

Customer Feedback Reviews : Clothing Business

ความสำคัญของ Customer Review ที่มีต่อธุรกิจ

การรีวิวจากลูกค้ามีความสำคัญอย่างยิ่งต่อธุรกิจในปัจจุบัน ด้วยเหตุผลหลักดังต่อไปนี้

1. ส่งผลต่อการตัดสินใจซื้อของลูกค้า

รีวิวจากลูกค้าถือเป็นข้อมูลที่มีน้ำหนักและมีอิทธิพลต่อการตัดสินใจซื้อสินค้าหรือบริการของลูกค้ารายใหม่ คนส่วนใหญ่มักอ่านรีวิวก่อนตัดสินใจซื้อเสมอ รีวิวที่ดีจะช่วยสร้างความมั่นใจและจูงใจให้ลูกค้าตัดสินใจซื้อ

2. เป็นช่องทางการสื่อสารกับลูกค้า

รีวิวเปิดโอกาสให้ลูกค้าแสดงความคิดเห็น ข้อเสนอแนะ หรือข้อร้องเรียนต่อธุรกิจ ได้อย่างตรงไปตรงมา เป็นประโยชน์ต่อการปรับปรุงสินค้าและบริการให้ตรงกับความต้องการของลูกค้ามากขึ้น

3. ส่งเสริมการรับรู้และภาพลักษณ์ของธุรกิจ

รีวิวที่ดีจะช่วยสร้างการรับรู้และภาพลักษณ์ที่ดีให้กับธุรกิจ ทำให้ลูกค้ารายใหม่รู้จักและเกิดความไว้วางใจในตราสินค้ามากขึ้น

4. เป็นแหล่งข้อมูลสำคัญในการวิเคราะห์และปรับปรุงธุรกิจ

รีวิวเป็นข้อมูลล้ำค่าจากมุมมองของลูกค้าที่มีธุรกิจสามารถนำไปวิเคราะห์เพื่อหาจุดแข็ง จุดอ่อน และโอกาสในการพัฒนาปรับปรุงสินค้า บริการ และกลยุทธ์ทางธุรกิจต่างๆ ได้อย่างตรงจุด

5. ส่งเสริมการบอกต่อและการซื้อซ้ำ

รีวิวที่ดีจะกระตุ้นให้ลูกค้ามีความประทับใจ อยากบอกต่อไปยังคนรู้จักหรือรอบข้าง และมีแนวโน้มที่จะซื้อซ้ำหรือใช้บริการจากธุรกิจนั้นอีกในอนาคต

ดังนั้น การให้ความสำคัญกับการรีวิวของลูกค้า การตอบสนองอย่างเหมาะสม และการนำข้อมูลจากรีวิวมาปรับปรุงธุรกิจอย่างต่อเนื่อง จึงเป็นกุญแจสำคัญที่จะส่งเสริมความสำเร็จของธุรกิจในระยะยาว

เป้าหมายของการทำ Customer Feedback Reviews Analysis

โปรเจกต์นี้จะเน้นสนใจ Rating, รีวิวด้านลบ และข้อความที่ลูกค้าเขียนรีวิวของสินค้าแต่ละประเภท

เพื่อจะได้นับมาปรับปรุงแก้ไขให้ถูกจุด เพื่อธุรกิจจะได้เติบโตอย่างยั่งยืน

เครื่องมือของการทำ Customer Feedback Reviews Analysis

- Python ใช้ในการทำ Data Preparation , Data Cleansing , Data Analysis
- My SQL ใช้ในการทำ Query เพื่อหา insight
- Power BI ใช้ในการทำ Visualization/ Dashboard
- Microsoft Word เพื่อทำ Report สรุปผล

แหล่งข้อมูล

<https://www.kaggle.com/datasets>

1. Data Preparation/ Data cleansing

1.1 นำเข้า Library

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import spacy
from collections import Counter
import warnings
warnings.filterwarnings("ignore")
```

1.2 นำเข้า dataset ชื่อว่า “Customer Feedback Reviews Dataset.csv”

```
df = pd.read_csv(r"C:\Users\muent\Desktop\my project for data analyst\Project-6\Dataset\Customer Feedback Reviews Dataset.csv")
df
```

	s.no	age	division_name	department_name	class_name	clothing_id	title	review_text	alike_feedback_count	rating	recommend_index
0	0	40	General	Bottoms	Jeans	1028	Amazing fit and wash	Like other reviewers i was hesitant to spend t...	0	5	1
1	1	62	General Petite	Tops	Blouses	850	Lovely and unique!	As is true of a bunch of the fall clothing pho...	12	5	1
2	2	47	General Petite	Bottoms	Skirts	993	Meh	I so wanted this skirt to work, love the desig...	3	1	0
3	3	45	General Petite	Bottoms	Pants	1068	Wow	Love love this! i was hesitant to buy this at	0	5	1

1.3 พิจารณา dataset / จัดการ Missing Values

```
df.isnull().sum()
```

```
: s.no      0
   age      0
   division_name    14
   department_name  14
   class_name    14
   clothing_id     0
   title    3810
   review_text   845
   alike_feedback_count  0
   rating        0
   recommend_index    0
   dtype: int64
```

```
df = df.dropna(subset=['department_name', 'review_text', 'rating'])
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 22628 entries, 0 to 23485
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   s.no                  22628 non-null  int64
1   age                  22628 non-null  int64
2   division_name        22628 non-null  object
3   department_name      22628 non-null  object
4   class_name           22628 non-null  object
5   clothing_id          22628 non-null  int64
6   title                19662 non-null  object
7   review_text          22628 non-null  object
8   alike_feedback_count  22628 non-null  int64
9   rating               22628 non-null  int64
10  recommend_index       22628 non-null  int64
dtypes: int64(6), object(5)
memory usage: 2.1+ MB
```

1.4 เปลี่ยนชื่อ column s.no ไปเป็น no

```
#เปลี่ยนชื่อ columns
df = df.rename( columns= {'s.no' : 'no'})
df
```

	no	age	division_name	department_name	class_name	clothing_id	title	review_text	alike_feedback_count	rating	re
0	0	40	General	Bottoms	Jeans	1028	Amazing fit and wash	Like other reviewers i was hesitant to spend t...	0	5	
1	1	62	General Petite	Tops	Blouses	850	Lovely and unique!	As is true of a bunch of the fall clothing pho...	12	5	
								Lee wanted			

1.5 สร้างcolumn ใหม่ ชื่อว่า rating label

โดยที่ ถ้า rating = 1, 2, หรือ 3 กำหนดให้ rating label = Bad ซึ่งก็คือ รีวิวด้านลบ

และ ถ้า rating = 4 หรือ 5 กำหนดให้ rating label = Good ซึ่งก็คือ รีวิวด้านบวก

