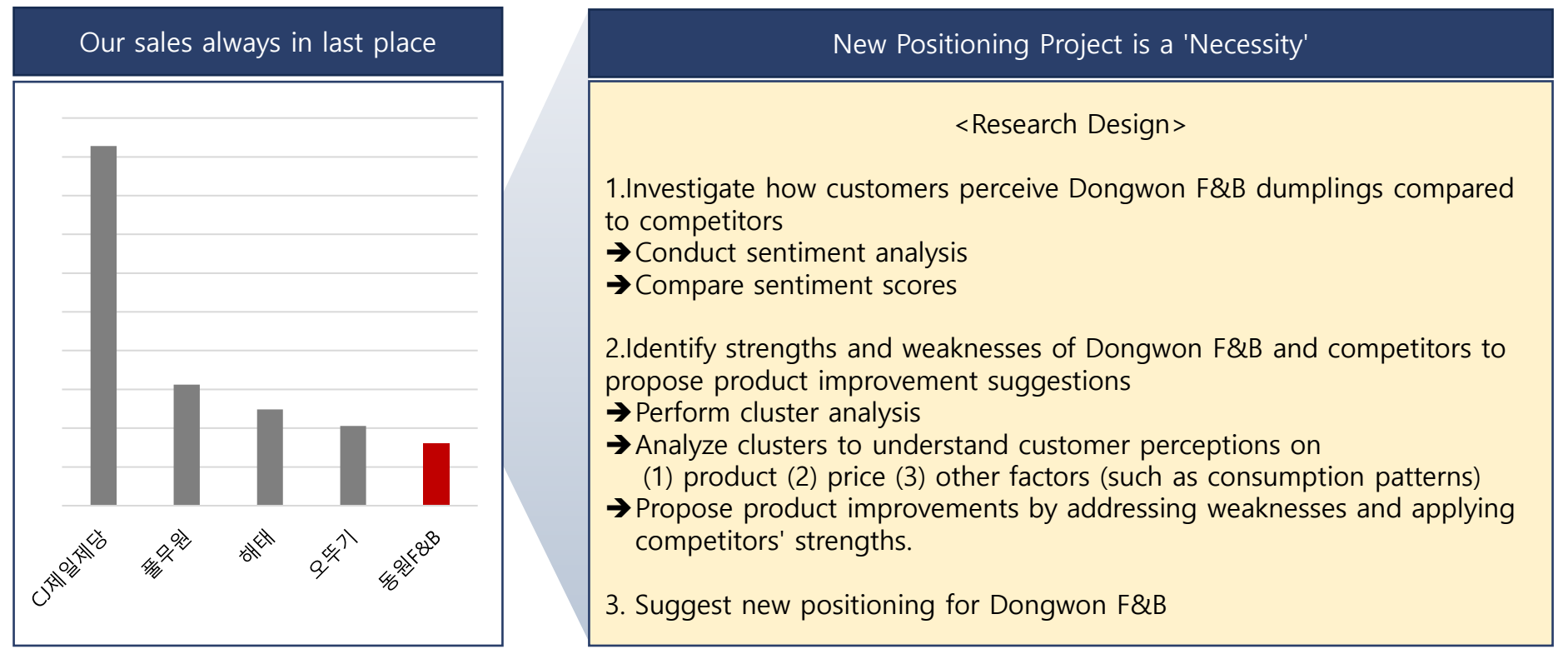


Strategic Direction Proposal for Dongwon F&B's 'Frozen Dumplings'

- Positioning Proposal through Text Big Data Analysis -

[Problem motivation] Our company sees a new positioning as essential, not optional.

The 400 billion KRW domestic frozen dumpling market, faces our company's sales decline due to an unclear positioning, lagging behind competitors. The imperative new positioning project utilizes text big data analysis of customer reviews to understand our customers and identify competitor preferences, aiming to seize market opportunities.



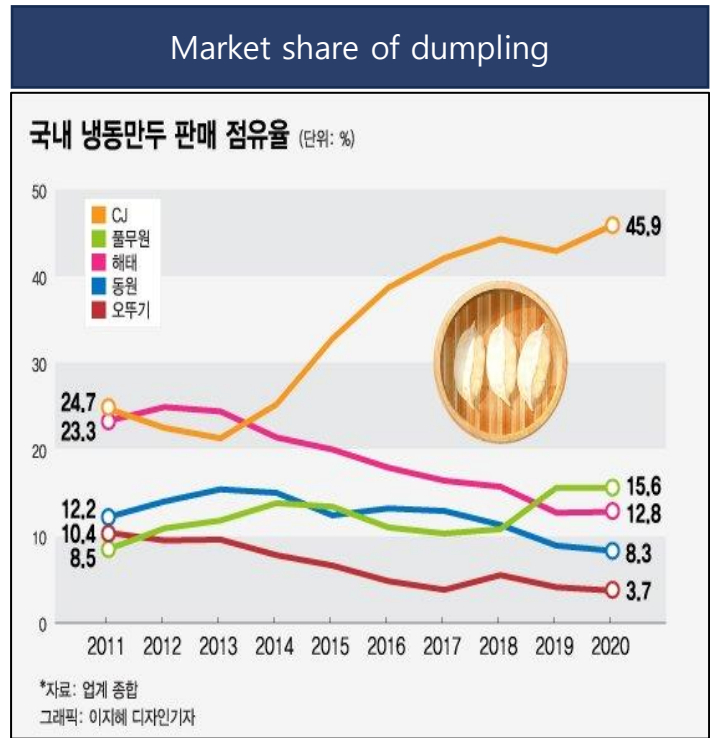
Numerous previous studies have collected online review data and proposed improvement suggestions to companies. In particular, Chi-Hoon Kim (2022) conducted a text mining analysis on 'TripAdvisor' reviews, performing topic modeling. Based on these analyses, the study suggested improvement points for hotels.

Literature Review *
<p>Purchase reviews are posts made by consumers who have purchased a specific product or used a service, sharing their user experiences or evaluations in the form of comments on online platforms (Lee, et al., 2013).</p> <p>With a survey indicating that 90% of consumers read reviews when making purchase decisions, reviews play a crucial role and have a significant impact (Saleh, 2015; Zhao et al., 2018).</p> <p>As consumer purchasing behavior shifts rapidly from offline to online, attributes of purchase reviews such as ratings, quantity of comments, and image information have emerged as representative forms of consumer participation in the digital era (Kim et al., 2013).</p>
Chi-Hoon, Kim(2022) **
<p><Text Mining Analysis of Hotel Reviews Focused on TripAdvisor Reviews></p> <p>The study attempted keyword analysis, network analysis, and topic modeling analysis on collected user reviews. Through this, we identified the most frequently mentioned keywords in reviews from domestic hotel users, examined how their networks are formed, and further explored how topic modeling is possible.</p>

*윤종필(2022),... 온라인 신선식품 전문몰에서의 구매 후기 분석, 고려대학교 대학원 **김치훈(2022). Text Mining Analysis of Hotel Reviews Focused on TripAdvisor Reviews,

[Statement of Research Objectives] Gathering idea from Coupang review data.

