

Apple or Samsung? What mobile phone should I buy?

Sentiment Analysis utilizing customer reviews on mobile phones

Author: Rosie Nguyen, MSBA student at Oklahoma State University

Mentors: Dr. Goutam Chakroborty, Oklahoma State University

Dr. Miriam McGaugh, Oklahoma State University

Nandan Moza, Senior Risk Analyst, Elevate

Abstract

The mobile phone market in the USA is usually regarded as the dominant segment of the world market because of its diversity and fast-paced technology development. A survey shows that nearly 200 million persons in the United States own a smartphone. The top two brands are Apple and Samsung with roughly 40% and 30% market share, respectively. A large number of sales of these two brands were made on Amazon.com, a popular e-commerce website with thousands of reviews for each brand. However, a potential buyer would find it impossible to review all of these customer reviews to decide which brand to buy. Therefore, consumers are very likely to refer to a small percentage of the comments as they attempt to make an informed decision. Our goal was to aim to leverage consumer opinion on Amazon's website to understand what features generated more positive feedback, and what aspects drew more positive impression after purchasing iPhone or Samsung Galaxy models. The two latest models from Apple (iPhone 6 & iPhone 7), and from Samsung (Galaxy S6 & Galaxy S7) were employed to understand customer insights on these two brand names.

With over 25,500 reviews to analyze, we aimed to categorize customer reviews made on Amazon. Text Cluster in SAS® Enterprise Miner were utilized to identify the best and worst features associated with each model. This analysis could be used as a reference when prospective customers are making their purchase decisions

Introduction

Amazon is commonly known as one of the largest electronic commerce and computing company in the world. Many customers turn to Amazon for a variety of products including mobile phones. Under each product, users can write reviews what they think about products or service they purchased there. Although those reviews could be a good indicator for potential buys when it comes to buying decision of the same products or similar, it is challenging to go through all the reviews to make the right decision upon their own interest. The paper aims at utilizing algorithm from SAS® Enterprise Miner to explore the comments and give recommendations to people when they want to decide what phone is more suitable.

Data Preparation



This data includes customer's reviews on unlocked mobile phones on amazon.com website. It was published on Kaggle.com at the beginning of this year. After accessing to this data, I narrow down to only two brand names: Apple and Samsung because of their high competitiveness in the American market. This comma separate value (csv) sheet is converted into a SAS dataset using SAS Enterprise Guide. The data has a total of 25,543 observations and 7 variables as follows:

#	Variable	Type	Description
3	Brand_Name	Char	Brand Names here are Apple and Samsung
1	ID	Num	Primary key to identify different reviews
4	Price	Num	Price to buy the product
2	Product_Name	Char	Product models of iPhone 6, 7, or Samsung Galaxy 6, 7 with specification
5	Rating	Num	Rating of how well customers think about the product
7	Review_Votes	Num	How many votes from other clients who read the reviews
6	Reviews	Char	Contents of what customers think about the products

Figure 1: Data information

CODES

```
libname phone ' C:\Users\huon\Desktop';  
  
DATA phone.phone_pos phone.phone_neg;  
SET WORK.APPLEVSSAMSUNG_0002;  
IF Rating >3 then output phone.phone_pos;  
IF Rating <=3 then output phone.phone_neg;  
DROP ID Review_Votes Rating Price Sentiment;  
RUN;
```

