Analyzing Amazon's Customer Reviews using SAS® Text Miner for Devising Successful Product Launch Strategies

Manideep Mellachervu, Oklahoma State University;
Anvesh Reddy Minukuri, Comcast Corporation

ABSTRACT

Digital economy is showing a tremendous growth in the 21st century and it is having a massive impact in the current society. E-commerce is one element of the Internet of Things and its worldwide retail sales amounted to 2.3 trillion US dollars.[1] This shows the popularity of online shopping and it indicates an evolution of retailers in this industry. A recent study conducted by GE Capital Retail Bank has found that 81% of consumers perform online research before buying products.[2] This tells that consumers rely heavily on others' opinions and experiences in order to buy a product. Businesses need to understand customers' view of their products and competitors' products for strategic marketing. E-commerce businesses provide a platform to generate user-experience content through customer reviews, which are vital for a buyer to choose the best product out of numerous similar products available in the market. Companies need to analyze the customers' perspective through reviews for a better business, evaluate customer engagement, and devise strategies for the launch of their products. This paper focuses on analyzing the customer reviews primarily on Amazon using Python, SAS® Text Miner, SAS® Sentiment Analysis and SAS® Visual Studio. This project will determine which product features are given high-ratings/low-ratings, how the high-rating features of a best-selling product are performing compared to a similar product that is sold by a different vendor; and how to account for the customers' perception to product price of different brands while launching a similar new product.

INTRODUCTION

The rapid evolution of online market sale helps customers to find the good product, based on online reviews in e-commerce websites like Amazon. These reviews suggest to determine high and low rating features of any individual product and thereby provide access to compare high rating features of a bestselling product with that of other similar products available in the market. It is often important for any business to understand the customer opinion about a product. The reviews from e-commerce websites help companies to determine the customer engagement and plan strategies for future products to gain better business value. Textual data in the form of reviews from Amazon have been gathered to extract the features of a product.

For the purpose of analysis, five different Bluetooth speakers with a price range of $150-$200, each of which belongs to different vendors have been chosen. Reviews are obtained from the Amazon website for each of these products and these reviews are used for further analysis using SAS® Enterprise Miner.

The speakers which are used for analysis are:

- Beats Pill 2.0 Portable Speaker – Black

- JBL Charge 3 JBLCHARGE3BLKAM Waterproof Portable Bluetooth Speaker (Black)

- Ultimate Ears MEGABOOM Charcoal Wireless Mobile Bluetooth Speaker (Waterproof and Shockproof)

- Bose Sound Link Color Bluetooth Speaker (Black)

- Altec Lansing iM227 Orbit MP3 Speaker

DATA PREPARATION

Each product in Amazon is assigned to a unique 10 digits Hexadecimal ASIN code. ASIN is an acronym for Amazon Standard Identification Number. Reviews are collected by navigating to multiple pages from the product page. Each page consists of a maximum of 10 reviews. ASIN code and range of pages are used to extract reviews for all products using Python. The extracted amazon reviews for each product are stored in an Excel workbook. The excel data is then converted to a SAS data set and used for further

analysis. The variables included in the data set are Author, Comment Header, Posted Date, User Rating and Comment. But variables used for analysis are only User Rating and Comment.

DATA DICTIONARY

Below are the variables present in the data set after scrapping the data from Amazon.

Variable	Description	Datatype
Author	Name of the Reviewer	Nominal
Comment Header	Initial few words of comment	Text
Posted Date	Date when comment posted in Amazon	Ordinal
User Rating	Rating given by Reviewer on 1 to 5 scale	Interval
Comment	Full comment given by Reviewer	Text

Table 1. Variables Present in the Data Set

Sample extracted data can be seen below.

Author	Comment Header	Posted Date	User_rating	Comment
Genao	JBL Charge 3 along with Spotify's pla	14-Jun-16	5	JBL is in it, to win it. I own the JBL Xtreme, the Charge 2+ and now the Charge 3. If I
AlexF1	UE Boom 2 vs. JBL Charge 3	5-Feb-17	5	I wanted to get a portable travel speaker in addition to my home Sonos setup. Goog
Mike	I RECEIVED A COUNTERFEIT.	9-Jul-17	1	The box looked legit however upon opening, the inside box contents was crushed. I
Braeden D	The Only Charge 3 Review You Will E	17-Aug-17	5	While I am seasoned in writing Amazon reviews to help my fellow millennial shoppe
GP	Fantastic speaker!	26-Jun-16	5	This is a fantastic speaker. JBL has truly hit a home run with this one. I have a charg
Amazon C	Best in its Class	6-Jul-16	5	I did a lot more research that I probably should've before buying the JBL Charge 3, b
Blasphema	This speaker is a better deal than the	18-Dec-17	5	A bit bulky, but worth it. This speaker is a better deal than the UE Boom 2. Ive had b
MindSeed	This is my Baby	16-Nov-16	5	I'm here to help... I bought this at a Sprint store and let me give you my honest opin
Paul S	Fantastic Customer Service	12-Nov-17	5	The main reason I am writing this review is because of JBL/HKs fantastic customer s
PWH	Cheap quality.	21-Feb-18	1	Bought them as a gift exactly 90 days ago for a Christmas present for my girlfriend.
C-Low	Bose better step up their game...	31-Jul-17	5	I really like this speaker, but I have to say i wish it was louder. I was torn between th
John Whal	Using this to Watch Movies? Sound (	28-Dec-17	5	I've done exhaustive research here on Amazon and watched tons of review videos (
KCW	The JBL Charge 3 has a nice balanced	11-Aug-17	4	The JBL Charge 3 has a nice balanced sound to it. I cannot attest to whether the wa
Djtonydex	Engineering brilliance	4-Jan-18	5	First, a bit about the situation that this product is being used in. I am a Sanitation W
glenn	BUYERS BEWARE: NO RETURNS afte	17-Oct-17	1	Yes, only 1 star. Buyers beware: After 30 days, there are no returns. We bought 3 ur
Gavaronee	Another excellent product from JBL	19-Jan-18	5	I got this for my son for Christmas and I wonder who listens to it more, me or him. I
J.P.M.	It sounds very good. Not quite "grea	1-Nov-16	5	I just received this today. I hit the power button on the top, then the bluetooth butt
Annabelle	sounds great, better than I expected	16-Aug-16	5	If you're looking for a great speaker, buy this one. The speaker sounds great, better
Ryan	LOVE this speaker!	11-Aug-17	5	This speaker is wonderful! It is a breeze to connect, and I'm no audiophile, but I thin
Warren	Great, Loud speaker for constructior	6-Jul-18	5	I recently switched tool brands as I work in the HVAC trade and I didn't want a tool
The Nelsoi	Would be great but...	26-Feb-18	3	Sounds awesome. Deep bass, good mid/hi. Used in my garage, at the neighborhood
Nick	Absolutely incredible speaker. Large	25-Jul-18	5	SOUND: I bought two and I can't believe how awesome they sound. I had the origin
S. Davis	Wonderful speaker!!!	1-Jun-18	5	Used my Christmas money to purchase this after reading many reviews... Decided o
C. Rogers	Werks pretty guhd. Heavy bass for al	12-Jun-18	4	Werks pretty guhd. Heavy bass for all you base heads. It does distort for most songs

Figure 1. Partial Data Set Used for Analysis

METHODOLOGY

The SAS data set for each product is then imported into SAS Enterprise Miner 14.2, which is then partitioned into two data sets using the filter node, one for 1 & 2 ratings and other for 4 & 5 ratings. Here in the current paper, analysis outputs are displayed only for Beats speakers. Similarly, analysis for speakers of other 4 brands can be interpreted.

