

Establishment of Marketing Strategy for Online Shopping Mall through Customer Cluster Analysis

Seonghye Kim · Joonsoo Bae[†]

Department of Management of Technology, Jeonbuk National University

소비자 군집분석을 통한 온라인 쇼핑몰 마케팅 전략 수립

김성혜 · 배준수[†]

전북대학교 융합기술경영학과

This study aims to establish an online shopping mall marketing strategy based on big data analysis methods. The customer cluster analysis method was utilized to analyze customer purchase patterns and segment them into customer groups with similar characteristics. Data was collected from orders placed over one year in 2023 at 'Jeonbuk Saengsaeng Market', the official online shopping mall for agricultural, fish, and livestock products of Jeonbuk Special Self-Governing Province. K-means clustering was conducted by creating variables such as 'TotalPrice' and 'ElapsedDays' for analysis. The study identified four customer groups, and their main characteristics. Furthermore, regions corresponding to customer groups were analyzed using pivot tables. This facilitated the proposal of a marketing strategy tailored to each group's characteristics and the establishment of an efficient online shopping mall marketing strategy. This study is significant as it departs from the traditional reliance on the intuition of the person in charge to operate a shopping mall, instead establishing a shopping mall marketing strategy through objective and scientific big data analysis. The implementation of the marketing strategy outlined in this study is expected to enhance customer satisfaction and boost sales.

Keywords : Jeonbuk Saengsaeng Market, Consumer Cluster Analysis, Marketing Strategy, K-means Clustering, Big Data

1. 서 론

1.1 연구배경 및 목적

코로나 팬데믹 이후 이커머스 시장의 급격한 성장과 함께 소비자의 구매 패턴은 더욱 다양화되고 복잡해지고 있다. 통계청이 발표한 '온라인 쇼핑 동향' 조사에 의하면 온라인 쇼핑 거래액은 2019년 135조 원에서 2023년 229조 원으로 69.6% 증가하며 역대 최대치를 갱신했다. 특히 여

러 상품군 중 식품군(음·식료품, 농축수산물)의 거래액이 2019년 17조 원에서 2023년 41조 원으로 141.1% 증가하였고, 이는 온라인 식품 구매에 대한 소비자들의 수요가 급증하고 있음을 보여준다. 이러한 현상은 코로나 팬데믹 이후 일상이 된 비대면 거래와 급격한 IT 기술 발전으로 편리함을 추구하는 소비자들의 경향이 반영된 결과이다. 이러한 소비자들의 편리함 추구는 가정간편식(HMR) 수요 증가, 배달 서비스 이용 증가, 정기구독 서비스 수요 증가, 새벽 배송 서비스 확대, 모바일 간편 결제 활성화 등 새로운 소비 트렌드를 불러일으켰다. 이러한 변화는 소비자들의 소비 양상을 더욱 다변화시켰으며, 앞으로도 지속될 것으로 전망된다. 이에 따라 온라인 쇼핑몰들은 치열해진 경쟁 속에서 시장 경쟁력을 확보하기 위해 개인화된 마케팅

Received 31 July 2024; Finally Revised 8 September 2024;

Accepted 9 September 2024

[†] Corresponding Author : jsbae@jbnu.ac.kr

전략 수립으로 소비자의 다양한 요구(needs)에 대응하는 것이 중요해졌다. 한편, 이러한 변화 속에서 지방자치단체(이하 지자체) 온라인 쇼핑몰 또한 지역 특산물과 농수축산물의 온라인 판매를 촉진하기 위해 운영되고 있다. 각 지자체는 지역 농가 및 중소기업들의 소득증대와 지역경제 활성화를 목적으로 온라인 쇼핑몰을 운영하고 있으며, 이를 통해 소비자들에게 신선하고 질 좋은 지역 농특산물을 제공하고 있다. 이러한 지자체 온라인 쇼핑몰에는 전북특별자치도의 ‘전북생생장터’, 강원특별자치도의 ‘강원터물’, 전라남도의 ‘남도장터’, 경상남도의 ‘e경남물’ 등이 있으며, 엄격한 입점 상품 관리와 가격할인 지원을 통해 소비자가 신뢰할 만한 안전하고 저렴한 먹거리를 제공한다는 긍정적인 인식을 얻고 있다. 그러나 변화하는 시대 흐름 속에서 모든 소비자를 대상으로 하는 단순 일회성 할인 이벤트만으로는 경쟁력을 확보하는 데 한계가 있으므로 소비자의 다양한 요구를 반영할 수 있는 효과적인 마케팅 전략의 필요성이 대두되고 있다. 이에 본 연구에서는 소비자 군집분석을 통해 소비자를 세분화하고, 각 군집별 특성을 바탕으로 온라인 쇼핑몰 마케팅 전략을 제안하고자 한다. 이를 통해 소비자 만족도를 높이고, 매출 증대에 기여할 수 있을 것으로 기대된다.

1.2 연구방법

군집분석(Cluster analysis)은 데이터 마이닝의 일환으로 많은 양의 데이터들의 특성을 분석하여 유사한 특성을 가진 데이터들을 동일한 집단으로 분류하는 분석 기법이다. 대용량 데이터의 숨겨진 패턴이나 구조 등을 발견하고, 이를 바탕으로 의사결정 등을 위한 정보로 활용할 수 있어 다양한 분야에서 활용된다[13].

군집분석에는 여러 기법이 존재하며, 계층적 군집분석, DBSCAN, 가우시안 혼합분포 군집분석, K-평균 군집분석 등이 있다. 계층적 군집분석은 대상을 가까운 것끼리 순차적으로 묶어(혹은 먼 대상끼리 분리하여) 군집들이 트리 구조를 이루도록 하는 방법이며 데이터가 매우 큰 경우에는 계산 시간이 길고, 이상치에 민감하다는 한계가 있다. 가우시안 혼합분포 군집분석(Gaussian mixture model clustering)은 분포 기반의 군집분석 방법이며 수렴 속도가 느리고, 군집의 크기가 작을수록 추정 정확도가 저하된다는 단점이 있다. DBSCAN은 밀도 기반의 군집화 기법으로, 군집의 수를 미리 정하지 않아도 되지만 DBSCAN 수행을 위한 파라미터 결정이 어렵다는 단점이 있다[11]. 본 연구에서는 K-평균 군집분석(K-means clustering)을 선택하였는데, 계층적 군집분석과 동일하게 거리 척도를 사용하는 거리 기반의 방법으로서, 계산이 빠르고 원리가 간단하며 대다수의 통계 프로그램에 포함되어 있어 쉽게 적용할 수

있고, 알고리즘 구현이 쉽다는 장점이 있다[14, 16]. 특히, 일반적으로 소비자 가치를 평가하고 세분화하는 방법으로 흔히 사용되는 RFM(Recency, Frequency, Monetary) 분석과 비교할 때, K-평균 군집분석은 다차원 데이터를 처리하여 더 정교한 군집화를 가능하게 하며, 각 군집의 특성을 보다 다각도로 분석할 수 있다는 장점이 있다.