Exploratory Data Analysis

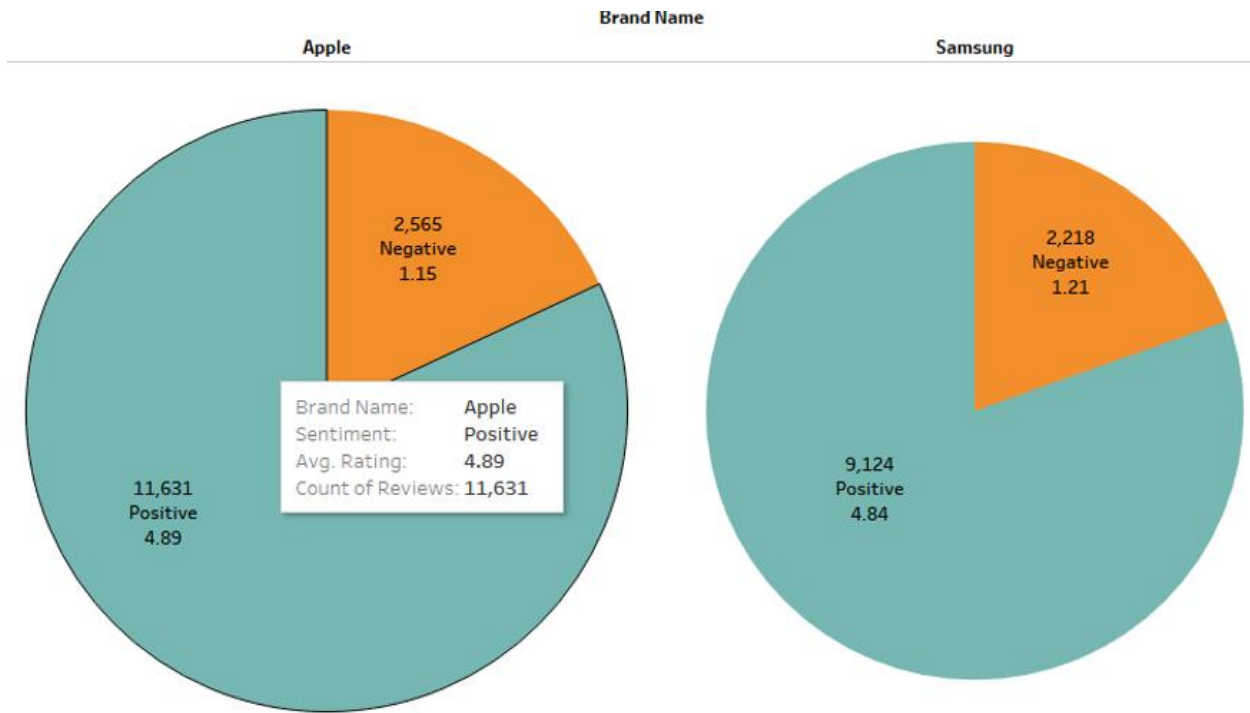


Figure 2: Brand Names with Sentiment

The two pie charts show that most of customers gave good feedback for mobile phones from Apple and Samsung. However, Apple seems to have higher proportion of regative reviews than Samsung have for its product models.

Product with Sentiment

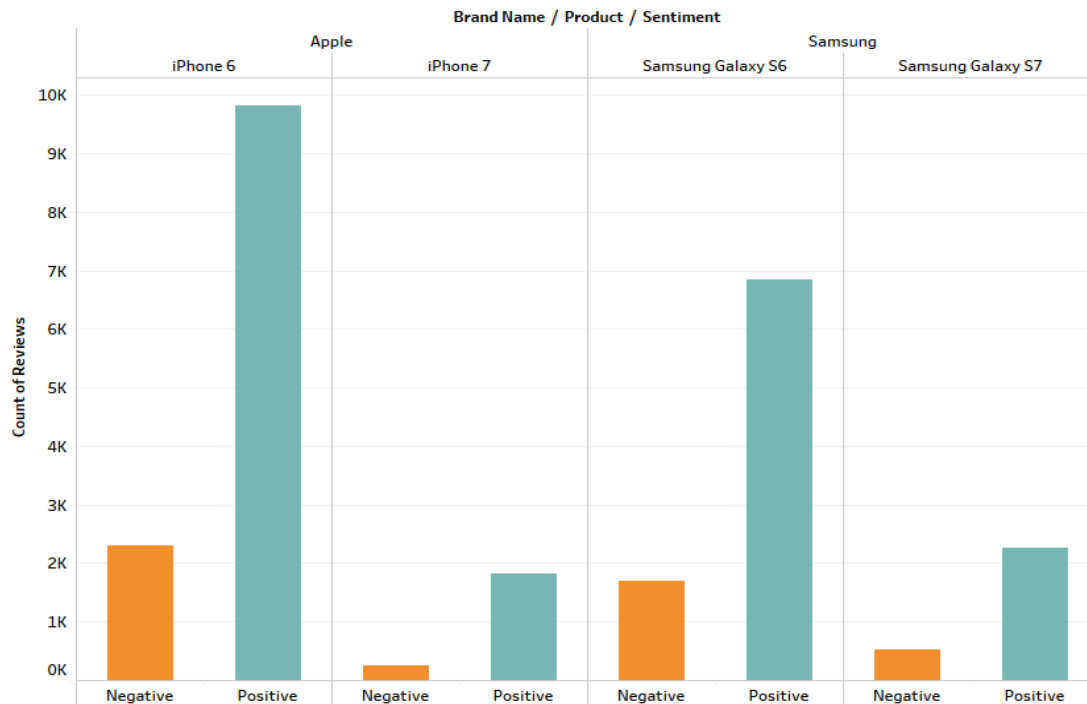


Figure 3: Product with Sentiment

To explore to model layers, we can see that iPhone 6 has higher number of negative reviews compared to the other products, but also drew highest number of positive comments. However, the other model of Apple (iPhone 7) seemed to generate the least negative feedbacks from customers. To gain the exact insights of those feedback, we will analyze text to see what features associated with which mobile phone models generate the most good or bad impression.

Text Analysis

In this paper, we performed three parts:

- Analyzing all the reviews for both iPhone 7 and Samsung Galaxy 7 to understand what customers think about those two products
- Dividing the data set into two groups by sentiment: Positive (Rating >3), and Negative (Rating ≤3);
- Analyzing Sentiment analysis of individual brand names

I. Analyzing both brand names

1. Diagram

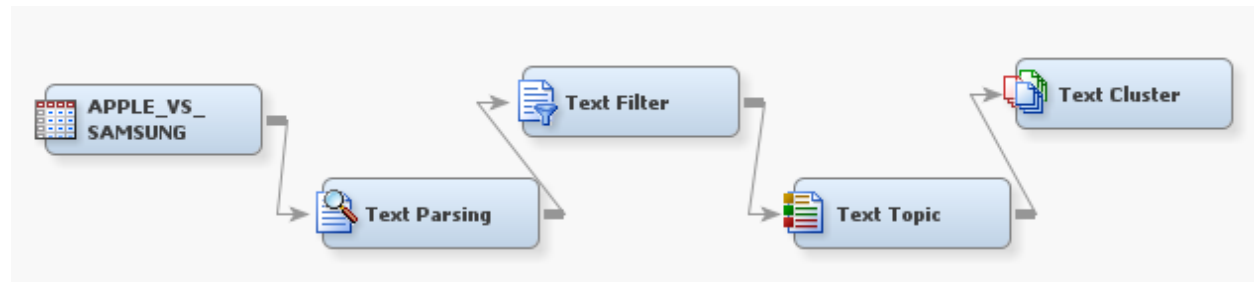


Figure 4: Text Mining diagram

2. Text import

The data source was in csv format, and then converted into sas format using Enterprise Guide. The new format is then imported into SAS® Enterprise Miner™ for text analysis