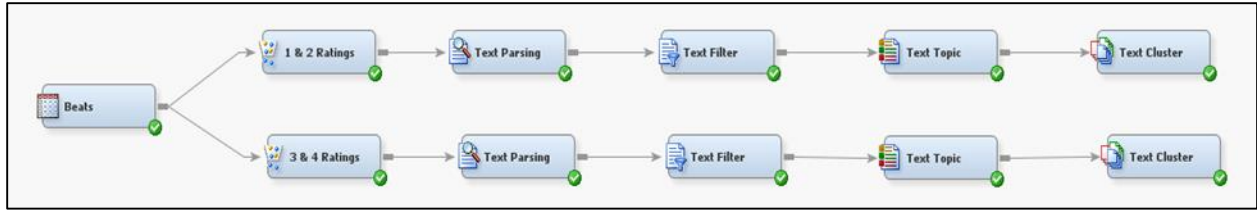The following figure shows the methodology used for the analysis:

Figure 2. Node Diagram to Generate Clusters

DATA SETS

SAS data sets have been created for each of the products and above methodology is applied individually for all the products. The data sets used in this analysis are:

- Beats.sas7bdat
- JBL.sas7bdat
- UltimateEars.sas7bdat
- Bose.sas7bdat
- AltecLansing.sas7bdat

TEXT PARSING

The text parsing node was connected to the data node and default settings are modified using properties panel before running the node. Following parameters are used in properties panel:

- The 'detect different parts of speech option' is set to 'yes' to be able to treat the same words of different parts of speech as different.
- 'Abbr', 'Aux', 'Conj', 'Det', 'Interj', 'Num', 'Part', 'Prep', 'Pron', 'Prop' parts of speech have been ignored.
- 'Num', 'Punct' types of attributes have been ignored.
- Then each of these data sets is attached to text parse node which is used for ignoring selected parts of speech. Here Abbr, Aux, Conj, Det, Interj, Num, Part, Prep, Pron, Prop were ignored.

The text parsing node generated a term by document matrix helps to identify the most frequently occurring words and the number of comments in which each word occurred. Figure 3. displays partial term by document matrix for Beats speakers having 1 & 2 ratings.

Term	Role	Attribute	Freq	# Docs	Keep	Parent/Child Status	Parent ID	Rank for Variable numdocs
not	...Adv	Alpha	361	202	N		2820	1
+ be	...Verb	Alpha	410	166	N	+	2867	2
+ do	...Verb	Alpha	159	111	N	+	2902	3
+ speaker	...Noun	Alpha	186	104	Y	+	794	4
+ have	...Verb	Alpha	178	93	N	+	2812	5
+ charge	...Verb	Alpha	88	73	Y	+	824	6
+ get	...Verb	Alpha	88	67	N	+	2779	7
+ product	...Noun	Alpha	104	66	Y	+	697	8
+ work	...Verb	Alpha	83	66	Y	+	1769	8
+ buy	...Verb	Alpha	87	60	Y	+	82	10
+ pill	...Noun	Alpha	91	60	Y	+	1485	10
+ sound	...Noun	Alpha	80	58	Y	+	1203	12
+ charge	...Noun	Alpha	58	54	Y	+	1948	13
+ quality	...Noun	Alpha	70	50	Y	+	426	14
no	...Adv	Alpha	60	47	N		2917	15
s	...Noun	Alpha	62	47	N		2778	15
very	...Adv	Alpha	51	47	N		2800	15
+ return	...Verb	Alpha	47	44	Y	+	156	18
+ plug	...Verb	Alpha	45	39	Y	+	1741	19
just	...Adv	Alpha	45	38	N		2799	20
+ month	...Noun	Alpha	42	37	Y	+	557	21
+ use	...Verb	Alpha	45	37	N		2814	21
even	...Adv	Alpha	38	36	N		2824	23
+ price	...Noun	Alpha	50	36	Y	+	2005	23
battery	...Noun	Alpha	40	35	Y		1929	25
+ beat	...Verb	Alpha	44	34	Y	+	990	26
sound	...Adj	Alpha	44	34	Y		926	26
+ beat	...Noun	Alpha	44	33	Y	+	34	28
better	...Adj	Alpha	45	33	Y		1782	28
now	...Adv	Alpha	36	33	N		2929	28
+ play	...Verb	Alpha	36	32	Y	+	120	31
+ time	...Noun	Alpha	36	32	Y	+	1039	31
+ purchase	...Verb	Alpha	39	30	Y	+	531	33

Figure 3. Text Parsing Results for Beats Speakers' Data Having 1 & 2 Ratings

Some of the most commonly used words by reviewers in the comments are speaker, not, charge, sound, quality, which is expected as these words relate to the features of speakers.

TEXT FILTER

Further, parsing node is then connected to text filter node as shown in figure 4. This node filters out words that occur least number of times specified in the properties panel. Following parameters are changed from default settings in properties panel:

- The minimum number of documents is set to 4.

- Check spelling option is set to 'yes', which enables SAS to create correctly spelled words in place of misspelled words.