```
#สร้างcolumn ใหม่ ชื่อว่า rating_label
```

```
df['rating_label'] = df['rating'].apply(lambda x : "Good" if x >=4 else "Bad")
```

```
df
```

no	age	division_name	department_name	class_name	clothing_id	title	review_text	alike_feedback_count	rating	recommend_index	rating_label
0	40	General	Bottoms	Jeans	1028	Amazing fit and wash	Like other reviewers i was hesitant to spend t...	0	5	1	Good
1	62	General Petite	Tops	Blouses	850	Lovely and unique!	As is true of a bunch of the fall clothing pho...	12	5	1	Good
2	47	General Petite	Bottoms	Skirts	993	Meh	I so wanted this skirt to work, love the desig...	3	1	0	Bad
							Love love thiel i was				

1.6 จัดการ columns age

พบว่าอายุของลูกค้าค่าน้อยสุด 18 ปี , มากสุด 99 ปี

จึงกำหนด range ได้ดังนี้

18-24 -->> 18 to 24

25-34 -->> 25 to 34

35-44 -->> 34 to 44

45-54 -->> 45 to 54

55-64 -->> 55 to 64

65-74 -->> 65 to 74

75-84 -->> 75 to 84

85-99 -->> 85 to 99

>= 99 -->> over 99

```
def age_range(i):
    lower = [18, 25, 35, 45, 55, 65, 75, 85]
    upper = [24, 34, 44, 54, 64, 74, 84, 99]
    for l, u in zip(lower, upper):
        if l <= i and i <= u:
            if u != 999:
                return f"{l} to {u}"
            else:
                return f"over {l}"

df['age_range_label'] = df['age'].apply(lambda x : age_range(x))

#เก็บค่าใส่ column ชื่อ ["age_range_label"]
```

```
df["age_range_label"].value_counts()
```

```
35 to 44    7428
25 to 34    4959
45 to 54    4940
55 to 64    3078
65 to 74    1166
18 to 24     866
75 to 84     160
85 to 99      31
Name: age_range_label, dtype: int64
```

1.7 Save File to_CSV เพื่อนำไปทำ Data Analysis ใน MySQL

```
#Save File to_CSV เพื่อนำไป Query ใน MySQL
df.to_csv("Customer Feedback Review.csv")
```

2. Data Analysis

โปรเจกต์นี้เราจะมุ่งสนใจตัวแปรที่ใช้วิเคราะห์เกี่ยวกับ Bad หรือ Good Review ซึ่งก็คือ

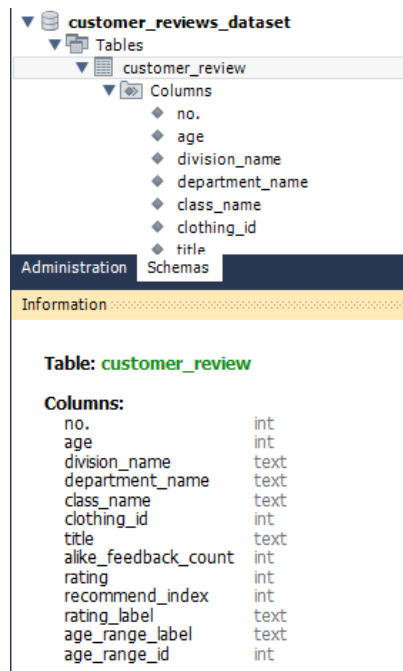
- age อายุของลูกค้า
- class_name ประเภทของสินค้า
- title หัวข้อที่ลูกค้ารีวิว
- recommend_index ข้อความรีวิวของลูกค้า
- rating คะแนน หรือดาวที่ลูกค้ารีวิว 1-3 คือ รีวิวด้านแย่ และ 4-5 คือ รีวิวด้านบวก
- rating_label Good สำหรับรีวิวด้านบวก, Bad สำหรับรีวิวด้านลบ
- age_range_label กลุ่มลูกค้าจัดเรียงตามอายุ

2.1 สร้าง Database และ Import File CSV เข้าใน MySQL

สร้าง Database ชื่อว่า “Customer_Reviews_Dataset”

```
CREATE DATABASE Customer_Reviews_Dataset;  
USE Customer_Reviews_Dataset;
```

นำเข้า data file ที่ save file มาจาก Python ตั้งชื่อไฟล์ “customer_review”



2.2 Analysis : Query on My SQL

Q1. เพิ่ม column ใหม่ ชื่อว่า “age_range_id” เพื่อกำหนดกลุ่มลูกค้า

```
/*ALTER TABLE customer_review  
DROP COLUMN age_range_id;*/  
  
SET SQL_SAFE_UPDATES = 0; #ปิดโหมด SET MODE  
ALTER TABLE customer_review  
ADD COLUMN age_range_id INT(10);  
  
UPDATE customer_review  
SET age_range_id = CASE  
    WHEN age_range_label = "18 to 24" THEN 1  
    WHEN age_range_label = "25 to 34" THEN 2  
    WHEN age_range_label = "35 to 44" THEN 3  
    WHEN age_range_label = "45 to 54" THEN 4  
    WHEN age_range_label = "55 to 64" THEN 5  
    WHEN age_range_label = "65 to 74" THEN 6  
    WHEN age_range_label = "75 to 84" THEN 7  
    ELSE 8  
END;
```

Q2. ลูกค้าที่เข้ามารีวิวสินค้ามีทั้งหมดกี่คน

```
2 • SELECT
3     COUNT(*) AS Transactions
4 FROM customer_review;
```

Transactions
22625

Q3. ลูกค้าให้คะแนนรีวิวเป็นอย่างไรบ้าง

```
4 • SELECT
5     rating,
6     ROUND(COUNT(rating)*100 / (SELECT COUNT(*) FROM customer_review),2) AS percentage_of_rating
7 FROM customer_review
8 GROUP BY rating
9 ORDER BY percentage_of_rating DESC ;
```

rating	percentage_of_rating
5	55.36
4	21.69
3	12.47
2	6.85
1	3.63

Q4. สัดส่วนของรีวิวด้านบวก และ สัดส่วนของรีวิวด้านลบ