3. Text parsing

The property panel is set up as follows:

- Ignore Parts of Speech: 'Aux', 'Conj', 'Det', 'Interj', 'Part', 'Prep', 'Pron', 'Prop'
- Ignore Type of Attributes: 'Num', 'Punct'
- 'Find entities' is set to 'Standard'
- 'Detect different parts of speech' is set to 'no' to be able to represent one word or terms as a whole

General	
Node ID	TextParsing
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
<input type="checkbox"/> Parse	
Parse Variable	
Language	English
<input type="checkbox"/> Detect	
Different Parts of Speech	No
Noun Groups	Yes
Multi-word Terms	SASHELP.ENG_MULTI
Find Entities	Standard
Custom Entities	
<input type="checkbox"/> Ignore	
Ignore Parts of Speech	'Abbr' 'Aux' 'Conj' 'Det' 'I...
Ignore Types of Entities	...
Ignore Types of Attributes	'Num' 'Punct'
<input type="checkbox"/> Synonyms	
Stem Terms	Yes
Synonyms	SASHELP.ENGSYNMS
<input type="checkbox"/> Filter	
Start List	...
Stop List	SASHELP.ENGSTOP
Select Languages	...

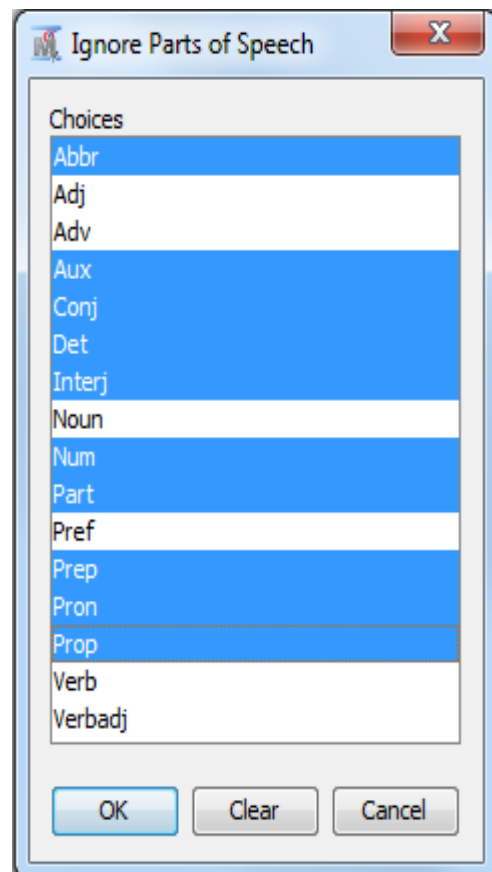
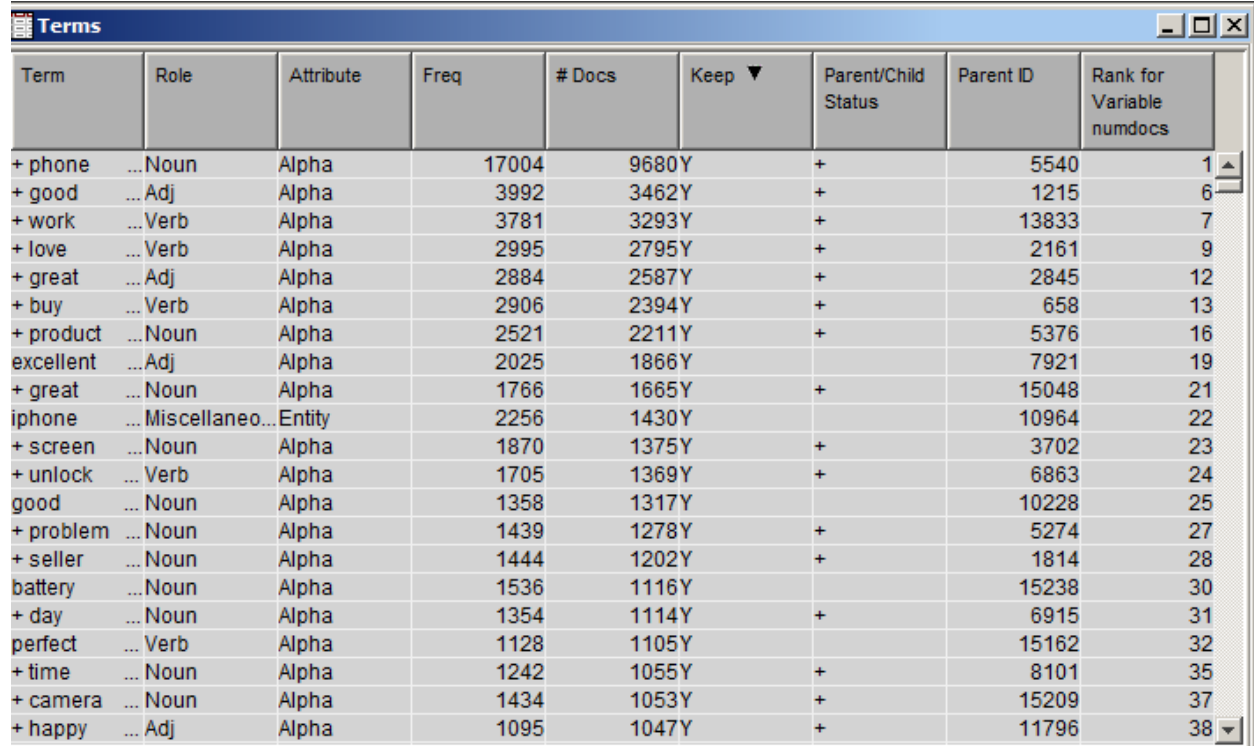


