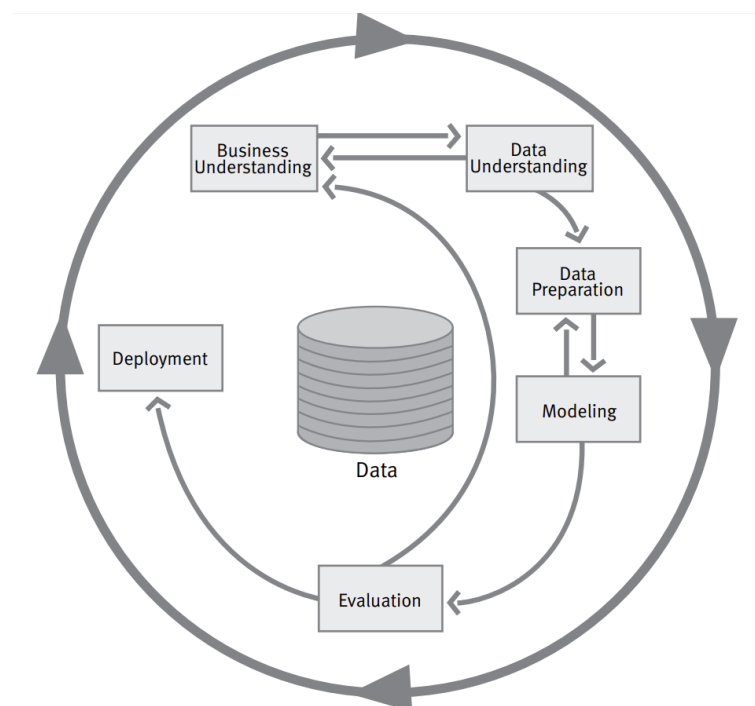


## **PROJECT ANALYSIS**

Moving forward, we will look at how text mining using topic modelling techniques can be adopted by Flashup to help them understand customers' needs and expectations, thereby improve their business branding by adopting the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework. Published by Daimler-Benz, ISL and NCR in 1999, the CRISP-DM framework was developed to provide a uniform framework and guideline for data miners (Shafique & Qaiser, 2014). It consists of six main phases namely Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment, and is widely used today for data mining, analytics, and data science projects (Hotz, 2024).

**Figure 7**

*CRISP-DM Framework*



### **1. Business Understanding**

Before beginning any analysis, it is important to understand what the business problem is about and establish a clear business and data mining objective that we hope to achieve through the modelling process.

As a new e-commerce startup brand, Flashup needs to ensure that the business has an overall better than average business branding to ensure its sustainability in the industry. Gaining insight from customer reviews is key to enhancing a company's branding, but this may prove challenging given the large volume of unstructured data in a variety of inconsistent formats that cannot be properly analysed using conventional methods. Recognising the potential struggles helps in determining reasonable goals that we hope to achieve through the results of the modelling process, thereby resolving our business problem.

Business Problem: The huge amount of unstructured online reviews complicates the processing of information, yet manual inspection of all reviews is impractical as it is very time-consuming.

Business Objective: To understand the reasons causing customers' negative shopping experience with Flashup so that suitable solutions can be implemented to resolve such issues, thus improving current and future shoppers' shopping experience and enhancing Flashup's business branding.

Data Mining Objective: Identify and uncover major and any latent themes among the many reviews posted by past customers using topic modelling, thus narrowing down the scope of changes that Flashup needs to focus on to improve customer satisfaction and ultimately, business branding.

## 2. Data Understanding

This stage focuses on gathering data, assessing its quality, and analysing it to gain insight and generate theories about information that may be buried. This stage helps to provide a clearer picture of what the dataset entails, what the features represent and if we need all or only some of them, and checking for data quality issues.

The original dataset was obtained from Kaggle titled "Womens Clothing E-Commerce Reviews". The dataset contains 23,486 records with 11 features – '#', 'Clothing ID', 'Age', 'Title', 'Review Text', 'Rating', 'Recommended IND', 'Positive Feedback Count', 'Division Name', 'Department Name', and 'Class Name'. Pre-processing of the raw dataset is required before scaling down to a sample-size dataset to facilitate faster modelling process. The explanation of the features are listed in Table.

**Table 2**

*Data dictionary of the original dataset*

Feature Name	Explanation
#	Review number
Clothing ID	Integer Categorical variable that refers to the specific piece being reviewed
Age	Positive Integer variable of the reviewers age
Title	String variable for the title of the review
Review Text	String variable for the review body
Rating	Positive Ordinal Integer variable for the product score granted by the customer from 1 Worst, to 5 Best
Recommended IND	Binary variable stating where the customer recommends the product where 1 is recommended, 0 is not recommended
Positive Feedback Count	Positive Integer documenting the number of other customers who found this review positive
Division Name	Categorical name of the product high level division
Department Name	Categorical name of the product department name
Class Name	Categorical name of the product class name

Firstly, some records were found to contain reviews at all. Since our analysis focus is on reviews, we remove these records that would not be able to provide any relevant insights entirely to prevent them from getting into our final dataset. Unlike structured data where we can sometimes use significant values to replace missing values, such as mean value or modal category, replacement of a missing unstructured data is not possible as they do not have any kind of distribution or basis that can be based on for substitution. In total, 845 of such records were removed.

### Figure 8

*845 records with no reviews*

	A	B	C	D	E	F	G	H	I	J	K
1		Clothing	Age	Title	Review T	Rating	Recomn	Positive	Division	Departm	Class Ne
94	92	861	23			5	1	0	General	PelTops	Knits
95	93	1081	31			5	1	0	General	Dresses	Dresses
100	98	1133	50			5	1	0	General	Jackets	Outerwear
137	135	861	35			4	1	0	General	PelTops	Knits
144	142	1126	35			5	1	0	General	Jackets	Outerwear
167	165	829	35			3	0	0	General	PelTops	Blouses
170	168	862	40			5	1	0	General	Tops	Knits
194	192	895	46			5	1	0	General	PelTops	Fine gauge
203	201	895	36			5	1	0	General	Tops	Fine gauge
250	248	1030	28			5	1	0	General	Bottoms	Jeans
267	265	878	63			4	1	0	General	Tops	Knits
300	298	1104	39			5	1	0	General	Dresses	Dresses
325	323	895	47			5	1	0	General	Tops	Fine gauge
344	342	836	32			5	1	0	General	Tops	Blouses
347	345	895	24			5	1	0	General	Tops	Fine gauge
402	400	1104	35			5	1	0	General	Dresses	Dresses
427	425	1083	50			5	1	0	General	Dresses	Dresses
457	455	234	33			5	1	0	Initmates	Intimate	Swim
495	493	1104	39			5	1	0	General	Dresses	Dresses
521	519	850	37			5	1	0	General	PelTops	Blouses
525	523	1104	39			1	0	0	General	Dresses	Dresses
529	527	850	25			5	1	0	General	PelTops	Blouses
538	536	984	53			5	1	0	General	Jackets	Jackets
544	542	984	56			5	1	0	General	Jackets	Jackets
556	554	1078	45			5	1	0	General	Dresses	Dresses
564	562	995	48			5	1	0	General	Bottoms	Skirts

Womens Clothing E-Commerce Revi

Ready 45 of 23486 records found Accessibility: Unavailable

Secondly, as the focus of this analysis is to understand customers' dissatisfaction, therefore we filter out records that do not recommend the product. We choose the records where the binary value in the 'Recommended IND' column is '0'. This leaves us with 4101 records, which constitutes to 18% of the remaining records.