```
2 • SELECT
3     rating_label,
4     COUNT(rating_label) AS count_rating_label,
5     ROUND(COUNT(rating_label)*100 / (SELECT COUNT(*) FROM customer_review),2) AS percentage_of_rating_la
6 FROM customer_review
7 GROUP BY rating_label;
```

rating_label	count_rating_label	percentage_of_rating_label
Good	17433	77.05
Bad	5192	22.95

รีวิวเฉพาะด้านบวก

```

10 • SELECT
11     rating_label,
12     COUNT(rating_label) AS count_rating_Good,
13     ROUND(COUNT(rating_label)*100 / (SELECT COUNT(*) FROM customer_review),2) AS percentage_of_rating_Go
14 FROM customer_review
15 WHERE rating IN (4,5)
16 GROUP BY rating_label;
17

```

rating_label	count_rating_Good	percentage_of_rating_Good
Good	17433	77.05

รีวิวเฉพาะด้านลบ

```

19 • SELECT
20     rating_label,
21     COUNT(rating_label) AS count_rating_Bad,
22     ROUND(COUNT(rating_label)*100 / (SELECT COUNT(*) FROM customer_review),2) AS percentage_of_rating_Ba
23 FROM customer_review
24 WHERE rating IN (1,2,3)
25 GROUP BY rating_label;

```

rating_label	count_rating_Bad	percentage_of_rating_Bad
Bad	5192	22.95

Q5. ดูค่าแต่ละกลุ่มอายุรีวิวด้านบวก และด้านลบอย่างไร

```

4 • WITH A AS (
5     SELECT
6         rating,
7         SUM(CASE WHEN recommend_index = 1 THEN 1 ELSE 0 END) AS recommend_Yes,
8         SUM(CASE WHEN recommend_index = 0 THEN 1 ELSE 0 END) AS recommend_No
9     FROM customer_review
10    GROUP BY rating
11 )
12 SELECT
13     A.*,
14     A.recommend_Yes + A.recommend_No AS Total_customer_by_rating,
15     ROUND(A.recommend_Yes *100/(A.recommend_Yes + A.recommend_No),2) AS percentage_recommend_Yes,
16     ROUND(A.recommend_No *100/(A.recommend_Yes + A.recommend_No),2) AS percentage_recommend_No
17 FROM A
18 ORDER BY percentage_recommend_Yes DESC;

```

Result Grid

Filter Rows:

Export:

Wrap Cell Content:

	rating	recommend_Yes	recommend_No	Total_customer_by_rating	percentage_recommend_Yes	percentage_recommend_No
5	12502	23	12525	99.82	0.18	
4	4744	164	4908	96.66	3.34	
3	1169	1653	2822	41.42	58.58	
2	94	1455	1549	6.07	93.93	
1	15	806	821	1.83	98.17	

Q6. ลูกค้านิยมซื้อสินค้าประเภทไหนมากที่สุด โดยเรียงจากมากที่สุดไปหาน้อยสุด

```

4 • SELECT
5     class_name,
6     COUNT(class_name) count_class_name,
7     ROUND(COUNT(class_name)*100/ (SELECT COUNT(*) FROM customer_review),2) AS percentage_of_sold_by_class_name
8 FROM customer_review
9 GROUP BY class_name
10 ORDER BY COUNT(class_name) DESC;

```

class_name	count_class_name	percentage_of_sold_by_class_name
Dresses	6145	27.16
Knits	4626	20.45
Blouses	2983	13.18
Sweaters	1379	6.10
Pants	1350	5.97
Jeans	1102	4.87
Fine gauge	1059	4.68
Skirts	903	3.99
Jackets	683	3.02
Lounge	669	2.96
Swim	332	1.47
Outerwear	319	1.41
Shorts	304	1.34
Sleep	214	0.95
Legwear	158	0.70
Intimates	147	0.65
Layering	132	0.58
Trend	118	0.52
Casual bot...	1	0.00
Chemises	1	0.00

Q7. พฤติกรรมการซื้อแบ่งตามกลุ่มอายุ ลูกค้ากลุ่มไหนซื้อสินค้ามากที่สุด

```

4 • SELECT
5     age_range_label,
6     age_range_id,
7     COUNT(age_range_label) count_age_range_label,
8     ROUND(COUNT(age_range_label)*100/ (SELECT COUNT(*) FROM customer_review),2) AS percentage_of_buy_of_age_range
9 FROM customer_review
10 GROUP BY age_range_label,age_range_id
11 ORDER BY age_range_id ASC;

```

age_range_label	age_range_id	count_age_range_label	percentage_of_buy_of_age_range
18 to 24	1	866	3.83
25 to 34	2	4959	21.92
35 to 44	3	7426	32.82
45 to 54	4	4940	21.83
55 to 64	5	3077	13.60
65 to 74	6	1166	5.15
75 to 84	7	160	0.71
85 to 99	8	31	0.14

Q8. พฤติกรรมการซื้อแบ่งตามกลุ่มอายุ , แต่ละกลุ่มอายุนิยมซื้อสินค้าประเภทไหนมากที่สุด 3

อันดับแรก

```
4 • SELECT
5     distinct age_range_label,
6     class_name,
7     count(class_name) AS count_class_name,
8     RANK () OVER(PARTITION BY age_range_label ORDER BY count(class_name) DESC) AS ranking
9 FROM customer_review
10 group by age_range_label, class_name;
```

age_range_label	class_name	count_class_name	ranking
18 to 24	Dresses	268	1
18 to 24	Knits	162	2
18 to 24	Blouses	94	3
18 to 24	Sweaters	54	4
18 to 24	Pants	42	5
18 to 24	Fine gauge	40	6
18 to 24	Skirts	38	7
18 to 24	Jeans	33	8
18 to 24	Jackets	25	9
18 to 24	Lounge	23	10
18 to 24	Swim	20	11
18 to 24	Intimates	15	12
18 to 24	Outerwear	14	13
18 to 24	Shorts	11	14
18 to 24	Sleep	11	14
18 to 24	Trend	6	16
18 to 24	Layering	5	17
18 to 24	Legwear	5	17
25 to 34	Dresses	1480	1
25 to 34	Knits	1001	2
25 to 34	Blouses	602	3
25 to 34	Pants	249	4
25 to 34	Sweaters	234	5
25 to 34	Skirts	225	6
25 to 34	Jeans	217	7

ยกตัวอย่าง เช่น กลุ่มอายุ 18 to 24 , กลุ่มอายุ 25 to 34 , กลุ่มอายุ 35 to 44 ,กลุ่มอายุ 45 to 54, กลุ่มอายุ 55 to 64 และ กลุ่มอายุ 65 to 74 นิยมซื้อสินค้า Dresses , Knits และ Blouses เป็น 3 อันดับแรก

กลุ่มอายุ 75 to 84 นิยมซื้อสินค้า Dresses , Knits และ Sweaters เป็น 3 อันดับแรก

กลุ่มอายุ 85 to 99 นิยมซื้อสินค้า Dresses , Knits , Jeans , Pants และ Sweaters เป็น 3 อันดับแรก

Q9. รีวิวด้านลบ

