

『KPI 도출 비즈니스 전략 아이디어 경진대회』 제안서

| ① 팀 명 | 거니의 아이들 | ② 제 목 | 고객 확보 어떻게 할 거니? |
|-------|---------|-------|-----------------|
|-------|---------|-------|-----------------|

분석 결과

1. 분석 배경

- McKinsey Quarterly에 따르면, 고객 경험을 개선하면 매출이 2~7% 증가하고 수익성이 1~2% 향상된다고 한다. 이는 고객 경험 개선이 기업 성과 향상과 직결됨을 시사한다.
- 고객 등급에 따라 차별화된 서비스와 혜택을 제공하는 것이 고객 충성도를 높일 수 있다고 한다. 즉, 고객 등급을 세분화하여 타겟마케팅을 실시함으로써 기업의 판매 실적을 향상시킬 수 있다고 한다.

2. 분석 목적

- 본 분석은 주요 성과 지표(KPI, Key Performance Indicator)를 도출하여 기업의 성과를 정량적으로 평가할 수 있는 기준을 마련함으로써, 비용 효율적인 관리 체계를 구축하고 기업의 지속 가능한 성장을 도모하는 것을 목표로 한다.
- 분석 배경의 내용을 고려해 기존 고객에 대한 만족도를 개선함으로써 기업의 이익을 확보하는 것이 목표이다. 따라서 고객 만족도를 기반으로 분석하여 적절한 솔루션을 마련하고자 한다.
- 이커머스 환경에서 고객의 구매 패턴을 깊이 이해하고, 이를 통해 고객 세분화를 실시하여 각 고객 그룹에 맞춤형 마케팅 전략을 제안하여 수익성을 확보하고자 한다.

3. 사용 데이터

- Brazilian E-Commerce company Dataset: 브라질 이커머스 기업의 데이터셋에 다음과 같은 정보가 포함되어 있다.
 - ✓ 고객 정보
 - ✓ 브라질 내 지역 및 위치 정보
 - ✓ 주문 품목 정보
 - ✓ 주문 및 배송 정보
 - ✓ 주문에 따른 지불 정보
 - ✓ 제품 정보
 - ✓ 고객 리뷰 정보
 - ✓ 판매자 정보

4. 데이터 탐색 및 분석 과정

1. 고객 만족도 개선을 위한 데이터 탐색 및 분석

1-1. 고객 만족도 점수에 대한 고객 분포 확인 (표 1-1, 그림 1-1 참조)

- 전체 고객 중 약 21% 정도가 3점 이하의 평점을 부여하는 것으로 확인된다.
- 이에 따라, 고객 만족도에 대한 전반적인 분석을 진행하였다.

1-2. 전처리 수행

❖ 고객 만족도에는 시간과 관련된 변수가 영향을 미칠 것으로 판단하여 해당 변수를 생성하였다.

- '리뷰 기간', '답변 기간', '제품 준비 기간', '실제 배송 기간', '예상 배송 기간', '실제 수령 총기간', '예상 수령 총기간', '수령 차이' 변수를 추가하였다.
- ○○ 기간 등의 변수들은 날짜 간의 차이를 계산하여 추가하였다.
- 배송 기간 변수는 물류 처리 시점으로부터 실제 수령까지의 기간이며, 수령 기간 변수는 구매 시점으로부터 실제 수령까지의 기간이다.
- '수령 차이' 변수는 '실제 수령 총기간'과 '예상 수령 총기간'의 차이로, 음수가 나오는 것은 예상보다 더 빠른 배송이 이루어졌음을 의미한다.
- 각 기간을 초 단위로 변환한 후 3600(초/시간)과 24(시간/일)을 곱한 값으로 나누어 일 단위로 변경하였다.
- '예상 배송 기간'이 음수인 경우는 물류 처리 날짜가 기대 배송 날짜보다 늦은 경우로, 이후 배송 지연율을 계산하기에 부적합하므로 제거한다.

1-3. 고객 만족도 상관분석

- 상관분석은 변수 간의 관계의 정도를 측정하는 통계적 기법으로, 고객 만족도와 큰 관련성이 있는 변수를 식별하여 고객 만족도를 개선하기 위한 발판을 마련하고자 한다.
 - [1-2]에서 추가된 변수를 포함하여 고객 만족도와 관계의 파악하였다. (표 1-2, 그림 1-2 참조)
- ▷ 고객 만족도와 실제 배송 기간, 실제 수령 총기간 등 기간과 관련된 변수들이 관계가 있음을 확인할 수 있다.

1-4. 고객 만족도에 대한 실제 배송 기간, 실제 수령 총 기간, 수령 차이 분포 확인 (표 1-3, 그림 1-3 참조)

- 배송이 일찍 이루어질수록 고객 만족도가 높아지는 경향이 보인다.
- 여기서 모든 점수에 대해 수령 차이가 음수인 이유는 배송 지연이 발생한 경우가 정상 배송된 경우보다 적기 때문이다.

1-5. 주별 고객 만족도, 배송 기간, 수령 기간 (그림 1-4, 5, 6 참조)

❖ [1-4]에서 주별 수령 차이에 대한 분포를 확인한 결과, 주마다 이상치가 있음을 확인할 수 있었다. 이는 데이터의 정확성을 저해하며, 추후 솔루션 도출 과정에 부정적인 영향을 미칠 수 있으므로 제거하고 진행하였다. (그림 1-7, 8 참조)

- <표 1-4>에서 확인할 수 있듯이, 고객 만족도가 낮은 주와 배송 기간(수령 기간)이 긴 주가 대부분 겹치는 것을 확인할 수 있다.
- 즉, 평균 점수와 배송 기간 (수령 기간)은 어느 정도 상관관계가 있다고 볼 수 있다.
- 몇몇 북부와 북동부 끝에 위치한 주들은 배송 기간이 길지만, 평균 이상의 만족도를 보인다.
- 이는 접근하기 어려운(배송 시간이 오래 걸리는) 지역이라고 반드시 만족도가 낮지 않음을 알 수 있다.
- ▷ 이를 기반으로, 고객 주에 대한 고객 만족도와 다른 변수들이 관계가 있을 것이라는 가설을 세우고 상관분석을 진행하였다.
- ▷ 27개의 주로 묶어서 분석했을 때, 해당 변수들이 [1-3]에서 보다 더 높은 상관관계를 나타내는 것을 확인하였다. (표 1-5 참조)