As overall, We would like to analyze the review data of **top 3 dumpling brands in Korea and our company** by clustering method. The main objective of this research is **to identity the key factors for evaluating products in dumpling market**. At the end, We would like to suggest possible strategies to increase sales and market value of Dongwon's F&B dumpling products.



Pulmuwon "알피 짬뽕 고기만두"

냉동
알피 짬뽕 고기만두
9,310원 (1kg당 9,310원) 400g x 2개 (냉동) 400g, 2개
냉동지: 냉동 냉동식품 제조
★★★★★ 43,000개 이상 리뷰
이 브랜드의 다른 상품

다 맛은 좋은데
가격 부담 - 주원 400g x 2개
400g

2개 18,620원 4개 37,240원
1kg당 18,620원

Competitor

CJ "비비교 왕교자"

비비교 왕교자 1,480원 x 2 1.4kg, 2개
냉동지: 냉동 냉동식품 제조
★★★★★ 46,000개 이상 리뷰

20,730원 (1kg당 14,800원)
400g x 2개 (냉동) 400g, 2개

다 맛은 좋은데
가격 부담 - 주원 1.4kg x 2개
500g 400g 1.05kg 1.360kg 1.4kg 1.905kg 1.825kg

2개 29,760원 1kg당 14,880원
3개 44,640원 1kg당 14,880원
4개 59,520원 1kg당 14,880원 (냉동지: 냉동)

Competitor

Haitai "고향만두"

냉동
고향만두 (냉동) 1.5kg, 1개
냉동지: 냉동 냉동식품 제조
★★★★★ 43,000개 이상 리뷰
이 브랜드의 다른 상품

다 맛은 좋은데
가격 부담 - 주원 1.5kg x 1개
1.5kg

1개 13,560원 4개 54,240원
22,120원 4개 88,480원
1kg당 13,560원

Competitor

Dongwon F&B "개성 왕만두"

냉동
개성 왕만두 (냉동) 2.24kg, 1개
냉동지: 냉동 냉동식품 제조
★★★★★ 43,000개 이상 리뷰
이 브랜드의 다른 상품

17,360원 (1kg당 7,750원) 2.24kg, 1개
400g x 2개 (냉동) 400g, 2개

다 맛은 좋은데
가격 부담 - 주원 2.24kg x 1개
400g 1.2kg 2.24kg

1개 17,360원 4개 69,440원
22,120원 4개 88,480원
1kg당 17,360원

Our Company

[Statement of Research Objectives] Gathering idea from Coupang review data.

The reviews were gathered from Coupang, chosen due to its dominant position in the e-commerce market. To understand how each product is received by the market, a significant number of reviews is crucial. And Coupang is considered the platform where the most definitive reactions can be observed.

Data Source

To understandhow each product is received by the market, a significant number of reviews is crucial

Coupang is considered the platform where the most definitive reactions can be observed.

주요 온라인몰에서의 식료품 구매 경험

	EVER 구매 경험 (N=1076)	GAP (CO/19)	최근 3개월 내 구매 경험 (N=1075)	GAP (CO/19)	주요 온라인몰 (N=1050)	GAP (CO/19)
주요	70.3	+5.5	48.1	+5.7	23.1	+1.2
이마트	58.8	+4.5	35.2	+5.5	16.1	+1.3
이마트24	40.6	+1.3	24.6	+7.4	7.8	+1.2
이마트24 소점*	47.2	N/A	27.5	N/A	7.0	N/A
홈플러스	39.1	-2.2	21.3	-1.3	6.9	-8.0
이마트	45.5	-3.9	23.8	-3.8	6.0	-2.0
리온	40.7	-2.4	21.2	-0.4	5.7	-0.4
이마트	46.1	-5.0	24.5	-7.9	5.5	-3.9
신세계	40.7	+7.6	30.3	+7.4	5.0	+2.2
롯데마트	31.7	-2.6	15.9	+0.1	4.8	-1.4
11번가	39.5	-2.6	19.0	-3.7	4.5	-1.7
홈	27.8	-1.1	12.7	-0.9	3.1	-1.0
오아시스 마켓*	5.2	N/A	3.5	N/A	1.9	N/A
CJ	31.2	+0.3	9.6	+1.2	1.2	-0.4
GS25	27.9	+0.5	10.7	-0.4	1.0	-0.3
이마트	8.0	+0.1	3.3	+0.7	0.4	+0.1
셀마켓	6.6	+1.3	2.6	+0.5	0.2	+0.1
인라인	15.8	+1.0	3.1	-1.8	0.1	-0.2
영광	0.1	-0.4	0	-1.1		

[Base: 한 달에 한 번 이상 온라인 식료품 구매자, 복수응답, %]
*주요 온라인몰 기준 경험 / * 내이버 소점, 오아시스 마켓, 2020년 신규 리스토

[Base: 주요 온라인몰 EVER 구매자, 복수응답, %]
[Base: 주요 온라인몰 3개월 내 구매자, 단수응답, %]

Research Goal

1. Collect and analyze Coupang review data
2. Confirm differences in customer perceptions between our company and competitors through sentiment analysis.
3. Utilize cluster analysis to precisely identify features and strengths/weaknesses of our products and those of competitors.
4. Based on these findings, present positioning insights

[Data and Methodlogy] 1. Data collection

The process of big data investigation and analysis involves data collection, refinement, and analysis. The textual data used in this study is unstructured, given that shopping mall reviews are written in a free format. Rigorous examination and processing of the data are necessary.

Web scraper is used.
Scraper iteratively gathers relevant reviews.

Used library:
beautifulsoup, selenium, pandas

	Title	Review	Reviewer	Date	Star	Images	Question1	Answer1	Question2	Answer2	Help
0	\n\n글>이랑 같...	공 x은 적당히 익었으나 예는 말라서 잘긴 밑가루 씹는 느낌이있어요 라면에 넣을 때...	오리행일이	2021.09.24	2	0	맛 만족도	괜찮아요	간편함	조리하기 불편해요	2
1	-1	얇은 피 만두 계속 먹다가 횡만두에서 사봤는데 피가 두꺼워서 그런지 속이랑 따로 놓...	시마랑	2022.09.18	2	0	맛 만족도	예상보다 맛 있어요	간편함	조리가 간편해요	1
2	\n\n백달 알맘...	냉동식품인데 드라이아이스 작은 거 2봉 넣고 거의 다 해동되어서 좋아요	양*희	2020.07.07	2	0	맛 만족도	괜찮아요	간편함	조리가 간편해요	2
3	-1	만두피 두껍네요 가격도 비싼 편입니다	김*린	2023.01.29	2	0	맛 만족도	예상보다 맛 있어요	간편함	보통이에요	1
4	\n\n맛있긴한데 ...		정*힐	2022.10.29	2	0	맛 만족도	예상보다 맛있어요	간편함	조리가 간편해요	-1