**Figure 9**

*4,101 records that do no recommend the product*

	A	B	C	D	E	F	G	H	I	J	K
1		Clothing	Age	Title	Review T	Rating	Recomn	Positive	Division	Departm	Class Ne
4	2	1077	60	Some major	I had such	3	0	0	General	Dresses	Dresses
7	5	1080	49	Not for the	I love tracy	2	0	4	General	Dresses	Dresses
12	10	1077	53	Dress looks	Dress runs	3	0	14	General	Dresses	Dresses
24	22	1077	31	Not what it	First of all,	2	0	7	General	Dresses	Dresses
27	25	697	31	Falls flat	Loved the n	3	0	0	Initmates	Intimate	Lounge
28	26	949	33	Huge disap	I have been	2	0	0	General	Tops	Sweaters
35	33	949	36	Mehh	I ordered th	2	0	0	General	Tops	Sweaters
58	56	368	33		I am pregna	2	0	3	Initmates	Intimate	Intimates
59	57	862	31	Boring front	This tank fit	3	0	0	General	Tops	Knits
63	61	368	36	Itchy tags	3 tags sewn	1	0	0	Initmates	Intimate	Intimates
70	68	862	37	I wanted to	I really love	2	0	0	General	Tops	Knits
71	69	1078	56	Great sumr	I really want	3	0	1	General	PelDresses	Dresses
73	71	822	36	Short and b	Why do des	2	0	0	General	Tops	Blouses
79	77	850	28	Zipper brok	The zipper t	2	0	0	General	PelTops	Blouses
87	85	866	58		I was really	1	0	2	General	PelTops	Knits
90	88	845	38	Huge	Really cute	2	0	4	General	PelTops	Blouses
96	96	845	44		I usually we	1	0	0	General	PelTops	Blouses
97	97	861	44	Huge	I was very e	1	0	0	General	PelTops	Knits
98	99	861	33	Pernette he	I am in need	3	0	17	General	PelTops	Knits
102	103	822	23	Not a fan	The fabric fi	2	0	9	General	Tops	Blouses
103	104	863	51		Runs big an	2	0	0	General	Tops	Knits
109	110	861	44	Poor quality	This is so th	1	0	11	General	PelTops	Knits
113	114	1082	32	Disappoint	The	2	0	0	General	Dresses	Dresses
117	118	850	37	Awkward fit	First, the fa	2	0	0	General	Tops	Blouses
123	124	923	50	Lovely top,	I was so exc	3	0	2	General	Tops	Sweaters
141	144	829	39	Beautiful, b	My firned	3	0	9	General	Tops	Blouses

Womens Clothing E-Commerce Revi

Ready 41 of 22641 records found Accessibility: Unavailable

Lastly, to scale down the dataset, 650 records would be chosen for the modelling process. However, it is important to ensure that the chosen records are randomised to prevent any bias that may arise due to the order of the records. To do so, the records are assigned a random number between 0 and 1 using the RAND() function on Excel. They are then sorted in ascending order based on the assigned random number, before choosing the first 650 records. This helps to eliminate any possible biasness in the records.

**Figure 10**

*Arranged records in ascending order based on their assigned random number*

	Clothing ID	Age	Title	Review Text	Rating	Recommen	Positive	Fer	Division	Na	Departmen	Class Name	
22979	1068	53	Buttons at t	Let me star	2	0	13	General	Pet	Bottoms	Pants	0.000265	
20658	830	39	Lampshade	If you like tc	1	0	8	General	Tops	Blouses		0.000369	
4313	1081	29	Meh...	I have the	3	0	1	General	Dresses	Dresses		0.000855	
2494	1078	40	Not quite t	On the hanj	4	0	0	General	Pet	Dresses	Dresses	0.001003	
19332	871	47	Like the ide	Like the ide	3	0	1	General	Tops	Knits		0.001049	
17004	927	53	Just ok cow	Gorgeous c	3	0	0	General	Tops	Sweaters		0.001235	
9498	895	37	Odd fit	The color oi	2	0	2	General	Tops	Fine gauge		0.001441	
20923	854	40	Disappoint	I love tiny bi	1	0	2	General	Pet	Tops	Knits	0.001462	
8670	774	52	Not enough	I thought i r	3	0	1	Initmates	Intimate	Lounge		0.002279	
5107	1035	37	Pilcro has g	Pilcro jeans	2	0	7	General	Bottoms	Jeans		0.002609	
16090	1068	37	Not flatteri	The picture	1	0	0	General	Pet	Bottoms	Pants	0.002723	
330	844	41	Awkward fit	I had high h	3	0	0	General	Pet	Tops	Blouses	0.002757	
9436	1094	32	Muumuu	I was surpri	3	0	0	General	Pet	Dresses	Dresses	0.002938	
9254	820	35	Pretty but n	I was initial	3	0	0	General	Tops	Blouses		0.003173	
20781	1078	29	Eh	I love the cc	3	0	8	General	Pet	Dresses	Dresses	0.003402	
3102	907	48	Frumpy	Did not wor	2	0	0	General	Tops	Fine gauge		0.003881	
8704	829	40	Beautiful to	This top is g	2	0	17	General	Pet	Tops	Blouses	0.00407	
11895	854	39	Too low-cut	The pattern	2	0	4	General	Tops	Knits		0.004335	
2706	1095	33	Burlap mati	No, this is c	1	0	1	General	Pet	Dresses	Dresses	0.005171	
1886	1059	36	Super cute	I am just ov	4	0	0	General	Pet	Bottoms	Pants	0.005244	
12348	868	67	April fools?	I do not eve	1	0	0	General	Tops	Knits		0.005707	
2467	820	58	The fit is a li		3	0	20	General	Tops	Blouses		0.005928	
17059	821	44	Wanted to l	I saw the cc	2	0	0	General	Tops	Blouses		0.005995	
1695	862	40	The materi		1	0	0	General	Tops	Knits		0.006073	
13929	1047	69	Hawthorne	These are h	3	0	3	General	Bottoms	Pants		0.006388	
2748	868	36	Not good	I loved this	2	0	5	General	Tops	Knits		0.006461	
22222	819	49	This blouse		1	0	2	General	Tops	Blouses		0.006978	

### 3. Data Preparation

This stage emphasises on choosing and preparing the final data set. It also includes the cleaning and transformation of data to ensure that it is ready for modelling analysis.

Since we are only looking at reviews, we can remove all other columns other than 'Review Text'. Additionally, since we would be using *tm* package's *VCorpus* function on R Studio which only recognises the headers 'doc\_id' and 'text', we renamed the headers accordingly. This gives us our final scaled down sample-size dataset which is ready for analysis.

