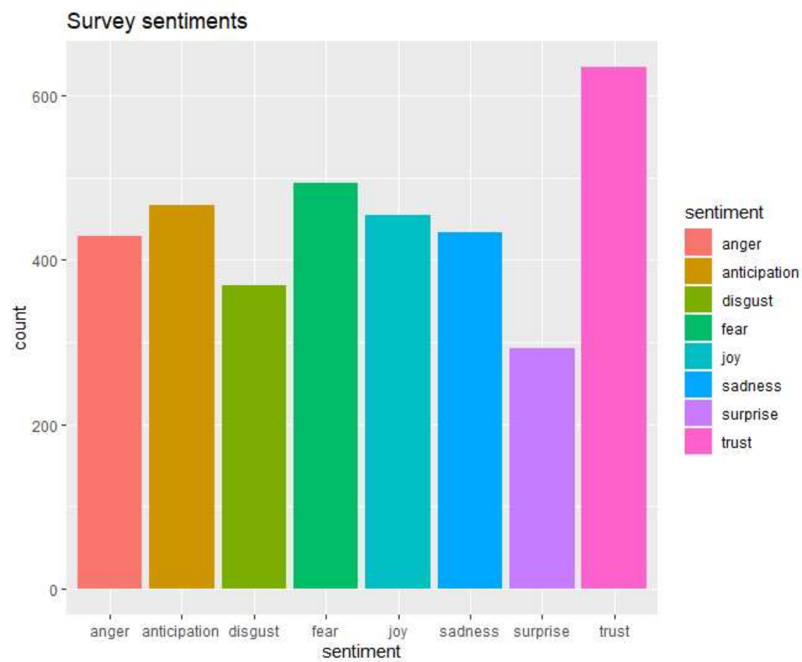
output [5]

Setiment nrc gives us scores of our qualitative information from our text data as show in output [6].

```
> d<-get_nrc_sentiment(Reviews)
> # head(d,10) - to see top 10 lines of the get_nrc_sentiment dataframe
> head (d,10)
  anger anticipation disgust fear joy sadness surprise trust negative positive
1     0            0       0    0   0       0        0     0        0        0
2     0            0       0    0   0       0        0     0        0        2
3     1            0       1    1   1       1        1     0        1        0
4     5           21       5   10  23       3       11    20       11       49
5   423          445     363  482 431     430      280   614     1155     1197
```

output [6]

We transform the data to visualize the overall sentimental survey results with a bar plot and determine the significance of each emotion from our text input as shown in Plot [3].



