

# NLP based Question & Answers for E-commerce

**Presented by:-**  
Ashish Gore

**mentored By:-**  
Kartik  
Parth

# BUSINESS OBJECTIVE

In any website or in ecommerce business there are very few FAQ or limited user review. All the questions cannot be shown in website as website will look clumsy. Due to limited exposure of the questions, there is a possibility that user might not find question and the answer He/ She is searching for or the question is not present or never have been asked. Considering different scenarios, we are proposing a model which will take user question and will bring out relevant top five question and answer from Question Bank.

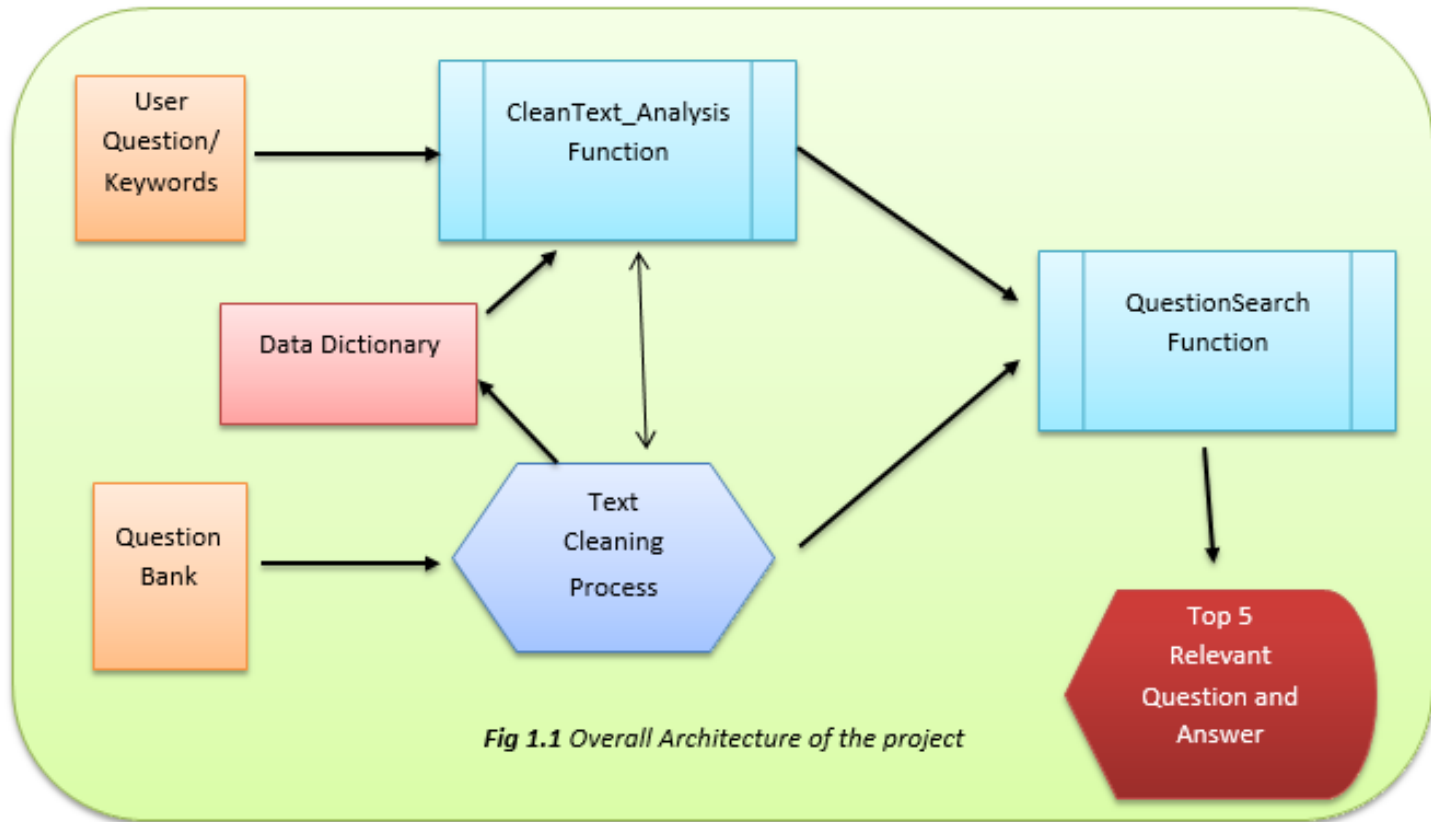
Our model ensures that

- User will get relevant answer for the question he/ she is searching for
- Website will not be clumsy
- Since all the process is happening at server side and only top 5 result is achieved, website will not be clumsy.

# PROBLEM STATMENT

An e-commerce company wants to build an algorithm to retrieve top 5 Question and answers based on the user given Keyword.