Pandas



BeautifulSoup

[Data and Methodlogy] 2. Text Preprocessing - Data refinement

This code removes predefined stopwords from the morphological analysis results in dataframes, storing the cleaned text as strings in a new column named 'Morphs_Cleaned.' The stopwords, defined in the variable, are excluded from the morphological analysis to refine the textual data for further analysis.

```
# 불용어 목록 정의
stopwords = ['의', '가', '이', '은', '들', '는', '좀', '잘', '강', '과', '도', '를', '으로', '자', '에', '와', '한', '하다']

# dfs에 있는 모든 데이터프레임에 대해 불용어 제거 및 문자열 형태로 저장
for df_name, df in dfs.items():
    print(f"Processing {df_name}...")
    if 'Morphs' in df.columns: # 형태소 분석 결과가 있는지 확인
        # 불용어 제거 결과를 저장할 리스트 초기화
        df['Morphs_Cleaned'] = None # 새로운 열을 생성

    for idx, row in df.iterrows():
        morphs = row['Morphs']
        # 불용어 제거
        morphs_cleaned = [word for word in morphs if word not in stopwords]
        # 결과를 문자열로 변환하여 해당 행에 저장
        df.at[idx, 'Morphs_Cleaned'] = " ".join(morphs_cleaned)
```

Procedure

The operations are applied iteratively to all dataframes in the dictionary.

This code refines the textual data by removing unnecessary words, leaving only the essential words in each review after morphological analysis.

[Data and Methodology] 2. Text Preprocessing - Data refinement

In the data filtering and refinement process, special characters were removed, and reviews were segmented to a maximum length of 200 characters. Typographical errors were corrected using the Hanspell library, and spacing issues were addressed with the pykosspacing library.

```
from hanspell import spell_checker
from pykosspacing import Spacing
import re

# 띄어쓰기 조정을 위한 함수 초기화
spacing = Spacing()

for df_name, df in dfs.items():
    print(df_name)
    if 'Review' in df.columns:
        for idx, row in df.iterrows():
            review = str(row['Review'])

            # 특수 문자 제거
            review = re.sub(r'^가-힣0-9a-zA-ZWs$', '', review)

            # 리뷰 길이가 지나치게 길 경우 분할
            max_length = 200 # 적절한 길이 설정
            review_parts = [review[i:i+max_length] for i in range(0, len(review), max_length)]

            corrected_parts = []
            for part in review_parts:
                # 오타 수정
                try:
                    spelled_review = spell_checker.check(part)
                    corrected_review = spelled_review.checked
                    corrected_parts.append(corrected_review)
                except Exception as e:
                    print(f'Error processing review part: {part[:30]}... Error: {e}')

            # 모든 종류의 공백 문자 제거 및 띄어쓰기 조정
            corrected_text = ''.join(corrected_parts)
            corrected_text = spacing(corrected_text)

            # 데이터프레임 업데이트
            df.at[idx, 'Review'] = corrected_text
```

Procedure

Used reg exp 're.sub(r'^가-힣0-9a-zA-ZWs\$', '', review)' to eliminate non-Korean, non-English, and spaces.

Set a maximum length of 200 characters to manage lengthy reviews. Segmented reviews into parts using the code 'review_parts = [review[i:i+max_length] for i in range(0,len(review), max_length)]'

Used the Hanspell library for spell-checking with the code 'spelled_review = spell_checker.check(part)'. Corrected typos in the reviews to enhance the overall text quality.

Utilized the pykosspacing library to adjust spacing. Applied the code 'corrected_text = spacing(corrected_text)' to ensure proper spacing in the refined text.

[Data and Methodology] 3. Sentiment Analysis

The code applies VADER sentiment analysis to assess sentiment in customer reviews, adjusting for strong negativity in labeled '1' reviews. It further integrates KoBERT for deep learning-based sentiment analysis, offering nuanced insights into sentiments, especially in complex language structures.

```
from nltk.sentiment.vader import SentimentIntensityAnalyzer
if text_cleaned == '':
    senti_scores['compound'] = -1
    senti_scores['pos'] = -1
    senti_scores['neg'] = -1
    senti_scores['neu'] = -1

for df_name, df in dfs.items():
    print(f"Processing {df_name}...")
    if 'Review' in df.columns: # 이미 불용어가 처리된 텍스트 데이터가 있는 열을 확인합니다.
        df['sent'] = None # 새로운 열을 생성하여 감정 점수를 저장할 준비를 합니다.

    for idx, row in df.iterrows():
        text_cleaned = row['Review'][:30] # 처음 30자까지만 사용

        # 문장을 KoBERT 모델 입력 형식으로 변환
        inputs = tokenizer(text_cleaned, return_tensors='pt', truncation=True, padding=True, max_length=64)

        # 모델로부터 감정 점수 계산
        outputs = model(**inputs)
        logits = outputs.logits
        probabilities = torch.softmax(logits, dim=-1)

        # 긍정 및 부정 감정 점수 추출
        pos_score = probabilities[0][1].item()
        neg_score = probabilities[0][0].item()

        # 레코드 값이 '1'인 경우 'compound' 값을 -1로 설정
        if text_cleaned == '':
            pos_score = -1
            neg_score = -1

        # 감정 점수를 해당 행에 저장
        df.at[idx, 'sent_pos'] = pos_score
        df.at[idx, 'sent_neg'] = neg_score
```

Procedure
Star rating of '1' often corresponds to a strongly negative sentiment in reviews, setting the sentiment scores to -1 is intended to portray such reviews as expressing intense negativity.
Text is tokenized and converted into input format of KoBERT model.
Calculate sentiment scores, probabilities for positive and negative sentiments.
Positive and negative sentiment scores are then stored.

[Data and Methodlogy] 4. Clustering Analysis – K means

Study used K-Means clustering, utilizing TF-IDF vectorization to represent reviews numerically based on word importance. Clustering analyzed reviews with ratings 1 or 2 and 4 or 5, revealing strengths, weaknesses, and sentiments for product improvement.

```
# TF-IDF 벡터화
vectorizer = TfidfVectorizer()
X = vectorizer.fit_transform(review_texts)

# K-Means 군집화
n_clusters = 3 # 군집의 개수
kmeans = KMeans(n_clusters=n_clusters, random_state=42)
kmeans.fit(X)
labels = kmeans.labels_

# 각 군집별 주요 단어 추출
top_words = get_top_features_cluster(X, labels, 20, vectorizer.get_feature_names_out())
```

Procedure

Utilized TfidfVectorizer to transform text data into numerical values using the TF-IDF format, representing each review with numerical importance assigned to words.

The variable `n_clusters` specifies the number of clusters.

After clustering, the `get_top_features_cluster` function was utilized to extract pivotal words from each cluster, facilitating a deeper understanding of their unique characteristics.

Clustering analyzed reviews with ratings 1 or 2 and 4 or 5, revealing strengths, weaknesses for product improvement.