본 연구에서는 K-평균 군집분석을 적용하여 소비자 군집분석을 실시하였다. 먼저 전북특별자치도 농수축산물 공식 온라인 쇼핑몰인 ‘전북생생장터’에서 2023년 1월 1일부터 2023년 12월 31일까지 1년간 발생한 주문 데이터를 수집하였다. 이 데이터는 주문자 ID, 주문일시, 주문내역 등을 포함한다. 데이터 정제 후 분석에 필요한 추가 변수들(총 주문금액, 1회당 평균 주문금액 등)을 생성하고, K-평균 군집분석을 수행하였다. 그 결과 유사한 구매 패턴을 가진 소비자 군집을 도출하였으며, 각 소비자 군집별로 주요 특성을 파악하였다. 또한, 피벗 테이블을 활용하여 소비자 군집에 따른 지역별 특성을 추가 분석하였다. 이를 바탕으로 각 군집의 특성에 맞는 마케팅 전략을 제시하여 효율적인 온라인 쇼핑몰 운영 전략을 수립하였다. 마지막으로 소비자 설문조사를 통해 제안된 전략의 타당성을 검증하였다.

2. 이론적 배경

2.1 소비자 세분화의 중요성

마케팅 관점에서 소비자 세분화(Customer Segmentation)는 중요한 의미를 지닌다[3]. 오늘날 소비자의 요구는 점차 다양하고 복잡해지고 있으며, 이는 소비자의 구매 행동에 큰 영향을 미치고 있다. 소비자들의 다양한 요구에 대응하기 위해서는 모든 소비자를 동일한 대상으로 보던 관점에서 벗어나 보다 개인화된 마케팅 전략이 필요하다[3]. 이를 위해 소비자 데이터베이스를 방대한 DB로만 가지고 있으면 안 된다[8]. 단순 데이터 수집에 그치지 않고 유사한 특성을 가진 소비자들을 하나의 그룹으로 묶는 소비자 세분화가 필수적이다. 세분화된 소비자 그룹 또는 더 나아가서는 개개인의 특성에 맞는 차별화된 마케팅을 수행함으로써 소비자 충성도를 높일 수 있을 것이다. 본 연구에서는 대량의 소비자 데이터베이스를 이용하여 소비자를 세분화하는 방법으로, 데이터마이닝의 일환인 군집분석을 활용하였다[3].

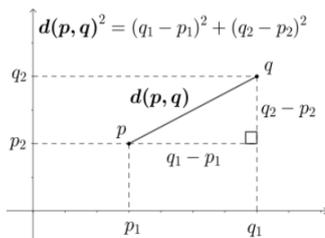
2.2 K-평균 군집분석의 개념 및 방법론

K-평균 군집분석(K-means clustering)은 대용량의 데이

터를 유사한 특성을 가진 k 개의 군집(cluster)으로 구분하는 기법이다. 초기 단계에서 최적의 군집 수(k)를 정하고, k 개의 중심점을 임의의 위치로 잡는다. 각 데이터 포인트를 가장 가까운 군집 중심점에 할당한다. 가까운 군집 중심점은 유클리드 거리(Euclidean Distance, 두 점 사이의 직선거리)를 사용하여 계산하며, 이를 통해 각 데이터 포인트는 k 개의 군집 중 하나에 속하게 된다. 식 (1)은 두 점 p 와 q 가 각각 (p_1, p_2, \dots, p_n) , (q_1, q_2, \dots, q_n) 좌표를 가질 때, 두 점 사이의 거리를 유클리드 거리 공식으로 표현한 식이며, 다차원이 아닌 2차원 공간에서 유클리드 거리는 <Figure 1>과 같다[6].

$$d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2} \quad (1)$$

$$= \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$



<Figure 1> Euclidean Distance in two-dimensional Space

그다음 각 군집에 속한 데이터 포인트들의 평균 지점을 계산하여 중심점을 이동한다. 각 데이터 포인트를 이동된 새로운 중심점을 기준으로 가장 가까운 중심점에 다시 할당하고, 그런 다음 데이터 포인트들의 평균 지점을 다시 계산하여 중심점을 이동한다. 이 과정을 반복하여 더 이상 이동이 발생하지 않는 위치를 찾으면 각 중심점을 기준으로 k 개의 군집이 형성된다. K-평균 군집분석은 k 값을 사전에 설정해야 한다는 단점이 있지만, 최적의 k 값을 찾는 데 엘보우 방법이나 실루엣 분석을 사용하여 이를 보완할 수 있다. 엘보우 방법(Elbow Method)은 군집 수(k)의 변화에 따른 왜곡의 변화를 그래프로 표현하면 그래프가 팔꿈치(Elbow) 모양으로 꺾이는 지점이 나타나는데, 그 지점의 k 를 최적의 k 로 선택하는 방법이다[10]. 그래프를 통해 꺾이는 지점을 시각적으로 쉽게 확인할 수 있어 직관적인 이해가 가능하다는 이점이 있다. 실루엣 분석(Silhouette analysis)은 군집화 품질을 평가하는 도구이다. 좋은 군집화란 서로 다른 군집들과는 거리가 멀고, 동일한 군집 내 데이터들은 가깝게 모여 있는 것을 말한다. 이는 실루엣 계수(Silhouette Coefficient)를 계산하여 나타낼 수 있는데, 아래 식 (2)으로 정의된다.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2)$$

군집 내 데이터들의 응집도(Cluster Coherence)인 $a(i)$ 와 군집 간 분리도(Cluster Separation)인 $b(i)$ 를 이용하여 실루엣 계수 $s(i)$ 을 계산한다. 즉, 데이터 i 가 동일한 군집 내의 데이터와 얼마나 가깝게 군집화되어 있고, 다른 군집의 데이터와는 얼마나 멀리 분리되어 있는지를 말한다. 실루엣 계수는 -1에서 1 사이의 값을 가지며, 1에 가까울수록 좋은 군집화를 의미한다[9]. 이를 통해 군집화 품질을 정량적으로 평가할 수 있으며, 군집 개수를 다르게 하여 각각의 실루엣 값을 비교함으로써 최적의 군집 수(k)를 결정할 수 있다는 장점을 지닌다[10].

2.3 선행연구 사례

최근 다양한 분야에서 여러 빅데이터 분석 기법을 활용하여 마케팅 전략을 수립하는 연구가 활발히 이루어지고 있다.

Bae[1]는 의사결정나무(Decision Tree) 분석을 이용하여 로컬푸드 직매장 시장을 세분화하고, 운영 전략 수립을 위한 시사점을 도출하였다. 예를 들어, 로컬푸드 직매장을 가장 많이 이용하는 기혼자이면서 회사원, 전문직, 전업주부를 대상으로 이들의 주중 및 주말 쇼핑 빈도를 고려한 할인 행사를 기획할 수 있다.