Figure 5: Text parsing Setting

Figure 6: Ignore Parts of Speech Setting

This node aims at creating terms by frequency document matrix, which illustrates terms most frequently mentioned and the number of documents. Also, the node shows what terms are dropped or kept in the analysis.



Term	Role	Attribute	Freq	# Docs	Keep ▼	Parent/Child Status	Parent ID	Rank for Variable numdocs
+ phone	... Noun	Alpha	17004	9680Y		+	5540	1
+ good	... Adj	Alpha	3992	3462Y		+	1215	6
+ work	... Verb	Alpha	3781	3293Y		+	13833	7
+ love	... Verb	Alpha	2995	2795Y		+	2161	9
+ great	... Adj	Alpha	2884	2587Y		+	2845	12
+ buy	... Verb	Alpha	2906	2394Y		+	658	13
+ product	... Noun	Alpha	2521	2211Y		+	5376	16
excellent	... Adj	Alpha	2025	1866Y			7921	19
+ great	... Noun	Alpha	1766	1665Y		+	15048	21
iphone	... Miscellaneo...	Entity	2256	1430Y			10964	22
+ screen	... Noun	Alpha	1870	1375Y		+	3702	23
+ unlock	... Verb	Alpha	1705	1369Y		+	6863	24
good	... Noun	Alpha	1358	1317Y			10228	25
+ problem	... Noun	Alpha	1439	1278Y		+	5274	27
+ seller	... Noun	Alpha	1444	1202Y		+	1814	28
battery	... Noun	Alpha	1536	1116Y			15238	30
+ day	... Noun	Alpha	1354	1114Y		+	6915	31
perfect	... Verb	Alpha	1128	1105Y			15162	32
+ time	... Noun	Alpha	1242	1055Y		+	8101	35
+ camera	... Noun	Alpha	1434	1053Y		+	15209	37
+ happy	... Adj	Alpha	1095	1047Y		+	11796	38

Figure 7: Text Parsing Node Output

The most frequent terms are phone, good, product, iphone, screen, unlock, battery and camera. This shows that most of customers seemed happy about the products and those last four terms are the features users show the most care about.

4. Text Filter

Text filter node serves as eliminating words with small frequency and little importance. Users can manually change “Minimum number of documents” in which a term occur to decide which one should be kept or dropped. “Maximum number of terms” could be identified to restrict terms to be analyzed. Spelling check option is set to “Yes” to autocorrect any misspelling words.

Term	Role	Attribute	Status	Weight	Imported Frequency	Freq	Number of Imported Documents	# Docs	Rank	Parent/Child Status	Parent ID
+ phone ...		Alpha	Keep	0.142	7705	7734	4583	4603	1+		9617
+ be ...		Alpha	Drop	0.000	7945	7947	3725	3725	2+		13063
+ good ...		Alpha	Keep	0.161	3624	3636	3091	3099	3+		2452
+ great ...		Alpha	Keep	0.169	3222	3226	2801	2803	4+		6283
+ work ...		Alpha	Keep	0.196	2516	2520	2180	2182	5+		4148
+ very ...		Alpha	Drop	0.000	2385	2396	2055	2063	6+		13038
+ love ...		Alpha	Keep	0.201	2299	2306	2045	2050	7+		4368
not ...		Alpha	Drop	0.000	2878	2878	1724	1724	8		13080
+ have ...		Alpha	Drop	0.000	2906	2906	1704	1704	9+		13051
+ new ...		Alpha	Drop	0.000	1454	1457	1276	1279	10+		12996
+ excellent ...		Alpha	Keep	0.250	1254	1269	1219	1234	11+		5431
+ perfect ...		Alpha	Keep	0.253	1238	1256	1184	1202	12+		937
+ get ...		Alpha	Drop	0.000	1422	1424	1124	1124	13+		13079
+ product ...		Alpha	Keep	0.261	1171	1175	1116	1120	14+		4799
s ...		Alpha	Drop	0.000	1433	1433	1117	1117	15		12997
+ do ...		Alpha	Drop	0.000	1699	1705	1046	1048	16+		13078
no ...		Alpha	Drop	0.000	1228	1228	1032	1032	17		13156
+ buy ...		Alpha	Keep	0.279	1179	1181	997	999	18+		4913
+ come ...		Alpha	Drop	0.000	1128	1128	959	959	19+		13074
+ use ...		Alpha	Drop	0.000	1394	1397	946	946	20+		13033
so ...		Alpha	Drop	0.000	1168	1168	934	934	21		13190
+ fast ...		Alpha	Keep	0.302	951	954	824	824	22+		8461
+ just ...		Alpha	Drop	0.000	957	959	821	823	23+		13108