# Project Architecture



# DATA SET DETAILS

	questionType	asin	answerTime	unixTime	question	answerType	answer
1	yes/no	0594033926	Dec 27, 2013	1388131200	Is this cover the one that fits the old nook color? Which I bel...	Y	Yes this fits both the nook color and the same-shaped nook ...
2	yes/no	0594033926	Jan 5, 2015	1420444800	Does it fit Nook GlowLight?	N	No. The nook color or color tablet
3	open-ended	0594033926	2 days ago	NA	Would it fit Nook 1st Edition? 4.9in x 7.7in ?		I don't think so. The nook color is 5 x 8 so not sure anything...
4	yes/no	0594033926	17 days ago	NA	Will this fit a Nook Color that's 5 x 8?	Y	yes
5	yes/no	0594033926	Feb 10, 2015	1423555200	will this fit the Samsung Galaxy Tab 4 Nook 10.1	N	No, the tab is smaller than the 'color'
6	yes/no	0594033926	Jan 30, 2015	1422604800	does it have a flip stand?	N	No, there is no smaller than the pocket in the front flap. ...
7	yes/no	0594033926	Jan 30, 2015	1422604800	does this have a flip stand	?	Hi, no it doesn't 'color'
8	open-ended	0594033926	Dec 22, 2014	1419235200	also fits the HD+?		It should. They are the same size and the charging port is in ...
9	yes/no	0594033926	Nov 16, 2014	1416124800	Does it have 2 positions for the reader? Horizontal/vertical T...	Y	Yes
10	open-ended	0594033926	Aug 7, 2014	1407394800	Is there a closure mechanism? Bands, magnetic, etc.?		No- it is more like a normal book would be. It doesn't flop o...
11	open-ended	0972683275	Dec 3, 2013	1386057600	how far out does the arm extend?		18 inches on our TV.
12	yes/no	0972683275	Oct 27, 2013	1382857200	can you mount a sound bar using the external "stick" arms v...	Y	Yes, I can but need sound bar mount.
13	yes/no	0972683275	Nov 27, 2012	1354003200	My TV has a VESA 200x200 bolt pattern and I need it to ext...	Y	200mm x200mm. Yes Fully extended from wall to TV suppor...
14	yes/no	0972683275	Feb 1, 2015	1422777600	my vizio has 200 ht x 600 width mounting holes. will this mo...	?	I'm sorry mine is mounted already so I could not measure fo...

- Data is given in NDJson format
- NDJson file is converted in .csv file
- Data has 314263 rows and 7 columns
- Null values
  - In UnixTime column-9390
  - In answer type column-148665
  - In Answer column-32

# **EXPLORATORY DATA ANALYSIS (EDA) AND FEATURE ENGINEERING**

# EXPLORATORY DATA ANALYSIS (EDA)

## Data cleaning steps

- UnixTime column converted into standard date format,so dropped Answer column
- Null values in UnixTime column are replaced by median date
- Text Analysis on Question and answer column
  - Remove stopwords
  - Add QuestionWords and AnswerWords columns
  - Remove url's
  - Remove words with end .com
  - Remove any words which come after or before /
  - Remove numbers and punctuations
  - Remove single letter words, whitespace and certain word which were not removed by the stop words.
  - Do word stemming using the “textstem” package
  - text analysis is converted into a function : Function For Text Analysis.r

# FEATURE ENGINEERING

- Added new column called “QuestionTypeOE”
- correct the misspelled words using hunspell package
- some words are manually changed
- Create table for wrong and correct words
  - If any words are left to be corrected, they can be replaced using the Findreplace function in DataCombine package
- Add columns Question length, Question Word Count, Answer length, Answer Word Count.
- Sentiment analysis
  - Average sentiment and standard deviation of question and answer column.
  - Replacing the Na values with median
  - Create a column for sentiment words and there count

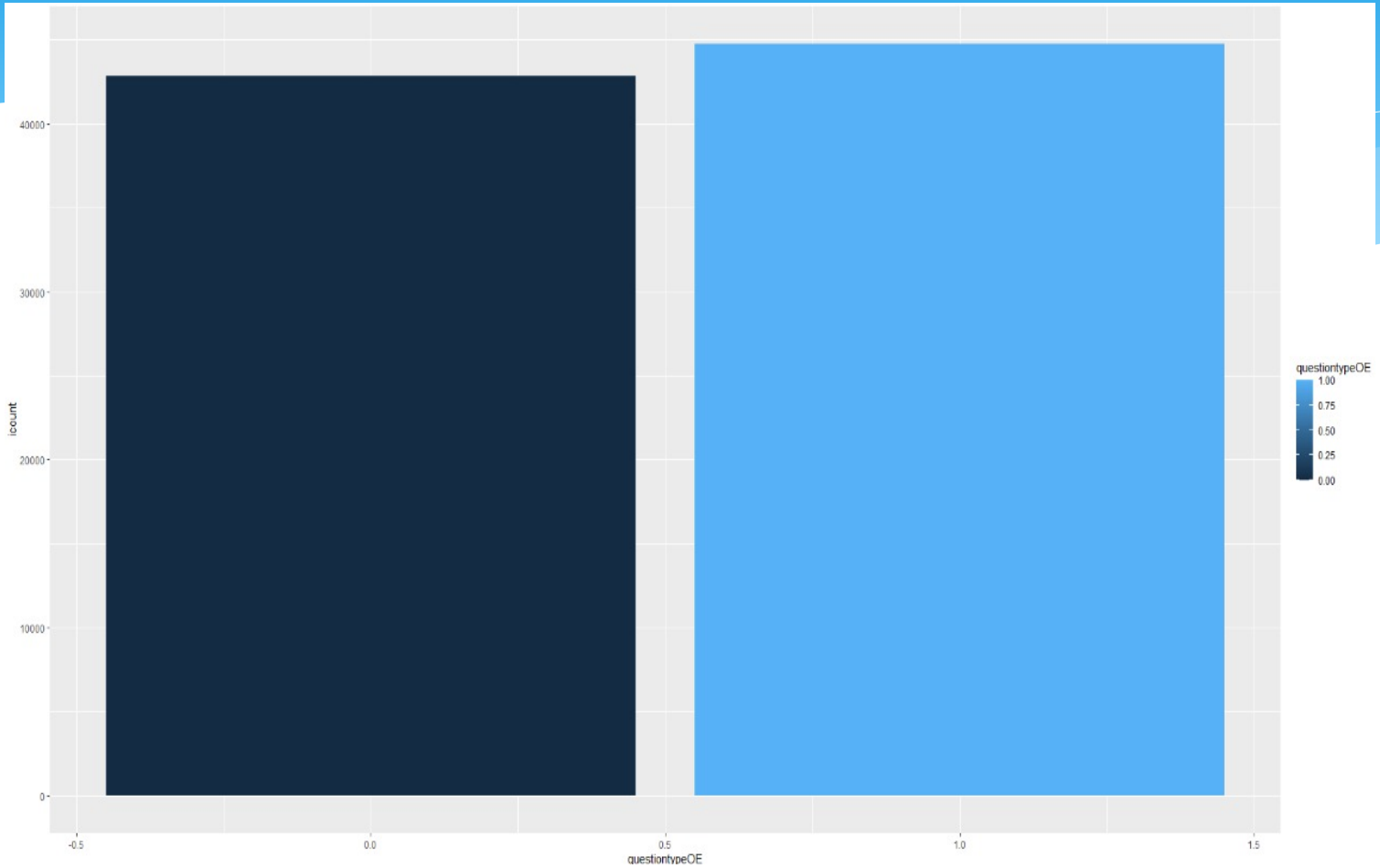
# Feature Engineering (continue)