```
5 • SELECT
6     age_range_label,
7     age_range_id,
8     class_name,
9     COUNT(rating) Bad_rating_by_age_range,
10    RANK() OVER(PARTITION BY age_range_label ORDER BY age_range_id ASC,COUNT(rating) DESC) AS ranking
11 FROM customer_review
12 WHERE rating IN (1,2,3)
13 GROUP BY age_range_label,age_range_id,class_name;
14
15
```

Result Grid					
Filter Rows: <input type="text"/>					
Export: Wrap Cell Content:					
	age_range_label	age_range_id	class_name	Bad_rating_by_age_range	ranking
▶	18 to 24	1	Dresses	49	1
	18 to 24	1	Knits	26	2
	18 to 24	1	Blouses	20	3
	18 to 24	1	Sweaters	11	4
	18 to 24	1	Skirts	7	5
	18 to 24	1	Lounge	7	5
	18 to 24	1	Fine gauge	7	5
	18 to 24	1	Jeans	6	8
	18 to 24	1	Jackets	4	9
	18 to 24	1	Pants	4	9
	18 to 24	1	Shorts	3	11
	18 to 24	1	Intimates	3	11
	18 to 24	1	Sleep	3	11
	18 to 24	1	Trend	2	14
	18 to 24	1	Swim	2	14
	18 to 24	1	Outerwear	2	14
	18 to 24	1	Layering	1	17
	25 to 34	2	Dresses	408	1
	25 to 34	2	Knits	256	2
	25 to 34	2	Blouses	176	3
	25 to 34	2	Sweaters	69	4
	25 to 34	2	Pants	60	5
	25 to 34	2	Skirts	52	6
	25 to 34	2	Fine gauge	46	7
	25 to 34	2	Jeans	39	8
	25 to 34	2	Swim	35	9
	25 to 34	2	Jackets	33	10

Q10. รีวิวด้านบวก

```
5 • SELECT
6     age_range_label,
7     age_range_id,
8     class_name,
9     COUNT(rating) Good_rating_by_age_range,
10    RANK() OVER(PARTITION BY age_range_label ORDER BY age_range_id ASC,COUNT(rating) DESC) AS ranking
11 FROM customer_review
12 WHERE rating IN (4,5)
13 GROUP BY age_range_label,age_range_id,class_name;
14
15
```

Result Grid					
Filter Rows:					
Export: Wrap Cell Content: IA					
	age_range_label	age_range_id	class_name	Good_rating_by_age_range	ranking
▶	18 to 24	1	Dresses	219	1
	18 to 24	1	Knits	136	2
	18 to 24	1	Blouses	74	3
	18 to 24	1	Sweaters	43	4
	18 to 24	1	Pants	38	5
	18 to 24	1	Fine gauge	33	6
	18 to 24	1	Skirts	31	7
	18 to 24	1	Jeans	27	8
	18 to 24	1	Jackets	21	9
	18 to 24	1	Swim	18	10
	18 to 24	1	Lounge	16	11
	18 to 24	1	Intimates	12	12
	18 to 24	1	Outerwear	12	12
	18 to 24	1	Sleep	8	14
	18 to 24	1	Shorts	8	14
	18 to 24	1	Legwear	5	16
	18 to 24	1	Layering	4	17
	18 to 24	1	Trend	4	17
	25 to 34	2	Dresses	1072	1
	25 to 34	2	Knits	745	2
	25 to 34	2	Blouses	426	3
	25 to 34	2	Pants	189	4
	25 to 34	2	Jeans	178	5
	25 to 34	2	Skirts	173	6
	25 to 34	2	Sweaters	165	7
	25 to 34	2	Fine gauge	124	8
	25 to 34	2	Jackets	124	8

Q11. Percentage แสดงสัดส่วนรีวิวด้านบวก vs รีวิวด้านลบ

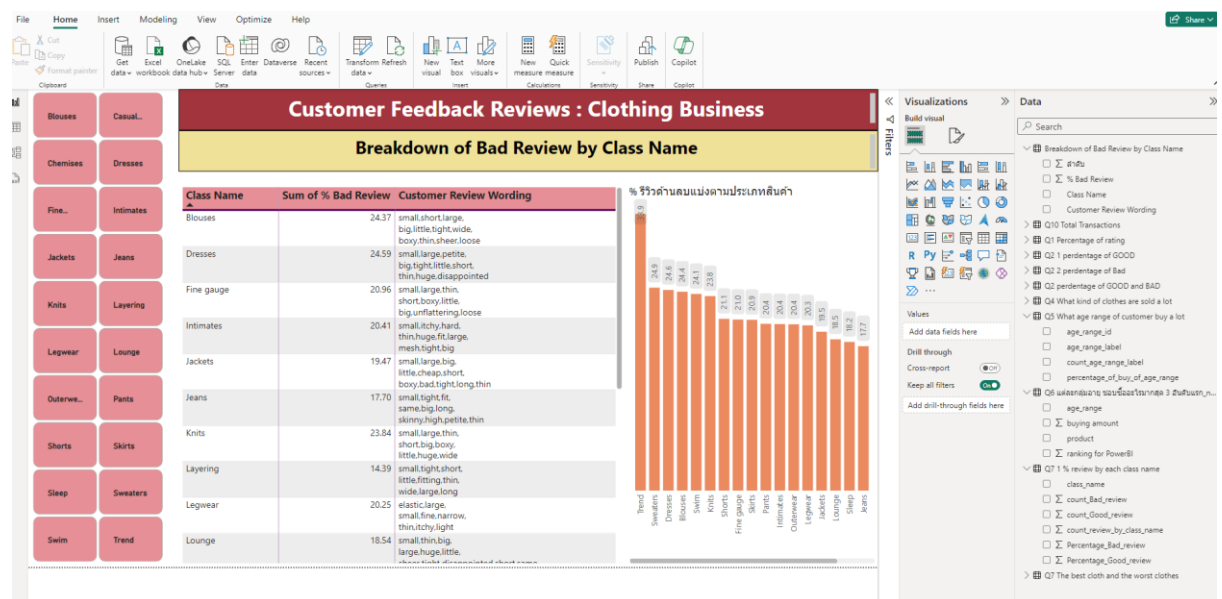
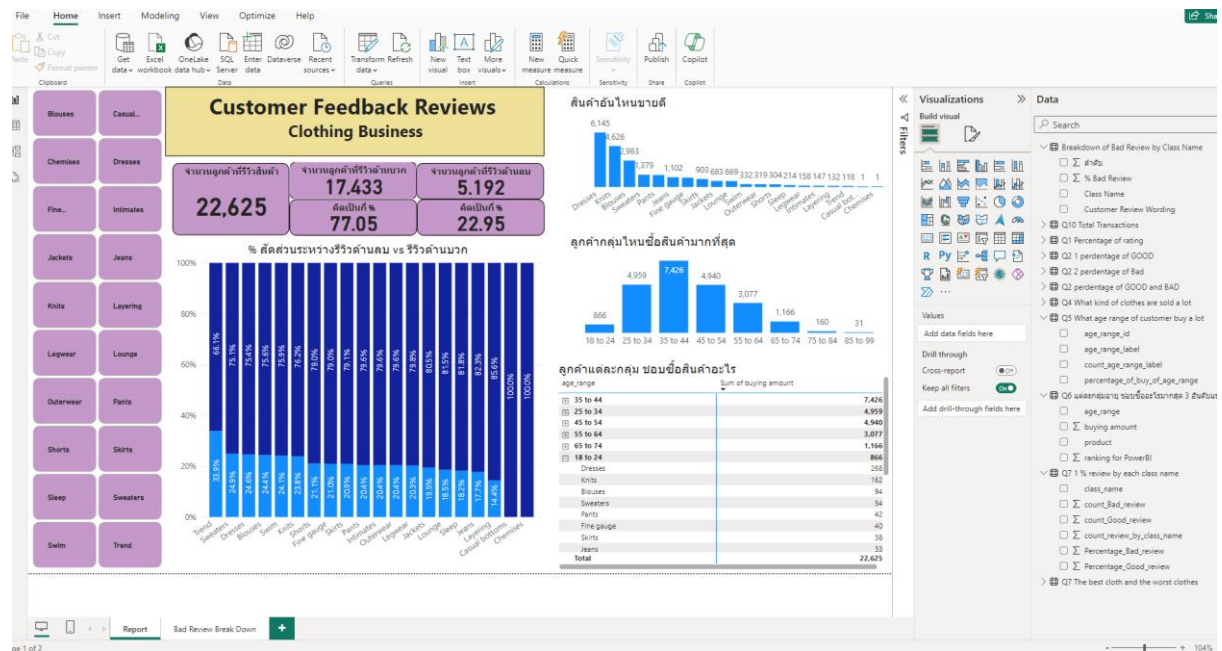
```