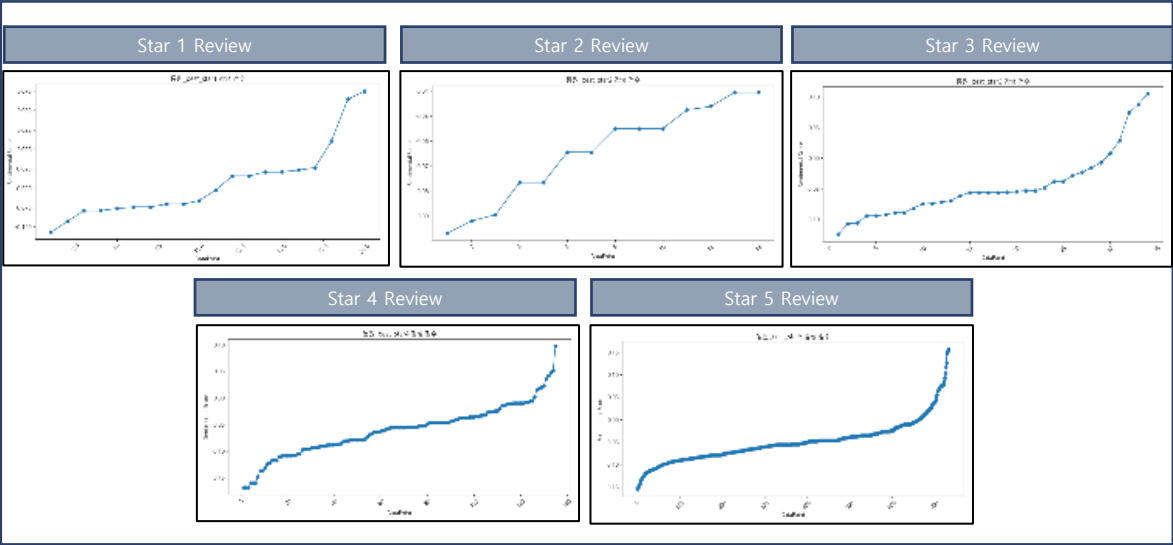
By identifying distinctive words and topics, businesses gained insights into diverse customer opinions, enhancing products across different rating ranges.

This approach provided a nuanced understanding of customer feedback.

[Analysis Result] Sentiment Analysis

Based on the sentiment analysis results, the average sentiment score for our company is -0.0572, significantly lower compared to competitors. This indicates that our product is perceived much more negatively by customers in comparison to competitors. Therefore, it is crucial to improve products.

	Mode		Median		Std Deviation	
Dongwon F&B		-0.056552839		-0.057240099		0.042653809
CJ		-0.056655204		-0.051786143		0.049678045
Pulmuwon		-0.040484864		-0.053668845		0.04945674
Haitai		-0.047643089		-0.051822263		0.048997711



The need for improvement

Through sentiment analysis, the average score for our company is -0.0572 (standard deviation 0.04265). Comparing this to the average scores of competitors, which are -0.0517, -0.0536, and -0.0518, **our score is noticeably lower.**

This indicates that negative sentiments are more pronounced in customer reviews for our products. **Therefore, it further emphasizes the necessity for product improvement through this project.**

[Analysis Result] Clustering Analysis

The cluster analysis revealed clusters related to various features of each company's products. These clusters were further interpreted by categorizing them into (1) Product (filling, wrapper), (2) Price, and (3) Others (consumption methods, etc.). Additionally, insights were derived based on the research results to apply a new positioning for Dongwon F&B.

			cluster	Analysis Result	Insight
p r o d u c t	F i l l i n g	Dong won F&B	[군집 1-2] 물컹물컹, 식감, 느 낌, 떨어지다	Although our dumplings are generously filled and meet customer needs with a substantial quantity, there is an identified issue with the dumpling filling being too mushy, leading to a compromised texture.	Complaints about the filling can be categorized into ㉠ texture (mushy), ㉡ unpleasant odor, and ㉢ excessive use of strong seasonings (too salty). Therefore, there is a need to develop a filling that maintains an optimal texture and eliminates the meaty odor. In contrast to competitors such as our company, Pulmuone, and Haitai, CJ does not seem to have negative evaluations regarding the dumpling filling. CJ adopts a processing method of 'chopping' rather than 'grinding' the filling to eliminate unpleasant odors, indicating the necessity for process R&D to preserve the texture.
		CJ		Since CJ is the only one without negative feedback on taste, it is crucial to incorporate reflections on CJ's dumpling production technique.	
		Pul Mu Won	[군집 1-1] 별로, 느끼하다, 맛 없다 [군집 1-3] 짜다, 고기, 냄새, 맛 없다 [군집 2-3] 짜다, 고기	The meaty odor, greasiness, and excessive saltiness are confirmed to be compromising the taste.	
		Haitai	[군집 1-1] 조미료, 강하다	The strong seasoning is adversely affecting the overall flavor. Haitai's dumpling products receive more reviews stating a relatively bland taste compared to other brands.	
	W r a p p e r	Dong won F&B	[군집 1-3] 만두피, 두껍다, 힘 들다	The thick dumpling wrapper causes inconvenience for customers during consumption.	Considering the trend of 'thin dumpling wrappers' following the success of Pulmuone's 'Yalpi Mandu,' R&D efforts are required to reduce the thickness of the dumpling wrapper. Moreover, in terms of the product, considering CJ's emphasis on 'good dumpling filling' and Pulmuone's focus on 'thin dumpling wrappers' meeting customer needs, our company needs to strategize for a distinct product identity.
		CJ			
		Pul Mu Won	[군집 2-1] 얇다, 좋다 [군집 2-2] 얇다, 정말, 최고, 만 두피	Pulmuone promotes 'Yalpi Mandu' as a product featuring thin dumpling wrappers, and the cluster analysis indicates a preference for thin dumpling wrappers.	
		Haitai			