Plot 3

Appendix for Sentimental Analysis

```
# Load
library("magrittr") # needs to be run every time you start R and want to use %>%
library("dplyr")
library("tm")
library("SnowballC")
library("wordcloud")
library("RColorBrewer")
library("syuzhet")
library("ggplot2")

# Read the text file from local machine , choose file interactively
Reviews <- read.csv("C:/Users/Phani/Downloads/tourist_accommodation_reviews.csv", header= TRUE)
setwd("C:/Users/Phani/Downloads/Sentiment Analysis Data")
reviewed <- Reviews %>% slice_sample(n=30)
View(reviewed)
reviewed <- reviewed["Review"]
View(reviewed)
#Saving the random sample file for SAS EM usage
write.csv(reviewed,"C:/Users/Phani/Downloads/sentisample.csv", row.names = FALSE)

# Load the data as a corpus
reviewed_c <- Corpus(VectorSource(reviewed))

#Replacing "/", "@" and "|" with space
toSpace <- content_transformer(function (x , pattern ) gsub(pattern, " ", x))
reviewed_c <- tm_map(reviewed_c, toSpace, "/")
reviewed_c <- tm_map(reviewed_c, toSpace, "@")
reviewed_c <- tm_map(reviewed_c, toSpace, "\\|")
# Convert the text to lower case
reviewed_c <- tm_map(reviewed_c, content_transformer(tolower))
# Remove numbers
reviewed_c <- tm_map(reviewed_c, removeNumbers)
# Remove english common stopwords
reviewed_c <- tm_map(reviewed_c, removeWords, stopwords("english"))
# Remove your own stop word
# specify your custom stopwords as a character vector
reviewed_c <- tm_map(reviewed_c, removeWords, c("review", "restaurant"))
# Remove punctuations
reviewed_c <- tm_map(reviewed_c, removePunctuation)
# Eliminate extra white spaces
reviewed_c <- tm_map(reviewed_c, stripWhitespace)
# Text stemming - which reduces words to their root form
reviewed_c <- tm_map(reviewed_c, stemDocument)
# Build a term-document matrix
reviewed_c_dtm <- TermDocumentMatrix(reviewed_c)
dtm_m <- as.matrix(reviewed_c_dtm)
# Sort by descearing value of frequency
dtm_v <- sort(rowSums(dtm_m),decreasing=TRUE)
dtm_d <- data.frame(word = names(dtm_v),freq=dtm_v)
# Display the top 5 most frequent words
head(dtm_d, 5)
#Plot the most frequent words
barplot(dtm_d[1:5,]$freq, las = 2, names.arg = dtm_d[1:5,]$word,
        col ="lightgreen", main ="Top 5 most frequent words",
        ylab = "Word frequencies")
```

```
#generate word cloud
set.seed(1234)
wordcloud(words = dtm_d$word, freq = dtm_d$freq, min.freq = 5,
          max.words=100, random.order=FALSE, rot.per=0.40,
          colors=brewer.pal(8, "Dark2"))
findAssocs(reviewed_c_dtm, terms = c("good","food","great"), corlimit = 0.25)

findAssocs(reviewed_c_dtm, terms = findFreqTerms(reviewed_c_dtm, lowfreq = 10), corlimit = 0.25)

#regular sentiment score using get_sentiment() function
# please note that different methods may have different scales
syuzhet_vector <- get_sentiment(Reviews, method="syuzhet")
# see the first row of the vector
head(syuzhet_vector)
# see summary statistics of the vector
summary(syuzhet_vector)
# bing
bing_vector <- get_sentiment(Reviews, method="bing")
head(bing_vector)
summary(bing_vector)
#affin
afinn_vector <- get_sentiment(Reviews, method="afinn")
head(afinn_vector)
summary(afinn_vector)
#compare the first row of each vector using sign function
rbind(
  sign(head(syuzhet_vector)),
  sign(head(bing_vector)),
  sign(head(afinn_vector))
)
# run nrc sentiment analysis to return data frame with each row classified as one of the following
# emotions, rather than a score:
# anger, anticipation, disgust, fear, joy, sadness, surprise, trust
# It also counts the number of positive and negative emotions found in each row
Reviews <- as.character(Reviews)
d<-get_nrc_sentiment(Reviews)
# head(d,10) - to see top 10 lines of the get_nrc_sentiment dataframe
head (d,10)
td<-data.frame(t(d))
#The function rowSums computes column sums across rows for each level of a grouping variable.
td_new <- data.frame(rowSums(td[,]))
#Transformation and cleaning
names(td_new)[1] <- "count"
td_new <- cbind("sentiment" = rownames(td_new), td_new)
rownames(td_new) <- NULL
td_new2<-td_new[1:8,]
#Plot One - count of words associated with each sentiment
quickplot(sentiment, data=td_new2, weight=count, geom="bar", fill=sentiment, ylab="count")+ggtitle("Survey sentiments")
 #Plot two - count of words associated with each sentiment, expressed as a percentage
barplot(
  sort(colSums(prop.table(d[, 1:8]))),
  horiz = TRUE,
  cex.names = 0.7,
  las = 1,
  main = "Emotions in Text", xlab="Percentage")
```

Task 5: Shiny Dashboard implementation in R

Our Shiny Dashboard just display top 5 frequency of terms that are text mined from our data. Picture [1] shows our results.