Bae[2]는 대형할인매장의 매출거래 데이터에서 연관규칙(Association Rules) 분석을 이용하여 연관규칙을 찾아내고, 이 규칙에 나타난 특성을 파악하여 그에 적절한 촉진 전략 방향성을 제시하였다. 예를 들어, 면류와 과자류를 사는 소비자가 우유를 함께 구매하는 경우가 크므로 면류와 과자류의 적극적인 촉진을 실시하고, 이들 상품과 우유의 매대 동선을 짧게 진열하여 우유의 구매를 촉진시킬 수 있다.

Chun, Lee, and Park[4]은 소비자의 항공사 선호도에 따른 마케팅 전략을 수립하기 위해 뉴스 데이터를 크롤링하고, 이를 텍스트 마이닝(Text Mining) 기법을 통해 분석하였다. 예를 들어, 최저가 마케팅 전략을 활용하여 정상가보다 크게 저렴한 특정 좌석을 제공하여 소비자들의 클릭 수를 높이고, 이를 통해 광고효과를 극대화할 수 있다.

이와 같이 다양한 빅데이터 분석 기법들이 마케팅 전략 수립에 활용되고 있으며, 본 연구에서는 K-평균 군집분석을 선택하였다. K-평균 군집분석은 분류기준이 알려져 있지 않을 때 다변량 자료의 분류에 주로 이용되는 비계층적(Non-hierarchical) 방법이다. 연구자가 군집의 수를 지정하여 분석할 수 있어 다양한 군집을 선택할 수 있으며, 각 군집의 결과를 비교하여 군집의 특성이 가장 잘 파악되는 최적의 k 개의 군집을 선택할 수 있는 장점이 있기 때문에

다양한 분야에서 폭넓게 사용되고 있다[5].

3. 데이터 분석방법

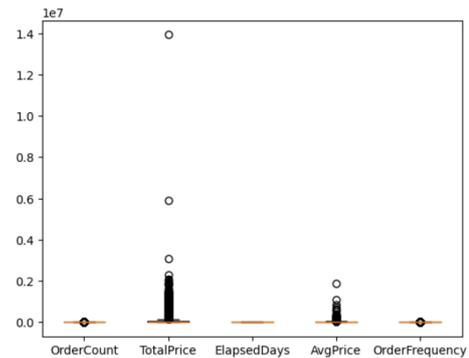
3.1 데이터 수집 및 전처리

본 연구에서 자료는 ‘전북생생장터’에서 2023년 1월 1일부터 2023년 12월 31일까지 1년간 발생한 주문 데이터를 수집하였다. 해당 데이터는 전북생생장터 주문 데이터 베이스에서 분석에 필요한 주문일시, 상품코드, (고유) 주문번호, 상품명, 주문자 주소(행정구역 단위), 주문자 ID, 주문수량, 상품단가를 추출하여 활용하였다. 주문자 ID의 경우 일부 소비자가 일반 회원 로그인, 카카오 계정 연동 간편 로그인 등 다양한 로그인 방식을 사용하여 여러 개의 ID로 분류되는 경우가 있다. 이를 해결하기 위해 동일 소비자의 여러 주문자 ID를 하나로 통합하였다.

데이터 전처리와 분석에는 Python 프로그래밍 언어와 Jupyter Notebook 환경을 사용하였으며, Pandas, NumPy, Matplotlib, Scikit-learn 등의 라이브러리를 활용하였다. 데이터 전처리는 데이터 정제, 분석용 데이터 생성, 데이터 분포 조정 단계로 수행하였다. 데이터 정제를 위해 결측치(Null) 값은 데이터 분석에 사용할 수 없으므로 제거하고, ‘주문수량’과 ‘상품단가’가 음수인 것도 제거하였다. 추가로 중복된 행 데이터를 제거하여 총 42,548개의 주문 데이터를 확보하였다. 다음으로 분석용 데이터를 생성하였다. ‘주문수량’과 ‘상품단가’를 곱해 ‘주문금액’을 계산하여 새로운 열(column)을 추가한 후 ‘주문자 ID’를 기준으로 그룹을 만들고, <Table 1>과 같이 분석에 필요한 변수들을 추가하여 새로운 소비자 데이터를 생성하였다. 각 변수는 소비자별 (‘고유’ 주문번호)의 개수를 세어 ‘총 주문횟수’를, 소비자별 ‘주문금액’의 총액을 계산하여 ‘총 주문금액’을, 소비자별 ‘주문일시’ 중에서 가장 마지막 주문 일자를 찾은 후 기준 날짜(2023.12.31.)에서 마지막 주문 일자를

빼 ‘마지막 주문일로부터의 경과일’을, 소비자별 ‘총 주문금액’을 ‘총 주문횟수’로 나누어 ‘1회당 평균 주문금액’을, 소비자별 ‘총 주문횟수’를 12로 나눠 ‘월평균 주문횟수’를 구하였다. 추가로 소비자별 ‘주문자 주소(행정구역 단위) 열(column)을 추가하여 소비자 군집분석에 사용할 총 14,019개의 소비자 데이터를 구성하였다.

마지막으로는 데이터 분포 조정을 수행하였다. 이는 분석 결과의 정확성과 신뢰성을 높이기 위해 데이터가 한쪽으로 치우치지 않았는지 확인하고, 데이터의 분포를 고르게 조정하는 과정이다. 먼저 박스플롯(boxplot) 그래프를 활용하면 <Figure 2A>와 같은 데이터 분포를 확인할 수 있다. 검은색 점으로 표시된 이상치(Outlier)가 많은 것으로 보아, 데이터가 고르게 분포하지 않고 치우쳐 있다는 것을 알 수 있다. 특히 ‘총 주문금액(TotalPrice)’과 ‘1회당 평균 주문금액(AvgPrice)’ 변수에서 눈에 띄게 큰 값의 이상치가 두드러진다.



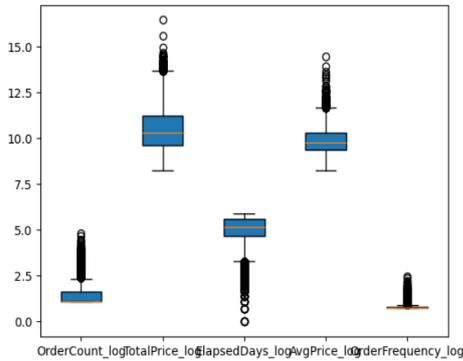
<Figure 2A> Boxplot of Customer Data

따라서 데이터 분포를 고르게 조정하기 위해 로그 함수를 적용하였으며, 이때 로그 변환 시 1을 더하여 0 또는 음수 값으로 인한 오류를 방지하였다. 로그 변환된 데이터의 박스플롯은 <Figure 2B>와 같다. 여전히 일부 이상치(Outlier)가 보이지만 전체적으로 중앙값 근처에 데이터 분

<Table 1> Customer Data Variables

Variable Name	Description
CustomerID	Customer ID
OrderCount	Total sum of number of orders which placed by each customer
TotalPrice	Total sum of orders which spent by each customer
ElapsedDays	The number of days since the last order, calculated by subtracting the last order date from the base date (2023.12.31)
AvgPrice	Average amount spent per order, calculated by dividing the total amount by the number of orders
OrderFrequency	Monthly average order count, calculated by dividing the ‘OrderCount’ by 12
City	Address(administrative district unit) for each customer

포가 밀집된 것으로 보아, 어느 정도 정규화된 분포를 가지게 되었음을 알 수 있다.