4 • WITH table1 AS(
5     SELECT
6         class_name,
7         SUM(CASE WHEN rating IN (4,5) THEN 1 ELSE 0 END) count_Good_review,
8         SUM(CASE WHEN rating IN (1,2,3) THEN 1 ELSE 0 END) count_Bad_review
9     FROM customer_review
10    GROUP BY class_name
11 )
12 SELECT
13     table1.*,
14     (table1.count_Good_review) + (table1.count_Bad_review) AS count_review_by_class_name,
15     ROUND(table1.count_Good_review *100/((table1.count_Good_review) + (table1.count_Bad_review)),2) AS Percentage_Good_review,
16     ROUND(table1.count_Bad_review *100/((table1.count_Good_review) + (table1.count_Bad_review)),2) AS Percentage_Bad_review
17 FROM table1
18 ORDER BY Percentage_Bad_review DESC;
19

```

class_name	count_Good_review	count_Bad_review	count_review_by_class_name	Percentage_Good_review	Percentage_Bad_review
Trend	78	40	118	66.10	33.90
Sweaters	1036	343	1379	75.13	24.87
Dresses	4634	1511	6145	75.41	24.59
Blouses	2256	727	2983	75.63	24.37
Swim	252	80	332	75.90	24.10
Knits	3523	1103	4626	76.16	23.84
Shorts	240	64	304	78.95	21.05
Fine gauge	837	222	1059	79.04	20.96
Skirts	714	189	903	79.07	20.93
Pants	1074	276	1350	79.56	20.44
Intimates	117	30	147	79.59	20.41
Outerwear	254	65	319	79.62	20.38
Legwear	126	32	158	79.75	20.25
Jackets	550	133	683	80.53	19.47
Lounge	545	124	669	81.46	18.54
Sleep	175	39	214	81.78	18.22
Jeans	907	195	1102	82.30	17.70
Layering	113	19	132	85.61	14.39
Casual bottoms	1	0	1	100.00	0.00
Chemises	1	0	1	100.00	0.00

2.3 Analysis : นำผลลัพธ์จาก SQL ไปทำ Visualization ใน Power BI



2.4 Analysis : เข้าไปดูข้อความรีวิวจากลูกค้า ใน Python

เป้าหมายต้องการรู้ว่าลูกค้าพูดถึงสินค้าแต่ละประเภทในด้านลบอย่างไรบ้าง
เพื่อจะได้นำมาเป็นแนวทางในการปรับปรุงธุรกิจ

2.4.1 ตัดคำ Adjectives จาก column name review_text

```
# ตัดคำ Adjectives จาก column name review_text
import spacy

# โหลดโมเดลภาษาอังกฤษ
nlp = spacy.load("en_core_web_sm")

# สมมติว่า df เป็น Pandas DataFrame ที่มีคอลัมน์ 'review_text'
# สร้างคอลัมน์ใหม่ชื่อ 'adj_column' ในดาต้าเฟรม
df['adj_word'] = ""

for i, row in df.iterrows():
    test_str = row['review_text']