Figure 8: Text Filter Node output

With the Interactive Filter viewer, we can view how Spelling check works

	Parent # Docs	Term	# Docs	Parent	Role	Parent Role	Min Distance	Dictionary	Key	Parent ID
1	21.0	worthy	2.0	worth			6.0		1477.0	27.0
2	21.0	worth	2.0	worth	LOCATION		0.0		4039.0	27.0
3	13.0	30 day	1.0	30 days	TIME	TIME_PERIOD	10.0		1842.0	70.0
4	5.0	microsd	2.0	microsd	PROP_MISC		8.0		263.0	109.0
5	5.0	micro sd	2.0	microsd	NOUN_GROUP		14.0		1505.0	109.0
6	5.0	micro-sd	1.0	microsd			7.0		2421.0	109.0
7	5.0	microsd	3.0	microsd	PROP_MISC		0.0		3154.0	109.0
8	29.0	definite	1.0	definitely			8.0		774.0	112.0
9	7.0	zero	1.0	zero	PROP_MISC		0.0		48.0	121.0
10	5.0	touchscreen	1.0	touch screen		NOUN_GROUP	8.0		3909.0	123.0
11	22.0	mod	3.0	mode			10.0		1456.0	129.0
12	8.0	phone waterproof	1.0	phonewaterproof	NOUN_GROUP		6.0		229.0	136.0
13	8.0	boot	5.0	boost			12.0		1343.0	143.0
14	36.0	water	5.0	water	LOCATION		0.0		1928.0	152.0
15	15.0	america	7.0	american	LOCATION		4.0		2793.0	209.0
16	7.0	sort	4.0	short			12.0		2891.0	218.0
17	5.0	exynos processor	1.0	exynos processor	NOUN_GROUP	PROP_MISC	0.0		2817.0	228.0
18	11.0	cornings	2.0	corning	PROP_MISC	PROP_MISC	4.0		445.0	231.0
19	48.0	charger	1.0	charger	PROP_MISC		0.0		300.0	235.0
20	62.0	quilty	3.0	quality			8.0		2534.0	253.0
21	62.0	quality	4.0	quality	PROP_MISC		0.0		4121.0	253.0
22	155.0	screen	1.0	screen	PROP_MISC		0.0		1883.0	280.0
23	7.0	silicon case	4.0	silicon case	PROP_MISC	NOUN_GROUP	0.0		3721.0	302.0

Figure 9: Text Filter Spelling

Also, the interactive Filter View enables us to see how similar words are grouped together.

Terms							
	TERM	FREQ	# DOCS	KEEP ▼	WEIGHT	ROLE	ATTRIBUTE
<input type="checkbox"/>	unlock	182	134	<input checked="" type="checkbox"/>	0.368		Alpha
<input type="checkbox"/>	unlocked	6	6			Miscellaneous Pro...	Entity
<input type="checkbox"/>	unlocked	158	128				Alpha
<input type="checkbox"/>	unlocked	1	1				Alpha
<input type="checkbox"/>	unlock	14	14				Alpha
<input type="checkbox"/>	unlocking	3	2				Alpha
<input type="checkbox"/>	samsung	323	129	<input checked="" type="checkbox"/>	0.409	Company	Entity
<input type="checkbox"/>	samsung	311	124			Company	Entity
<input type="checkbox"/>	samsungâ	3	3			Miscellaneous Pro...	Entity
<input type="checkbox"/>	samsung	6	4			Miscellaneous Pro...	Entity
<input type="checkbox"/>	samsungs	3	3			Miscellaneous Pro...	Entity
<input type="checkbox"/>	time	149	126	<input checked="" type="checkbox"/>	0.367		Alpha
<input type="checkbox"/>	time	133	111				Alpha
<input type="checkbox"/>	timesa	1	1				Alpha
<input type="checkbox"/>	times	15	15				Alpha
<input type="checkbox"/>	problem	135	112	<input checked="" type="checkbox"/>	0.384		Alpha
<input type="checkbox"/>	problems	58	54				Alpha
<input type="checkbox"/>	problem	77	64				Alpha

Figure 10: Text Filter Spelling correction

Concept link

Concept link is a function in Interactive Filter view, which helps to see top associations between concepts. Based on the frequency of words mentions, there are four words that customers seem to care the most: Camera, charger, battery and screen. Camera has the strong relationship with terms “high” and “party”.