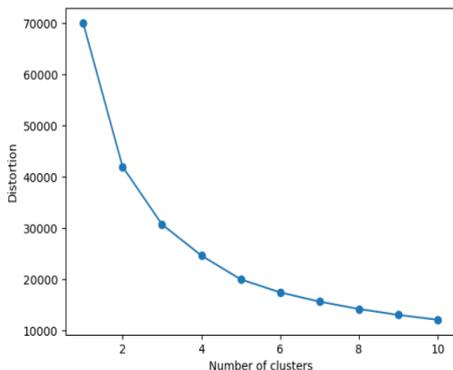


<Figure 2B> Boxplot of Log-Transformed Customer Data

추가적인 스케일링(Scaling)을 통해 데이터를 더욱 정규 분포에 가깝게 만드는 작업을 수행하였다. 스케일링은 데이터를 평균이 0이고 표준편차가 1이 되도록 변환하여, 각 변수의 스케일(scale)이 다른 경우 이를 동일한 스케일로 맞추는 데이터 표준화 과정이다. 이는 데이터가 다양한 스케일을 가지는 경우 K-평균 군집분석 수행 시 특정 변수가 과도한 영향을 미쳐 오류를 발생시키는 것을 방지하고, 데이터 분석 결과의 신뢰성을 높이기 위함이다. 이러한 데이터 전처리 과정들을 통해 소비자 데이터는 분석에 적합한 형태로 변환되었다.

3.2 소비자 군집분석

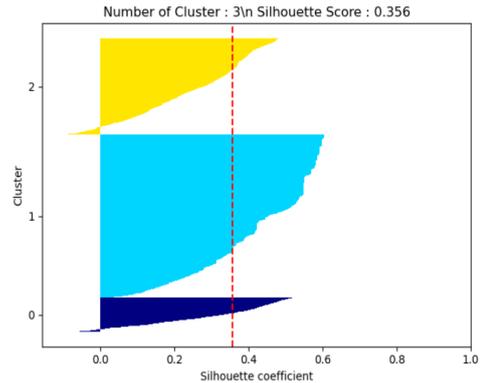
소비자 군집분석을 위해 K-평균 군집화 모델인 K-means를 사용하였다. 먼저 최적의 군집 개수(k)를 선택하기 위해 엘보우 방법을 적용하였다. k 를 1부터 10까지 변화시키면서 왜곡(Distortion) 값의 변화를 그래프로 나타내어 팔꿈치 모양으로 꺾이는 엘보우(Elbow) 지점을 도출



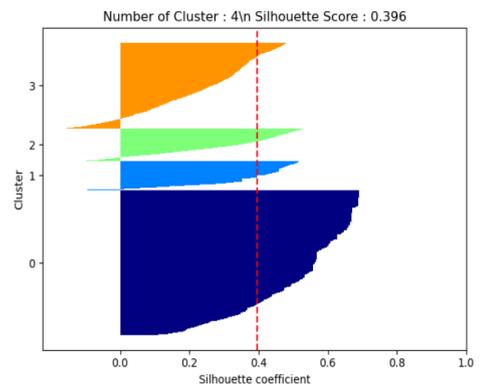
<Figure 3> Elbow Method for Optimal k

했다. 그래프는 <Figure 3>과 같다.

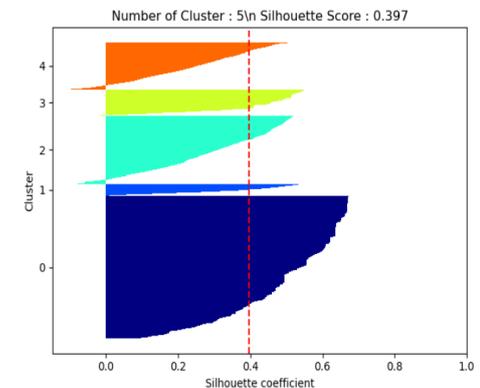
위 그래프를 보면 엘보우라고 할 수 있는 지점, 즉 k 가 3 또는 4, 5가 될 수 있음을 보여준다. 엘보우 방법만으로 최적의 군집 개수(k)를 확정하기에는 한계가 있어 실루엣 분석을 추가로 수행하였다. silhouetteViz 함수를 사용하여 실루엣 계수와 각 군집의 비중을 가로 바 차트로 시각화한 결과는 <Figure 4A>, <Figure 4B>, <Figure 4C>와 같다.



<Figure 4A> Silhouette Analysis for 3 Clusters

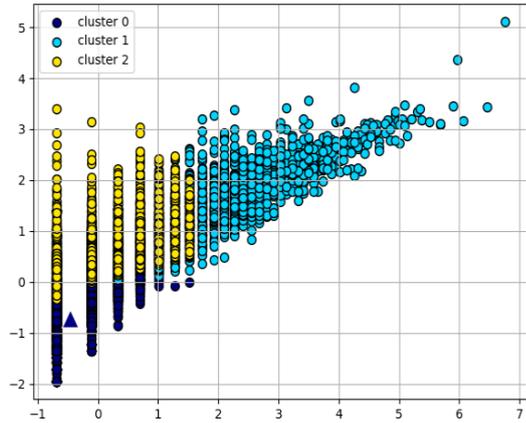


<Figure 4B> Silhouette Analysis for 4 Clusters

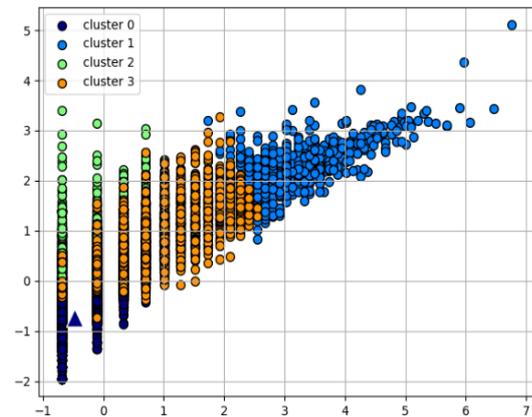


<Figure 4C> Silhouette Analysis for 5 Clusters

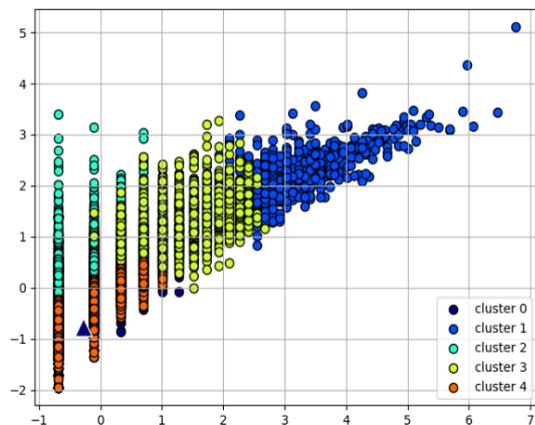
실루엣 계수는 -1과 1 사이의 값을 가지며, 값이 클수록 군집화가 잘 되었다는 것을 의미한다. 각 군집 비중이 한쪽으로 치우치지 않고 비슷하면서 실루엣 계수가 큰 것은 k 가 3, 4개인 경우로, 실루엣 계수는 각각 0.396, 0.397



<Figure 5A> Scatter Plot for 3 Clusters



<Figure 5B> Scatter Plot for 4 Clusters



<Figure 5C> Scatter Plot for 5 Clusters

이다. 또한, clusterScatter 함수를 사용하여 각 군집의 데이터 분포를 스캐터 차트로 시각화한 결과는 <Figure 5A>, <Figure 5B>, <Figure 5C>와 같다.