    # ประมวลผลข้อความด้วย SpaCy
    doc = nlp(test_str)

    # สร้างลิสต์เก็บคำคุณศัพท์
    adj = [token.text for token in doc if token.pos_ == "ADJ"]

    # เก็บลิสต์ของคำคุณศัพท์ลงในคอลัมน์ 'adj_column' ของดาต้าเฟรม
    df.at[i, 'adj_word'] = ', '.join(adj)
```

```
df.head()
```

department_name	class_name	clothing_id	title	review_text	alike_feedback_count	rating	recommend_index	rating_label	age_range_label	adj_word
Bottoms	Jeans	1028	Amazing fit and wash	Like other reviewers i was hesitant to spend t...	0	5	1	Good	35 to 44	other, hesitant, much, good, full, fresh
Tops	Blouses	850	Lovely and unique!	As is true of a bunch of the fall clothing pho...	12	5	1	Good	55 to 64	true, bright, vivid, unique, bib, soft, nice, ...
Bottoms	Skirts	993	Meh	I so wanted this skirt to work, love the desig...	3	1	0	Bad	45 to 54	small
				Love love this! I was						hesitant, kin, sure

2.4.2 เข้าไปนับจำนวนคำใน column adj_word ว่าแนวโน้มเป็นอย่างไร

ลูกค้ารีวิวว่าอะไรบ้าง เกี่ยวกับสินค้าแบบไหน

เพื่อแสดงให้เห็นว่าคำส่วนใหญ่ใน คอลัมน์ ["adj_word"] มีคำว่าอะไรบ้าง

เราสามารถนับความถี่ของคำที่ปรากฏและแสดงคำที่มีความถี่สูงสุดได้ดังนี้:

```
# นำเข้า counter จาก collections
from collections import Counter

# รวบรวมคำทั้งหมดจากคอลัมน์ 'adj_word'
all_words = []
for words in df['adj_word']:
    all_words.extend(words.split(' '))

# นับความถี่ของคำ
word_counts = Counter(all_words)

# แสดงคำที่มีความถี่สูงสุด
most_common = word_counts.most_common(20)

print(" รีวิวลูกค้าส่วนใหญ่ 20 อันดับแรก คือ : ")
for word, count in most_common:
    print(f"{word}: {count}")
```

```
รีวิวลูกค้าส่วนใหญ่ 20 อันดับแรก คือ :
great: 5729
small: 4514
perfect: 3657
little: 3632
soft: 3284
comfortable: 3019
nice: 2931
cute: 2809
flattering: 2793
large: 2792
beautiful: 2720
petite: 2045
```

2.4.3 Filter รีวิวด้านลบ แยกตาม Class Name

เพราะ โปรเจกต์นี้เราสนใจรีวิวด้านลบ เพื่อจะนำไปปรับปรุงสินค้า จึงต้องการทราบว่าในสินค้าแต่ละประเภทลูกค้าที่ให้ rating สินค้า 1- 3 คะแนน เขาได้พูดถึงสินค้าในด้านไหนบ้าง

วิเคราะห์ Bad Review ของแต่ละ class name

ลูกค้าส่วนใหญ่รีวิวว่าอย่างไรบ้าง ¶

```
from collections import Counter

class_names = ["Trend", "Sweaters", "Dresses", "Blouses", "Swim", "Knits", "Shorts", "Fine gauge", "Skirts", "Pants", "Intimates", "Outerwear", "Legwear"]
results = []

for class_name in class_names:
    # สร้าง DataFrame สำหรับ class_name นั้นๆ
    df_class = df[(df["rating_label"] == "Bad") & (df["class_name"] == class_name)]

    # รวบรวมคำทั้งหมดจากคอลัมน์ 'adj_word'
    all_words = []
    for words in df_class['adj_word']:
        all_words.extend(words.split(' '))

    # นับความถี่ของคำ
    word_counts = Counter(all_words)

    # เพิ่มผลลัพธ์ลงในรายการ results
    for word, count in word_counts.most_common(20):
        results.append({"class_name": class_name, "word": word, "count": count})

# สร้าง DataFrame จากรายการ results
df_bad_review_by_class_name = pd.DataFrame(results)
print(df_bad_review_by_class_name)
```


เก็บค่าใส่ Dataframe ชื่อว่า df_bad_review_by_class_name

```
df_bad_review_by_class_name
```

	class_name	word	count
0	Trend	small	12
1	Trend	large	10
2	Trend	beautiful	8
3	Trend	much	6
4	Trend	sure	6
...
355	Layering	large	2
356	Layering	yellow	2
357	Layering	long	2
358	Layering		2
359	Layering	most	2

360 rows × 3 columns