	TERM ▲	FREQ	# DOCS	KEEP	WEIGHT	ROLE	ATTRIBUTE
⊟	button	11	5	☑	0.738	Noun	Alpha
	bultton	1	1			Noun	Alpha
	button	1	1			Verb	Alpha
	button	5	3			Noun	Alpha
	buttons	4	3			Noun	Alpha
⊟	buy	87	60	☑	0.315	Verb	Alpha
	buying	13	13			Verb	Alpha
	buy	37	30			Verb	Alpha
	bought	37	33			Verb	Alpha
	buy	12	11	☑	0.586	Noun	Alpha
	buy electronics	2	2	☐	0.0	Noun Group	Alpha
⊟	buyer	3	3	☐	0.0	Noun	Alpha
	buyer	1	1			Noun	Alpha
	buyers	2	2			Noun	Alpha
⊟	buzz	2	2	☐	0.0	Verb	Alpha
	buzz	1	1			Verb	Alpha
	buzzing	1	1			Verb	Alpha
⊞	buzz sound	1	1	☐	0.0	Noun Group	Alpha
	by itself	2	2	☐	0.0	Adv	Alpha
⊟	cable	8	6	☑	0.709	Noun	Alpha
	cables	2	2			Noun	Alpha
	cable	6	4			Noun	Alpha
⊞	cable wire	1	1	☐	0.0	Noun Group	Alpha
	cake	1	1	☐	0.0	Verb	Alpha
⊟	call	10	7	☐	0.0	Verb	Alpha
	calling	1	1			Verb	Alpha
	called	7	5			Verb	Alpha
	call	2	2			Verb	Alpha
	call	1	1	☐	0.0	Noun	Alpha
	call company	1	1	☐	0.0	Noun Group	Alpha
⊞	can	8	7	☐	0.0	Noun	Alpha
⊟	cancel	1	1	☐	0.0	Verb	Alpha
	cancelled	1	1			Verb	Alpha
	cancel	0	0			Verb	Alpha

Figure 4. Grouping of Word Forms and Misspelled Words in Beats Speakers' Data Having 1 & 2 Ratings

Figure 4 from interactive filter viewer shows various forms of words 'button', 'buy', 'cable', 'call', which are formed into groups along with misspelled words with the help of English dictionary.

Concept Links

Concept links can be viewed in the interactive filter viewer from the properties panel of text filter node. Concept links are the type of association analysis between the terms used. They can be created for all the terms that are present in the comments, however, it is meaningful to create links for only a few important terms. Concept link diagram shows the word to be analyzed at the center and the words which

are associated with that word are connected to it using links. The thickness of the link explains the strength of association between the two words in reviewer comments. Below are the Concept links for some of the most frequent terms:

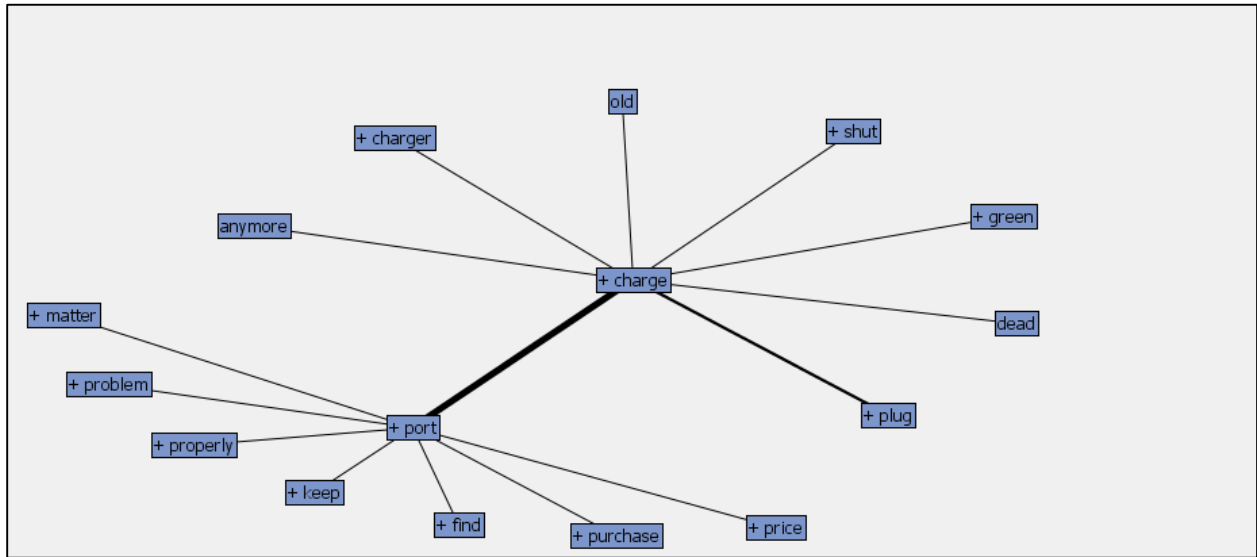***Concept Links for words in Beats Speakers' data having 1 & 2 Ratings***

Figure 5. Concept Link Diagram for Word CHARGE

Concept link diagram in figure 5 shows word CHARGE is highly associated with word PORT which is associated with word PROBLEM. Hence, one of the reasons for customer dissatisfaction could be charging port issues.
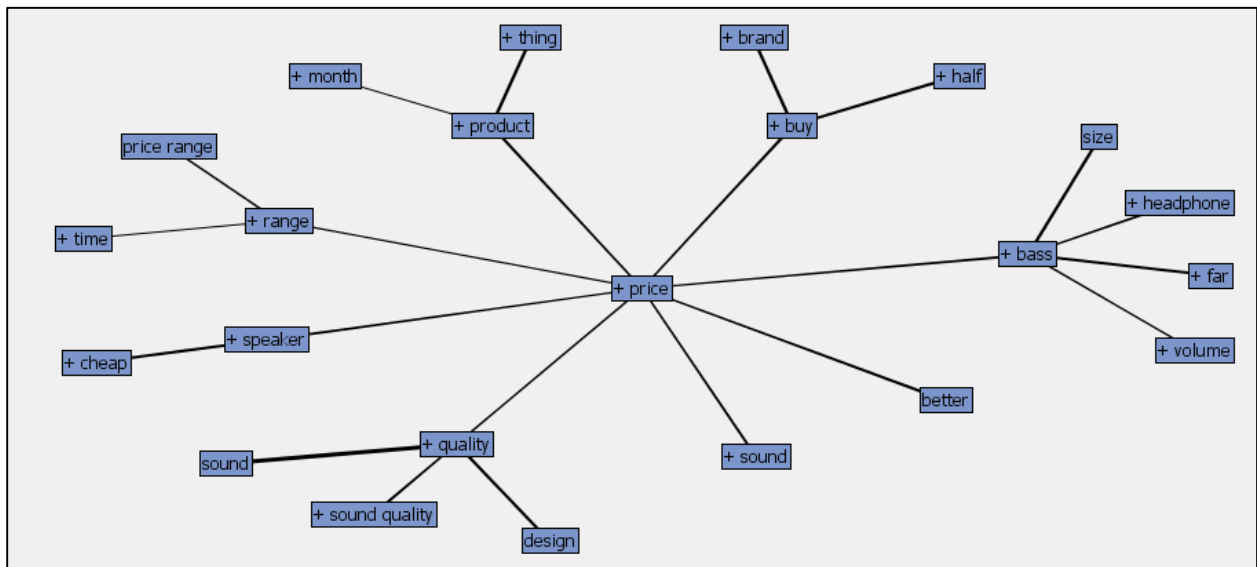

Figure 6. Concept Link Diagram for Word PRICE

Concept link diagram in figure 6 shows word PRICE is highly associated with words QUALITY, BASS, SPEAKER, SOUND which suggests that customers relate these parameters to the price of the product.