Picture [1]

Appendix for Shiny Dashboard

```
ui <- dashboardPage(
dashboardHeader(title = "Sentimental Analysis"),
dashboardSidebar(),
dashboardBody(plotOutput("plot1", "plot2" width = 8)))
server <- function(input, output) {
output$plot1 <- renderPlot(barplot(dtm_d[1:5,]$freq, las = 2, names.arg = dtm_d[1:5,]$word,
col ="lightgreen", main ="Top 5 most frequent words",
ylab = "Word frequencies"))
output$plot2 <- renderPlot(quickplot(sentiment, data=td_new2, weight=count, geom="bar", fill=sentiment,
ylab="count")+ggtitle("Survey sentiments"))
}
shinyApp(ui, server)
```

B. Sentimental Analysis implementation on SAS EM

We start by importing the text file saved after sample slicing file in R program from localhost. Figure [2] shows our node connection used to achieve opinion mining in SAS EM.

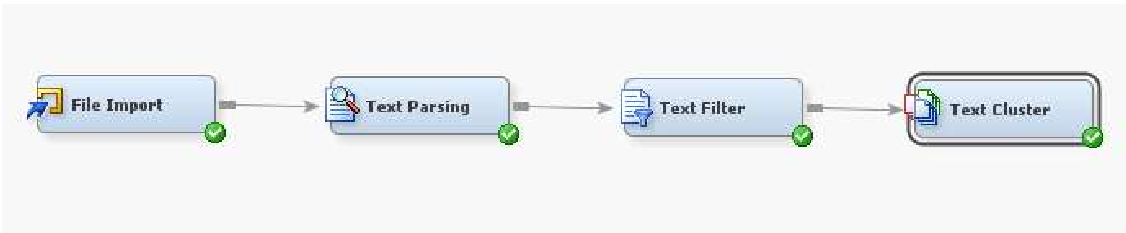


Figure [2]

After successful import of the text file, we connect to the Text Parsing node with custom selection of setting to Ignore Type of Attributes as show in Figure [3].

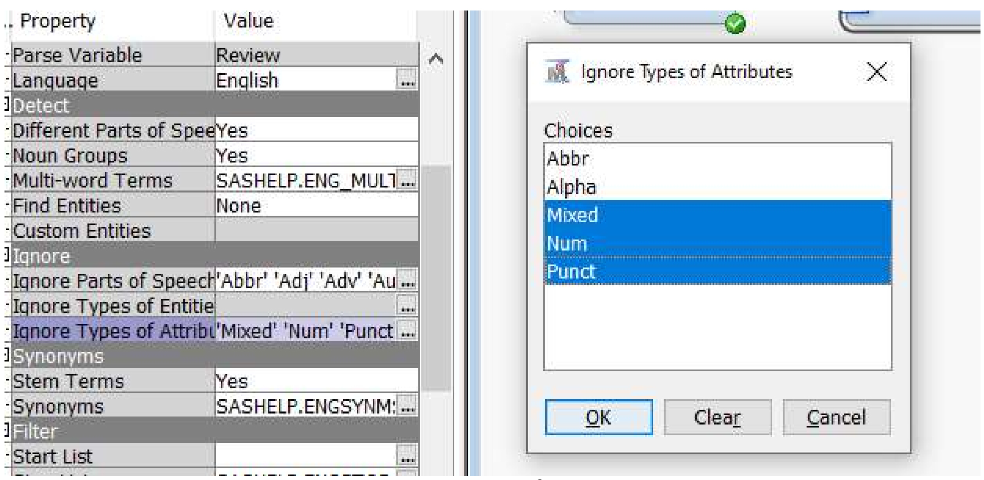
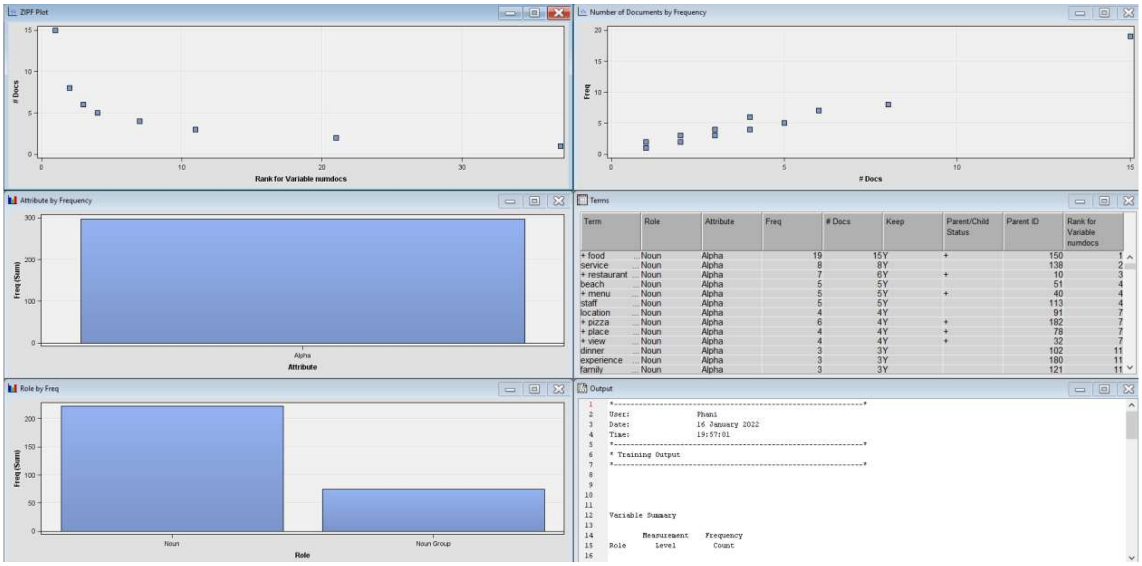