1-6. 배송 지연율 (표 1-6, 그림 1-9 참조)

- 주에 따라 배송 기간과 수령 기간이 다를 것이라는 가설이 어느 정도 타당하다고 생각하지만, 확실한 분석을 위해 배송 지연율이라는 새로운 파생 변수를 만들어 추가 분석을 진행하였다.
- 배송 지연율 : 지연된 거래 횟수 / 총거래 횟수
(여기서 배송 지연율은 총거래 횟수에 대한 지연된 거래 횟수의 비율임)
- 브라질의 전체 주에 대해 분석한 결과, 고객 만족도가 평균보다 낮은 주인 RR, AL, MA, SE, PA, BA, CE, RJ, PI, ES 등 배송 지연율이 평균 이상이며, 고객 만족도와 배송 지연율이 관련이 있음을 알 수 있다.
- 이는 배송 기간과 수령 기간이 고객 만족도와 관련이 있다는 가설을 더 타당하게 만들어 주며, 이후 솔루션을 제시할 때 충분히 고려될 수 있음을 시사한다.

1-7. 지역별 분석

- [1-5], [1-6]에서 여러 변수를 고려하여 분석한 결과, 약간의 특이점은 있지만 대체로 지역적 특성이 배송에 미치는 영향이 명확하게 드러난다고 판단하였다.
- 이에 따라, 고객 만족도를 남긴 구매자를 대상으로 주를 북부, 북동부, 중서부, 남동부, 남부로 5개의 지역으로 분류해 분석하였다.

1-8. 첫 번째 핵심 KPI 선정 – 배송 지연율 (표 1-7, 그림 1-10, 11, 12, 13, 14 참조)

- 북동부와 북부 지역에서 낮은 고객 만족도가 관찰되었으며, 이는 지역별 지연율과 반비례하는 관계를 보여주었다.
- 지역별 배송 기간과 지역별 수령 총기간 또한 대체로 반비례하는 형태를 보임을 확인하였다.
- 지역별 거래 횟수는 남동부 지역에 집중되어있는 것으로 나타났다.
- 전체 지역에서 약 6% 이상의 지연율을 보이는 가운데, 북동부 지역은 약 14%의 지연율을 보였다.
- ▷ 이와 같은 지역별 분석 결과를 통해 고객 만족도와 배송 지연율 사이에 뚜렷한 반비례 관계를 확인할 수

있었다.

- ▷ 지역별 배송 지연율을 주기적으로 모니터링하며, 이를 기반으로 만족도 향상을 위한 전략을 마련할 필요가 있다고 판단하였다.
- ▷ 따라서, 지역별 배송 지연율을 핵심 성과 지표(KPI)로 선정하여 지역별로 세밀한 관리와 개선 방안을 탐색하도록 하겠다.

2. RFM 분석 과정

❖ 고객관계관리(CRM)에 보편화된 RFM분석은 이커머스 사업에서 고객의 최근 구매 행동(Recency), 구매 빈도(Frequency), 구매 금액(Monetary)을 분석함으로써, 가치 있는 고객을 식별하고, 각 고객 등급에 맞춤형 마케팅 전략을 수립할 수 있다.

2-1. RFM에 따른 K-Means 클러스터링 및 집단 수 선정

- RFM 점수를 정규화하여 K-Means 클러스터링을 수행하였다.
- 클러스터링되는 군집의 개수는 Elbow Method를 통해 오차제곱합(SSE, sum of squared error)의 변화율이 급격히 줄어지는 부분인 4개를 최적 집단 개수로 선정하였다. (그림 2-3, 4 참조)

2-2. RFM 변동 계수에 따른 가중치 선정

- K-Means 클러스터링에 의해 그룹화된 집단에 대해 각각 R, F, M의 변동 계수를 구해, 자료의 변동성과 측정 단위가 다른 자료에 대해 적절한 가중치를 산정하기 위해 다음과 같은 과정을 수행하였다.
- 각 집단의 R, F, M의 변동 계수를 구하기에 앞서 변수에 대한 평균과 표준편차를 계산하였고, 결과는 다음과 같다. (표 2-1 참조)
- <표 2-1>을 이용하여, 4개의 집단으로부터 각 집단의 R, F, M의 변동 계수를 얻을 수 있었으며, 변동 계수는 그룹별 변수에 대한 표준편차에 평균을 나눠 계산하였다.
- 이를 표로 나타낸 결과는 다음과 같다. (표 2-2 참조)
- 변동 계수를 이용하여 집단별 변수의 특성을 최대한 가중치에 반영하되, 변동성이 작은 수치를 사용함으로써 각각의 평균으로부터의 큰 변동성을 갖지 않는 가중치를 부여할 수 있게 하였다. 취득된 가중치는 합이 1이 되도록 정규화하여 최종 결정되었다.
- 변동 계수로부터 취득한 각 집단의 R, F, M의 비 정규화 가중치와 정규화 가중치는 다음과 같다. (표 2-3 참조)
- 정규화된 가중치를 살펴보면 RFM Score의 영향력은 $M \rightarrow F \rightarrow R$ 순임을 확인해볼 수 있다.

2-3. 고객별 RFM 점수 분포 및 RFM Score에 따른 등급 분류

- 각 고객의 R, F, M에 앞서 구한 가중치를 적용하여 고객마다 RFM Score를 취득하였을 때, 고객들의 점수 분포에 따라 4개의 등급으로 분류하였고 그 결과는 다음과 같다. (표 2-4 참조)

2-4. ANOVA를 통한 통계적 유의성 검증

- 각 고객을 등급별로 나누었을 때, 거래 가격, R, F, M 각각에 대해 유의한 차이를 보이는 것을 확인할 수 있었으며, 검증 결과는 다음과 같다. (표 2-5 참조)