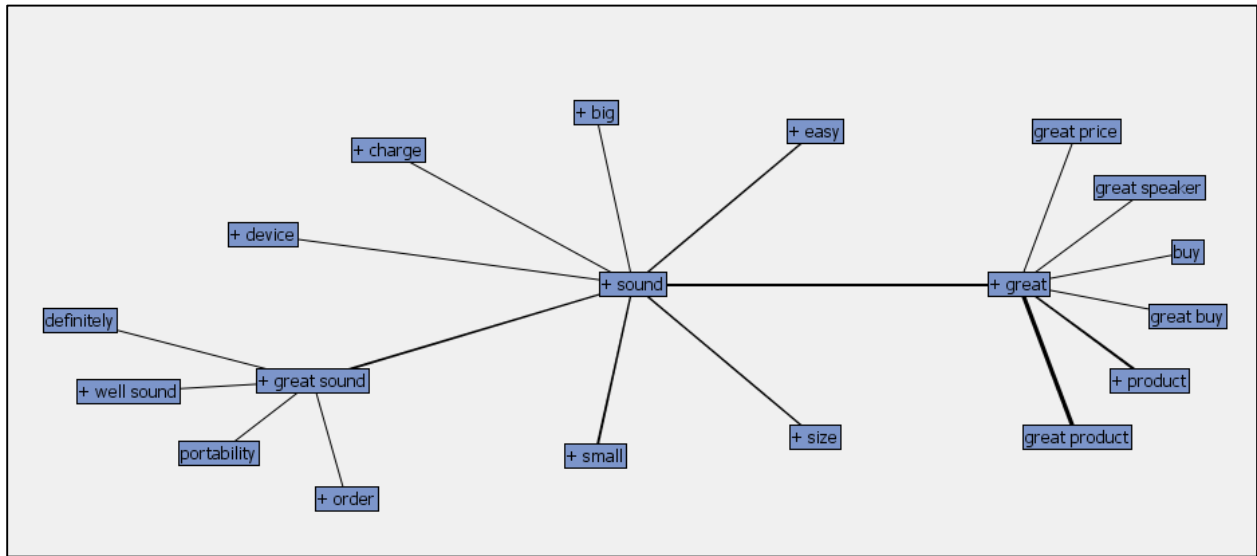*Concept Links for words in Beats Speakers' data having 4 & 5 Ratings*

Figure 7. Concept Link Diagram for Word SOUND

Concept link diagram in figure 7 shows word SOUND is strongly associated with word GREAT, indicates that Beats speaker is a great product with the good sound quality feature.
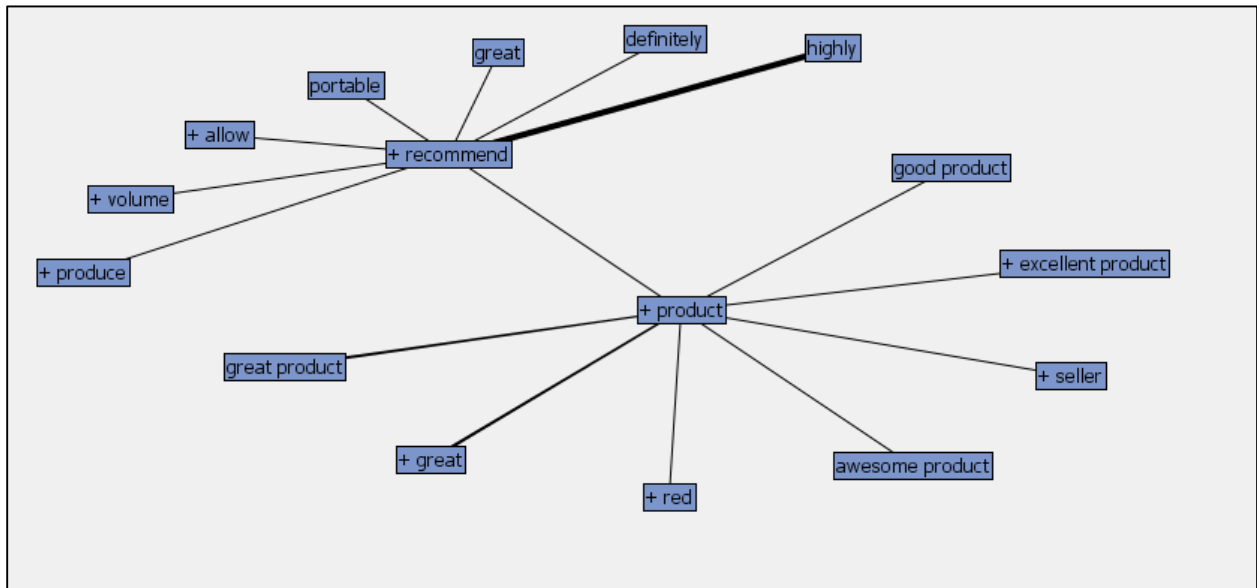

Figure 8. Concept Link Diagram for Word PRODUCT

Concept link diagram in figure 8 shows word PRODUCT is highly associated with words GREAT, RECOMMENDED, EXCELLENT, AWESOME. It tells that Beats speaker is a good product and reviewers highly recommend this product to other customers.

TEXT TOPIC

After connecting the Text Filter node in SAS Enterprise Miner to Text Parsing node, Text topic node is joined to the Text Filter node, which enables SAS to combine terms into topics for obtaining valuable insights from data. The number of Multi-Term Topics has been set to 20 in properties panel to understand data and get the features that reviewers are more interested to comment about the products.