**Figure 11**

*Final scaled-down dataset to be used for analysis*

doc_id	text
22	First of all, this is not pullover styling, there is a side zipper. i wouldn't have purchased it if i knew there was a side zipper because i have a large bust and side zip
33	I ordered this 3 months ago, and it finally came off back order. a huge disappointment. the fit wasn't so much the issue for me. the quality of the wool is si
56	I am pregnant and i thought this would be a great sleep bra. it's soft and fits okay, but it has zero support or shape. i would only buy if you are a b cup or smaller.
57	This tank fit well and i loved the ruffle in the back and how it layed. but the front was not a good look and i will be returning it.
68	I really loved this top online and wanted to love it in person. it is soft and the patter is okay in person. the neckline is higher than i am used to. also, there are two
69	I really wanted this to work. alas, it had a strange fit for me. the straps would not stay up, and it had a weird fit under the breast. it worked standing up, but the n
97	I was very excited to order this top in red xs. so cute, but it was huge, shapeless and support thin! it had to go back. i should've looked at other reviews.
114	The design/shape of the dress are quite flattering, flirty and feminine. but.... there is no way that the dress i received is new. the color is a faded washed out red
178	The colors are vivid and perfectly autumnal but the fit is a mess. it was overall too large, the waistline curves up in the front and then falls into small pleats whic
184	Online, this looks like a great sweater. i ordered an xxsp and found that this sweater is much wider in the middle than pictured. in fact, i'm pretty sure they pinn
232	The shirt is absolutely cute looking. but that's just that....looks. i'm 5'2" size 6 & wears m to l top. i tried on the medium in the store, it was rather flowing. i liked
314	I thought this top was adorable in the store and online. it just didn't work for me. although it fit, it flares out too much in the front and just wasn't flattering on m
330	I had high hopes for this top. really boxy, short
376	I was looking for a basic tee, but this one was just ok....the quality is okay, but it is not as soft as i would have liked. unfortunately, i will be returning this item.
475	I bought this in the white, size m (140lbs., 5'8", 34b) because i wanted a baggy fit, which i got, so this is tts. the white is more of an off-white rather than a bright
498	Cropped and wide- would look cuter on someone who is more petite. was too wide and cropped for me. very sheer.
578	Was drawn to this dress in the store. my location had it displayed well (one mannequin with the dress & a jean jacket...another with the faux fur cardigan). the c
579	I was really excited about this top since i have others from retailer that are similar in styling. unfortunately, it just didn't work out. i am petite with a short torso i
588	Fits very strange. odd cut. does not look well on.
646	I waited three months for these pants. when they finally arrived, i was mostly pleased. they seemed like the staple pant everyone is raving about. and they shou
657	So i recognized the fact that this dress was likely intended more for lounging than for wearing out and about, but i once had a lot of luck with dresses of this pric
699	Don't buy this dress unless you are normally a medium or larger. order it one or two sizes smaller than your normal size. i ordered an xs and it's more like a med
704	The colors and print are absolutely beautiful and i'm quite heartbroken that this suit just doesn't fit. i don't have a long torso but the v was just so plunging that i
706	The overall styling was great, and the dress is super-cute, if a little thick (it's made of denim, after all). i ordered a size down from my normal, and even then, i w
744	Love the color, but the fit is just odd and frumpy. the sleeves are shorter than the should be. when i lift my arms, the entire sweater gets lifted. going back.

The necessary packages and the dataset are loaded into R studio as shown in Figure 12.

**Figure 12**

*Loading packages and dataset onto R*

```
library(tm)
library(SnowballC)
library(topicmodels)
library(slam)
library(sentimentr)
library(tidytext)
library(dplyr)
library(ldatuning)
library(LDAvis)
library(servr)

#loading text file and creating corpus for data frame
txtpath<-file.path("C:/Users/peizh/OneDrive/Desktop/SUS

txt<-read.csv(txtpath)

comments_corpus<-VCorpus(DataframeSource(txt))
comments<-comments_corpus
```

Data transformation is required in text mining analysis in order to obtain a quantitative representation that accurately captures the text's properties. This allows for the application of conventional statistical techniques just like with structured data, to extract insights (Zhang, 2024). To do so, we refer back to Figure 1, where we focus on Text Parsing, Text Filtering, and Transformation/Vectorisation.

Text Parsing involves extracting, cleaning, and producing a dictionary of words from the corpus using Natural Language Processing (NLP) by recognizing sentences, breaking them up into individual tokens, figuring out speech components, eliminating stop words, and stemming words back to its root form, including spelling and grammatical errors. Despite the removal of stop words in the Text Parsing stage, there may still contain many unnecessary frequent occurring words which depending on the context of the business, may not provide any useful

information. For example, a dataset obtained from a fashion business may contain frequent terms like “cloth”, “pants”, “shirt”, “dress”, while a dataset obtained from a food business may contain frequent terms like “food”, “drinks”, “ingredients”, “tastes”. These terms are likely to appear very often in the context of their business domain that analysing them would not provide any useful information and be rather meaningless. Consequently, in order to exclude frequently appearing words according to their business domain, another round of term removal is required under the Text Filtering stage. This would necessitate prior domain knowledge in order to determine which words to delete. Simultaneously, common-occurring phrases may be identified by doing a term count and then removed based on a specified threshold. Therefore, stop words are frequently eliminated several times during the topic modelling analysis because, after reviewing each output, more stop words are typically added to the list for removal if they are deemed redundant or insignificant to the business problem (Zhang, 2024) (see Appendix A1 for code on text parsing and filtering in R for the dataset). In the code, the stop words that were first removed were using the common English stop words available in the tm package. Additional stop words that have a count of more than a hundred were later added into a custom stop words list for another round of removal.

Using techniques based on linear algebra, the text transformation phase entails numerical representation of the terms in relation to the document. The remaining terms are converted to a vector format commonly known as a document-term matrix or term-document matrix. The terms that appear often in a set of documents are outlined by this mathematical matrix. A term-document matrix has each row corresponding to a term and each column to a document, whereas a document-term matrix has each row corresponding to a document and each column to a term. Both formats are acceptable for use (see Appendix A2 for code on generating Document-Term matrix in R for the dataset).

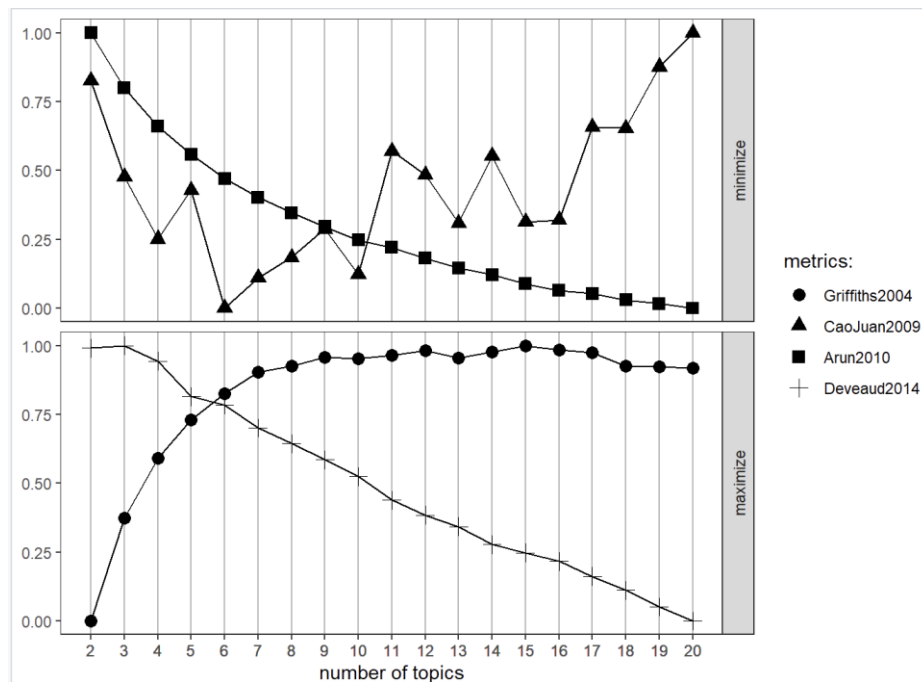
#### 4. Modelling

Modelling analysis begins here where we use the LDA algorithm to do topic modelling. A common challenge in executing topic modelling analysis is choosing the right k-value to generate the number of topics. Fortunately, certain R packages like ldatuning do provide us with metrics that we can follow to narrow down our range of k-values (see Appendix B1 for code on generating graph for k-values) instead of random trial-and-error which may be time-consuming and impractical (Nidhi, 2017).