[Analysis Result] Clustering Analysis

		Cluster	Analysis Result	Insight
P r i c e	Dong Won F&B	[군집 1-3] 비싸다, 가격 등 [군집 2-1] 가격, 가성비, 저렴하다	It can be inferred that our company's pricing positioning is not clearly defined, as there are clusters both indicating high prices and conflicting clusters suggesting cost-effectiveness and affordability.	In the case of CJ and Haitai, similar to frozen dumplings, the affordable pricing leaves a positive impression on customers, emphasizing the importance of pricing positioning. However, when the actual prices are compared, CJ's dumplings are positioned as so-called premium dumplings, being 10-20% more expensive than those perceived as affordable, such as Haitai. Therefore, efforts should be made to ensure that customers perceive the pricing as reasonable relative to the product quality.
	CJ	[군집 2-1] 많다, 저렴하다, 매달 [군집 2-2] 저렴하다 [군집 2-3] 가격, 좋다, 저렴하다, 용량, 착하다	With multiple clusters highlighting good prices, ample quantity, and affordability, it is evident that CJ's pricing is perceived as budget-friendly by customers, presenting a significant advantage.	
	Pul Mu Won			
	Haitai	[군집 2-1] 용량, 많다 [군집 2-2] 가격, 싸다, 저렴하다, 용량 [군집 2-3] 가격, 저렴하다, 가성비	Similarly, the presence of clusters emphasizing good prices, ample quantity, and affordability indicates that Haitai's pricing is perceived as budget-friendly by customers, serving as a significant advantage.	
O t h e r s	B r a n d	Dong Won F&B	[군집 2-2] 개성, 왕만두 [군집 2-3] 개성, 왕만두	While our brand and flagship product, '개성 왕만두' (Unique King Mandu), are clearly recognized by customers, we do not dominate the market as overwhelmingly as '비비고 왕교자' (Bibigo King Mandu), which is evident even in reviews from other brands. Therefore, benchmarking CJ's brand imprint strategy and attempting to enhance brand awareness are necessary.
		CJ	[군집 2-1] 비비고, 역시, 항상 [군집 2-2] 비비고, 왕교자 [군집 2-3] 비비고	
		Pul Mu Won	[군집 2-1] 풀무원, 비비고	
		Haitai	[군집 1-1] 옛날, 예전, 고향	

[Analysis Result] Clustering Analysis

			Cluster	Analysis Result	Insight
Others	Consumption Methods	Dong Won F&B	[군집 2-2] 찌다, 넣다, 끓이다, 만두국	The customer's dumpling consumption methods include ㉠ steaming and ㉢ using as an ingredient in other dishes (dumpling soup)	The customer's dumpling consumption methods include ㉠ steaming, ㉢ baking (using an air fryer, etc.), and ㉡ using as an ingredient in other dishes (dumpling soup, ramen). Considering the customer's consumption habits, there is a need for future product development that takes these methods into account.
		CJ	[군집 2-1] 만두국, 찌다, 굽다 [군집 2-3] 간식, 아이		
		Pul Mu Won	[군집 2-2] 만두국, 에프, 돌리다		
		Haitai	[군집 1-1] 라면, 만들다 [군집 2-1] 찌다, 넣다, 굽다		
	Others	Dong Won F&B	[군집 1-1] 배송, 머리카락, 버리다 등	Poor delivery conditions resulted in product breakage, and there have been instances where foreign substances were found, leading to the disposal of dumplings	As common delivery issues have been identified, corrective measures are necessary in packaging and other aspects. Furthermore, complaints regarding basic foreign materials have been noted in comparison to our company and competitors, emphasizing the need for efforts to improve quality.
		CJ	[군집 1-1] 늦어지다, 너무, 배송, 제품, 찢어지다, 냉동, 상태, 힘들다 [군집 1-2] 녹다, 배송, 오다, 너무, 아이스팩, 냉동식품, 인데 [군집 1-3] 부서지다, 상자, 배달, 신경, 스티로폼, 반품, 포장	delivery issues have been identified	
		Pul Mu won			
		Haitai	[군집 1-2] 배송, 녹다	delivery issues have been identified	

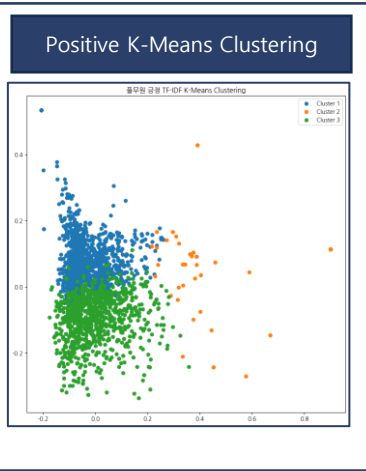
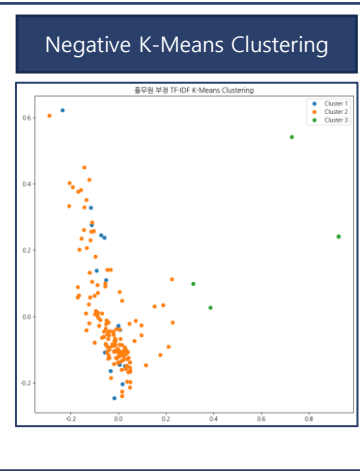
(Appendix) K-means Output