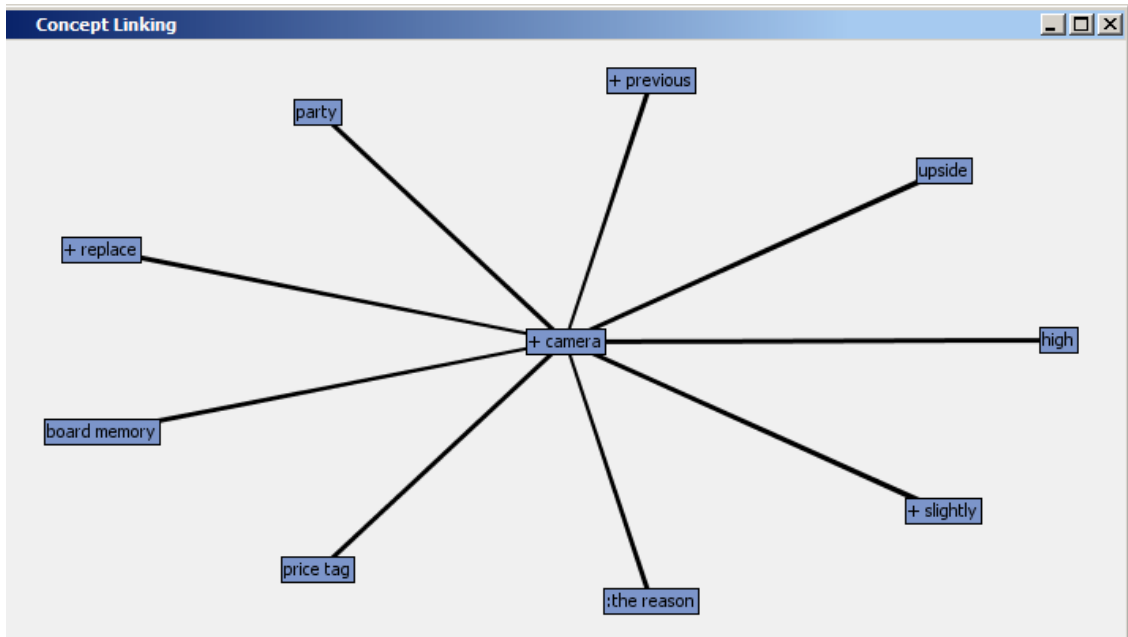


Figure 11: Camera concept link

With battery, “GSM Quality”, “S7” and “lag” are of the most significance to customers.

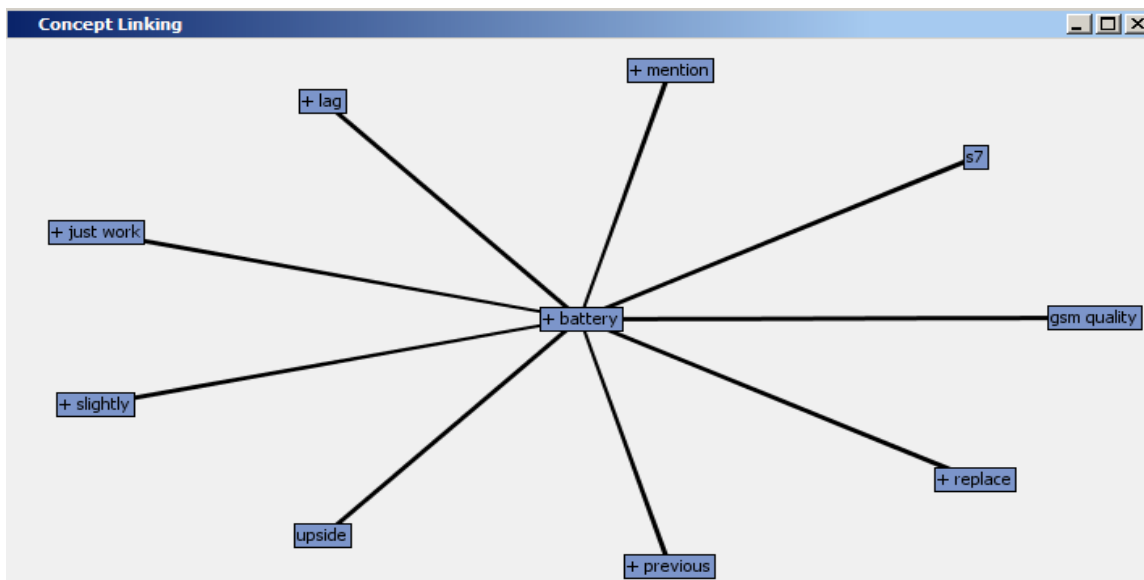


Figure 12: Battery concept link

Screen shows the strong connection with high, “slightly” and also “replace”