Category	Topic ID	Document Cutoff	Term Cutoff	Topic	Number of Terms	# Docs
Multiple	1	0.148	0.073	+stop,+working,+month,+week,+buy	13	24
Multiple	2	0.152	0.076	sound,+sound quality,+quality,close to,poor	15	30
Multiple	3	0.132	0.078	+battery,+year,+less,+minute,at all	34	39
Multiple	4	0.149	0.078	back,+box,+solution,+buy,+problem	42	18
Multiple	5	0.136	0.078	+charge,+long,+last,+month,+long charge	29	35
Multiple	6	0.152	0.079	better,+price,+bass,+sound,+speaker	34	38
Multiple	7	0.153	0.077	+order,+cord,+return,+work,+power	27	36
Multiple	8	0.136	0.078	+blow,+purchase,+product,+volume,+speaker	36	36
Multiple	9	0.147	0.078	+defective,+product,happy,+buy,christmas	33	41
Multiple	10	0.121	0.078	+order,+charger,music,+pill,+thing	32	25
Multiple	11	0.128	0.079	+thing,+phone,+connect,+work,+month	36	37
Multiple	12	0.121	0.079	+beat,+piece,crap,terrible,+volume	33	34
Multiple	13	0.127	0.078	+beat,+work,+pill,+fake,+pill	24	35
Multiple	14	0.137	0.078	money,+waste,+year,old,terrible	27	32
Multiple	15	0.137	0.077	+charge,+charger,+purchase,+port,+disappoint	25	48
Multiple	16	0.135	0.077	+break,+item,+return,+purchase,+arrive	27	32
Multiple	17	0.116	0.079	+turn,+pill,+red,+light,+green	31	26
Multiple	18	0.113	0.079	+sound,+work,+easy,+bass,+unit	37	18
Multiple	19	0.120	0.078	+plug,turn on,+unit,+charge,+thing	30	38
Multiple	20	0.140	0.079	+play,music,+hold,+look,+charge	33	38

Figure 9. Text Topic Node Output for Beats Speakers' Data Having 1 & 2 Ratings

Figure 9 shows 20 different topics with corresponding IDs. Topic 1 states that product stops working within few weeks of buying whereas Topic 2 suggests that sound quality of Beats speakers is poor but this topic is of less intense than Topic 3 which explains that battery discharges quickly. Similarly, the other topics can be interpreted. From topic 6, the other major concern that reviewers express is regarding price which could be better.

Category	Topic ID	Document Cutoff	Term Cutoff	Topic	Number of Terms	# Docs
Multiple	1	0.149	0.076	+gift,christmas,+purchase,+son,item	15	65
Multiple	2	0.171	0.074	+product,great product,+great,+seller,+recommend	11	58
Multiple	3	0.119	0.082	+pill,+music,+device,+phone,+price	31	81
Multiple	4	0.186	0.073	+great sound,+great,+sound,+easy,+size	5	55
Multiple	5	0.157	0.070	great,+work,+sound,great,+little	4	52
Multiple	6	0.153	0.070	+awesome,awesome sound,awesome product,+sound...	7	50
Multiple	7	0.133	0.071	+nice,item,+loud,+easy,+speaker	7	35
Multiple	8	0.133	0.071	good,+sound,bass,+look,+arrive	5	36
Multiple	9	0.138	0.075	+good,good sound,good product,good quality,+sound	12	74
Multiple	10	0.119	0.066	ðÿ,+recommend,highly,+work,+pay	3	9
Multiple	11	0.147	0.077	+quality,sound,sound quality,great,+recommend	15	68
Multiple	12	0.132	0.073	+love,daughter,+daughter love,+buy,christmas	7	47
Multiple	13	0.130	0.079	+buy,+happy,+purchase,+year,+old	19	80
Multiple	14	0.141	0.079	+speaker,+small,small speaker,+loud,daughter	18	64
Multiple	15	0.128	0.078	+amaze,+loud,+pill,+sound,+beat	18	55
Multiple	16	0.165	0.072	+love,daughter,christmas,+good,great	7	117
Multiple	17	0.127	0.075	buy,great buy,+great,+loud,+clear	16	24
Multiple	18	0.114	0.072	+perfect,item,+son,+purchase,beach	10	24
Multiple	19	0.136	0.078	+speaker,+son,+little,+powerful,+love	11	94
Multiple	20	0.117	0.077	+excellent,+sound,+product,+excellent product,+purch...	19	27

Figure 10. Text Topic Node Output for Beats Speakers' Data Having 4 & 5 Ratings

Figure 10 shows 20 different topics highlighting the best features of Beats speakers. From topics 2, 6, 11, 20, adjectives excellent, good, great, awesome can be seen which states that product is great with respect to sound quality.

TEXT CLUSTER

Once the Text Topic node generates the topics, Text Cluster node is connected to the Text Topic node to know the best and worst features of products based on ratings given by the reviewers. SAS Enterprise miner allows for grouping terms that closely relate to each other into separate clusters of related terms. After some trial-and-error, the properties settings for the Text Cluster node are set to generate well-separated clusters in the cluster space. Default settings are used in the properties panel which results in the solution with 40 as the maximum number of clusters and 15 as the number of descriptive terms to describe clusters using Expectation-Maximization Cluster Algorithm.

Cluster ID	Descriptive Terms		Frequency	Percentage
1	+blow +piece +product +buy company +purchase gift terrible +week +beat +year +return +long +pay +box	...	71	23%
2	+phone back +back +connect +order +'beat pill' customer service +pill +cord +box music +device +day +thing	...	44	14%
3	+charge +work +stop +month +item 'turn on' +plug +working +last +power +port +break +battery +red +problem	...	84	27%
4	size +bass +sound +price better +high +half good portable +volume +cheap +know +range +cost +want	...	42	14%
5	sound +'sound quality' +quality life +loud +unit +first +love +battery +device +good portable +high +disappoint +volume	...	33	11%
6	condition +play +year +full last +keep +hold +plug old +receive +waste +stay music +spend money		36	12%

Figure 11. Text Cluster Node Output for Beats Speakers' Data Having 1 & 2 Ratings
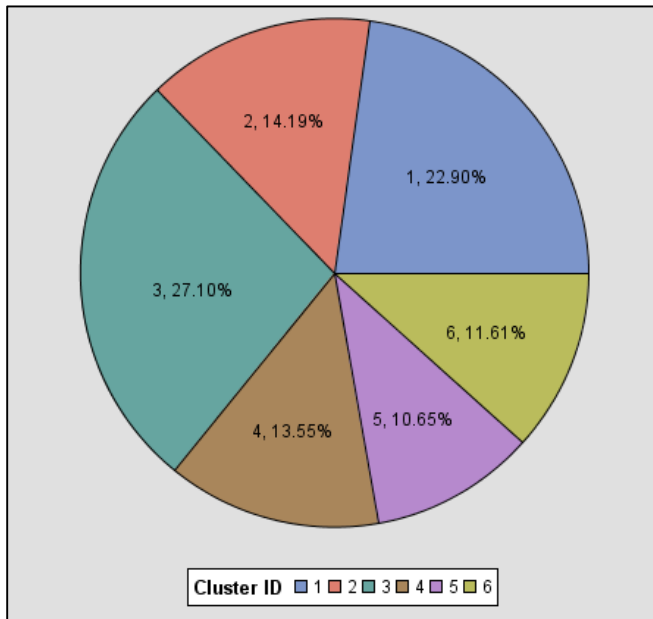

Figure 12. Pie Chart Distribution of Clusters Obtained from Figure 11

Text cluster node generates 6 clusters from Beats 1 & 2 reviews data set as shown in figure 11. Cluster 3 has the highest frequency from figure 12. Words in this cluster occur together, the maximum number of times in comments extracted from Amazon. The worst features of Beats speakers from above cluster analysis could be charge port affecting the battery and the price of the product.