2-5. 고객별 등급별 분석

- 등급을 나눌 때, 점수 분포에 따라 등급을 결정하였으며, Silver, Gold, Platinum, Diamond의 각 인원은 40239, 29757, 10083, 5026명이 존재하여, 높은 등급일수록 인원이 적다.
- 파레토 법칙에 따라, 상위 20%(Platinum, Diamond)의 고객이 대부분의 매출을 차지할 것이란 예상과 달리 Silver와 Gold 등급의 고객 매출 비중이 약 58.4%로 절반 이상의 매출이 낮은 등급의 고객들로부터 발생하는 것을 확인할 수 있다.
- 고객 별 1인당 평균 거래액은 등급 Silver와 Gold 등급의 고객은 서로 큰 차이를 보이지 않으나, Silver에서 Gold, Platinum에서 Diamond 고객은 서로 큰 차이를 보이는 것을 확인할 수 있다.
- 1인당 평균 거래액은 Diamond 등급의 고객보다 Platinum 등급의 고객이 더 높은 것을 확인할 수 있었으며, 이는 R 수치와 F 수치를 모두 반영한 결과로 보인다.
- 이를 표로 정리한 결과는 다음과 같다. (표 2-6 참조)

2-6. RFM Score 별 그룹화

- RFM Score 별로 그룹화한 결과는 다음과 같다. (표 2.7 참조)
- Bloomreach-RFM Segmentation을 참고하여 고객을 세분화하였다.
- 데이터 특성상 Frequency의 빈도수가 대부분 1이기 때문에 VIP 고객과 충성 고객을 제외하고 고객 등급을 세분화하였다.

2-7. 고객 그룹별 분석

- 고객 수, 매출 비중, 1인 거래 평균, 거래 가격 기준 상위 20% 고객 수를 확인하여 수익성이 높은 그룹을 파악하였다. (표 2-8, 그림 2-5, 6, 7, 8, 9 참조)

2-8. 핵심 그룹 선정

❖ <표 2-8>에서 거래 가격 기준 상위 20% 고객 수를 기준으로 핵심 그룹, 관심 그룹으로 분류하였다. 예외로, 신규 고객은 12명으로 매우 적지만 핵심 그룹으로 선정하였다. 총 신규 고객의 수가 많고, 해당 고객들을 만족시키면 충성도가 유지되고 수익이 증가할 가능성이 존재하기 때문이다.

- 핵심 그룹
 - ✓ 잠재 고객, 놓치면 안 될 고객 : 그룹별 분석을 통해 매출에 큰 비중을 차지하는 것을 알 수 있다.
 - ✓ 신규 고객, 휴면 예정 고객, 이탈 우려 고객, 관심 필요 고객 : 일반적으로 현재 고객의 만족도 유지가 신규 고객 확보보다 비용 효과적이다.
Bain & Company에 따르면, 기존 고객의 유지율이 5% 증가할 때마다 기업이 얻는 이익은 최소 25%에서 최대 95%까지 증가한다고 한다. 따라서, 기존 고객을 활성화하고, 충성도 높은 고객으로 전환 시킬 필요가 있다.
- 관심 그룹
 - ✓ 동면 고객, 이탈 고객, 잠재 충성 고객 : 고객 수가 적으므로 마케팅 비용을 절약할 필요가 있다.

2-9. 두 번째 핵심 KPI 선정 - 핵심 그룹 매출 비율

- 기업에서 모든 고객이 만족하는 마케팅을 진행하는 것은 현실적으로 어렵다.
- 따라서, 고객 세분화를 통해 상대적으로 유지비용이 낮으면서 수익성이 높은 고객들에게 집중하여 마케팅 효율을 높여야 한다.
- 핵심 그룹을 제외한 그룹의 마케팅 비용을 최소화한다.
- <표 2-8>에서 잠재 고객과 놓치면 안 될 고객의 매출 비중 합계가 83.0%로 상당히 높은 것을 확인할 수 있다. 1인 거래 가격의 평균이 타 그룹에 비해 월등히 높은 것을 알 수 있다.
- 상위 20% 고객이 전체 매출에서 약 62%를 차지하고 있다. (거래 가격 합계 : 18158674.25, 상위 20% 거래 가격 합계 : 11247259.52)
- 파레토 법칙에 따르면 평균적으로 매출의 80%가 20%의 고객으로부터 발생한다고 한다. 따라서, 핵심 고객 20%에 집중하여 해당 그룹이 차지하는 매출 비중을 80%까지 끌어올리는 것이 중요하다.

3. 시간대별 카테고리 AOV

- ❖ AOV는 고객 구매 패턴 파악 및 경험 개선을 통해 수익 최적화를 할 수 있는 지표로, 이를 통해 고객이 선호하는 제품 가격대를 파악하고, 맞춤형 마케팅 전략과 서비스를 제공할 수 있다.

3-1. 데이터 전처리

- ❖ 주어진 데이터에 따르면 소비자들의 구매 형태와 소비 패턴이 시간대에 따라 변화한다. 예를 들어, AOV는 22시부터 감소하고 6시부터 서서히 증가하는 추세를 확인할 수 있었다. (그림 3-1 참조)
이에 따라, 각 시간대의 주문 수와 AOV를 반영한 세부적인 마케팅 전략이 필요하다고 판단된다. 이를 위해서는 데이터 전처리 과정을 통해 필요한 정보를 추출해야 하며, 다음 항목들을 구하였다.

- $AOV = \text{총거래 가치} / \text{총판매량}$
- 시간대별 AOV : 시간대별로 거래 가치의 합과 판매량의 합을 구하여 AOV 계산.
- 시간대별 카테고리 AOV : 각 시간대의 카테고리별 거래 가치의 합과 판매량의 합을 구하여 AOV 계산.
- 시간대별 카테고리 & 지역 AOV : 시간대별로 각 카테고리의 지역별 거래 가치의 합과 판매량의 합을 구하여 AOV 계산.
- 시간대별 카테고리 & 지역 & 주 AOV : 시간대별로 카테고리, 지역, 주의 거래 가치의 합과 판매량의 합을 구하여 AOV 계산
- 시간대별 카테고리의 고객 등급 : 고객 등급별로 각 시간대의 카테고리를 구매한 고객의 수 계산

3-2. 위의 데이터를 활용하여 Power BI로 시각화 후 분석 진행.

- ❖ 앞서 언급했듯이 AOV는 시간에 따라 변동하기 때문에 이를 기반으로 기업은 시간별 판매전략을 세울 수 있다. 예를 들어 기업이 AOV가 높은 제품의 판매량을 증가시키고자 한다면 AOV가 높은 시간대를 파악한 후, 적절한 카테고리를 고객들에게 추천하거나 할인 행사를 진행하는 등의 행동을 취할 수 있다. 이러한 전략을 세우기 위해서는 주어진 자료를 시각화하는 과정이 필요하다고 판단하였다. 따라서 세부 정보를 직관적으로 파악할 수 있는 Power BI를 사용하여 시각화를 진행하였다.