### **Figure 13**



*Plot to determine range of k-values*



As shown in Figure 13, the top graph's goal is to minimise the score to get the optimal number for k-value, while the bottom graph's goal is to maximise the score for the same purpose. We see that other than Deveaud2014's metric which is a constant decrease as the k-value increases, the other three metrics generally plateau around a range of 6-10 for the k-value. Thus, we test out a range of k-values between 6-10 (see Appendix B2-B3 for code on generating number of topics, k). The topics generated by the model using Gibbs sampling method for the respective values of k are shown in the following figures.

**Figure 14**

*Topics generated when k=6*

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
[1,]	"howev"	"materi"	"disappoint"	"good"	"didnt"	"flatter"
[2,]	"beauti"	"thin"	"even"	"bit"	"cut"	"short"
[3,]	"normal"	"onlin"	"retail"	"seem"	"big"	"front"
[4,]	"bodi"	"thought"	"button"	"person"	"dont"	"pretti"
[5,]	"usual"	"nice"	"first"	"expect"	"bought"	"cute"
[6,]	"still"	"cute"	"wide"	"store"	"pattern"	"shoulder"
[7,]	"need"	"see"	"got"	"feel"	"bad"	"hip"
[8,]	"line"	"side"	"perfect"	"littl"	"sad"	"length"
[9,]	"area"	"felt"	"show"	"excit"	"sale"	"long"
[10,]	"blous"	"blue"	"littl"	"poor"	"though"	"wasnt"

**Figure 15**

*Topics generated when k=7*



	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7
[1,]	"blous"	"disappoint"	"beauti"	"materi"	"usual"	"cute"	"howev"
[2,]	"black"	"even"	"littl"	"retail"	"think"	"didnt"	"big"
[3,]	"still"	"pattern"	"felt"	"pretti"	"short"	"flatter"	"length"
[4,]	"store"	"thought"	"button"	"thin"	"see"	"bought"	"feel"
[5,]	"poor"	"first"	"seem"	"front"	"bodi"	"receiv"	"normal"
[6,]	"bad"	"onlin"	"good"	"bit"	"expect"	"even"	"medium"
[7,]	"nice"	"boxi"	"cut"	"line"	"need"	"rather"	"shoulder"
[8,]	"sheer"	"person"	"hope"	"long"	"nice"	"show"	"smaller"
[9,]	"hole"	"store"	"blue"	"sad"	"side"	"super"	"might"
[10,]	"dont"	"sale"	"dont"	"wide"	"though"	"larger"	"issu"

**Figure 16**

*Topics generated when  $k=8$*

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
[1,]	"thought"	"beauti"	"store"	"nice"	"think"	"materi"
[2,]	"good"	"flatter"	"onlin"	"cut"	"front"	"cute"
[3,]	"pattern"	"big"	"dont"	"blous"	"bit"	"disappoint"
[4,]	"bad"	"blue"	"pretti"	"length"	"short"	"button"
[5,]	"got"	"boxi"	"felt"	"hip"	"see"	"side"
[6,]	"didnt"	"hole"	"even"	"seem"	"sheer"	"receiv"
[7,]	"poor"	"say"	"first"	"perfect"	"sad"	"long"
[8,]	"odd"	"brand"	"person"	"see"	"complet"	"super"
[9,]	"wasnt"	"shoulder"	"though"	"differ"	"pair"	"sale"
[10,]	"slight"	"dri"	"hope"	"might"	"issu"	"still"

	Topic 7	Topic 8
[1,]	"retail"	"howev"
[2,]	"didnt"	"thin"
[3,]	"usual"	"even"
[4,]	"short"	"bodi"
[5,]	"bought"	"littl"
[6,]	"feel"	"medium"
[7,]	"item"	"definit"
[8,]	"need"	"normal"
[9,]	"sure"	"quit"
[10,]	"rather"	"smaller"

**Figure 17**

*Topics generated when  $k=9$*

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
[1,]	"usual"	"bit"	"didnt"	"medium"	"feel"	"even"
[2,]	"cute"	"bought"	"big"	"short"	"flatter"	"thin"
[3,]	"see"	"length"	"sad"	"littl"	"beauti"	"onlin"
[4,]	"seem"	"front"	"short"	"long"	"side"	"person"
[5,]	"dont"	"shoulder"	"materi"	"thing"	"got"	"good"
[6,]	"cant"	"super"	"blue"	"slip"	"bodi"	"nice"
[7,]	"thought"	"though"	"beauti"	"pretti"	"pretti"	"larger"
[8,]	"quit"	"expect"	"bad"	"think"	"complet"	"boxi"
[9,]	"wide"	"felt"	"stretch"	"bad"	"area"	"littl"
[10,]	"littl"	"pair"	"end"	"found"	"show"	"retail"

	Topic 7	Topic 8	Topic 9
[1,]	"materi"	"disappoint"	"howev"
[2,]	"store"	"retail"	"cut"
[3,]	"first"	"normal"	"blous"
[4,]	"line"	"still"	"button"
[5,]	"wasnt"	"pattern"	"hip"
[6,]	"felt"	"definit"	"need"
[7,]	"disappoint"	"item"	"perfect"
[8,]	"receiv"	"awkward"	"black"
[9,]	"excit"	"hope"	"flatter"
[10,]	"rather"	"leg"	"poor"