데이터가 섞이지 않고 같은 군집끼리 모여 있는 형태가 되면 군집화가 잘 된 것을 의미한다. k 가 4개일 때 데이터 분포가 시각적으로 더 명확하게 구분되는 것으로 나타났으며, 최종적으로 k 값은 4로 결정하였다.

3.3 소비자 군집에 따른 지역 분석

앞서 도출된 소비자 군집 데이터를 활용하여 전체 소비자의 지역별 분포 비율을 파악하고, 각 군집별로 지역별 분포를 비교 분석하였다. 이 분석은 피벗 테이블을 활용하여 수행하였다. 피벗 테이블은 대량의 데이터를 빠르게 요약하는 데 사용하는 도구로, 다양한 기준에 따라 데이터를 집계하여 분석하기에 용이하다는 장점이 있다.

4. 연구결과

4.1 소비자 군집분석

K-평균 군집분석(K-means clustering)을 통해 소비자 군집분석을 하여 총 4개의 군집을 도출하였다. 군집 번호(0, 1, 2, 3)는 임의로 할당되었으며, 각 군집별 주요 지표는 <Table 2>와 같다. 지표 중 ‘평균 총 주문횟수’와 ‘1회당 평균 주문금액’을 기준으로 <Table 3>과 같이 각 군집별 특성을 도출하였다. 군집 0에 속한 소비자는 7,626명으로, 가장 많은 소비자가 포함되어 있다. 이들의 평균 총 주문횟수는 1.4회, 1회당 평균 주문금액은 12,899원으로 단발성 소액 구매자들이다. 즉, 군집 0은 저렴한 상품을 일회성 구매한 소비자들로 이루어져 있다.

군집 1에 속한 소비자는 632명으로, 가장 적은 수의 소비자가 포함되어 있다. 이들의 평균 총 주문횟수는 19.6회, 1회당 평균 주문금액은 30,145원으로 고빈도 중액 구매자들이다. 이 군집의 소비자들은 중간 가격대의 상품을 자주 구매하여 평균 총 주문금액이 가장 높은, 즉 충성도가 높은 VIP 소비자들로 이루어져 있다. 군집 2에 속한 소비자는 3,324명이다. 이들의 평균 총 주문횟수는 1.7회, 1회당 평균 주문금액은 51,033원으로 저빈도 고액 구매자들이다. 이 군집의 소비자들은 연 1~2회 정도 여러 상품을 한 번에 구매하거나 한 상품을 한 번에 여러 개 구매하여 높은 금액을 소비하며, 재구매율은 매우 낮으나 1회당 평균 주문금액이 가장 크다는 특징이 있다. 군집 3에 속한 소비자는 2,437명이다. 이들의 평균 총 주문횟수는 5.7회, 1회당 평균 주문금액은 26,440원으로 중

<Table 2> Key Indicators by Cluster

Cluster Number	Indicator Name					
	Number of Data	Average Total Order Count	Average Total Order Amount	Monthly Average Order Count	Average Order Amount per Order	Average Days Since Last Order
0	7,626	1.4	17,726	0.12	12,899	198
1	632	19.6	585,057	1.64	30,145	50
2	3,324	1.7	79,789	0.14	51,033	194
3	2,437	5.7	149,938	0.47	26,440	96

<Table 3> Characteristics of Customer Clusters

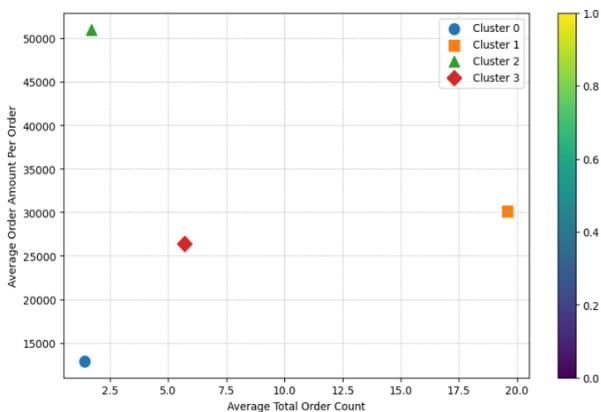
Cluster Number	Average Total Order Count	Average Order Amount per Order	Characteristic
0	Low Frequency (Single Order)	Small Amount	Single Small Purchase Customers
1	High Frequency	Medium Amount	High Frequency Medium Purchase Customers
2	Low Frequency (1~2 Orders)	Large Amount	Low Frequency High Purchase Customers
3	Medium Frequency	Medium Amount	Medium Frequency Medium Purchase Customers

빈도 중액 구매자들이다. 군집 3은 주기적으로 중간 가격대의 상품을 구매하는 일반적인 소비자로 이루어져 있으며, 향후 잠재적 VIP 소비자로 볼 수 있다. 한편, 소비자 군집을 시각화 결과는 <Figure 6>와 같다. 각 점의 색깔과 모양은 군집 번호를 나타내며, 좌측 상단의 초록색 삼각형은 군집 2를, 좌측 하단의 파란색 원형은 군집 0을, 중앙에서 좌측 하단의 빨간색 마름모는 군집 3을, 우측 중앙의 주황색 사각형은 군집 1을 나타낸다. 이를 통해 각 군집의 특성을 한눈에 파악할 수 있다. 이와 같이

도출된 군집의 특성을 바탕으로 각 군집별 차별화된 마케팅 전략 수립이 가능해졌다.

4.2 소비자 군집에 따른 지역 분석

군집별 지역(행정구역 단위별, 이하 도시별) 분포와 군집별 지역(권역별) 분포는 각각 <Table 4A>, <Table 4B>와 같다. <Table 4A>에 따르면 전북(4,515명), 경기(3,054명), 서울(2,433명) 순으로 많으며, 세 지역의 합은 전체(14,019명)의 약 71.3%(10,002명)이다. 이는 전북과 수도권 지역에 집중된 마케팅 활동이 필요함을 시사한다. <Table 4B>에 따르면 수도권(서울, 경기, 인천) 소비자가 전체의 43%(5,968명)를 차지하고 있으며, 이는 수도권 지역에서 ‘전북생생장터’의 인지도가 비교적 높음을 시사한다. 또한, 각 군집별로 수도권과 전북 지역 분포를 보면, 군집 2의 경우 수도권은 전체의 45%, 전북 지역은 전체의 28%의 소비자가 포함되어 있다. 다른 군집의 경우 평균적으로 수도권에 약 42%, 전북 지역에 약 32%의 소비자가 차지하고 있어, 군집 2가 다른 군집에 비해 수도권 소비자 비율이 더 높다. 앞선 소비자 군집분석 결과 군집 2에 속한 소비자들 고액 구매자들인 점을 고려할 때, 고액 구매자들이 상대적으로 수도권에 더 많이 분포해 있음을 알 수 있다. 이는 수도권 지역의 높은 구매력이 반영된 결과로 볼 수 있다.