- Power BI의 첫 페이지는 시간대별 AOV를 막대그래프, 시간대별 총판매량을 꺾은선 그래프로 나타낸다. (그림 3-2 참조)
- 원하는 시간대를 선택하여 '선택 시간대로 이동' 버튼 클릭 -> 해당 시간대의 카테고리별 AOV 및 총판매량, 고객 등급 비율, 고객 지역별 AOV, 고객 주별 AOV를 나타낸 상세 페이지로 이동. (그림 3-3 참조)
- 동시에 '고객 지역별 AOV 그래프'에서 지역을 선택하면 해당 지역의 고객 등급 비율과 고객 주별 AOV를 확인할 수 있다. (그림 3-4 참조)
- '카테고리별 AOV 및 총판매량 그래프'의 카테고리를 선택하면 해당 카테고리의 고객 등급 비율과 고객 지역별 AOV, 고객 주별 AOV를 확인할 수 있다. 동시에 '고객 지역별 AOV 그래프'에서 지역을 선택하면 해당 카테고리, 해당 지역의 고객 등급 비율과 고객 주별 AOV를 확인할 수 있다. (그림 3-5, 6 참조)

3-3. 세 번째 핵심 KPI 선정 – 시간대별 카테고리 AOV

❖ 수익성 및 소비자 구매 패턴 분석

- 시간대별 카테고리 AOV는 상품 카테고리별 평균 거래 가치를 나타내기 때문에, 이를 통해 어떤 상품이 기업의 수익에 기여하고 있는지를 알 수 있다. 시간대별 카테고리 AOV를 적절히 활용한다면, 제한된 시간에만 진행되는 프로모션과 같은 마케팅 전략을 통해 고객의 즉각적인 구매를 유도함으로써 기업의 수익성 증대를 기대할 수 있다.
- 또한 카테고리별로 고객 등급 비율과 지역별 AOV는 어떤 유형의 고객이 해당 상품을 구매하는지를 나타낸다. 따라서, 이를 분석한다면 소비자 구매 성향을 알 수 있으며, 특정 카테고리에서 활발한 구매를 보이는 등급 및 거주 지역에 집중하여 특정 시간에만 진행되는 할인 프로모션 등의 마케팅 전략을 세울 수 있다고 판단했으므로 시간대별 카테고리 AOV를 세 번째 KPI로 선정하게 되었다.

| | |
|-----------|--|
| 분석 기반 KPI | 지역별 배송 지연율, 핵심 그룹 매출 비율, 시간대별 카테고리 AOV |
| 솔루션 제안 | |

※ 솔루션 및 전략

1. 지역별 배송 지연율

1-1. 지역별 배송 지연율 솔루션

- 배송 지연율을 해결하기 위한 솔루션으로, 지역마다 물류 운송 거점센터 설립을 고려하여 배송 지연을 방지하도록 한다.
- 지역별 지연율이 높은 소비자와 판매자의 위치 정보 데이터, 지역 별 거래량이 높은 소비자와 판매자의 위치 정보 데이터를 사용하여 5개의 지역 데이터로 나누어 준다. (관점에 따라 다른 데이터 적용할 수도 있음)
- Elbow Method를 사용하여 지역별 최적의 클러스터 수를 도출한 후 (그림 1-15 참조), K Means Clustering을 진행한다.
- 최적의 클러스터 수로 분류한 지역별 거점센터 후보 좌표를 지도에 표시한 후 (그림 1-16 참조), 브라질의 지리적 특성, 교통, 치안 등 다양한 도메인 지식을 활용해 적절한 위치를 선별한다.

1-2. 솔루션 적용 및 검토 (그림 1-17 참조)

- 북부 지역은 첫 번째 거점을 선택하였다. 이 위치는 발지칸스 국제공항 부근이며, 바다로의 접근성 또한 우수해 선정하였다.
- 북동부 지역의 경우 두 번째 거점을 선택하였는데, 근처에 항구 도시가 있어 바다로의 접근성이 우수하며 인프라가 충분해 선정하였다.
- 중서부 지역은 세 번째 거점을 선택하였다. 치안과 환경이 좋고, 교통 편의성이 뛰어난 고이아니아를 중심으로 브라질리아와의 접근성을 고려하여 물류 거점을 선정했다.
- 남동부 지역은 두 번째 거점을 선택하였다. 상파울루 주의 경제 중심지이며 고객 밀도가 높고 상공업이 발달해 선정하였다.
- 남부 지역은 첫 번째 거점을 선택하였는데, 대중교통 체계가 발달해 쿠리치바 주에서 효율적인 물류 운송을 기대할 수 있어 선택하였다.
- 추가로, 아마존 분지의 열대우림 지대에 위치한 마나우스 주에 특수 물류 운송 거점센터 설립을 추진한다. 이유는 아마존에서의 주문량도 갈수록 증가하는 추세이며, 뛰어난 문화시설을 갖추고, 항구 도시인 마나우스 주가 배송 면에서 뛰어난 역할이 가능하다고 판단하였기 때문이다.

1-3. 기대효과

- ▷ 각 지역의 특성을 고려하여 최적의 물류 거점을 선정함으로써, 배송 지연율을 최소화하고 배송 서비스 품질 향상 및 고객 불만족을 최소화할 기대할 수 있다. 이는 고객 만족도 향상으로 이어져 기업의 수익성 확보에 도움이 될 것으로 판단된다.

2. 핵심 그룹 매출 비율 (표 2-7 참조)

2-1. 핵심 그룹 매출 비율 솔루션 및 기대효과

- 잠재 고객
 - ✓ 잠재 고객은 최근에 거래하였으나 거래 빈도가 낮고 큰 금액을 소비한다. 거래 빈도가 낮은 이유를 파악하고, 더 많은 거래를 유도할 필요가 있다.
 - ▷ 리뷰데이터 분석을 통한 서비스 개선
글로벌 시장 조사기관 포레스터(Forrester)에 따르면, 고객 경험 관리에 신경 쓰는 기업일수록 그렇지 않은 기업에 비해 고객 유지율이 1.9배 높은 것으로 나타났다.
잠재 고객은 큰 금액을 소비하지만, 거래 빈도수가 낮으므로 서비스에 불만족을 느꼈을 가능성이 있다. 고객의 불만 사항을 확인할 수 있는 창구 중 하나인 리뷰데이터를 분석하여 더 나은 서비스를 제공할 필요가 있다.
 - ▷ 크로스 셀링 유도를 위한 연관 상품 추천 및 마케팅
포브스(Forbes) 조사에 따르면, 86%의 고객들이 좋은 경험을 제공하는 기업에 더 많은 돈을 지불할 의향이 있는 것으로 나타났다. 그만큼 고객들은 '경험'을 중요하게 생각하고 고객 집단이 아닌 '개인'으로 인식되기를 바란다.
고객들에게 좋은 경험을 제공하기 위해 개인화된 연관 상품 추천 및 마케팅이 도움이 될 것이다.