**Figure 18**

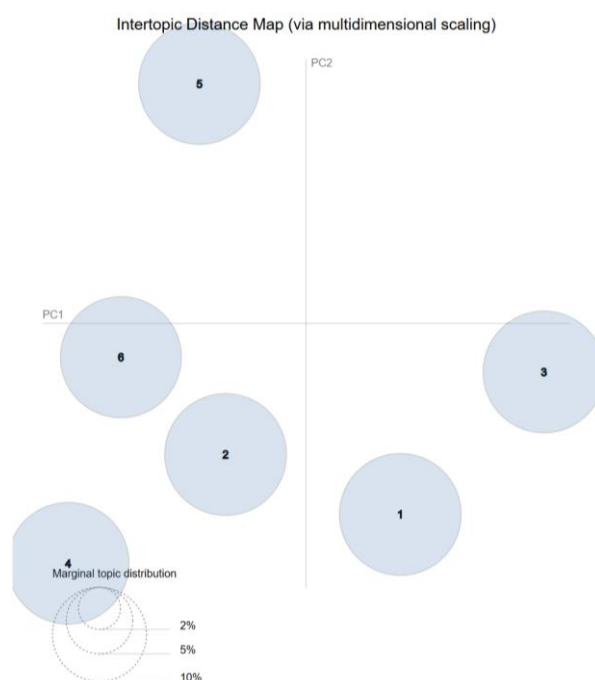
*Topics generated when  $k=10$*

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
[1,]	"beauti"	"materi"	"short"	"even"	"disappoint"	"nice"
[2,]	"usual"	"cute"	"person"	"bought"	"store"	"feel"
[3,]	"cute"	"long"	"thought"	"bad"	"sad"	"think"
[4,]	"seem"	"pretti"	"hip"	"medium"	"first"	"side"
[5,]	"dont"	"use"	"bodi"	"sheer"	"blous"	"need"
[6,]	"see"	"doesnt"	"cant"	"issu"	"shoulder"	"thing"
[7,]	"though"	"quit"	"see"	"shapeless"	"expect"	"bit"
[8,]	"wide"	"felt"	"good"	"super"	"flatter"	"took"
[9,]	"light"	"hit"	"length"	"second"	"stretch"	"wish"
[10,]	"smaller"	"arriv"	"usual"	"tent"	"pair"	"pattern"
	Topic 7	Topic 8	Topic 9	Topic 10		
[1,]	"materi"	"retail"	"didnt"	"howev"		
[2,]	"big"	"thin"	"onlin"	"cut"		
[3,]	"front"	"line"	"button"	"item"		
[4,]	"good"	"wasnt"	"littl"	"perfect"		
[5,]	"flatter"	"felt"	"got"	"black"		
[6,]	"bit"	"boxi"	"disappoint"	"normal"		
[7,]	"larger"	"definit"	"still"	"flatter"		
[8,]	"pretti"	"hope"	"might"	"slip"		
[9,]	"rather"	"expect"	"length"	"sure"		
[10,]	"wide"	"receiv"	"poor"	"show"		

Although we managed to narrow down our range of  $k$ -values, we still need to decide on one final  $k$ -value to obtain the latent number of major topics uncovered from the corpus. In the paper written by Quarshie et al. (2024), they used an inter-topic distance map to justify the  $k$ -value to use. The map helps users to visualise how similar or different each topic is from each other based on the distance between their bubble, and the size of each bubble represents its relative prevalence in the corpus, where a larger bubble indicates that the topic appears more frequently across the documents. We apply the same technique to Flashup's data and generate the inter-topic distance map for the range of  $k$ 's from 6-10 (refer to Appendix C1-3 for the codes on generating inter-topic distance maps for  $k$ -values of 6-10).

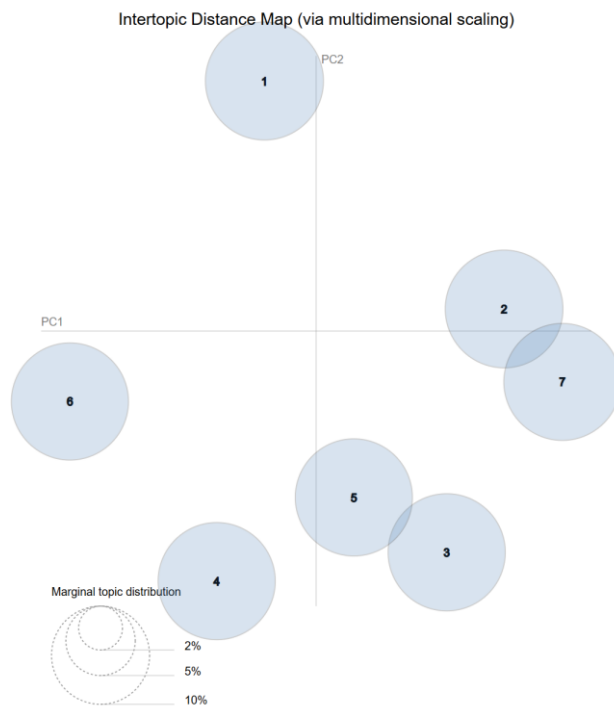
**Figure 19**

*Inter-topic distance map when  $k=6$*



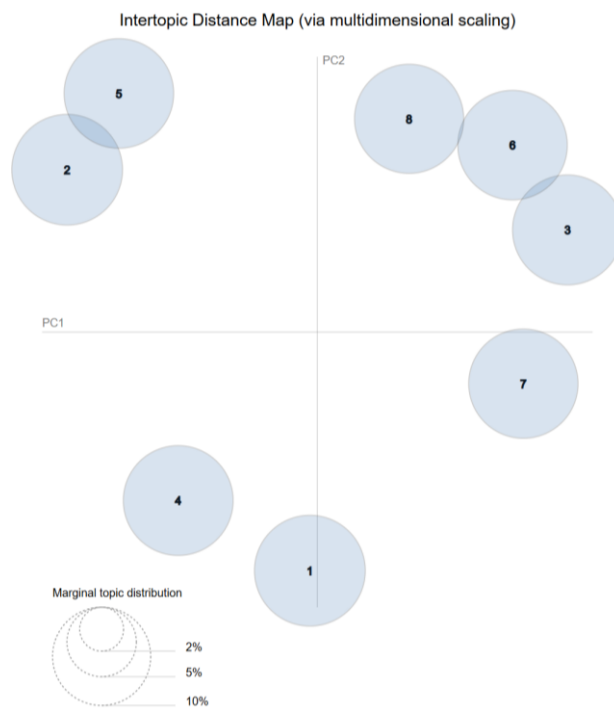
**Figure 20**

*Inter-topic distance map when  $k=7$*



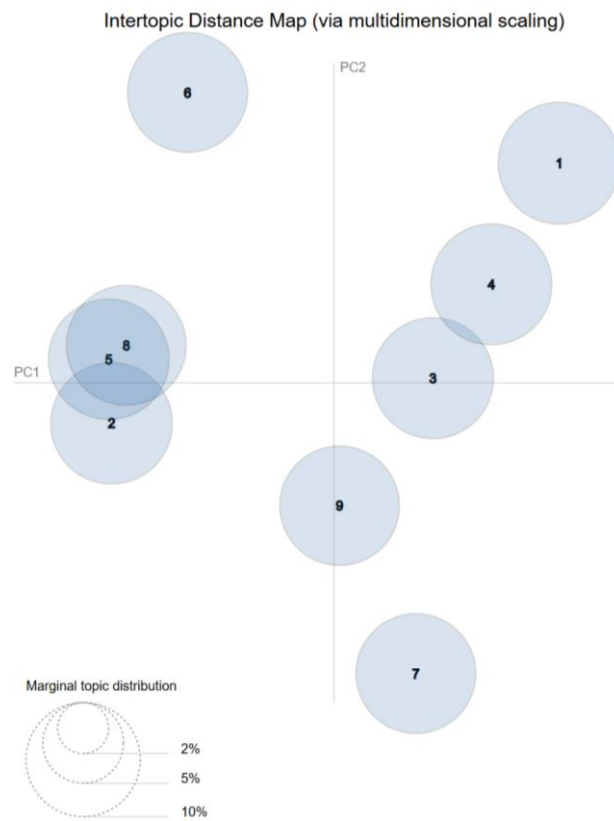
**Figure 21**

*Inter-topic distance map when  $k=8$*



**Figure 22**

*Inter-topic distance map when  $k=9$*



**Figure 23**

*Inter-topic distance map when  $k=10$*



Looking at the figures above, it is clear that when the k-value is 9 and 10, there are significant overlapping of a few topics. Considering that the purpose of topic modelling is to uncover as many latent themes as possible, we choose our final k-value to be 8.