<Figure 6> Customer Segmentation Based on 'Average Total Order Count' and 'Average Order Amount Per Order'

<Table 4A> Distribution of Cities by Cluster

City	Cluster Number				Total
	0	1	2	3	
Gangwon	91 (1%)	10 (2%)	40 (1%)	24 (1%)	165 (1%)
Gyeonggi	1,605 (21%)	131 (21%)	776 (23%)	542 (22%)	3,054 (22%)
Gyeongnam	215 (3%)	27 (4%)	147 (4%)	93 (4%)	482 (3%)
Gyeongbuk	135 (2%)	8 (1%)	59 (2%)	33 (1%)	235 (2%)
Gwangju	146 (2%)	12 (2%)	57 (2%)	64 (3%)	279 (2%)
Daegu	148 (2%)	9 (1%)	68 (2%)	63 (3%)	288 (2%)
Daejeon	207 (3%)	12 (2%)	73 (2%)	42 (2%)	334 (2%)
Busan	284 (4%)	29 (5%)	147 (4%)	115 (5%)	575 (4%)
Seoul	1,311 (17%)	109 (17%)	593 (18%)	420 (17%)	2,433 (17%)
Sejong	72 (1%)	6 (1%)	23 (1%)	21 (1%)	122 (1%)
Ulsan	90 (1%)	9 (1%)	38 (1%)	27 (1%)	164 (1%)
Incheon	253 (3%)	26 (4%)	119 (4%)	83 (3%)	481 (3%)
Jeonnam	114 (1%)	9 (1%)	81 (2%)	48 (2%)	252 (2%)
Jeonbuk	2,629 (34%)	195 (31%)	942 (28%)	749 (31%)	4,515 (32%)
Jeju	22 (0%)	1 (0%)	13 (0%)	2 (0%)	38 (0%)
Chungnam	191 (3%)	17 (3%)	87 (3%)	67 (3%)	362 (3%)
Chungbuk	113 (1%)	22 (3%)	61 (2%)	44 (2%)	240 (2%)

<Table 4B> Distribution of Areas by Cluster

Cluster Number	Metropolitan area	Gangwon Area	Chungcheong Area	Honam Area	Yeongnam Area	Jeju	Total
0	3,169 (42%)	91 (1%)	583 (8%)	2,889 (38%)	872 (11%)	22 (0%)	7,626
1	256 (42%)	10 (2%)	57 (9%)	216 (34%)	82 (13%)	1 (0%)	632
2	1,488 (45%)	40 (1%)	244 (7%)	1,080 (32%)	459 (14%)	13 (0%)	3,324
3	1,045 (43%)	24 (1%)	174 (7%)	861 (35%)	331 (14%)	2 (0%)	2,437
Total	5,968 (43%)	165 (1%)	1,058 (8%)	4,046 (35%)	1,744 (13%)	38 (0%)	14,019

5. 마케팅 전략 수립

5.1 군집별 및 지역별 마케팅 전략

본 연구에서는 소비자 군집분석을 통해 도출된 각 군집별 특성과 지역 분석 결과를 바탕으로 맞춤형 마케팅 전략을 제시하여, 효율적인 쇼핑몰 운영 전략을 수립하고자 한다.

각 군집별 특성에 맞는 마케팅 전략은 아래와 같다.

1) 군집 0에 속한 소비자들은 첫 구매 후 재방문하지 않는 경향이 있는 단발성 구매자로, 가격에 민감한 알뜰 소비자이다. 카페 및 커뮤니티를 통한 바이럴

(Viral) 마케팅과 포인트 적립 제도 운영으로 이들의 재방문과 재구매를 유도하는 것이 필요하다. 구체적으로 포인트 적립 제도는 구매 금액의 일정 비율을 포인트로 적립해 향후 사용할 수 있도록 하여 재구매를 유도할 수 있으며, 리뷰 작성 시 포인트를 지급하여 소비자 커뮤니티를 활성화할 수 있다.

2) 군집 1에 속한 소비자들은 충성도가 높은 소비자들로, VIP 프로그램을 통해 충성도를 유지하는 것이 중요하다. 본 논문에서는 VIP 프로그램의 일환으로, 앞서 제시한 포인트 적립 제도를 멤버십 프로그램과 연계하여 운영할 것을 제안한다. 이는 소비자들의 구매

실적에 따라 등급을 나누고, 그 등급에 따라 포인트 적립과 같은 혜택에 차등을 주는 것이다. 예를 들어 연간 50만 원 이상 주문한 회원들을 대상으로 VIP 등급을 부여하고, 모든 등급 중 가장 높은 포인트 적립률과 VIP 전용 5% 할인쿠폰 등의 특별 혜택을 제공할 수 있다.

- 3) 군집 2에 속한 소비자들은 가끔(연 1~2회 정도) 여러 상품을 한 번에 구매하여 높은 금액을 소비하거나, 한 상품을 한 번에 여러 개 구매하여 높은 금액을 소비하는 소비자이다. 선물하기 서비스를 도입하여 이들의 추가 구매를 유도하는 것을 제안한다. 이는 기존 소비자가 다른 사람에게 선물 및 메시지를 보낼 수 있는 기능으로, 카카오커머스의 ‘카카오톡 선물하기’ 서비스가 대표적인 예이다. 이를 통해 감사선물, 명절선물 등 여러 용도로 마음을 전할 수 있도록 하여 추가 구매를 이끌 수 있으며, 소비자 편의성을 증대시켜 만족스러운 구매 경험을 제공할 수 있다.
- 4) 마지막으로 군집 3에 속한 소비자들은 주기적으로 상품을 구매하는 일반적인 소비자이자, 잠재적 VIP다. 따라서 주기적인 프로모션 운영과 정기구독 서비스를 통해 구매 빈도를 높이는 전략을 제시한다. 정기구독 서비스는 소비자가 필요한 상품을 매주 또는 매월 원하는 주기에 맞춰 자동으로 배송해주는 서비스이다. 이를 통해 소비자의 편리성과 만족도가 높아지며, 지속적인 소비자 관계를 구축하여 안정적인 매출로 이어질 수 있다.