연관 상품 추천을 통해 재거래를 유도하고, 맞춤형 제안을 제공하여 고객들의 관심을 끌고 더 많은 거래를 유도할 수 있을 것이다.

■ 놓치면 안 될 고객

- ✓ 놓치면 안 될 고객은 최근에 거래하지 않았으나 큰 금액을 소비했다. 수익성이 높은 고객이기 때문에 재활성화 시킬 필요가 있다.

▷ 재활성화 마케팅

Profitwell에 따르면 재활성화 마케팅은 이탈한 B2B 고객의 20% 이상을 다시 가져올 수 있다. 또한, 이미 이루어진 비용 투자를 토대로 하여 신규 고객을 확보하는 비용을 피할 수 있다. 해당 고객은 첫 거래 이후 오랜 시간이 지난 그룹이므로 사이트에 대한 흥미가 떨어진 고객이라 할 수 있다. 따라서, 사이트의 존재를 다시금 알리고 관심을 가질 수 있게끔 유도하는 것이 필요해 보인다.

과거 데이터를 바탕으로 개인화된 제안과 웰컴백 쿠폰 발행을 알리는 SMS, 이메일, 앱 알림 등의 마케팅 메시지를 활용한다면 긍정적인 효과가 있을 것으로 보인다.

웰컴백 쿠폰은 다른 쿠폰들과는 달리 사용하지 않으면 빠르게 사라지기 때문에 거래를 유도하는데 더욱 효과적일 것이다.

■ 신규 고객

- ✓ 신규 고객은 최근 유입되어 거래 빈도가 낮고, 거래 금액이 적다. 고객 확보 비용보다 신규 고객을 유지하는 것이 비용 효과적이다.

▷ 멤버십 첫 달 무료 체험

전체 이커머스 데이터의 재거래 평균 주기는 32.4일로 나타났다. 이는 한 달 안에 재거래를 발생시키지 못하면 고객이 이탈할 확률이 높아진다는 것을 의미한다.

멤버십 첫 달 무료 체험을 통해 '락인' 효과를 통해 신규 고객을 유지하는 것이 효과적일 것이다. 락인 효과란 특정 상품이나 서비스를 이용하기 시작하면 대체재가 있어도 기존의 것을 계속 유지하려는 경향성을 의미한다.

이를 통해, 신규 고객의 재거래를 유도하고 고객에게 긍정적인 경험을 제공한다면 고객 유지에 효과적일 것이다.

■ 휴면 예정 고객

- ✓ 휴면 예정 고객은 최근성, 빈도, 거래 금액이 모두 평균 이하이다. 휴면 고객으로 전환되기 전, 해당 이커머스에 다시 관심을 가지도록 유도하는 것이 필요하다.

▷ 타임세일

사람들은 같은 크기의 이익과 손실이라 해도 이익에서 얻는 효용(기쁨)보다 손실에서 느끼는 비효용(고통)을 더 크게 느껴, 사람들이 손실(고통)을 줄이려고 하는 성향이 있다. 이를 손실 회피성이라 한다. 타임세일은 지금 당장 거래하지 않으면 손해라는 심리를 부추겨, 소비자의 거래 욕구를 자극한다.

휴면 예정 고객의 거래 목록을 분석하고, 해당 목록에 맞는 타임세일을 진행할 때 SNS 채널을 활용하여 개인화된 CRM 메시지를 발송한다면 고객의 관심을 얻을 수 있을 것이다.

■ 이탈 우려 고객

- ✓ 이탈 우려 고객은 최근 거래하지 않았으나, 자주 거래했고 많은 금액을 소비했다. 수익성과 거래 빈도

가 높으므로 이러한 고객이 이탈한다면 기업 매출에 큰 타격이 될 것이다.

▷ 할인 쿠폰

Forrester Research의 연구에 따르면 온라인 쇼핑객의 68%는 온라인 쿠폰이나 프로모션 코드를 제공하는 브랜드에 대한 충성도가 더 높다고 답했으며, 온라인 쇼핑객의 59%는 제안하는 브랜드에서 거래할 가능성이 더 높다고 응답했다.

고객의 거래 패턴을 분석하여 할인 쿠폰을 제공하고 합리적인 만료 날짜를 설정하여 긴급성을 높인다면 더 효과적일 것이다.

■ 관심 필요 고객

- ✓ 최근성, 거래 빈도, 거래 금액이 모두 평균 이상이다. 해당 고객을 유지한다면 앞으로의 기업 성장에 큰 도움이 될 것이다.

▷ 재거래 주기에 맞춘 상품 광고

고객 유지율을 5% 증가시키면 수익이 25%~95% 증가할 수 있다.

고객의 재거래 주기에 맞춰 적절한 상품을 추천하고, 할인 쿠폰까지 제공한다면 고객 유지에 효과적일 것이다.

2-2. 그 외 그룹 솔루션 및 기대효과

■ 관심 그룹

- ✓ 잠재 충성 고객, 동면 고객, 이탈 고객은 수익성이 낮은 그룹이기 때문에 마케팅 비용을 절감하는 것이 필요하다. 이벤트 진행 시 SNS 채널을 활용한 메시지를 보내는 것이 비용 측면에서 유리할 것이다.