Cluster ID	Descriptive Terms		Frequency	Percentage
1	ðŸ highly +look +love +recommend +work +speaker +sound	...	24	4%
2	+pill +big +charge +device +music +phone portable +little +size bass battery +speaker +volume +listen +easy	...	273	41%
3	+good +quality sound good +'good sound quality' 'sound quality' +cheap overall 'on time' +happy +beat +package +listen +look +product	...	97	14%
4	+love +gift christmas daughter +niece +son +buy great +'daughter love' +perfect +purchase grandson +nice husband +old	...	161	24%
5	great +product +'great sound' 'great product' +'excellent product' +awesome +seller +sound shipping 'awesome product' +fast 'awesome sound' 'great buy' +amaze +order	...	114	17%

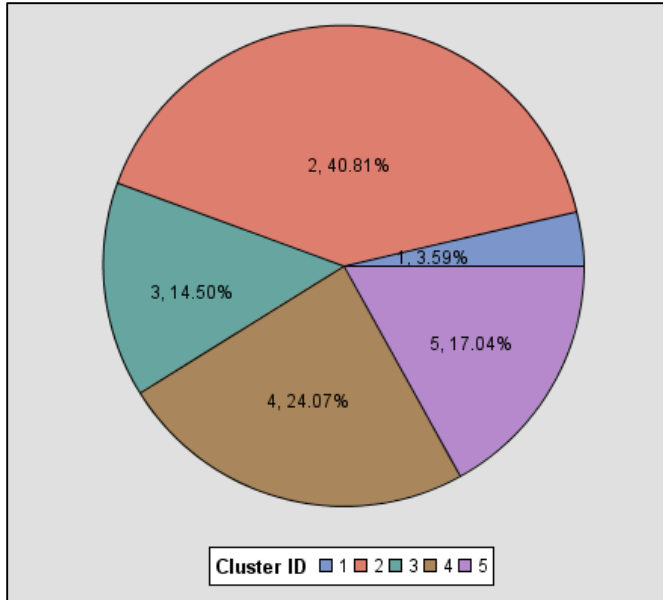**Figure 13. Text Cluster Node Output for Beats Speakers' Data Having 4 & 5 Ratings**

Figure 14. Pie Chart Distribution of Clusters Obtained from Figure 13

Text cluster node generates 5 clusters from Beats 4 & 5 reviews data set as shown in figure 13. Cluster 2 has the highest frequency from figure 14. This cluster states that the speaker is portable due to its size. Clusters 3 and 5 talks about the good sound quality of the speaker. Hence, the best features of Beats speakers from above cluster analysis could be portability and sound quality.

For remaining four speakers namely JBL, Ultimate Ears, Bose and Altec Lansing, the above process has been followed starting from Text Filter node till Text Cluster node for both 1&2 as well as 3&4 rating data. Below screen capture shows the diagram in SAS Enterprise Miner used for whole analysis.

Figure 15. Node Diagram

JBL Speakers Data Analysis:

Cluster ID	Descriptive Terms		Frequency	Percentage
1	'sound quality' sound +quality +bad 'bluetooth connection' +lose +day +phone connection good +people +low +good +cheap music	...	24	15%
2	+box fake +instruction +open +real +notice research +receive +counterfeit +original 'jbl charge' +item +buy +pay different	...	16	10%
3	+long +month +work +button +unit +contact +fail +warranty +recommend +first fact great +device +year +happen	...	31	19%
4	+song +bass +add +small +waterproof fact impressive +disappoint +high immediately +base +design music +good +'jbl product'	...	8	5%
5	+'serial number' +number +serial +counterfeit +support +box +sell +add +contact +defective +extremely +tooth +'jbl product' +suppose version	...	9	6%
6	+break +charger crap +port service +item +static +week jbl money great +month +base +time +issue	...	13	8%
7	+cord +full +charge +hour +problem +design +know battery +thing 'bluetooth connection' +amaze +happen +people different +right	...	13	8%
8	+half +volume +loud +bluetooth +low +first +bass +clear +order +high good +'bluetooth speaker' music +support charge	...	24	15%
9	+update +connect +'jbl product' version +purchase +tooth +sell +stop +working +time +point +read +buy +product battery	...	25	15%

Figure 16. Text Cluster Node Output for JBL Speakers' Data Having 1 & 2 ratings

Nine clusters are generated for JBL speakers 1 & 2 rating data as shown in figure 16. Cluster 3 has the highest frequency of 31. Cluster analysis shows that JBL has the bad Bluetooth connection

Cluster ID	Descriptive Terms		Frequency	Percentage
1	ðÿ å +good +sweet +talk +favorite +boy +old good +dont +house money always +son +'amazing sound'	...	56	4%
2	+great +sound +'great sound' awesome +life 'battery life' +'great speaker' +battery +'awesome sound' +speaker +easy +compact portable size +'amazing sound'	...	220	14%
3	+'bluetooth speaker' best +price +find +bluetooth +review +look +purchase +range always money +feature +speaker +time +sweet	...	245	15%
4	amazing +'amazing speaker' +satisfy +sound +little +'amazing sound' +clear +'clear sound' bass great +nice +box +speaker +want +bring	...	99	6%
5	+phone +device +charge +charger +connect portable +time music power +hour +play +big +pool +waterproof +compact	...	102	6%
6	+perfect +pair +amaze +play +song +device +waterproof awesome +pool easily highly +small +boy +recommend +easy	...	74	5%
7	sound +quality +product +excellent +'sound quality' +'good sound quality' 'great sound quality' +'great product' +great +gift +good +expect +amaze +connection +easy ...	...	192	12%
8	+love +buy +son +gift christmas +good +'well speaker' +husband +daughter +'good sound' highly +product +recommend +proof +old	...	174	11%
9	+volume +high 'a bit' +bass loud +low +loud +beat +battery +life 'battery life' +room +clarity +song +hear	...	185	12%
10	+day +work +last +charge +hour +year +long +review +battery music +favorite +listen +crisp +'good sound' +thing	...	159	10%
11	great +connect +nice +input easily +work +easy +sound +charger power +connection +'good sound' +decent better +daughter	...	86	5%

Figure 17. Text Cluster Node Output for JBL Speakers' Data Having 4 & 5 Ratings

Eleven clusters are formed from JBL speakers 4 & 5 rating data as shown in figure 17. Cluster 3 has the highest frequency of 245. Cluster analysis shows that JBL speakers have awesome waterproof, great sound quality and longer battery life.