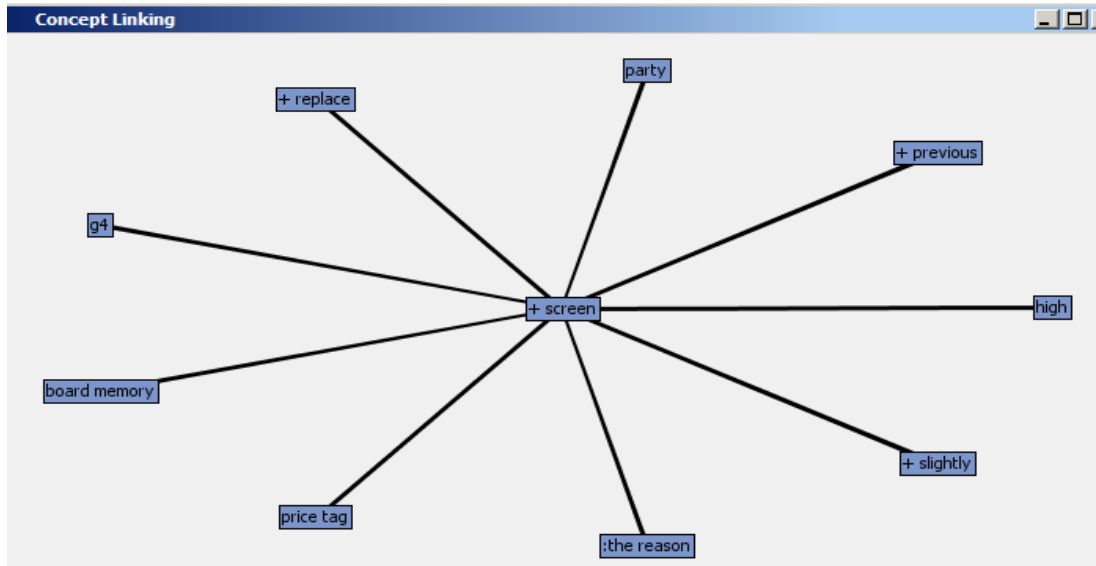


Figure 13: Screen concept link

II. Exploring positive and negative reviews separately

1. Positive reviews

a. Diagram

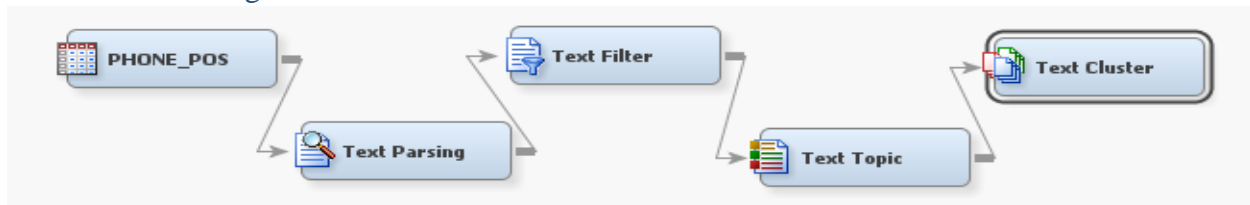


Figure 14: Positive Reviews diagram

The setting for Text Parsing, Text Filter are the same as the diagram above for both brand names.

b. Text Clustering

This node aims at grouping documents with the similar contents under the expectation - maximum algorithm. As a result, we have eight clusters with the following contents:

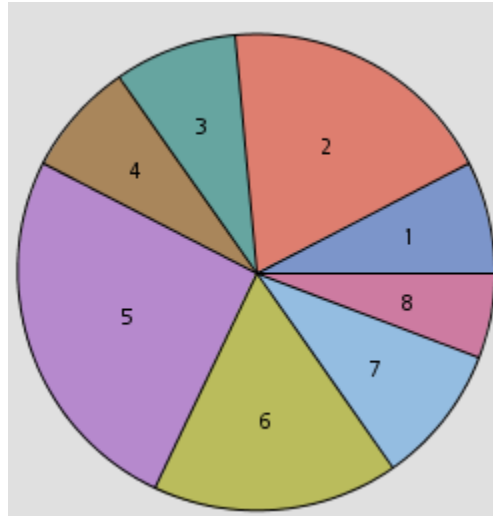
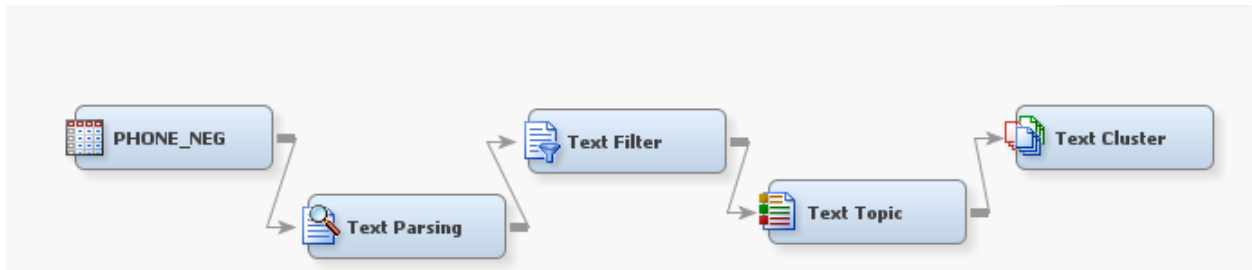