3. 시간대별 카테고리 AOV

3-1. 시간대별 카테고리 AOV 솔루션

- 부가 상품 판매 : 특정 시간에 AOV가 높다는 것은 고가의 상품이 주로 판매된다는 것으로 해석될 수 있으며, 그 시간대의 AOV가 높은 카테고리의 상품과 부가 상품들을 추천하여 세트 구매를 유도한다면 매출 성장에 효과적일 것이다.
 - ✓ AOV가 가장 높은 시간대는 오전 9시로, 해당 카테고리인 PCS, portable_home_oven_and_coffee, home_appliances_2, cinema_photo, bedroom_furniture 등에 관련된 부가 상품들을 추천한다.
 - ✓ 전략 예시)
 - 카테고리: home_appliances_2(가전제품)
 - 기본 상품: 전자레인지
 - 부가 상품: 전자레인지 용기, 스팀 클리너, 제습제
 - 설명: 이러한 부가 상품들은 전자레인지의 사용성을 높여주고, 음식 조리과 관련된 다양한 활동을 더욱 편리하게 해줄 뿐만 아니라 부가 상품을 함께 판매함으로써 총매출 상승을 기대할 수 있다.
- 묶음 할인 혜택 : AOV가 낮은 시간대는 저가 상품을 많이 구매한다고 판단할 수 있으며, 그 시간대의 AOV가 낮고, 판매량이 높은 카테고리에 대해 묶음 할인 혜택을 제공하는 것이 효과적일 것이다.

- ✓ AOV가 가장 낮은 시간대인 오전 2시에는 AOV가 낮고, 판매량이 높은 카테고리인 bed_bath_table, furniture_decor, sports_leisure, health_beauty, watches_gifts 등에 대해 묶음 할인 혜택을 제공하여 더 많은 구매를 유도한다.
- ✓ 전략 예시)
 - 카테고리: sports_leisure
 - 기본 상품: 등산화
 - 묶음 상품: 등산화 + 배낭 or 등산 스틱
 - 예시 설명: 패키지 제안으로 등산을 처음 시작하는 사람들이나 장비를 새로 마련하고자 하는 사람들에게 매력적인 옵션을 제공할 뿐만 아니라 묶음 할인 판매로 재고 처리 가능 및 수익 증가를 기대할 수 있다.

3-2. 기대효과

- ▷ 고객 등급, 지역을 고려하여 프로모션을 진행한다면 고객 만족도 및 충성도 향상과 판매량 증가를 동시에 기대할 수 있다.

『부록』

<표1-1. 고객 만족도 분포>

| 점수 | 리뷰 개수 | 비율 |
|----|-------|------|
| 1 | 8746 | 0.10 |
| 2 | 2716 | 0.03 |
| 3 | 7348 | 0.08 |
| 4 | 17331 | 0.20 |
| 5 | 51732 | 0.59 |

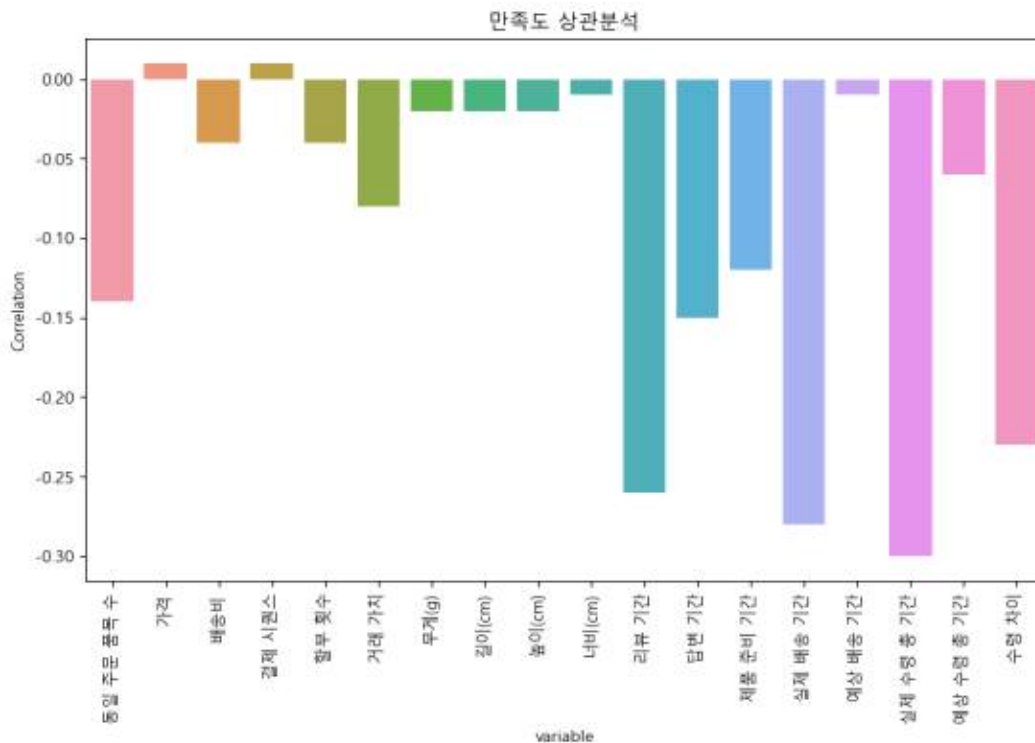
<그림 1-1. 고객 만족도 점수에 대한 고객 분포 시각화>



<표1-2. 고객 만족도 상관분석>

| 변수 | 상관관계 | 유의확률 |
|------------|-------|-------|
| 실제 수령 총 기간 | -0.30 | 0.000 |
| 실제 배송 기간 | -0.28 | 0.000 |
| 리뷰 기간 | -0.26 | 0.000 |
| 수령 차이 | -0.23 | 0.000 |
| 답변 기간 | -0.15 | 0.000 |
| 동일 주문 품목 수 | -0.14 | 0.000 |
| 제품 준비 기간 | -0.12 | 0.000 |
| 거래 가치 | -0.08 | 0.000 |
| 예상 수령 총 기간 | -0.06 | 0.000 |
| 할부 횟수 | -0.04 | 0.000 |
| 배송비 | -0.04 | 0.000 |
| 무게(g) | -0.02 | 0.000 |
| 길이(cm) | -0.02 | 0.000 |
| 높이(cm) | -0.02 | 0.000 |
| 너비(cm) | -0.01 | 0.016 |
| 예상 배송 기간 | -0.01 | 0.004 |
| 결제 시퀀스 | 0.01 | 0.031 |
| 가격 | 0.01 | 0.028 |