Finally, we also look at the keywords generated when k=8 by the VEM sampling method to see if there are any differences in the topics produced between the two sampling methods.

**Figure 24**

*Topics generated when k=8 using VEM sampling method*

```
> terms_vem
      Topic 1      Topic 2      Topic 3      Topic 4      Topic 5      Topic 6      Topic 7
[1,] "retail"      "materi"    "materi"  "usual"    "big"      "littl"  "didnt"
[2,] "flatter"     "flatter"  "short"   "nice"     "flatter"  "materi" "beauti"
[3,] "feel"        "short"    "howev"   "materi"   "nice"     "howev"  "good"
[4,] "onlin"       "disappoint" "cute"    "bit"      "disappoint" "beauti" "person"
[5,] "side"        "thin"     "feel"    "retail"   "bad"      "even"   "even"
[6,] "disappoint"  "cut"      "think"   "dont"     "front"    "bought" "cute"
[7,] "howev"       "big"      "onlin"   "cut"      "onlin"    "feel"   "disappoint"
[8,] "shoulder"    "cute"     "beauti"  "big"      "black"    "short"  "thought"
[9,] "seem"        "didnt"    "button"  "blous"    "good"     "cute"   "dont"
[10,] "slip"       "bought"   "didnt"   "pretti"   "seem"     "see"    "sad"

      Topic 8
[1,] "littl"
[2,] "even"
[3,] "howev"
[4,] "materi"
[5,] "didnt"
[6,] "retail"
[7,] "cute"
[8,] "normal"
[9,] "usual"
[10,] "big"
```

## 5. Evaluation

“Even though topic modelling is an unsupervised machine learning technique, human efforts are still required to compare results generated by different model specifications to get meaningful topics.” (Zhang, 2024). Using the keywords listed, the following topic names for the 8 topics generated by both sampling methods are shown in Table 3.

**Table 3**

*Topic names*

Topic	VEM	Gibbs
1	Retail Experience	Product Impressions
2	Material Quality	Aesthetic Appeal
3	Product Expectation	Shopping Experience
4	Customer Satisfaction	Fit and Comfort
5	Product Features	Product Expectation
6	Minor Complaints	Material Quality
7	Perceived Value	Retail Comparison

8	General Observations	Sizing Variability
---	----------------------	--------------------

Gibbs method is chosen as the final model as the topics generated seems to be more specific towards the fashion domain. A topic distribution table, gammaDF, for each document is generated.

**Figure 25**

*GammaDF table*

	1	2	3	4	5	6	7	8
22	0.12692308	0.11153846	0.09615385	0.14230769	0.11153846	0.18846154	0.12692308	0.09615385
33	0.14236111	0.10069444	0.08680556	0.11458333	0.15625000	0.15625000	0.14236111	0.10069444
56	0.11328125	0.16015625	0.12890625	0.11328125	0.16015625	0.09765625	0.11328125	0.11328125
57	0.13181818	0.11363636	0.11363636	0.11363636	0.16818182	0.13181818	0.11363636	0.11363636
68	0.11885246	0.13524590	0.13524590	0.10245902	0.11885246	0.13524590	0.13524590	0.11885246
69	0.12719298	0.14473684	0.12719298	0.12719298	0.12719298	0.10964912	0.10964912	0.12719298
97	0.14473684	0.12719298	0.10964912	0.14473684	0.10964912	0.14473684	0.10964912	0.10964912
114	0.13214286	0.11785714	0.13214286	0.11785714	0.08928571	0.13214286	0.14642857	0.13214286
178	0.11956522	0.10507246	0.11956522	0.14855072	0.13405797	0.11956522	0.09057971	0.16304348
184	0.13983051	0.10593220	0.13983051	0.12288136	0.13983051	0.10593220	0.12288136	0.12288136
232	0.12313433	0.09328358	0.15298507	0.09328358	0.10820896	0.15298507	0.16791045	0.10820896
314	0.16228070	0.12719298	0.14473684	0.10964912	0.12719298	0.10964912	0.10964912	0.10964912
330	0.11792453	0.11792453	0.13679245	0.11792453	0.13679245	0.11792453	0.13679245	0.11792453
376	0.13181818	0.11363636	0.11363636	0.11363636	0.13181818	0.11363636	0.15000000	0.13181818
475	0.18150685	0.11301370	0.08561644	0.11301370	0.16780822	0.11301370	0.11301370	0.11301370
498	0.13425926	0.11574074	0.11574074	0.13425926	0.13425926	0.11574074	0.13425926	0.11574074
578	0.13805970	0.10820896	0.13805970	0.12313433	0.12313433	0.12313433	0.13805970	0.10820896
579	0.11507937	0.13095238	0.09920635	0.09920635	0.13095238	0.11507937	0.17857143	0.13095238
588	0.13942308	0.12019231	0.12019231	0.13942308	0.12019231	0.12019231	0.12019231	0.12019231
646	0.13214286	0.14642857	0.14642857	0.14642857	0.10357143	0.10357143	0.13214286	0.08928571

To evaluate if the topic names are representative of the documents, we look at the actual top 5 reviews for each discovered topic, which can be deduced based on the gamma values (see Appendix D1-8 for the actual Gamma values for the top 5 reviews).

**Figure 26**

*Top 5 texts for Topic 1*

doc_id	text
1483	I wanted this tank to work so badly! it was a great deal and so beautiful. unfortunately the top fitted portion ends right below my chest then the shirt flares out so i looked like a cupcake. my belly also showed slightly because the shirt is short. if you have a short torso, this would work great. if not, unfortunately, its not great - tear
2331	Color is beautiful but looks like a maternity top on. i found i could tuck it in slacks and it looked fine. now for the cons: it is supposed to be peacock design but i had only one peacock, under my arm, which could be seen, i guess, if i waved. the rest were parts of the bird. i would love to have the one in the photos! next, who thought cheap white buttons on the back was a good idea? if i had gotten a nice patterned top, i would have replaced the buttons. blouses that i like never seem to b
2734	The print of this skirt is so fun. i love how it's styled on the model, and i could see it being very versatile. i decided not to keep it because, mainly, it is pricey and i'm afraid it'll get easily damaged. if i would wear it, i think i'd be worried about sitting on a dirty chair or spilling something on my lap and it being really visible-- and then having to upkeep it through dry cleaning. the fabric's nice and good quality. but honestly, if the skirt were cheaper and the fabric were cotton a
10322	I had mixed feelings about this dress. the material is lovely but the skirt is really odd. it has two slips, one that fits close and a second that's a bit too puffy. so the skirt didn't sit well and i thought it made my hips look bigger. the odd length made it unflattering. it might be nicer without the puffy slip but i didn't want to risk it.
19106	Too much of a good thing is...still too much. there's too much flounce on this jacket for me--the waist sits slightly high, and the incredible amount of fabric beneath it (when you twill around, it's basically a full circle) makes my hips look huge. if there were about 1/3rd (not 1/2, but 1/3) the amount of fabric to the skirt, i'd be totally in love. as it is, i'll be returning it.