Figure 13: Text Cluster for Positive Reviews

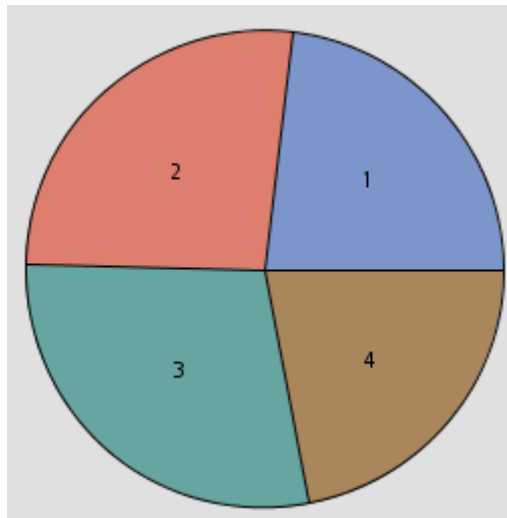
Cluster ID	Descriptive Terms	Percentage	Description
1	good excellent a++ awesome version good product wonderful great gucci 'excellent service +force +tweak	7%	product
2	battery +camera +screen +charge samsung best +life +thing s6 +quality +feature +edge back +case 6s	19%	camera, screen, battery
3	+great +phone +'great phone' nice +price +condition +expectation +'great price' awesome +'great condition' amazing +service 'amazing phone' +'great product' 'awesome phone'	8%	Good condition after shipping
4	iphone +nice +want +expect +apple +display +photo +enjoy 6s beautiful +big +model +year amazon +screen	8%	Display and screen
5	phone +work +buy +purchase +happy +card +unlock +seller +problem +arrive +brand +box sim +charger great	25%	Sim card and charger
6	+good +love perfect +phone +condition awesome +'well phone' excelent +'perfect condition' +'good phone' +amaze +'good condition' cell absolutely +'excellent phone'	17%	product
7	excellent +product 'excellent product' excelente +'good product' +'great product' +fast 'on time' delivery shipping 'excelent product' +'fast ship'	9%	Shipping
8	+great +work perfectly phone fine 'on time' well +well +expect +look +arrive +problem +advertise far	6%	Phone

2. Negative Reviews

a. Diagram



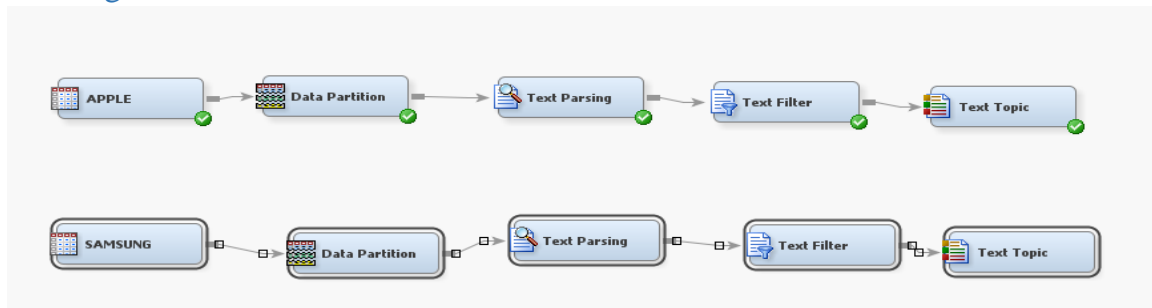
b. Text Clustering



Cluster ID	Descriptive Terms	Percentage	Description
4	return +charger battery +charge +receive +box +open +bad +look +package +hour +refurbish phone +good +disappoint	22%	Battery
3	+phone +unlock +card iphone sim +receive +apple +'sim card' +lock sprint +pay +activate +order +item +fake	28%	Sim card
2	buy +month samsung +work +product working +warranty +purchase +seller +repair international +stop +contact +sell +year	27%	Warranty
1	camera water +big +feature quality +day +break great +case black +drop +know +screen back +week	23%	Camera, case

III. Analyzing iPhone models and Samsung Galaxy models separately

1. Diagram



2. Recommendations

Below is the recommendation (x) if features on the left side is what customers look into a phone.

It shows that if someone loves camera and headphone, iPhone 6, or iPhone 7 would be a good choice. At the same time, if phone users are interested in battery and screen with dual sim, Samsung Galaxy S6 or Galaxy S7 is the top choice for them.

	Iphone	Samsung Galaxy
Camera	x	
Package	x	
Unlock	x	
Headphone	x	
Waterproof	x	
Delivery	x	
Feature	x	
Dual Sim		x
Screen		x
Case		x
Adapter		x
Warranty		x
Battery		x

IV. Conclusion

With the rapid development to technology, mobile phones become an essential tool in human life. With many commercials and ever expanding media exposure, sometimes, customers are confused what to choose for their suitable mobile phones. From the data set that we have from Kaggle.com and support of SAS® Enterprise Miner, we have found out that when it comes to a phone purchasing decision, customers care most about screen, battery, camera and design. For those who want a reliable battery and nice screen, Samsung Galaxy S6/S7 seems to be a top option

for them. For those who are interested in design and camera, iPhone 6/7 would serve them the best.

V. Future Scope

As part of future scope, Text Rule Builder Node in SAS® Enterprise Miner can be used to predict the sentiment of the review. Also, unlocked phone review corpus can be analyzed using SAS® Sentiment Analysis Studio.

VI. References

- [Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS®](#) by Goutam Chakraborty, Murali Pagolu, Satish Garla.
- [Gathering customer information from a mobile application](#) by James Adams, SAS Institute Inc.
- [Sentiment Analysis of Opinions about Self Driving Cars](#) by Nachiket Kawitkar and Swapneel Deshpande

VII. Acknowledgement

We appreciate Dr. Goutam Chakraborty , SAS® professor of marketing analytics and Dr. Miriam McGaugh, lecturer at Oklahoma State University and Nandan Moza, Sr. Data Analyst at Elevate for their guidance and support throughout the research.