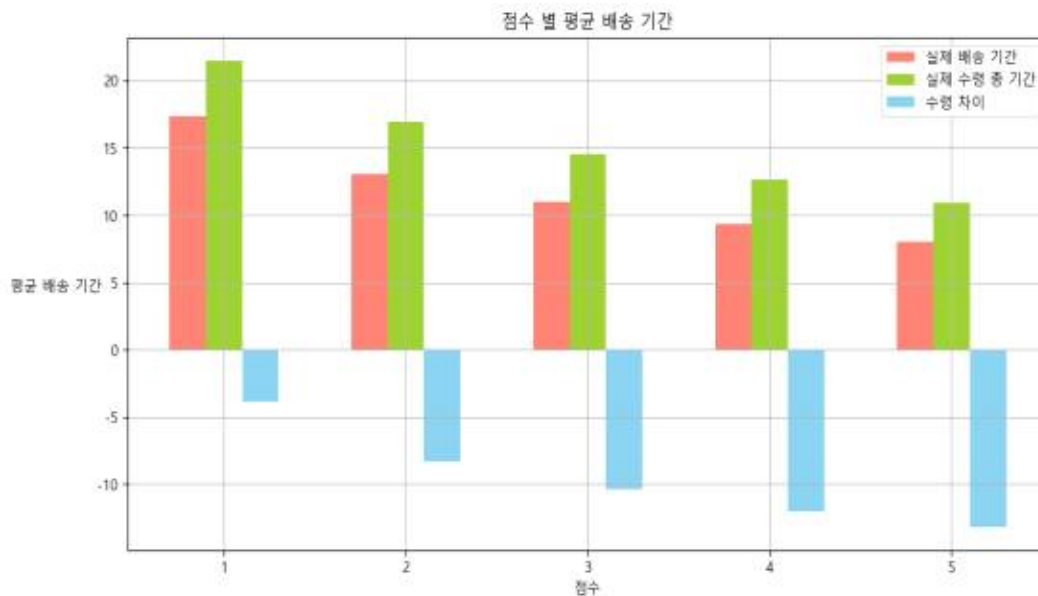
<그림 1-2. 고객 만족도 상관분석 시각화>



<표1-3. 고객 만족도 기간, 수령 차이 분포>

| 점수 | 실제 배송 기간 | 실제 수령 총 기간 | 수령 차이 |
|----|----------|------------|---------|
| 1 | 17.325 | 21.496 | -3.820 |
| 2 | 13.059 | 16.927 | -8.316 |
| 3 | 11.015 | 14.579 | -10.386 |
| 4 | 9.385 | 12.614 | -12.064 |
| 5 | 8.035 | 10.948 | -13.156 |

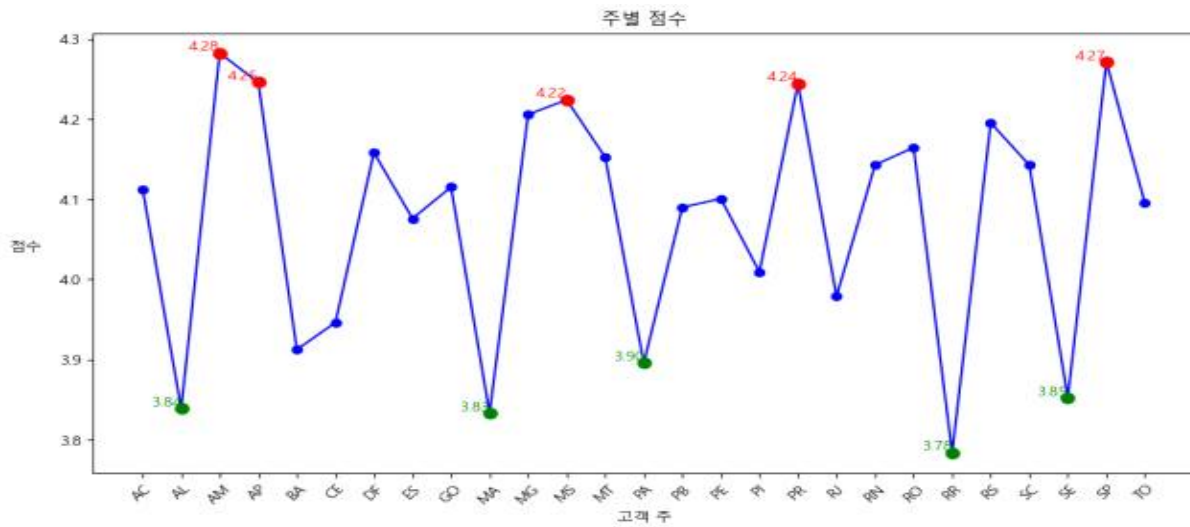
<그림 1-3. 고객 만족도에 대한 실제 배송 기간, 실제 수령 총기간, 수령 차이 분포 시각화>



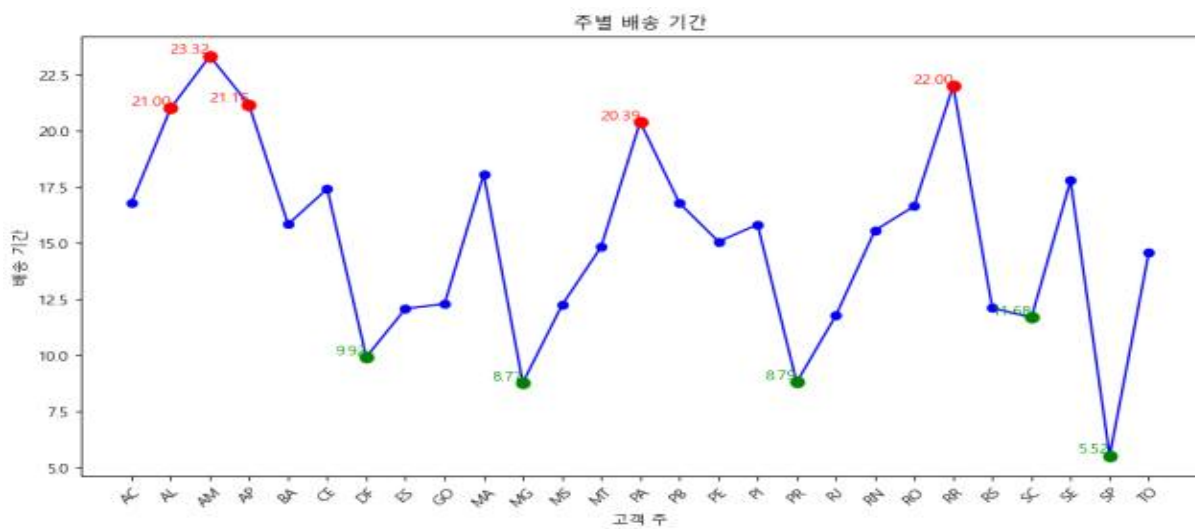
<표 1-4. 주별 고객 만족도가 평균 이하인 주와 배송 및 수령 기간이 평균 이상인 주>

| | |
|------------------|--|
| 고객 만족도가 평균 이하인 주 | RR, MA, AL, SE, PA, BA, CE, RJ, PI, ES |
| 평균보다 배송 기간이 긴 주 | RR, AM, AP, AL, PA, SE, CE, MA, AC, PB, RO, PI, BA, RN |
| 평균보다 수령 기간이 긴 주 | RR, AM, AP, AL, PA, SE, MA, CE, AC, PB, PI, RO, BA, RN |