**Figure 27**

*Top 5 texts for Topic 2*

doc_id	text
5540	I bought this top in the blue motif and loved it so so very much until i washed it. it came out of the wash (delicate/cold cycle) with 3 small holes. i am devastated because it was destined to be my favorite top for summer. it is flattering, lightweight, gorgeous color, cool raw edges, v-back detail...i could go on and on about all the things i love about this top, but now it has
7527	This sweater was so itchy i immediately was turned off. it looks cute in pictures but another problem was that the knitting in the shoulders looks like a kid made it. it was uneven and stretched out. it looked like it was not well made to me. it also shed so much that just from me trying it on my pants were covered in orange fuzz. i returned this one.
12602	I really wanted to love this dress, but i couldn't. it ran large and heavy. rather than a sleek cocktail dress, it felt more like a heavy sack once i put it on. there was one upside: the drape in the back did not appear to drop quite as low, so finding a bra to wear under the dress might be easier than if it draped as low as it appears in the picture.
16476	I looked at this dress for a long time before i purchased it because of the money. i wear a size 2p. was entirely disappointed in the style and construction of this dress. for one thing the lining is bulky and you could never wear it in warmer weather without burning up. the dress as pictured looks very feminine and light weight silk but there is just too much material going on--- thick
20404	I am 5'2" and i typically wear a 2 or a 2p. lately i've been finding that petite is just a little too short to wear with wedges so i often order a 2 as i did in this dress. it fits well and it's flattering. my two complaints are that it gets wrinkly fast and looks sloppy. yes, linen is a wrinkly look and that's normal for linen but somehow this dress just looks sloppy to me. more importantly, while wearing for the first time, i realized that the seams on the side were very lightly sewn and there

**Figure 28**

*Top 5 texts for Topic 3*

doc_id	text
794	I was excited to see this top in person, but once i got a close up look at it (and also tried it on) i was disappointed. first, the color is much more persimmon orangey in person and that's a hard color to pull off for a lot of people. it just pulls any pink or red in your skin and highlights it.
6390	second, the top was poorly finished. i had to pick up two to try on because the lace was wonky across the front of the first one i The dress is really frumpy standing up. something about the bottom tier of lace just turns it into a frumpy maw-maw nightgown. the lace is really pretty, though
12045	This top was so pretty online! unfortunately it did not work for me- the front was a bit more shapeless in person, though the back was gorgeous, with a button gather at center. the color appeared yellow online, but was a bit more green in person. also, the armholes were very low, so it would need a camisole or no bra.
15978	I bought a pair of these in stone and after the first wash they become all splotchy. i chalked it up to perhaps there was bleach leftover in the my washer. although nothing else in the load was blotchy. i then, because i loved the cut of these pants, bought another pair in the darker blue color. the same thing happened after the first wash. again, nothing else in that wash load was the least bit ruined -- just my new retailer pants! such a bummer-- definitely a dye issue. i called the stores i pur
17359	I own the essential slims and love them, so i figured i could order these in my normal size and they'd fit great. i was way off. they were uncomfortably tight in the calves (which is understandable - i'm a runner so my calves are disproportionately large) and uncomfortable in the waist. i decided to order one size bigger and pretty much had the same problem in the calves. the high waist was also still unflattering and uncomfortable. these might be great for hour glass shape with skinny calves. t

**Figure 29**

*Top 5 texts for Topic 4*



doc_id	text
4627	I purchased the blouse in both the ivory and black. the black i decided to keep because it will go with anything. i was hoping the ivory blouse would work but it's thin and see through. if someone could please tell me where i could find a casual, white or ivory blouse that is not entirely see through i would greatly appreciate it!
9646	Ugh. i wanted to love this dress so much. i ordered it for a fall wedding and was so excited about it. unfortunately the cut is so very different from how it is pictured on the model. the color and print are the same, the detailing around the next are even more beautiful than pictured. the length, however is so much longer then pictured. and it's not just an issue of the model being tall. the slip is the perfect length (and it's actually the perfect slip), but the overlay comes down way past my
9729	I have loved the pilcrow hyphen chinos, and have purchased many pairs. however, the last 3 pairs i've ordered are all a different cut now. they are no longer the chino i loved, although they have the exact same name. they are now "boy cut", and are much more narrow through the hips and thighs. nowhere on the description does it say "boy cut". it's not until the package arrives in the mail that i'm saddened to find this changed product with the same number. it really should say this int he descri
11827	I ordered this blouse as a gift for my mom. it is too pricey to live up to my expectations. it is quite sheer and the hem of the sleeves are unfinished which you can't see in the pictures. this makes a very expensive blouse look cheap. it is made of viscose and dry clean only. the material used is not on par with the price in my opinion. if it had been real silk or another
21223	I love the designs of tiny blouses. but i wish they could make them more durable. i have owned several blouses in the past where i have to use a little pin to secure the button line from pulling at the breast area, or that after a few delicate washings have come apart at a seam. and ironing or steaming them is a must. this blouse was no exception. it went back because of pulling at the bust. i'm willing to take some extra care of my clothing out of my busy schedule, but i shouldn't have to! 5'7"

**Figure 30**

*Top 5 texts for Topic 5*

doc_id	text
2617	I tried this top on with a pair of white shorts, half tucked, and it looked really classy and nice. the fit was loose but flattering, it is a tiny bit sheer, but i was wearing a magenta bra, so perhaps i am biased. overall, i think this is a good basic addition, to a summer wardrobe, can be worn with a nice skirt, or with jeans as shown here. the 0 (or xs, can't remember) fit me at 115 lbs,
5803	I couldn't wait to try this on this sweater. sadly, it was too large in all ways. the sleeves are adorable, but the ruffle overwhelmed my arms. this was much longer on me than the model. it stopped mid-thigh. the quality seemed fine, but it is a
17277	The white is completely sheer. i am so tired of see through shirts!! even at the sale on top of sale prices, i wouldn't recommend this shirt. if the fabric weren't sheer it would be a very comfy shirt with a cute back. but it just looks silly with a tank top under
21673	Saying that these pants are comfortable would be redundant. they are over & above in comfort. i wear a size 8; and they are right on in sizing. i want to comment on the color/print. to see the green, i had to take the pants by a window with natural light and look closely. that being said, i think the pants take on the color of top one chooses. i walked through my closet with the pants in hand. they match with light grey, periwinkle blue, light sky blue, navy, white, your choice. i do prefer a so
22523	The design of these ankles pants is very cool and the rise is perfect. however, the fabric, despite 2% spandex, has absolutely no give from the knee down. the extensive stitching probably contributes to the lack of stretch. while i was able to pull them on (size 27 for 110 lb/5.4'), i could not bend my legs at the knee comfortably. with longer wear, i think they would cut off the circulation.