- \* Change the variables of answer type : NO to N, YES to Y, NaN to NotApp
- \* Create a normalize function and normalizing all the columns
- \* Partition of data into train and test datasets on basis of AnswerType
- \* sampling using the Caret package
- \* Choosing the accurate model



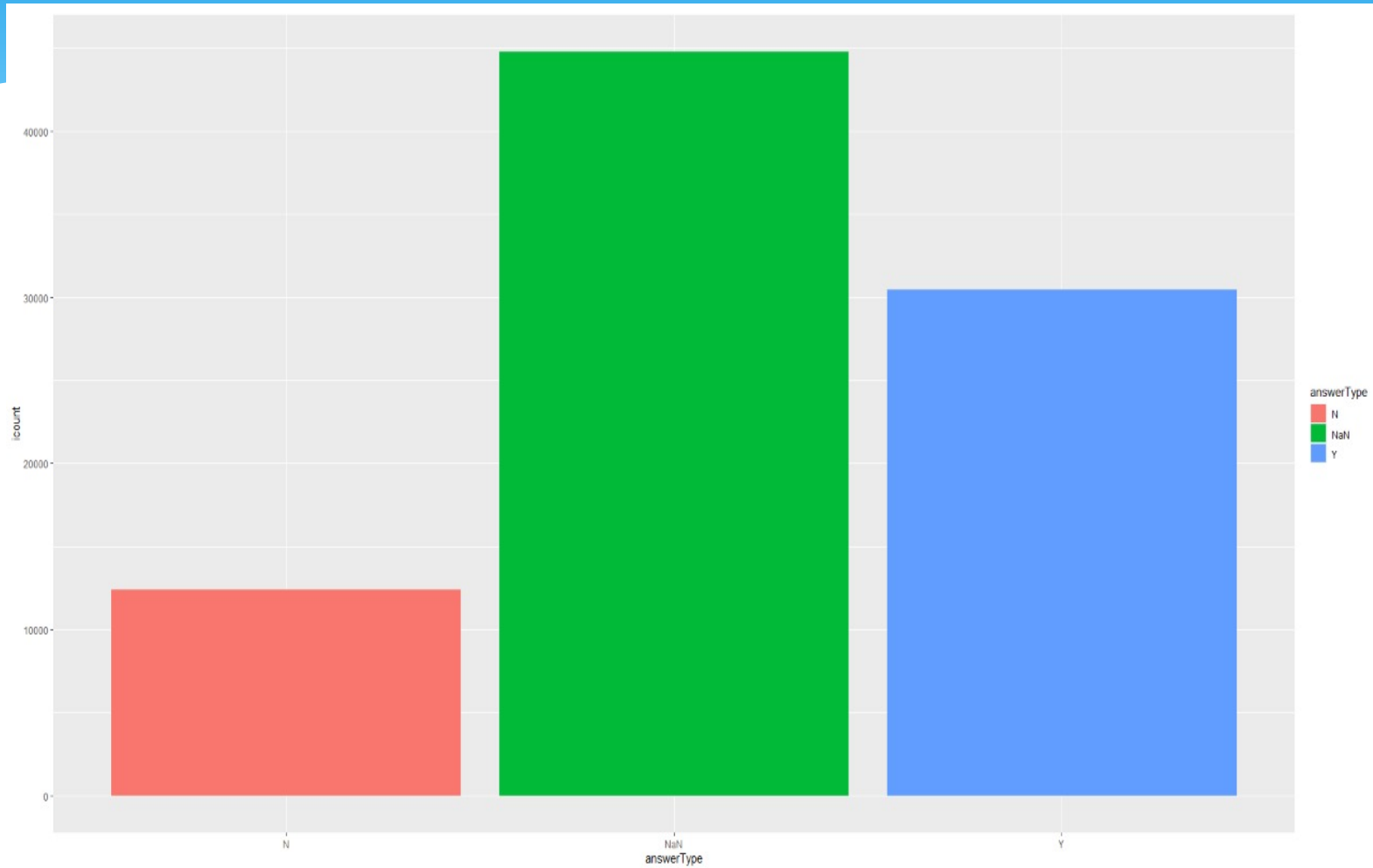
# VISUALIZATION

Plots for QuestionTypeOE Column



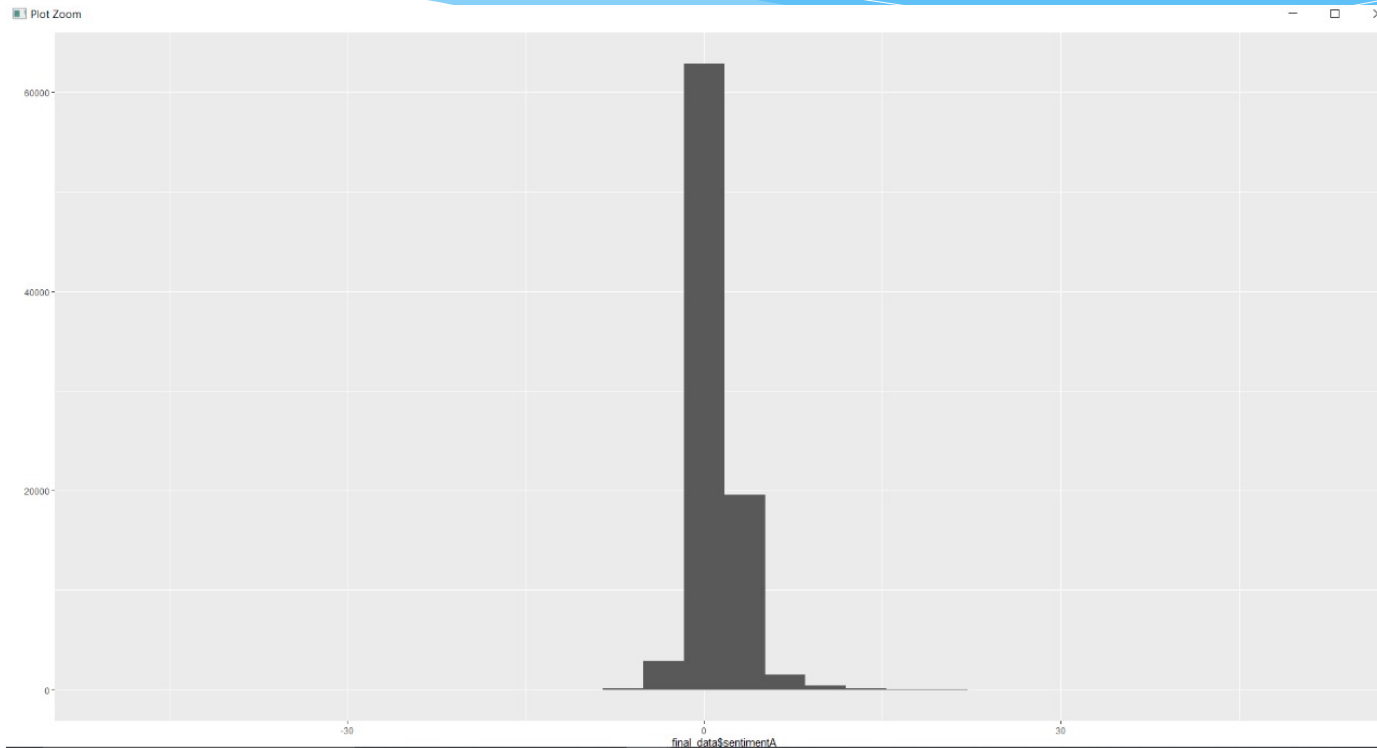
# VISUALIZATION

## Plots for AnswerType Column



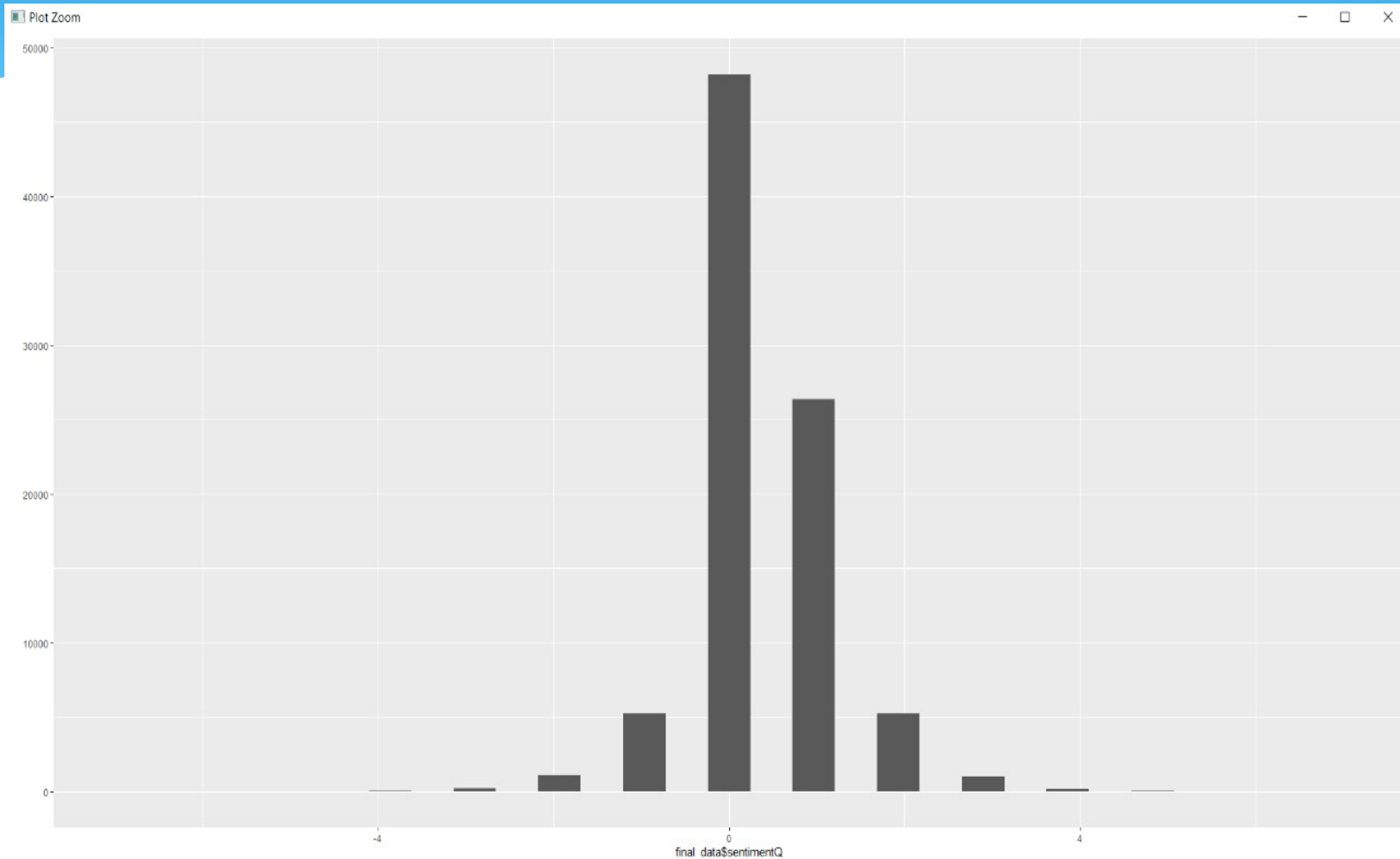
# VISUALIZATION (CONTINUE)