다음으로 소비자 군집에 따른 지역 분석 결과를 바탕으로 전북 지역과 서울 및 경기 지역에 집중한 마케팅 활동 전략을 아래와 같이 제시한다. 먼저 전북 지역의 경우 도내에서 생산되는 신선하고 품질 좋은 로컬 제품임을 강조하는 마케팅이 필요하다. 또한, 지역 내 브랜드 인지도를 높이기 위해 전북 지역에서 열리는 축제 및 행사에 참여하여 브랜드를 홍보하거나 지역 매체(전북 신문, 라디오, JTV 등)를 통한 광고 진행을 추천한다. 서울 및 경기 지역의 경우 다양한 소비자층을 대상으로 한 디지털 광고나 O2O 마케팅이 필요하다. 지하철 전광판이나 소셜 미디어를 활용한 디지털 광고를 통해 광범위한 도달 범위를 확보하여 브랜드 인지도를 높일 수 있다. O2O(Online to Offline)는 마케팅 분야에서 온라인과 오프라인의 연결되는 현상을 의미하며, 예를 들어 스타벅스의 ‘사이렌 오더’, 무신사의 ‘무신사 스탠다드 매장’, 올리브영의 ‘오늘드림’ 등이 있다. 본 논문에서는 O2O 마케팅으로 유동인구가 많은 주요 상권에서 한시적으로

팝업 스토어나 직거래 장터 운영하는 것을 제안한다. 이를 통해 소비자들은 제품을 직접 확인하고 구매하여 신뢰성을 높이고, 브랜드 인지도 제고와 매출 증대를 도모할 수 있다.

또한, 추가로 온라인 쇼핑물 전반에 걸친 큐레이션 서비스(Curation Service) 도입을 제안한다. 오늘날 정보 과잉 시대에 소비자들은 모든 정보를 소비할 수 없으므로 원하는 정보만을 받아서 소비하고 싶어 한다. 큐레이션(Curation)은 사용자에게 의미 있고 가치 있는 정보를 선별하여 제공해 주는 것을 말하며, 이 용어는 미술관이나 박물관에서 전시품을 선택하고 배치하는 큐레이터(Curator)의 역할에서 유래되었다. 마케팅 분야에서 큐레이션 서비스는 개인의 취향을 분석하여 적절한 콘텐츠를 추천해주는 것을 의미한다[7, 12]. 대표적인 예로 소비자의 과거 구매 이력, 검색 기록, 장바구니 항목, 관심 상품 등을 분석하여 소비자가 관심을 가질 만한 상품이나 홍보 정보를 제공하는 서비스가 있다. 이러한 서비스를 통해 소비자에게 보다 개인화된 맞춤형 정보를 적시 제공함으로써 소비자 만족도를 높일 수 있으며, 이는 궁극적으로 브랜드 신뢰도와 매출에 긍정적인 영향을 미칠 것으로 기대된다[7].

5.2 소비자 설문조사

본 논문에서는 앞서 제시한 전략에 대한 타당성을 검증하고자 소비자 설문조사를 실시하였다. 설문조사는 2024년 7월 28일부터 29일까지 이틀간 진행되었으며, 총 46명의 응답을 받았다. 소비자별 설문 응답 결과는 <Table 5>와 같다. 설문자의 구매 패턴(주문횟수 및 금액)에 대한 응답 정보를 바탕으로 설문자를 각 군집으로 분류하였고, 그 결과 <Table 2>의 소비자 군집분석 결과와 약 91% 일치하였다. 이는 소비자 군집분석 결과의 정확성이 높음을 시사한다. 또한, 제시한 마케팅 전략에 대한 선호도 설문 응답 결과는 <Table 6>과 같다. 군집 0의 설문자는 총 21명(46%)이며, 각각 포인트 적립(Points) 18명, 선물하기(Gifting) 2명, 정기구독(Subscription) 1명이었다. 군집 1의 설문자는 총 5명(11%)이며, 각각 멤버십(Membership) 4명, 선물하기 1명이었다. 군집 2의 설문자는 총 6명(13%)이며, 각각 선물하기 4명, 포인트 적립 2명이었다. 군집 3의 설문자는 총 13명(28%)이며, 각각 정기구독 7명, 포인트 적립 4명, 멤버십 1명, 선물하기 1명이었다. 그리고 나머지 설문자 1명은 군집 2 또는 3으로 분류하였으며, 포인트 적립에 응답하였다. 이러한 선호도 응답 결과는 앞서 각 군집별로 제시한 마케팅 전략과 높은 일치도를 보였다.

<Table 5> Customer Survey Response Results

No.	Annual order frequency	Average order amount per order (unit: KRW)	Cluster	Most preferred service
1	10+times	50,000+	1	Gifting
2	1-2 times	≤15,000	0	Points
3	1-2 times	50,000+	2	Points
4	3-9 times	15,000-30,000	3	Points
5	1-2 times	15,000-30,000	0	Gifting
6	1-2 times	15,000-30,000	0	Points
7	3-9 times	15,000-30,000	3	Points
8	3-9 times	30,000-40,000	3	Points
9	3-9 times	15,000-30,000	3	Subscription
10	1-2 times	≤15,000	0	Points
11	3-9 times	15,000-30,000	3	Subscription
12	1-2 times	≤15,000	0	Points
13	3-9 times	30,000-40,000	3	Subscription
14	10+times	30,000-40,000	1	Membership
15	1-2 times	≤15,000	0	Subscription
16	1-2 times	15,000-30,000	0	Points
17	1-2 times	≤15,000	0	Points
18	1-2 times	≤15,000	0	Points
19	1-2 times	50,000+	2	Points
20	1-2 times	≤15,000	0	Points
21	3-9 times	15,000-30,000	3	Points
22	10+times	30,000-40,000	1	Membership
23	1-2 times	≤15,000	0	Points
24	1-2 times	≤15,000	0	Points
25	3-9 times	50,000+	2 or 3	Points
26	1-2 times	≤15,000	0	Points
27	1-2 times	≤15,000	0	Points
28	1-2 times	≤15,000	0	Points
29	10+times	30,000-40,000	1	Membership
30	10+times	30,000-40,000	1	Membership
31	3-9 times	15,000-30,000	3	Membership
32	1-2 times	≤15,000	0	Points
33	1-2 times	50,000+	2	Gifting
34	3-9 times	15,000-30,000	3	Subscription
35	1-2 times	50,000+	2	Gifting
36	3-9 times	30,000-40,000	3	Subscription
37	1-2 times	15,000-30,000	0	Points
38	1-2 times	≤15,000	0	Gifting
39	3-9 times	30,000-40,000	3	Gifting
40	1-2 times	15,000-30,000	0	Points
41	1-2 times	15,000-30,000	0	Points
42	1-2 times	50,000+	2	Gifting
43	3-9 times	30,000-40,000	3	Subscription
44	3-9 times	30,000-40,000	3	Subscription
45	1-2 times	50,000+	2	Gifting
46	1-2 times	≤15,000	0	Points

<Table 6> Preference Results for Marketing Strategies

Cluster	Preferred Marketing Strategy	Number of Respondents	Rate
0	Gifting	2	10%
	Subscription	1	5%
	Points	18	86%
	Total	21	46%
1	Membership	4	80%
	Gifting	1	20%
	Total	5	11%
2	Gifting	4	67%
	Points	2	33%
	Total	6	13%
3	Membership	1	8%
	Gifting	1	8%
	Subscription	7	54%
	Points	4	31%
	Total	13	28%
2 or 3	Points	1	100%
	Total	1	2%
Grand Total		46	100%