**Figure 31**

*Top 5 texts for Topic 6*

doc_id	text
	<p>First of all, this is not pullover styling. there is a side zipper. i wouldn't have purchased it if i knew there was a side zipper because i have a large bust and side zippers are next to impossible for me.</p> <p>second of all, the tulle feels and looks cheap and the slip has an awkward tight shape underneath.</p>
22	<p>not at all what is looks like or is described as. sadly will be returning, but i'm sure i will find something to exchange it for!</p> <p>I loved this skirt when i tried it on. the material is soft with a bit of stretch. i'd say this runs a little large. for reference, i'm usually a size 2/4 in retailer and the 2 was still a little large towards the end of the day as it stretched out.</p>
5302	<p>however, my disappointment was with the buttons.i wore it for a day when all of a sudden the top button popped off. fine, that button is the most used. i sewed it back on myself and even reinforced the top three buttons that i use to take on and of</p>
9082	<p>I ordered a small, which is my usual retailer size, but this was like a tent on me. it's fine if you like things super super blousey but that wasn't what i was looking for. also it is very long, a very weird amount of long. also there is only eyelet on the front. the whole back of the shirt is just plain fabric. i like the idea of this shirt but all in all it was very disappointing.</p>
15448	<p>I have blazers in size 8 from this designer and they fit with room to spare. i understand this is a more fitted style, but it is still way too small. size 10 and size 12 were not much bigger. great colors and cute design, but just too small even if you do not plan on buttoning it. i hope they bring back this jacket but make it true to size.</p>
15630	<p>I was hoping that this would be a nice neutral layering staple but it simply leaves a lot to be desired. it fits tts, length is fine...the inside is some, but the outside darker material is stiff and cheap (kinda canvas feeling?). the material is super stiff which makes it flare out like like a tent in a super unflattering way. it also is on the thinnish side - it looks in the model shot like</p>

**Figure 32**

*Top 5 texts for Topic 7*

doc_id	text
6696	<p>I love deletta and regularly search for their items here on retailer. however, this one is a miss as i think it runs small or you need to be taller and thinner then i for the gathering, opening to hit you right. i did really like the color ( ordered the brown) and</p>
8223	<p>I usually love bordeaux so i was very disappointed with this shirt when it arrived. the fabric smelled plastic-y which is a first for me with retailer products. when i put the shirt on, i could see it had that sheen that comes from rayon spandex blend shirts, i hadn't noticed that it had spandex in it when i ordered it. this fabric looks and feels exactly like the cheap blends sold at</p>
9000	<p>I read the existing reviews, before purchasing, and a few mentioned that this sweater ran large. this has been my experience with most tops i have purchased from retailer. i figured by ordering a medium, i would receive a slightly loose to moderately loose fitting sweater. this was not the case! the sweater is short, does not look anything like it does when you view the model from behind (and i assume the model is taller than i...as models are usually tall). for reference, i am 5'6", medium build</p>
10175	<p>This hoodie is beautiful, soft and flattering. but the soft fabric pilled badly after one wearing and it's probably going to look ratty after a few washings. i might accept pilling in a \$30 sweatshirt but not at this price. i don't really want to take it back but i will. (i've complained to retailer employees in the past about other clothes pilling and poor quality, and they've always said i could</p>
19099	<p>I've been wearing these shorts for years, i work outside and it was great to have comfortable jean shorts that were just the perfect length. the fabric quality on these shorts (and many many other retailer items) has gone down. sure, they're great when you try them on in the store..soft, comfortable, flattering, but when these are worn for a day they get so stretched out that i feel like i need a belt to keep them on. i wore these several times, sat down on a relatively smooth concrete wall, and a</p>

**Figure 33**

*Top 5 texts for Topic 8*

doc_id	text
8293	Just piping in here -- ordered my usual size of small petite. the slip that came with the dress is about a size negative zero, it could hardly squeeze over my body and the dress itself is a bright pale aqua and it is a shift. and because of the smocking it hangs very strangely. i think it looks very cheap and is ill fitting. i would say if you are a person on which shift dresses look
10219	I wanted to love this because it's all my favorites: simple, chic design; pure silk; gorgeous color. in reality, the dress was 100% see-through, very thin, and poor quality silk. i posted pictures of me wearing this dress with a nude slip on underneath that i brought along for shopping elsewhere for business dresses. despite the thin quality of the silk, i was still considering spending this dress (despite it's unreasonably high \$\$\$ for the quality), but after trying it on i was really disappoint
12544	I tried wearing it, however, i thought i needed courage to wear this tank. unfortunately, i did not purchase this tank although i liked color and design. first, it runs small than the regular size. it feels even smaller because my size is xs. in addition, it tends to exaggerate a body line. i am not quite sure why this happens, but maybe it is just the way this tank is. if you are confident about your upper body posture, then yes, it would look very cool. however, if you have thin postur
14333	I'm disappointed in these pants, which looked so beautiful online and even have pockets! in person the pattern and colors aren't as nice, the fabric is too thin, and the pants look cheap. the pockets are pretty shallow. the fit isn't good either. i ordered up (i'm between sizes) and they're still not as flowy as i'd have liked. these are too pricy as lounge pants for me to be on the fence about them, so they're going back. if you buy them knowing about the thin fabric and snug fit, they might wo
18925	I thought it would be heavier than it is. it's thin and more delicate than what i would expect for a winter sweater (it contains wool so not really suitable for spring/summer). it's also shorter than it appears on the model and a little wider/boxier too.

After studying the actual reviews, while some of the topic names are representative of the reviews, others may not be so accurate and may need longer names to further elaborate how it represents the reviews as a whole. Nonetheless, Flashup can still proceed to adopt these topic names to implement strategies that can promote their business branding.

## 6. Deployment

Based on the earlier findings, the following solutions are suggested to Flashup to help improve their business branding:

- (1) Improving on product material, packaging, and details etc., would help to further improve customers' impression of the product as a whole.
- (2) Staying relevant to current fashion trends, prioritising quality and materials, and a good combination of colour palette etc., can increase the aesthetic appeal of the products to attract even more customers.
- (3) Improving website interface and service of online Q&As, and providing more details about the product online would greatly enhance consumers' shopping experience on the website, encouraging them to return in the future.
- (4) Study should be done to generalise general clothing sizes to fit the consumers based on their demographics as consumer body type may vary from region to region, thus sizes that fit the target audiences' demographic is key to improving fit and comfort.

- (5) Flashup should attempt to incorporate more product details such as materials, close up photos and videos on the listing so that consumers can have more realistic expectations of the product, thus avoiding disappointment when receiving the actual product.
- (6) Flashup should ensure that the materials they use for their products are safe and comfortable to use and fit regulations as comfort is one of the most important factors in terms of clothing.
- (7) Ensuring that their products and services are on par with physical retail shops is essential to convince customers that online shopping has no difference from physical shops at all.
- (8) Flashup can consider offering more variety of sizes on their products on top of the standard sizes to fit 'outlier' customers. That way, more customers would be intrigued to purchase from Flashup.

## **SUMMARY**

To conclude, topic modelling is an excellent way of revealing latent themes among a huge corpus of documents, which makes it easier to process unstructured data as compared to manual inspection. Flashup should focus on improving customers' satisfaction based on the 8 topics that were uncovered and monitor to see if customers' satisfaction improves a not by comparing the number of negative recommendations given by customers before and after the implementations.

There are definitely limitations to the experiment. The corpus that was analysed comprises of negative recommended reviews as a whole across Flashup, not taking into account the different types of clothing that Flashup offers. The difference in clothing category may have an impact on the keywords uncovered for each topic in the analysis, which may as a result produce a different combination of bag of words for each topic uncovered, thereby influencing the final topic naming convention. If topic modelling was done for each category of clothing that Flashup provides, it may be able to give rise to more targeted insights for each category, thereby improving customers' dissatisfactions closer to the root.