Dongwon F&B	군집 1-1 주요 단어:		군집 1-2 주요 단어:		군집 1-3 주요 단어:		군집 2-1 주요 단어:		군집 2-2 주요 단어:		군집 2-3 주요 단어:	
	단어	점수	단어	점수	단어	점수	단어	점수	단어	점수	단어	점수
	0	만두 0.018536	0	음용물청 0.431570	0	비싸다 0.473079	0	좋다 0.005157	0	먹다 0.182759	0	맛있다 0.401505
	1	먹다 0.017443	1	식감 0.431570	1	견입 0.473079	1	배송 0.002775	1	만두 0.140882	1	먹다 0.099602
	2	있다 0.013574	2	매우 0.431570	2	가격 0.445965	2	이다 0.002651	2	좋다 0.085409	2	만두 0.063583
	3	없다 0.011981	3	느낌 0.431570	3	두피 0.445965	3	포장 0.002558	3	있다 0.072582	3	개성 0.051984
	4	정말 0.011267	4	말리지다 0.406835	4	두집다 0.393221	4	가격 0.002425	4	맛있다 0.058170	4	보다 0.043920
	5	날다 0.011006	5	먹다 0.299122	5	힘들다 0.000000	5	빠르다 0.002386	5	짜다 0.045951	5	원만 0.041465
	6	아니다 0.008757	6	마련 0.000000	6	만두 0.000000	6	감사하다 0.002237	6	자다 0.044666	6	너무 0.034675
	7	나다 0.008187	7	로켓 0.000000	7	로켓 0.000000	7	양도 0.002138	7	완만 0.042249	7	짜다 0.033619
	8	인데 0.008121	8	로켓 0.000000	8	리치 0.000000	8	만두 0.001870	8	보다 0.040027	8	크다 0.031876
	9	이렇다 0.008041	9	리치 0.000000	9	마라 0.000000	9	완만 0.001817	9	개성 0.035562	9	좋다 0.030958
	10	제품 0.007630	10	마라 0.000000	10	마련 0.000000	10	가성 0.001803	10	너무 0.035079	10	자다 0.027232
	11	그냥 0.007256	11	힘들다 0.000000	11	마찬가지 0.000000	11	아니다 0.001792	11	날다 0.034741	11	이다 0.024623
	12	배송 0.006983	12	마찬가지 0.000000	12	마르 0.000000	12	맛있다 0.001787	12	쿨이다 0.031371	12	많다 0.023574
	13	머리카락 0.006987	13	리치 0.000000	13	만들다 0.000000	13	자다 0.001786	13	만족국 0.030857	13	있다 0.023515
	14	보다 0.006422	14	만두 0.000000	14	만두소 0.000000	14	보다 0.001716	14	주분 0.030052	14	같다 0.020337
	15	두집다 0.006359	15	만두소 0.000000	15	리치 0.000000	15	많다 0.001709	15	같다 0.029877	15	좋아하다 0.020279
	16	버리다 0.006295	16	만들다 0.000000	16	만지다 0.000000	16	저렴하다 0.001694	16	이다 0.029349	16	생각 0.019147
	17	돼지 0.006214	17	만지다 0.000000	17	만들 0.000000	17	좋다 0.001660	17	좋아하다 0.028469	17	진짜 0.018782
	18	좋아하다 0.006176	18	만족 0.000000	18	많다 0.000000	18	시키다 0.001615	18	많다 0.026672	18	많다 0.018195
	19	같다 0.005887	19	많다 0.000000	19	많이 0.000000	19	지퍼 0.001542	19	알다 0.025960	19	만족국 0.017576
CJ	군집 1-1 주요 단어:		군집 1-2 주요 단어:		군집 1-3 주요 단어:		군집 2-1 주요 단어:		군집 2-2 주요 단어:		군집 2-3 주요 단어:	
	단어	점수	단어	점수	단어	점수	단어	점수	단어	점수	단어	점수
	0	오다 0.668655	0	눅다 0.018335	0	참다 0.390461	0	맛있다 0.913972	0	먹다 0.093485	0	좋다 0.626957
	1	눅다 0.509589	1	만두 0.018055	1	부서지다 0.353699	1	많다 0.052316	1	만두 0.078650	1	가격 0.064069
	2	만두 0.171983	2	배송 0.017720	2	상자 0.353699	2	비비고 0.029179	2	맛있다 0.062387	2	저렴하다 0.057759
	3	놓여지다 0.116937	3	먹다 0.012564	3	쓸다 0.332195	3	저렴하다 0.014607	3	비비고 0.054036	3	간식 0.042085
	4	너무 0.074112	4	오다 0.011457	4	배달 0.324016	4	역시 0.013989	4	있다 0.038888	4	음향 0.036122
	5	배송 0.060322	5	있다 0.011034	5	신경 0.305103	5	만두 0.013784	5	자다 0.036204	5	맛있다 0.031817
	6	제품 0.056955	6	너무 0.009849	6	스티로폼 0.305103	6	먹다 0.013600	6	보다 0.033683	6	많다 0.029255
	7	놓여지다 0.046337	7	아이스크림 0.009186	7	반쯤 0.268341	7	너무 0.007180	7	좋다 0.033383	7	너무 0.027477
	8	냉통 0.045985	8	냉통 0.008641	8	포장 0.255506	8	황고 0.007173	8	구매 0.025558	8	양도 0.025208
	9	상태 0.032215	9	되다 0.008308	9	만두 0.171148	9	짜다 0.007107	9	너무 0.024238	9	이라 0.024838
	10	말랑말랑하다 0.000000	10	보다 0.007920	10	눅다 0.169477	10	있다 0.006738	10	황고 0.023956	10	제품 0.024750
	11	말랑하다 0.000000	11	없다 0.007876	11	완주다 0.000000	11	좋다 0.005290	11	좋아하다 0.023951	11	비비고 0.022481
	12	맛있다 0.000000	12	그냥 0.007207	12	말기다 0.000000	12	가격 0.004210	12	보다 0.019302	12	보다 0.019302
	13	명하다 0.000000	13	날다 0.006995	13	말다 0.000000	13	만족국 0.003959	13	역시 0.022231	13	착하다 0.018608
	14	힘들다 0.000000	14	없다 0.006850	14	말고 0.000000	14	굽다 0.003482	14	짜다 0.021978	14	역시 0.018509
	15	맛다 0.000000	15	냉통식물 0.006724	15	매다 0.000000	15	자다 0.002698	15	같다 0.021411	15	자주 0.016048
	16	말다 0.000000	16	아니다 0.006562	16	명하다 0.000000	16	맞주다 0.000000	16	이다 0.020737	16	아주 0.016397
	17	완주다 0.000000	17	인데 0.005932	17	매번 0.000000	17	매다 0.000000	17	저렴하다 0.020143	17	아이 0.016383
	18	말기다 0.000000	18	하나 0.005812	18	맛있다 0.000000	18	매달 0.000000	18	냉통 0.019390	18	구매 0.016129
	19	말다 0.000000	19	좋다 0.005644	19	말랑하다 0.000000	19	매번 0.000000	19	주분 0.018381	19	상품 0.015869

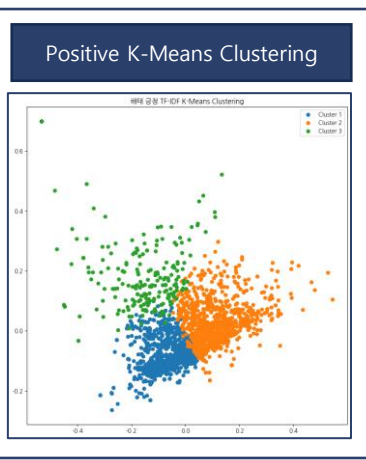
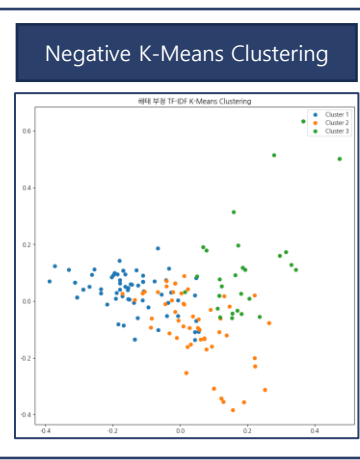