## Histogram Plots for sentimentA

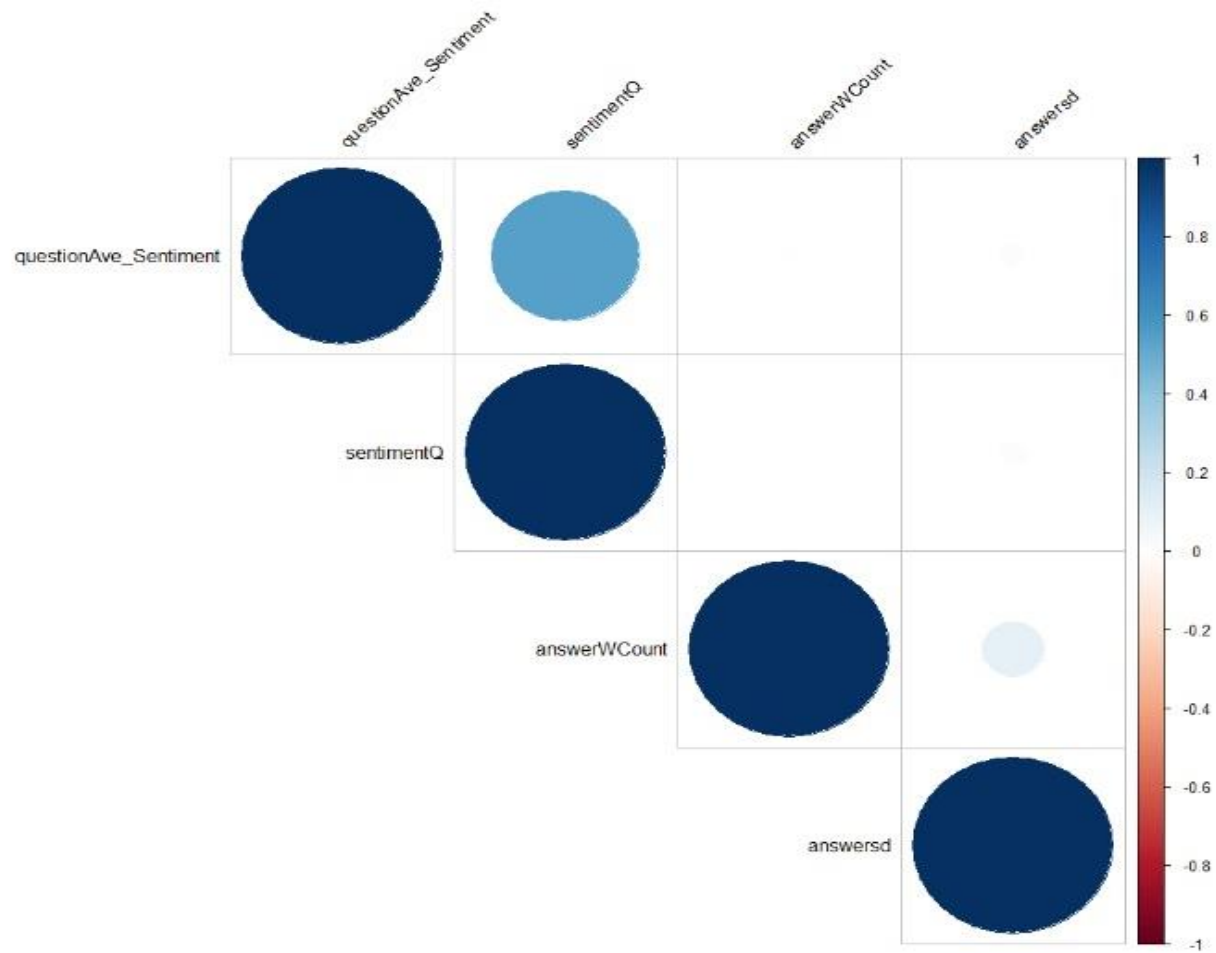


# VISUALIZATION (CONTINUE)

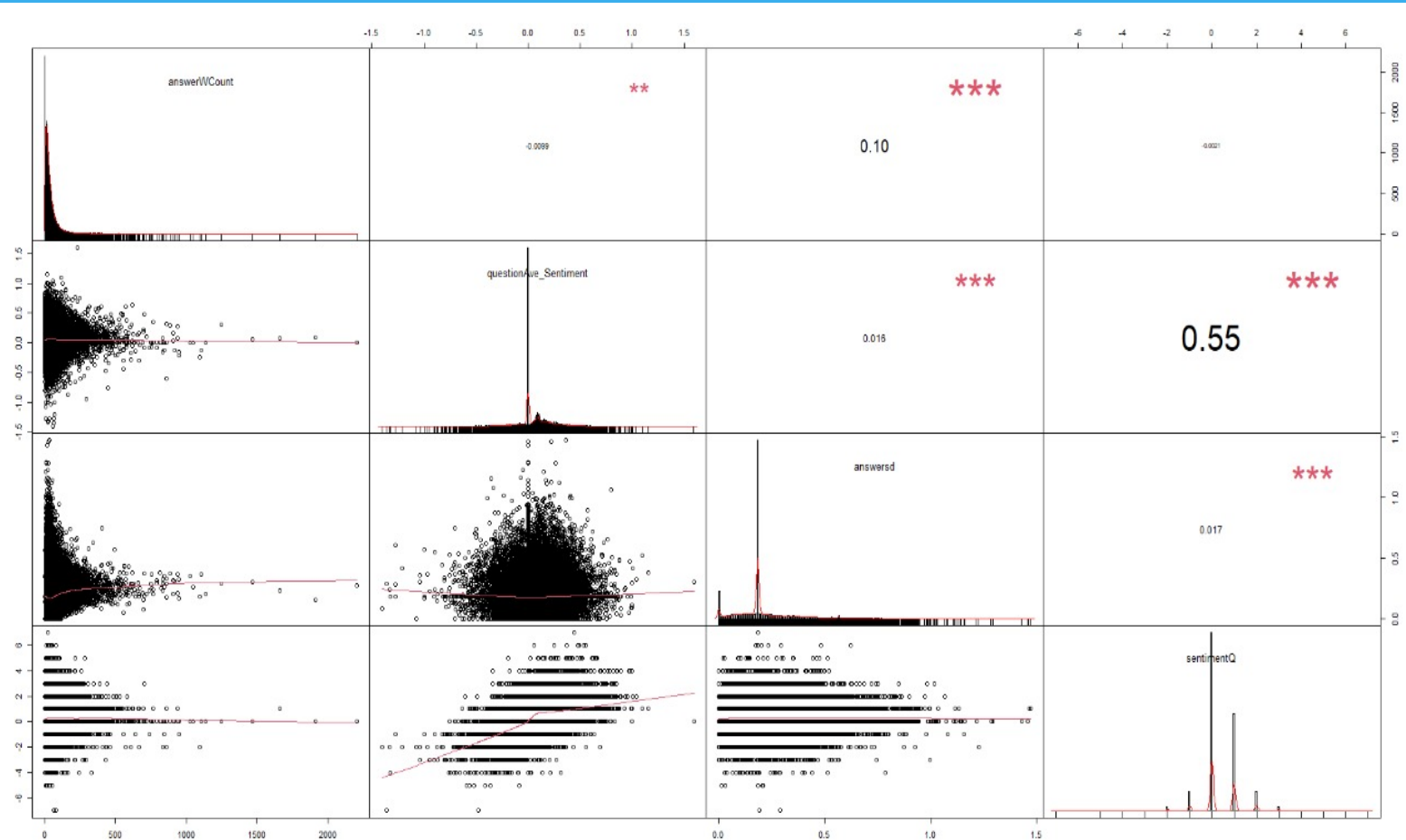
## Histogram Plots for sentimentQ



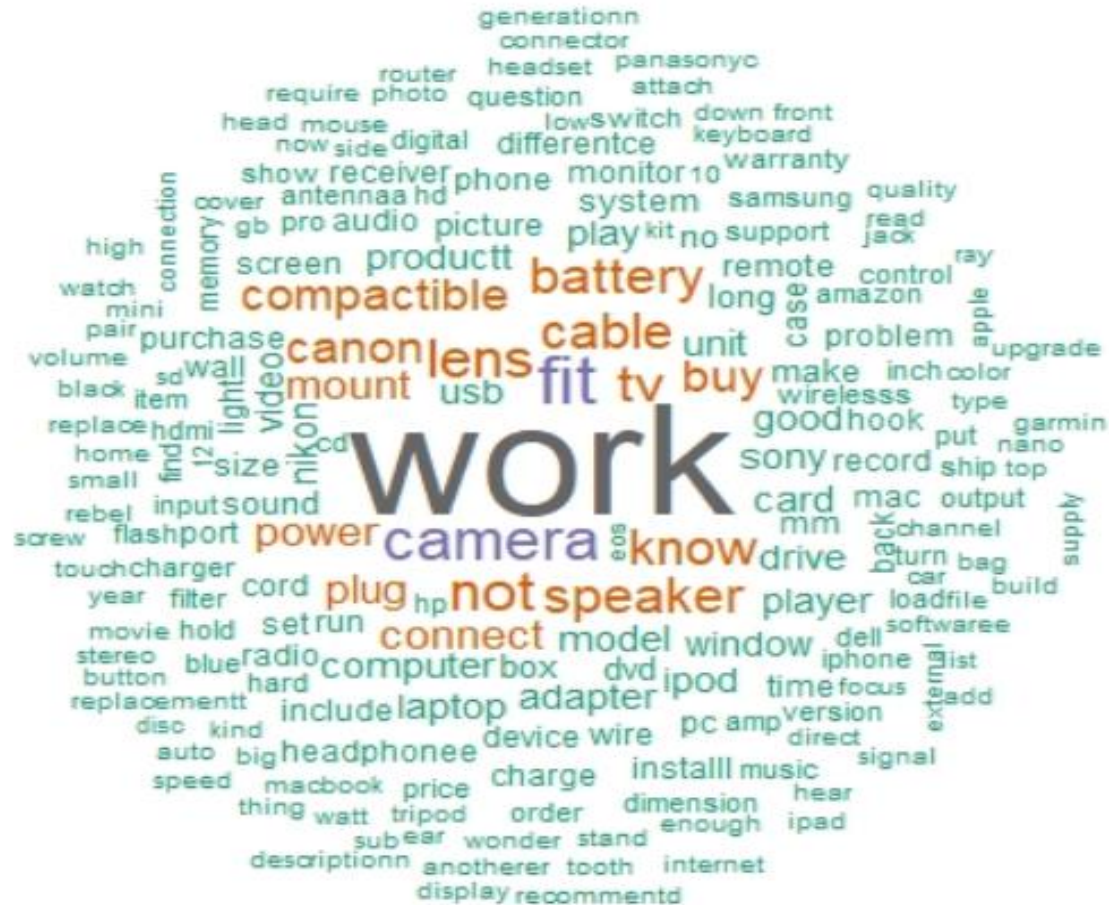
# VISUALIZATION (CONTINUE)



# VISUALIZATION (CONTINUE)



## Word Cloud for Question Column



# Final Data Set For Model

rownum	asin	unixTime	questiontypeOE	question	answer	questionWords
1	0594033926	2013-12-27	0	Is this cover the one that fits the old nook color? Which I bel...	Yes this fits both the nook color and the same-shaped nook ...	cover fit nook color
2	0594033926	2015-01-05	0	Does it fit Nook GlowLight?	No. The nook color or color tablet	fit nook glow light
3	0594033926	2014-12-29	1	Would it fit Nook 1st Edition? 4.9in x 7.7in ?	I don't think so. The nook color is 5 x 8 so not sure anything...	fit nook 1st edition 9in 7in
4	0594033926	2014-12-29	0	Will this fit a Nook Color that's 5 x 8?	yes	fit nook color