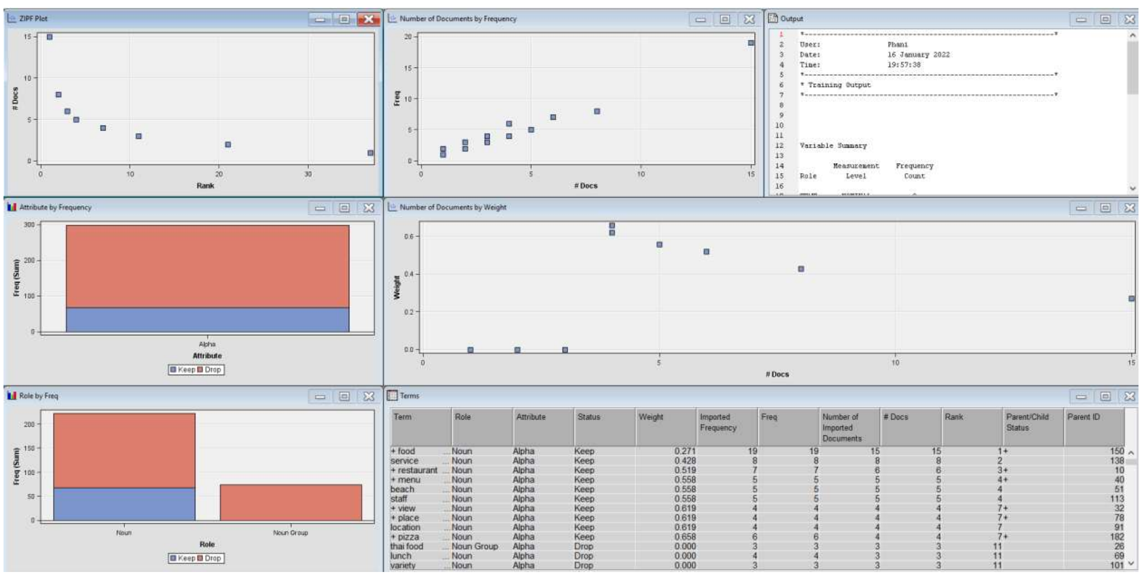
Figure [3]

Commencing text parsing yields us following results Output [7]. We have Rank of Variable numdocs, number of Documents by Frequency, Attribute by Frequency, Terms, Role by Freq, and overall Output. Term count in SAS EM output is evidently higher and proves deeper text mining than that in R.