(Appendix) K-means Output

Pulmuwon	물무원 군집 1-1 주요 단어:		물무원 군집 1-2 주요 단어:		물무원 군집 1-3 주요 단어:		물무원 군집 2-1 주요 단어:		물무원 군집 2-2 주요 단어:		물무원 군집 2-3 주요 단어:	
	단어	점수	단어	점수	단어	점수	단어	점수	단어	점수	단어	점수
	0	그냥 0.006320	0	쓰다 0.336219	0	먹다 0.166467	0	먹다 0.136247	0	맛있다 0.701567	0	좋다 0.030164
	1	없다 0.006187	1	생각 0.272687	1	만두 0.157934	1	만두 0.131238	1	없다 0.110665	1	맛있다 0.023749
	2	그렇다 0.005574	2	나다 0.269376	2	짜다 0.157478	2	맛있다 0.092531	2	만두 0.060475	2	먹다 0.015943
	3	별로 0.004966	3	짜다 0.212681	3	너무 0.080202	3	없다 0.069568	3	먹다 0.044320	3	없다 0.015234
	4	고기 0.004638	4	먹다 0.185467	4	없다 0.069027	4	있다 0.066409	4	진짜 0.039984	4	있다 0.013366
	5	있다 0.004454	5	가지다 0.182060	5	보다 0.055593	5	좋다 0.052253	5	좋다 0.039576	5	만두 0.013290
	6	배송 0.004342	6	기울 0.182060	6	없다 0.052977	6	보다 0.050330	6	정말 0.039087	6	보다 0.009736
	7	너무 0.004021	7	기억 0.182060	7	맛있다 0.047032	7	맛있다 0.045146	7	만토록 0.033856	7	자다 0.009694
	8	아니다 0.003976	8	결절 0.182060	8	이다 0.037732	8	고기 0.042011	8	김치 0.032930	8	짜다 0.009654
	9	맛없다 0.003875	9	정말 0.182060	9	고기 0.036929	9	물무원 0.036983	9	애프 0.031538	9	가격 0.009334
	10	좋다 0.003822	10	장기 0.182060	10	구매 0.036172	10	너무 0.036172	10	이다 0.029141	10	비비고 0.009253
	11	보내다 0.002716	11	이미지 0.182060	11	같다 0.032950	11	자다 0.035148	11	사다 0.026332	11	없다 0.009163
	12	적다 0.002587	12	먼저 0.182060	12	별로 0.030992	12	좋아하다 0.034065	12	물리다 0.026327	12	괜찮다 0.009061
	13	날다 0.002571	13	물랑 0.182060	13	날새 0.030433	13	구매 0.033946	13	최고 0.020156	13	배송 0.009005
	14	느끼하다 0.002569	14	미지다 0.182060	14	많이 0.027175	14	김치 0.031968	14	숙이다 0.020145	14	고기 0.008704
	15	오다 0.002551	15	싫어하다 0.174203	15	김치 0.026413	15	같다 0.031355	15	구매 0.018487	15	숙이다 0.008522
	16	먹다 0.002541	16	순간 0.174203	16	있다 0.026300	16	날다 0.031103	16	해트 0.015578	16	구매 0.008074
	17	시키다 0.002494	17	폐기 0.168109	17	맛없다 0.025682	17	고기만 0.030679	17	재다 0.015357	17	많이 0.008051
	18	눅다 0.002387	18	살다 0.168109	18	물무원 0.024668	18	만토록 0.029956	18	두피 0.014901	18	이다 0.007472
	19	박스 0.002304	19	회수 0.168109	19	그렇다 0.024134	19	좋다 0.028361	19	같다 0.013562	19	없다 0.006319



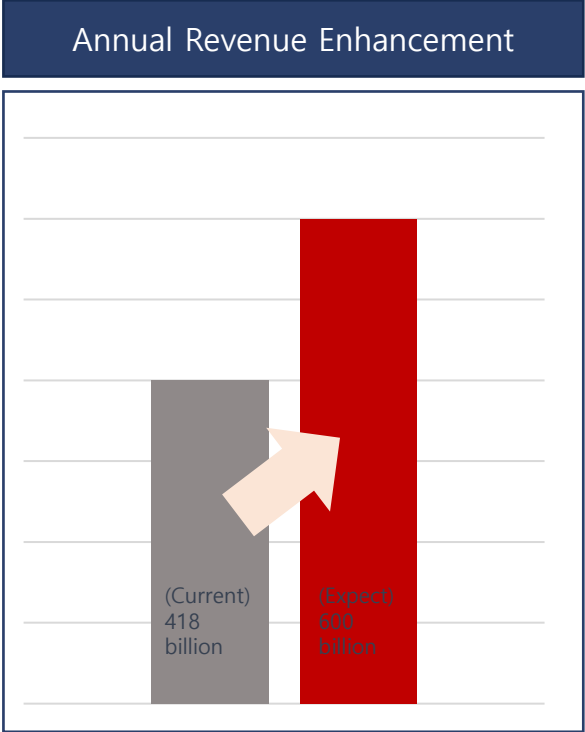
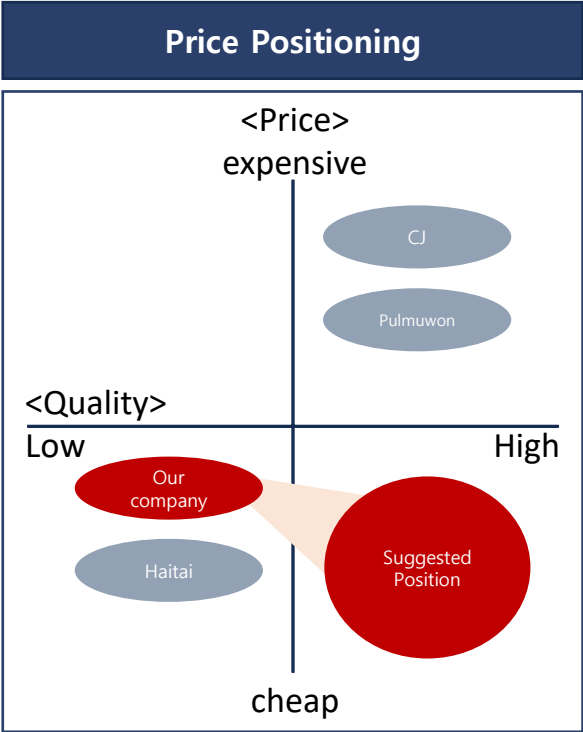
Haitai	군집 1-1 주요 단어:		군집 1-2 주요 단어:		군집 1-3 주요 단어:		군집 2-1 주요 단어:		군집 2-2 주요 단어:		군집 2-3 주요 단어:	
	단어	점수	단어	점수	단어	점수	단어	점수	단어	점수	단어	점수
	0	물결 0.330897	0	만두 0.022290	0	맛없다 1.0	0	만두 0.166133	0	먹다 0.013223	0	좋다 0.264736
	1	비고 0.330897	1	먹다 0.020634	1	힘들다 0.0	1	먹다 0.156258	1	없다 0.011972	1	없다 0.096456
	2	조미료 0.330897	2	없다 0.016326	2	마르 0.0	2	고향 0.082630	2	가격 0.011311	2	먹다 0.066599
	3	강하다 0.310823	3	배송 0.014277	3	많이 0.0	3	맛있다 0.077695	3	배송 0.011266	3	먹기 0.064816
	4	나서다 0.262263	4	있다 0.012838	4	많다 0.0	4	없다 0.049495	4	만두 0.011106	4	만두 0.063341
	5	숙이다 0.256432	5	별로 0.011547	5	만들 0.0	5	짜다 0.043922	5	자다 0.010615	5	가격 0.063339
	6	이라 0.256432	6	고기 0.009877	6	만족 0.0	6	자다 0.040460	6	보다 0.010274	6	양도 0.052576
	7	예전 0.246497	7	아니다 0.009834	7	만에 0.0	7	보다 0.039018	7	있다 0.010036	7	있다 0.036536
	8	사다 0.227946	8	오다 0.009693	8	만들어지다 0.0	8	날다 0.037404	8	크다 0.009416	8	맛있다 0.036413
	9	좋다 0.216898	9	보다 0.009556	9	만들다 0.0	9	좋다 0.034342	9	맛있다 0.009061	9	너무 0.032153
	10	너무 0.212180	10	눅다 0.008864	10	만토록 0.0	10	좋아하다 0.031316	10	싸다 0.008078	10	배송 0.032101
	11	가다 0.205851	11	그냥 0.008730	11	만두소 0.0	11	구매 0.029882	11	주은 0.007989	11	크기 0.031286
	12	고향 0.202040	12	같다 0.008723	12	만두 0.0	12	같다 0.029238	12	빠르다 0.007666	12	고향 0.030849
	13	아니다 0.189192	13	너무 0.008411	13	마지 0.0	13	너무 0.028112	13	이다 0.007733	13	저절하다 0.029970
	14	먹다 0.162507	14	냉동 0.007423	14	말다 0.0	14	없다 0.027753	14	사다 0.007632	14	물랑 0.028811
	15	만두 0.156311	15	고향 0.006966	15	마지막 0.0	15	같다 0.027059	15	같다 0.006601	15	빠르다 0.027347
	16	라서 0.000000	16	없다 0.006529	16	마음 0.0	16	이다 0.026737	16	저절하다 0.006713	16	보다 0.026027
	17	만들다 0.000000	17	가다 0.006216	17	마르다 0.0	17	물랑 0.025973	17	물랑 0.006686	17	아이 0.023680
	18	라만 0.000000	18	좋다 0.006008	18	리소 0.0	18	먹시 0.023703	18	반다 0.006610	18	대비 0.023265
	19	라면 0.000000	19	그렇다 0.005832	19	로켓 0.0	19	없다 0.023351	19	냉동 0.006234	19	가성 0.022808