5	0594033926	2015-02-10	0	will this fit the Samsung Galaxy Tab 4 Nook 10.1	No, the tab is smaller than the 'color'	fit samsung galaxy tab nook 10

answ	QuestionLength	AnswerLength	questionWCount	answerWCount	questionAve_Sentiment	answerAve_Sentiment	questionsd	answersd
yes fit nook color shape nook ttablet	75	65	16	13	0.00000000	0.2218800785	0.000000000	0.183333300
no nook color color ttablet	27	34	5	7	0.17888544	0.000000000	0.102062100	0.000000000
nook color not small stay lock close	46	112	11	25	0.08911020	-0.3038635721	0.115470054	0.393750000
yes	40	3	11	1	0.14142136	0.800000000	0.102062100	0.183333300
no tab small color	48	39	11	8	0.14142136	0.000000000	0.102062100	0.183333300

sentimentQ	sentimentA	answerType
0	0	Y
0	0	N
0	0	NaN
0	0	Y
0	0	N



# MODEL BUILDING

# MODEL BUILDING

- Load the necessary libraries and the Final Data
- Load stop words to correct user input and removing yes and no from the list

```
stopwd <- read.table("stop.txt")
stopwd <- as.character(stopwd$V1)
```
- Load correct spellcheck for user input

```
correct3 = read.csv(correct3_WithNumbers.csv',
stringsAsFactors = FALSE)
```
- convert the Question Word column into list and get its length

```
QuestionList <- as.list(FinalData$questionWords)
QuestionLength <- length(QuestionList)
```
- define the function

```
QuestionSearch <- function(UserQuestion)
```
- Do a text analysis of the user question

```
UserQuestion <-
CleanText_Analysis(UserQuestion, correct3)
```

# MODEL BUILDING (CONTINUE)

- combine the user question and question list

```
TotalQuestions <- VectorSource(c(QuestionList,  
UserQuestion))
```

- Function will create a corpus and remove the white space and stem the document

```
QuestionCorpus <- VCorpus(TotalQuestions) %>%  
tm_map(stemDocument)  
tm_map(stripWhitespace)
```

- Create a term document matrix

```
term.doc.matrix.stm <- TermDocumentMatrix(QuestionCorpus,  
control=list( weighting=function(x)  
weightSMART(x,spec="ltc"),  
wordLengths=c(1,Inf)))
```