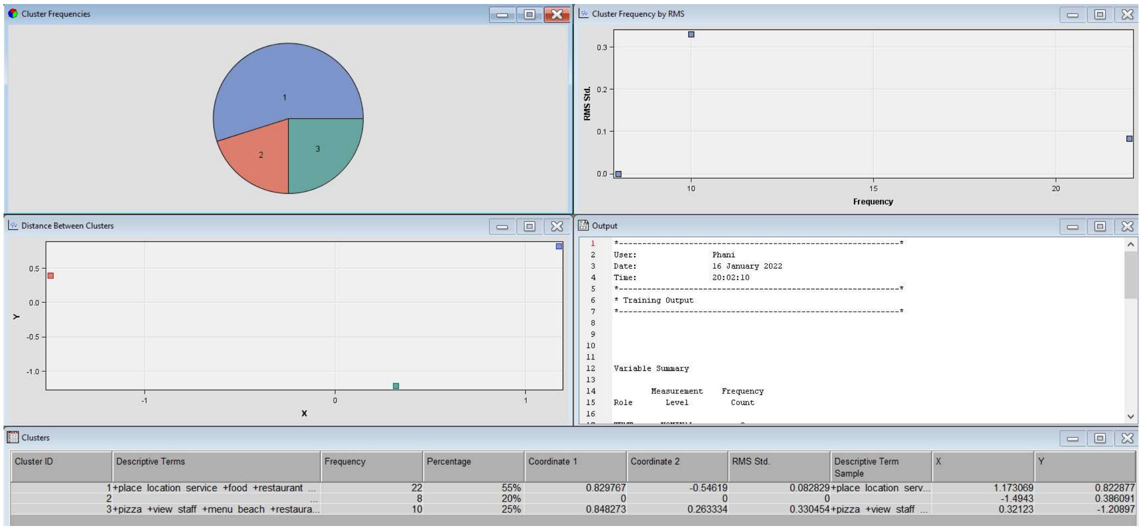
Output 7

Connecting to the next node Text Filter gives us the following Output 8. In Term results of text filter, we see actual filtering is taking place and status is displayed whether the computing keeps or drops the word from the output.



Output 8

By connecting Text Filter node to Text Cluster node, we have Output 9 computed and this gives Cluster Frequencies with 3 cluster partitions, by RMS as well, and Distance Between Clusters, all 3 Cluster's statistics like coordinates of clusters, RMS standard, location coordinates of clusters on X-Y plane.



Output 8

5. Conclusions and Comparison of R and SAS EM outputs.

Extracting sentiment from hotel review is successfully implemented in both SAS Enterprise Miner and R studio. Our R programming results are compared with SAS results in this section. When it comes to datamining and extracting sentiment from text R tends towards stricter rule and term creations in producing results and wouldn't be possible without Corpus functions. Plot 3 shows the extracted sentiment measures visually and our review text qualitative results are promising with higher percentage of positive term frequency which conclude, we have higher percentage of positive reviews.

SAS EM creates more terms and term frequency when compared to R. SAS also needs different techniques to be implemented to yield results. Observing all outputs from SAS tells us this has more statistical information than that of R such text clustering and distance of those clusters and co-ordinates on a X-Y plot plane. Descriptive terms are mined in SAS which is a pointer that SAS is notch higher in producing results. By now, we can easily draw a conclusion that SAS text mining is evidently superior in producing accurate and segmented output, But R can be used for a very strict mining and just mining sentiment from the text.

References

1. Cambria, E., Das, D., Bandyopadhyay, S. and Feraco, A. eds., 2017. A practical guide to sentiment analysis (pp. 1-196). Cham, Switzerland: Springer International Publishing.
2. Agarwal, Apoorv, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca J. Passonneau. "Sentiment analysis of twitter data." In Proceedings of the workshop on language in social media (LSM 2011), pp. 30-38. 2011.
3. Gonçalves, P., Araújo, M., Benevenuto, F. and Cha, M., 2013, October. Comparing and combining sentiment analysis methods. In Proceedings of the first ACM conference on Online social networks (pp. 27-38).
4. Medhat, W., Hassan, A. and Korashy, H., 2014. Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal, 5(4), pp.1093-1113.
5. Hutto, C. and Gilbert, E., 2014, May. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 8, No. 1).
6. Jain, V.K. and Kumar, S., 2017. Improving customer experience using sentiment analysis in e-commerce. In Handbook of Research on Intelligent Techniques and Modeling Applications in Marketing Analytics (pp. 216-224). IGI Global.