[Expected original contribution]

Through the current project, we can identify the strengths and pain points that customers associate with Dongwon products, providing valuable insights for product improvement. This approach will serve as a key factor in fostering revenue growth.

Product Improvement Proposal		
Product	Filling	- Adopt a processing method of 'Chopping'
	Wrapper	- Reduce the thickness - Strategize for a distinct product identity
Price		- Pricing positioning
Others	Brand	- Benchmark CJ's brand strategy to enhance brand awareness
	Delivery	- Enhance packaging



[Reference List]

윤종필(2022). 온라인 신선식품 전문몰에서의 구매 후기 분석, 고려대학교 대학원

김치훈(2022). Text Mining Analysis of Hotel Reviews Focused on TripAdvisor Reviews, 건국대학교 대학원

채진해 등(2020). 텍스트 빅데이터 분석을 통한 COVID-19 전후 서울시 주요 도시공원의 시민 이용행태 및 관심도 변화, 2020서울연구논문공모전

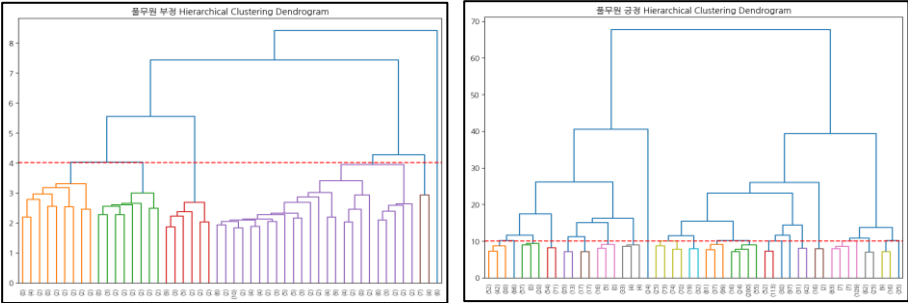
조우리 등(2021). 텍스트마이닝을 활용한 SNS 내 미술 전시 감상 후기 분석, 조형교육 79집

중소기업벤처진흥공단 (2019). KOSME 산업분석 리포트-음식류: 간편식(HMR)-

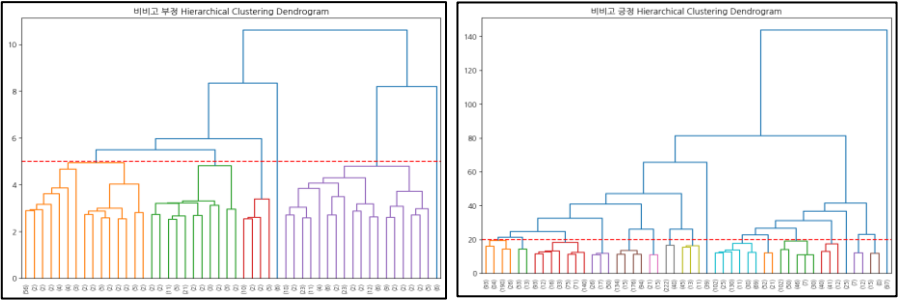
(Appendix) Hierarchical Clustering Output

To understand customer perceptions of our company and competitors' product features, we conducted not only k-means cluster analysis but also went further to include Hierarchical Clustering and Topic Modeling. While our primary intention was to derive insights, we found that k-means cluster analysis was the most effective in achieving this goal. Therefore, we centered our utilization of the research results around the outcomes obtained from k-means cluster analysis.

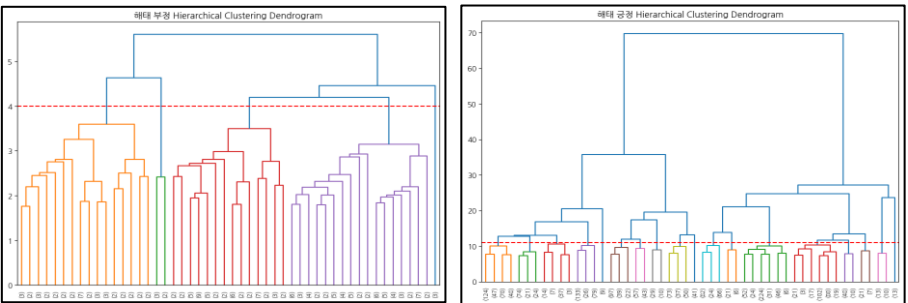
"Pulmuone" Hierarchical Clustering Dendrogram



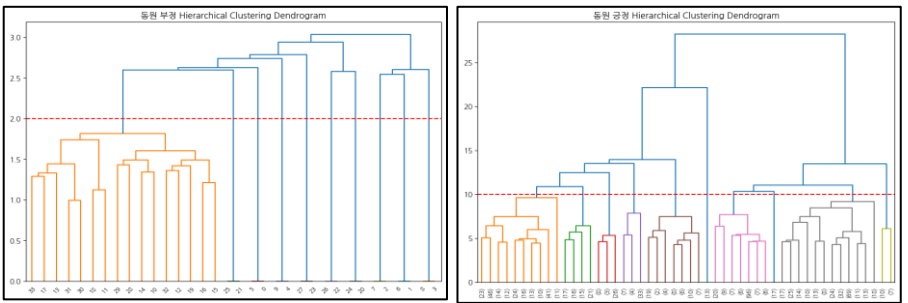
"Bibigo" Hierarchical Clustering Dendrogram



"Haitai" Hierarchical Clustering Dendrogram



"Dongwon F&B" Hierarchical Clustering Dendrogram



"Pulmuone"