6. 결 론

본 연구에서는 소비자 군집분석을 위해 ‘전북생생장터’에서 2023년 1년간 발생한 주문 데이터를 수집하였다. 이후 데이터 전처리와 K-평균 군집분석(K-means clustering)을 수행하였다. K-means 군집화 모델의 신뢰성을 높이기 위해 최적의 군집 수(k)를 결정하는 단계에서 엘보우 방법과 실루엣 분석을 적용하였고, 그 결과 k 는 4로 결정되었다. 이를 통해 4개의 소비자 군집이 도출되었다. 각 군집을 설명할 수 있는 여러 지표 중 ‘평균 총 주문횟수’와 ‘평균 1회당 주문금액’을 기준으로 각 군집별 주요 특성을 도출하여 맞춤형 마케팅 전략을 제시하였다. 군집 0에 속한 소비자는 일회성 소액 구매자이며, 바이럴 마케팅과 포인트 적립 제도 운영으로 이들의 재방문과 재구매를 유도할 필요가 있다. 군집 1에 속한 소비자는 VIP 소비자이며, 멤버십 프로그램 운영을 통해 VIP 등급을 부여하여 특별 혜택을 제공하는 것이 필요하다. 군집 2에 속한 소비자는 저빈도 고액 구매자로 가끔 여러 상품을 한 번에 구매하여 높은 금액을 소비하거나, 한 상품을 한 번에 여러 개 구매하여 높은 금액을 소비하는 소비자이다. 따라서 선물하기 서비스 기능을 도입하여 이들의 추가 구매를 유도하는 것을 제안한다. 군집 3에 속한 소비자는 주기적인 구매 패턴을

보이는 일반 소비자이자 잠재적 VIP이다. 이 군집 대상으로는 주기적인 프로모션 운영과 정기구독 서비스를 도입하여 지속적인 소비자 관계를 구축하고 충성도를 높이는 전략을 추천한다. 또한, 소비자 군집에 따른 지역 분석을 추가로 수행하였으며, 그 결과 전북과 서울 및 경기 지역에 집중된 마케팅 활동이 필요함을 알 수 있었다. 이러한 결과를 바탕으로 전북 지역에 국한된 광고 노출이 아닌 수도권까지 권역을 넓혀 홍보 전략을 수립하는 것이 필요하다. 본 연구는 담당자의 직감에 의존하여 운영해오던 기존 관행과는 달리 빅데이터 분석방법을 적용함으로써 보다 객관적으로 마케팅 전략을 제시한 데에 의의가 있다. 또한, 본 연구의 타당성 검증을 위한 소비자 설문조사에서는 연구의 한계로 볼 수 있는 비교적 작은 표본이지만 이 제한된 표본을 통해서도 군집별 소비자 특성과 앞서 제시한 마케팅 전략 간의 명확한 연관성을 확인할 수 있다는 점에서 연구 결과의 신뢰성을 충분히 확보할 수 있다고 판단된다. 그러나 이러한 한계를 극복하기 위해서는 더 큰 표본 크기를 바탕으로 추가적인 연구가 필요하다고 판단된다. 따라서 향후 다양한 소비자층을 대상으로 설문조사를 확대함으로써 더 많은 표본을 확보하고, 연구 결과의 일반화 가능성을 높일 계획이다. 또한 이러한 전략들을 실제 실행하지 못한 한계가 있어 향후 실제 사례에서 본 연구에서 제안된 마케팅 전략을 적용하고, 그 효과를 실증적으로 평가하는 연구를 진행하여 제시된 전략의 타당성을 더욱 강력하게 입증하고자 한다.

Acknowledgment

This study has been supported by MOTIE funding program “Advanced Graduate Education for Management of Convergence Technology.”

References

- [1] Bae, H.W., Analysis of Local Food Direct Store Market Segmentation Using Decision Tree Model [master's thesis], [Seoul, Korea]: Chung-Ang University, 2021.
- [2] Bae, S.G., An Exploratory study ; Developing A Campaign Management by using Association Rule - based on C Hyper Market [master's thesis], [Seoul, Korea]: Hanyang University, 2003.
- [3] Cho, H.J., A Comparison of Datamining Techniques for Customer Segmentation [master's thesis], [Busan, Korea]: Dong-A University, 2001.
- [4] Chun, Y.H., Lee, S.J., and Park, S.H., For airline preferences of consumers Big Data Convergence Based Marketing Strategy, *Journal of Industrial Convergence*, 2019, Vol. 17, No.3, pp. 17-22.
- [5] Chung, M.R. and Na, C.K., A Study on Environmental Worldview of Kindergarten Teachers: Classification by K-Mean Cluster Analysis and Each Cluster's Tendency, *Journal of Korean Society for Environmental Education*, 2019, Vol. 32, No.2, pp. 111-126.
- [6] HeyTech, <https://heytech.tistory.com/357>
- [7] Jang, H.S., A Study on the analysis of customer behavior toward brand attitude and purchase intention through personalized curation service in mobile app [master's thesis], [Seoul, Korea]: Hongik University, 2018.
- [8] Kim, S.H., Customer segmentation for the internet shopping mall using data mining technique [master's thesis], [Seoul, Korea]: Dongguk University, 1999.
- [9] Lee, D.H., Detection of inappropriate contents repeatedly posted on SNS using K-means clustering technique [master's thesis], [Seoul, Korea]: Korea University, 2022.
- [10] Lee, J.Y., *Data Science-based Python Big Data Analysis*, 1st ed., Seoul, Hanbit Academy, 2020, pp. 359-382.
- [11] Lee, S.H., A Comparative Analysis of Clustering Algorithms for Assessing College Students' Key Competencies [master's thesis], [Seoul, Korea]: Seoul Women's University, 2024.
- [12] Lee, S.Y., The Effect of Fashion Curation Service Quality Attributes on Satisfaction, Trust, and Purchase Intention [master's thesis], [Seoul, Korea]: Yonsei University, 2016.
- [13] Park, H.C. and Cho, K.H., K-means Clustering of Social Indicator Survey Data, *Journal of the Korean Data Analysis Society*, 2005, Vol. 7, No.2, pp. 465-476.
- [14] Song, J.W., The Effect of Nonresponse Rates on K-Means Cluster Analysis with Missing Data, *Journal of the Korean Data Analysis Society*, 2017, Vol. 19, No.3, pp. 1272-1282.
- [15] Wikipedia, <https://ko.wikipedia.org/wiki/O2O>.
- [16] Yang, D.G., Myung, J.S., Lee, S.H., and Song, J.W., Comparison of k-means clustering methods for data with missing values, *Journal of the Korean Data Analysis Society*, 2023, Vol. 25, No.6, pp. 2131-2142.

ORCID

Seong Hye Kim | <http://orcid.org/0009-0006-0761-7953>

Joon Soo Bae | <http://orcid.org/0000-0001-8872-5169>