Ultimate Ears Speakers Data Analysis:

Cluster ID	Descriptive Terms		Frequency	Percentage
1	+stop working +purchase +damage +connect +second +month +replacement +first +buy +return +bluetooth money +back +full	...	65	18%
2	+arrive +pair +unit +order +review box +fine +boom +back +charger +plug +work +turn +little +issue	...	66	19%
3	+sound +loud +bass +distort +volume music +good +blow +disappoint +high +little +price +play first money	...	74	21%
4	+bottom +charge +month +fix +problem 'at all' +battery +speaker +update great +plug +work +charger +last +year	...	79	22%
5	+customer +'customer service' +service support +hour +quality +response 'sound quality' +company +phone +battery +week +last +unit +product ...	...	72	20%

Figure 18. Text Cluster Node Output for Ultimate Ears Speakers' Data Having 1 & 2 Ratings

Five clusters are generated for Ultimate Ears speakers 1 & 2 rating data as shown in figure 18. Cluster 4 has the highest frequency of 79. Cluster analysis shows that Ultimate ears speakers come with a bad charger and many people returned them as charger has been damaged within a few days of delivery.

Cluster ID	Descriptive Terms		Frequency	Percentage
1	+cable +'charge cable' 'usb cable' +charge +unit +include +fit +port +miss +hard +case +issue +month +instruction +find	...	34	3%
2	+tone +incredible +option +advertise +low +clear bass +update +happy able +megaboom +issue best well +product	...	56	4%
3	+speaker portable +'portable speaker' +good +'bluetooth speaker' +recommend +bluetooth +battery best +last +waterproof great loud highly +small	...	284	21%
4	+phone +connect music +device +play +button +turn +cool quickly +listen +hard +easy +house able +water	...	87	7%
5	sound +quality +carry +amaze 'sound quality' wireless +little +case +purchase excellent +price size +pair +find +instruction	...	180	14%
6	+love +product 'great product' +awesome +'excellent product' +quality sound +gift +great +buy +'awesome speaker' +unbelievable wife 'sound quality' excellent	...	133	10%
7	+perfect +expectation +'good bass' +alone +backyard +exceed +distortion +happy +full wife +camp +include +high river +day	...	22	2%
8	+volume +hour +pool range +feature beach outdoor +bass +device +distortion +system penny +full +backyard able	...	117	9%
9	+great +battery +'great sound' life +'battery life' +'great speaker' +sound +speaker +long connectivity design +easy outdoor highly +'excellent product'	...	107	8%
10	+high +full better +low +test +boom +big +keep back +little +add +extremely +feature +find +price	...	86	7%
11	great 'ue boom' +big +sound +good +boom +megaboom +thing well +party +expect +house music +unit 'usb cable'	...	81	6%
12	+charge +plug +charger +work +play +long +hour +time +first +house +problem +day +blast +include 'usb cable'	...	77	6%
13	+nice +work +cool +fine exactly +instruction +expect +advertise great +alone +option +extremely +system +product +package	...	35	3%
14	excellent +'excellent sound' quickly +complaint +voice +float river +blast penny +issue +'excellent product' wireless +absolutely +area +arrive	...	22	2%

Figure 19. Text Cluster Node Output for Ultimate Ears Speakers' Data Having 4 & 5 Ratings

Fourteen clusters are formed from Ultimate Ears speakers 4 & 5 rating data as shown in figure 19. Cluster 3 has the highest frequency of 284. Cluster analysis tells that Ultimate ears speakers have the good waterproof capability, excellent sound quality, good Bass, longer Battery life.

Bose Speakers Data Analysis:

Cluster ID	Descriptive Terms	Frequency	Percentage
1	+pair +phone +connect +defective +device +work +know back +easy +item +crap +back +bluetooth +problem +receive ...	52	16%
2	+stop +working warranty +month +year +'bose product' +product +less +week return +buy christmas +charge +die +gift ...	89	27%
3	+port power +cord +'charge port' +love +cable +issue +charger +plug +dead +charge +receive +look +bad +unit ...	67	21%
4	sound 'sound quality' +feel +loud +bass +quality +volume +high +good +sound +price +brand +thing +review +keep	80	25%
5	+'rechargeable battery' +rechargeable +battery +hold +read +last +crap +long +charge +fully +hour +little great +first +late ...	38	12%

Figure 20. Text Cluster Node Output for Bose Speakers' Data Having 1 & 2 Ratings

Five clusters are generated for Bose speakers 1 & 2 rating data as shown in figure 20. Cluster 2 has the highest frequency of 89. Cluster analysis shows these speakers have issues with battery and Bluetooth.

Cluster ID	Descriptive Terms	Frequency	Percentage
1	+cable +play +review +charger +plug +charge +hour +volume +time +big +issue +thing +range +last +long	174	7%
2	+device +connect +pair +carry +phone +bluetooth easily +time +easy +'bluetooth speaker' +house +connection +room +charge +problem	311	12%
3	+battery +life 'battery life' +long +last +hour +charge +connection great +issue +range +portable +day connectivity 'great sound quality'	139	5%
4	+unit fantastic +happy +ease 'small unit' +'bluetooth connectivity' +brother +user connectivity +'high volume' +issue +sync +problem great +high	68	3%
5	+sound +great +love +'great sound' +gift husband +speaker +'amazing sound' +son +'great speaker' +easy christmas +small +'awesome sound' 'set up'	630	24%
6	+quality sound +product 'sound quality' +excellent +'great product' 'great sound quality' +price +'bose product' +expect +awesome +great amazing size +good ...	367	14%
7	+system music +home home +listen +advertise office +room always +house +fill perfect +hear +purchase +want	116	4%
8	+work +day +listen +phone well christmas music +problem +great +want +plug +thing +loud +play	257	10%
9	+good +bass +price +loud +money +clear +portable +nice +recommend +speaker +high +little +big +volume +sound	580	22%

Figure 21. Text Cluster Node Output for Bose Speakers' Data Having 4 & 5 Ratings

Nine clusters are generated for Bose speakers 4 & 5 rating data as shown in figure 21. Cluster 5 has the highest frequency of 630. The highlighted feature for these speakers could be sound quality.