# MODEL BUILDING (CONTINUE)

- Convert it into Data frame using 'tidy' function and get grouped by the number of documents as per the term document matrix

```
term.doc.matrix <- tidy(term.doc.matrix.stm) %>%  
  group_by(document) %>%  
  mutate(vtrLen=sqrt(sum(count^2))) %>%  
  mutate(count=count/vtrLen) %>%  
  ungroup() %>%  
  select(term:count)
```

- \* Next term.doc.matrix mutate it, and filter it

```
docMatrix <- term.doc.matrix %>%  
  mutate(document=as.numeric(document)) %>%  
  filter(document<QuestionLength+1)  
QuestionMatrix <- term.doc.matrix %>%  
  mutate(document=as.numeric(document)) %>%  
  filter(document>=QuestionLength+1)
```

- \* The output will be in QuestionMatrix will now find terms of question and answers based on cosine similarity and return the top 5 question and answer in Wordsearch

# MODEL BUILDING (CONTINUE)

Here we can successfully retrieved top 5 Question and answers based on the user given Keyword

QuestionSearch("sony")

[1] "Question: Sony MDR-7506"

[1] "Answer: i do not know what the question is, but, the straps worked well enough to hold my manfrotto tripod and manfrotto ball head to the bottom of my manfrotto camera back pack, which the items that it was holding was about 5 lbs. i do hope this answers your questions, if not, then do please feel free to contact me a gain."

[1] "Question: Anyone tried this with the Sony HX50V?"

[1] "Answer: I use it with the Sony RX100 and it fits perfectly. If the specs are similar, it should work well. The RX100 doesn't have room to spare, though--it's a comfortable but snug fit. Hope this helps!"

[1] "Question: need help with my sony dw790: Is there something off here?"

[1] "Answer: Yup I HAD that problem as well when I was using RCA inputs but when I changed to Digital Optical, I found that the volume was way too loud and had to turn it down a lot just to fit viewing need. Well, the way I have my setup with the same HTIB(Home Theater in Box) is: PS3&gt;Reciever via Digital Optical for Audio PS3&gt;TV via Component(Plan to upgrade to HDMI once the cable comes) TV&gt;Reciever via RCA(Red and white plugs) For yours specifically Castellanos, I think we need a bit more info on the TV and the Toshiba, like what kind of Inputs and Outputs are on the back of both products or if not that way, give a link to see information on those items. Hopefully we can get this straighten out as with others in the same boat."

[1] "Question: A30 w/ a Sony KDL40V2500?"

[1] "Answer: TheA-30 is definitely a better bet. Just received one 2 weeks ago and no regrets at all. Even upconverts regular dvd's beautifully."

[1] "Question: Has anyone tried this on a Sony a99?"

[1] "Answer: No"

> |

# MODEL DEPLOYMENT

# MODEL DEPLOYMENT USING R SHINY

## Query Search

Enter your query



Search Results

# MODEL DEPLOYMENT USING R SHINY

User question with single word and punctuation

## Query Search

Enter your query

iphone 4s?

x

Q

## Search Results

Question 1 Can you use it with iPhone 4s?

Answer 1 Yes, should be able to use on any device with 1/4 stereo out (std head phone jack). I use on my andriod phone.

Question 2 can it use for iphone 4s

Answer 2 Yes It works with the iphone 4S

Question 3 can it use for iphone 4s

Answer 3 Yes it works with the iphone 4S

Question 4 Which one is for the iPhone 4S and which one is for the ipad??

Answer 4 The unit has two outputs. 2.1A and 1A, the 2.1A will charge faster, you can use it for any device. You should use the 2.1A for the iPad, or any tablet for that matter. I would only use the 1A if you need to charge two devices simultaneously.

Question 5 iphone 4s: is this compatible with the iPhone 4s?

Answer 5 There was a person saying it worked fine on theirs. i believe it was in the reviews shown under the product without expanding the reviews.



# MODEL DEPLOYMENT USING R SHINY

User question with incorrect spelling

## Query Search

Enter your query

xQ

## Search Results

Question 1	Will this cable work on my Sony RX10
Answer 1	This is a mechanical cable release. The Sony RX10 uses the Sony RM-VP1 wired remote, and some generic remotes with the same electrical connection.
Question 2	Will this work with my Sony A77?
Answer 2	Sony Alpha SLT-A77 Translucent Mirror Digital SLR Camera Yes! This will work with the Sony Alpha SLT-A77 Translucent Mirror Digital SLR Camera -. (I found this info on the Amazon.com page for this lens...so, if Amazon's info is accurate, this lens should work just fine.
Question 3	will this work with sonys f828
Answer 3	Yes that is the camera I have been using it on.
Question 4	will it work on a sony a58
Answer 4	Probably not. The lens linkage is specific to camera make. I would contact Sigma to see if they make a Sony configuration.
Question 5	will it work on a sony a58
Answer 5	Probably not. The lens linkage is specific to camera make. I would contact Sigma to see if they make a Sony configuration.

# CHALLENGES FACED?

- The data has missing values, has noise.
- Memory issue for loading data.
- Choosing right algorithm for model building and EDA.
- Extraction of correct search result as per the user defined keyword

# HOW DID YOU OVERCOME?

- Preprocessing data
- Memory issue we overcame by shortening the data
- Chosen algorithms by verifying which algorithm or parameters work best on the data.



**THANK YOU**