```
# save to csv
df_bad_review_by_class_name.to_csv("bad_review_by_class_name.csv")
```

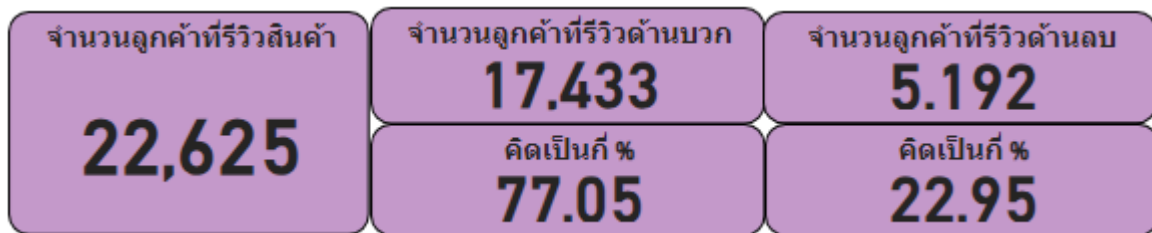
3. Conclusion

เมื่อนำข้อมูลไปหา insight โดยใช้ Python และ SQL รวมถึงการสร้าง Visualization ด้วย Power BI สรุปผลได้ดังนี้

3.1 ลูกค้าที่มีรีวิวสินค้ามีทั้งหมด 22,625 คน

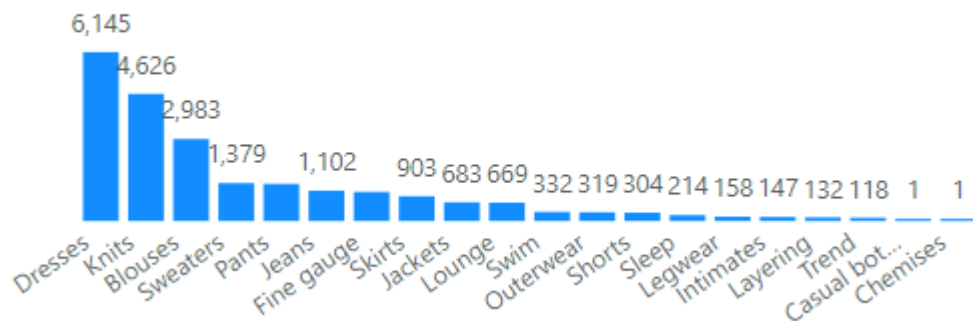
แบ่งเป็นลูกค้าที่รีวิวด้านบวก ลูกค้าชอบสินค้า 17,433 คน คิดเป็น 77.05 % ของลูกค้าที่มีรีวิว

และลูกค้าที่รีวิวด้านลบ ลูกค้ายังไม่ชอบสินค้า 5,192 คน คิดเป็น 22.95 % ของลูกค้าที่มีรีวิว



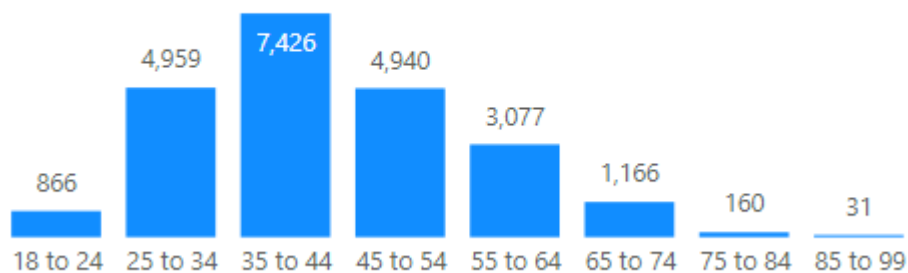
3.2 สินค้าที่ขายดี เรียงลำดับจากมากสุดไปน้อยสุด โดย 5 ลำดับแรกที่ขายดี ได้แก่ Dresses, Knits, Blouses, Sweaters และ Pants

สินค้าอันไหนขายดี



3.3 สินค้าเป็นที่นิยมในกลุ่มลูกค้าอายุระหว่าง 25- 64 ปี

ลูกค้ากลุ่มไหนซื้อสินค้ามากที่สุด



3.4 ลูกค้าแต่ละกลุ่มนิยมซื้อสินค้าประเภทไหน

- กลุ่มอายุ 18 to 24 , กลุ่มอายุ 25 to 34 , กลุ่มอายุ 35 to 44 ,กลุ่มอายุ 45 to 54, กลุ่มอายุ 55 to 64

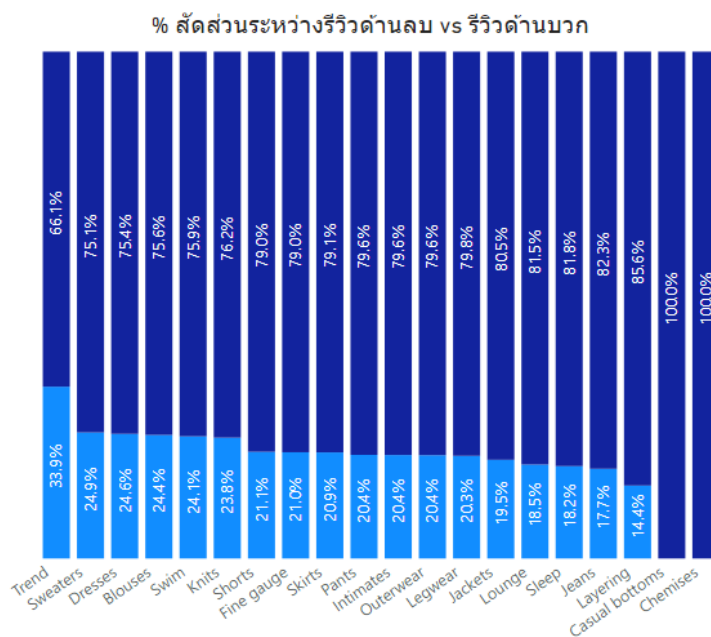
และ กลุ่มอายุ 65 to 74 นิยมซื้อสินค้า Dresses , Knits และ Blouses เป็น 3 อันดับแรก

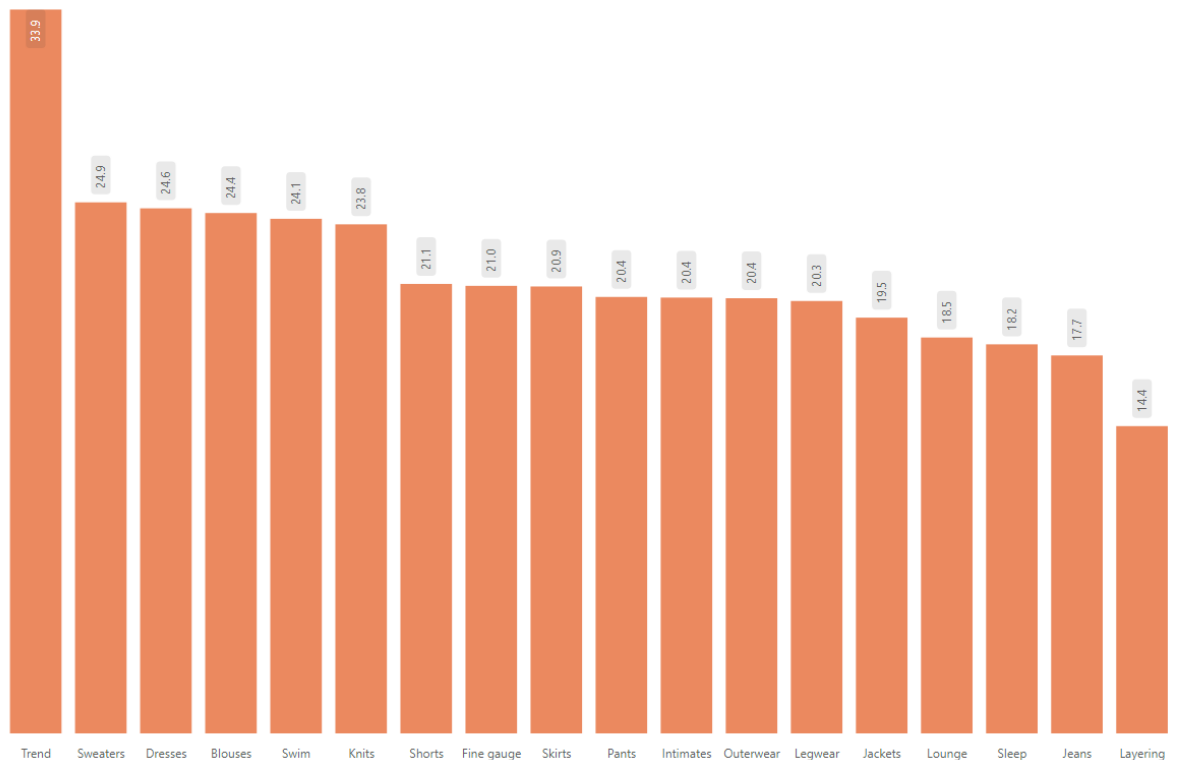
- กลุ่มอายุ 75 to 84 นิยมซื้อสินค้า Dresses , Knits และ Sweaters เป็น 3 อันดับแรก

- กลุ่มอายุ 85 to 99 นิยมซื้อสินค้า Dresses , Knits , Jeans , Pants และ Sweaters เป็น 3 อันดับแรก

age_range	Sum of buying amount
05 to 14	1,100
18 to 24	866
Dresses	268
Knits	162
Blouses	94
Sweaters	54
Pants	42
Fine gauge	40
Skirts	38
Jeans	33
Jackets	25
Lounge	23
Swim	20
Intimates	15
Total	22,625

3.5 ลูกค้ารีวิวด้านลบ โดยแบ่งตามประเภทสินค้าดังนี้





- สินค้าประเภท Trend ได้รับรีวิวด้านลบ 33.9 % จากรีวิวทั้งหมด
- สินค้าประเภท Sweaters ได้รับรีวิวด้านลบ 24.87 % จากรีวิวทั้งหมด
- สินค้าประเภท Dresses ได้รับรีวิวด้านลบ 24.59 % จากรีวิวทั้งหมด
- สินค้าประเภท Blouses ได้รับรีวิวด้านลบ 24.37 % จากรีวิวทั้งหมด
- สินค้าประเภท Swim ได้รับรีวิวด้านลบ 24.10 % จากรีวิวทั้งหมด
- สินค้าประเภท Knits ได้รับรีวิวด้านลบ 23.84 % จากรีวิวทั้งหมด
- สินค้าประเภท Shorts ได้รับรีวิวด้านลบ 21.05 % จากรีวิวทั้งหมด
- สินค้าประเภท Fine gauge ได้รับรีวิวด้านลบ 20.96 % จากรีวิวทั้งหมด
- สินค้าประเภท Skirts ได้รับรีวิวด้านลบ 20.93 % จากรีวิวทั้งหมด
- สินค้าประเภท Pants ได้รับรีวิวด้านลบ 20.44 % จากรีวิวทั้งหมด
- สินค้าประเภท Intimates ได้รับรีวิวด้านลบ 20.41 % จากรีวิวทั้งหมด
- สินค้าประเภท Outerwear ได้รับรีวิวด้านลบ 20.38 % จากรีวิวทั้งหมด
- สินค้าประเภท Legwear ได้รับรีวิวด้านลบ 20.25 % จากรีวิวทั้งหมด