Altec Lansing Speakers Data Analysis:

Cluster ID	Descriptive Terms	Frequency	Percentage
1	+distortion +volume +listen +high +low +control replacement music +'high volume' +'low volume' +blow +problem +metal +cheap +disappoint ...	29	14%
2	price +expect +cord couple +garbage +portable +hope +buy +last +good +bass money +replace +year +long	37	18%
3	'sound quality' sound +quality second good +travel +case +'altec lansing speaker' +love +old +positive +happen +purchase +plug music ...	14	7%
4	+test +right +ear +know +open +best +home late +set cable +unit +find box computer purchase	7	3%
5	+review +read +noise +contact +back +day +product +screech +disappoint +late money +order +annoy +minute +positive	32	15%
6	'battery compartment' +battery +compartment +hold life +fall 'battery life' always +batterie +lack +bass +well +metal +happy +thing	15	7%
7	+month +break +bad +wire +stay +year +switch +fine +working +item few great +stop +quit +start	29	14%
8	+device audio +laptop +connect cable +plug +inch desk issue +connection able horrible +second +move +less	10	5%
9	+button 'power button' +drain +push +power purchase +light +reason +junk +turn +thing +unit +pretty +little +switch	16	8%
10	'mp3 player' 'volume control' control +player mp3 life +phone +long 'battery life' +short +portable +hope +minute +kind loud	18	9%

Figure 22. Text Cluster Node Output for Altec Lansing Speakers' Data Having 1 & 2 Ratings

Ten clusters are generated for Altec Lansing speakers 1 & 2 rating data as shown in figure 22. Cluster 2 has the highest frequency of 37. From cluster analysis, these speakers have the noise problem, cheap material.

Cluster ID	Descriptive Terms	Frequency	Percentage
1	+quality sound +'sound quality' +portable highly +recommend +'portable speaker' +excellent +look +product +price +great size +easy +purchase ...	166	12%
2	+cord +unit +button power +short +long +switch +nice +problem +star +case +big 'a bit' +device +plug	194	14%
3	+'high volume' +star +distortion +level +clip +high +distort +volume +rechargeable +cord +hear +bang +loud +'extension cable' +compare ...	34	2%
4	+work +music +book +listen +hear +car great +phone +play +love +ipod +day +want +radio +buy	191	14%
5	+player +'mp3 player' mp3 +control +'volume control' +volume +device +phone +base first +car +rechargeable +problem +'battery life' +laptop	95	7%
6	+room +watch +fill +laptop +travel +batterie +music +turn +rechargeable +hour +plug +thing +'aaa battery' +first +big	160	12%
7	+cable +'extension cable' extension audio +short +device +cost +long +carry power +radio +switch +rechargeable +control +travel	38	3%
8	+great +small +'small speaker' +sound +friend +bass +speaker +'great sound' great +definitely +good perfect +little +price size	212	15%
9	+review +read +gift +product +first +purchase +base christmas second +buy +channel +problem first +ipod +year	84	6%
10	+battery +life +'battery life' +hour +year +'aaa battery' +last +batterie +rechargeable sound +'great sound' great highly +long +'portable speaker' ...	127	9%
11	+bang +recess +channel +buck cool impressive +switch second +clip +carabiner +gift first +christmas stereo +radio	18	1%
12	good pretty +compare 'a bit' +batterie perfect +'aaa battery' +distort +thing +'portable speaker' +star +clip first +case +carabiner	59	4%

Figure 23. Text Cluster Node Output for Altec Lansing Speakers' Data Having 4 & 5 Ratings

Twelve clusters are formed from Altec Lansing speakers 4 & 5 rating data as shown in figure 23. Cluster 8 has the highest frequency of 212. Cluster analysis shows that these speakers have great sound quality, longer battery life.

COMPARISION OF FEATURES OF 5 SPEAKERS

Product	High Rating Features	Low Rating Features	Average Rating	Price
Beats	Portability, Sound Quality	Charge Port , Price	3.6	170
JBL	Waterproof, Sound Quality, Battery life	Bluetooth	4.5	149
Ultimate Ears	Waterproof, Sound Quality, Bass, Battery life	Charger, Speaker Material	4	146
Bose	Sound Quality	Battery, Bluetooth	4.4	160
Altec Lansing	Sound Quality, Battery life	Noise, Outer Material	3.9	160

Table 2. Ratings and Features of all Speakers

CONCLUSION

- Bestselling product in the market is JBL with average rating 4.5 whereas Beats has the least average rating of 3.6. This would let us understand the features that we need to incorporate while launching a similar product in the market to withstand the competitor strategies.

- Sound Quality, Battery life, Waterproofness, Bass are the features responsible for the higher rating by the customers. This kind of topics extraction for high ratings will let us know how to message the products to customers, and this would quickly establish the positive emotional connection with customers.

- Noise and Price are the primary causes for lower rating by the consumers. We could clearly be cautioned that we should try to overcome these lower rating features that avoid customer dissatisfaction on the new launches.

- Further, Price could be one of the attributes for customer satisfaction as in most of the products, the Average rating is inversely correlated to the Price and vice-versa

According to the analysis above, customers focus on buying speakers with a good Bluetooth connection, has a good battery life, better noise control, includes speaker material, is affordably priced while having excellent sound quality, is waterproof and has good bass. Hence, this feature analysis helps retailers and companies to understand customer expectations to create future products that meet the needs of their customers and gain business value. Additionally, this analysis prototype also benefits the amazon retailers to constantly view and evaluate the customer's reviews and increase their brand loyalty by following the above recommendations.

REFERENCES

1. Retail e-commerce sales worldwide from 2014 to 2021, Statista 2018. Accessed August 2018. "https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/".

2. Study: 81% research online before making big purchases: CSA STAFF. Accessed July 12, 2013. https://sourceforge.net/projects/winmerge/postdownload.

3. Goutam Chakraborty, Murali Pagolu, Satish Garla, *Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS® by SAS Institute Inc 2014*, Accessed 20 Mar 2017.

4. Getting Started with SAS® Text Miner 13.2. Cary, NC: SAS Institute Inc., Accessed 20 Mar 2017.

5. Goutam Chakraborty, Jiawen Liu and Mantosh Kumar Sarkar. April 2013. "Feature-Based Sentiment Analysis on Android App Reviews Using SAS® Text Miner and SAS® Sentiment Analysis Studio." Proceedings of the SAS Global Forum.

6. Jin Meizi, Jiawen Liu, Goutam Chakraborty. March 2014. "A Way to Fetch User Reviews from iTunes Using SAS®." Proceedings of the SAS Global Forum, Washington D.C.

7. Goutam Chakraborty, Jiawen Liu and Mantosh Kumar Sarkar. 2016. "Feature-Based Sentiment Analysis on Android App Reviews Using SAS® Text Miner and SAS® Sentiment Analysis Studio." Proceedings of the SAS Global Forum, Las Vegas.

ACKNOWLEDGEMENTS

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Manideep Mellachervu
Oklahoma State University
(405) 614 9964
manideep.mellachervu@okstate.edu
www.linkedin.com/in/mm2310

Anvesh Reddy Minukuri
Comcast Corporation
(405) 780-5346
anveshreddy_minukuri@comcast.com
https://www.linkedin.com/in/anvesh-reddy-minukuri-4a5a6160/