- สินค้าประเภท Jackets ได้รับรีวิวด้านลบ 19.47 % จากรีวิวทั้งหมด
- สินค้าประเภท Lounge ได้รับรีวิวด้านลบ 18.54 % จากรีวิวทั้งหมด
- สินค้าประเภท Sleep ได้รับรีวิวด้านลบ 17.70 % จากรีวิวทั้งหมด
- สินค้าประเภท Jeans ได้รับรีวิวด้านลบ 19.47 % จากรีวิวทั้งหมด
- สินค้าประเภท Layering ได้รับรีวิวด้านลบ 14.39 % จากรีวิวทั้งหมด

3.6 ลูกค้าได้รีวิวเกี่ยวกับสินค้าแต่ละประเภทอย่างไรบ้าง

Class Name	% Bad Review	Customer Review Wording
Trend	33.90	small ,large, big,ridiculous, wide, skinny, disappointed, oversized, sad, flat
Sweaters	24.87	small,big, itchy, short, large, long, little, huge, boxy, petite
Dresses	24.59	small,large,petite, big,tight,little,short, thin,huge,disappointed
Blouses	24.37	small,short,large, big,little,tight,wide, boxy,thin,sheer,loose
Swim	24.10	small,little,large, big,thin,tight,long, short,disappointed
Knits	23.84	small,large,thin, short,big,boxy, little,huge,wide
Shorts	21.05	short,small, big,little,large, curvy,fit,long,tight
Fine gauge	20.96	small,large,thin, short,boxy,little, big,unflattering,loose
Skirts	20.93	small,large,elastic, little,cheap,thick, long,petite,short,big,high

Class Name	% Bad Review	Customer Review Wording
Pants	20.44	small,large,little, big,tight,petite, short,fit,thin,huge
Intimates	20.41	small,itchy,hard, thin,huge,fit,large, mesh,tight,big
Outerwear	20.38	small,large,tight, unlined,bulky,big, unflattering,huge,petite, cheap,bad,disappointing
Legwear	20.25	elastic,large, small,fine,narrow, thin,itchy,light
Jackets	19.47	small,large,big, little,cheap,short, boxy,bad,tight,long,thin
Lounge	18.54	small,thin,big, large,huge,little, sheer,tight,disappointed,short,same
Sleep	18.22	large,disappointed, thin,short,small,tight, loose,cozy,big,tall,whole
Jeans	17.70	small,tight,fit, same,big,long, skinny,high,petite,thin
Layering	14.39	small,tight,short, little,fitting,thin, wide,large,long

ตามข้อความรีวิวของลูกค้าที่ให้คะแนน rating อยู่ที่ 1- 3 คะแนน วิเคราะห์ออกมาได้ 2 ประเด็นหลัก คือ ขนาดไซส์ และ ลักษณะเนื้อผ้า

- ขนาดไซส์ : ใหญ่เกินไป เล็กเกินไป สั้นเกินไป ยาวเกินไป ใส่ไม่พอดี

ลูกค้าไม่ชอบสินค้าในเรื่องไซส์ที่ใส่ไม่พอดี ธุรกิจอาจจะต้องหากกลยุทธ์เพื่อตอบรับประเด็นหลักนี้ เช่น สามารถเปลี่ยนไซส์ได้ภายใน 7 วันหลังจากได้รับสินค้า

- ลักษณะเนื้อผ้า : ใส่ไม่สบาย ผ้าขยับ ใส่แล้วคัน

อาจจะต้องดูแลเรื่องการแพทเทิร์นการตัดเย็บให้ละเอียดมากขึ้นเพื่อให้เสื้อผ้าทรงสวย รวมถึงเนื้อผ้าต้องใส่สบาย ไม่ระคายผิว