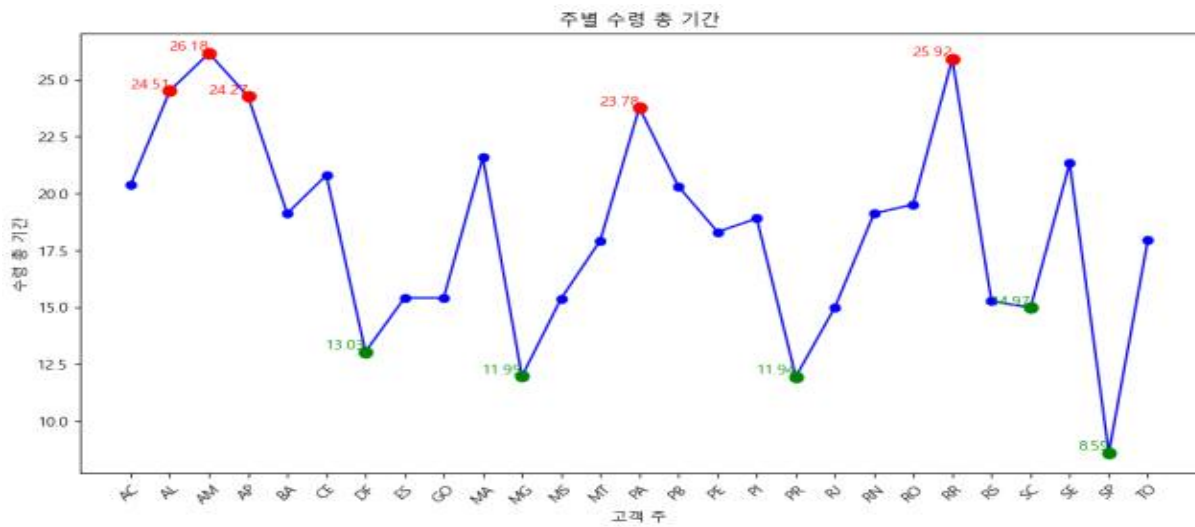
<그림 1-4. 주에 대한 고객 만족도 (리뷰 점수) 시각화>



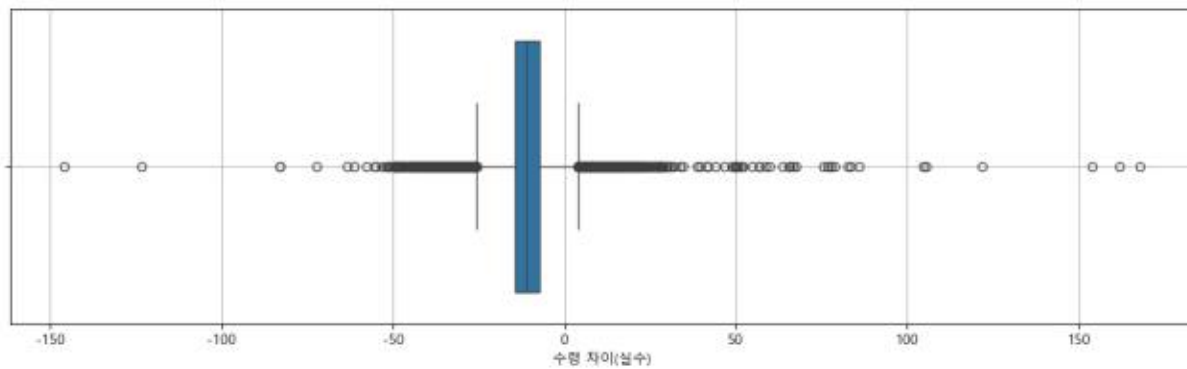
<그림 1-5. 주에 대한 배송 기간 시각화>



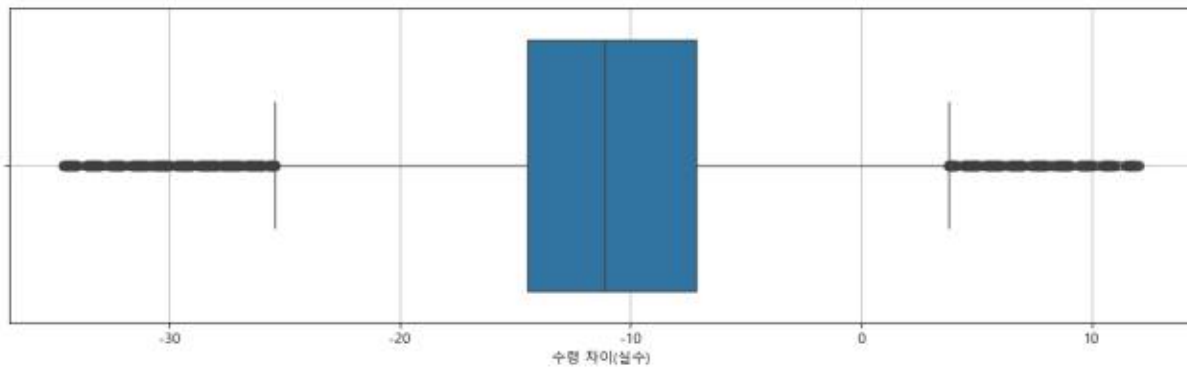
<그림 1-6. 주에 대한 수령 기간 시각화>



<그림 1-7. 주에 대한 수령 차이 (이상치 제거 전) 시각화>



<그림 1-8. 주에 대한 수령 차이 (이상치 제거 후) 시각화>



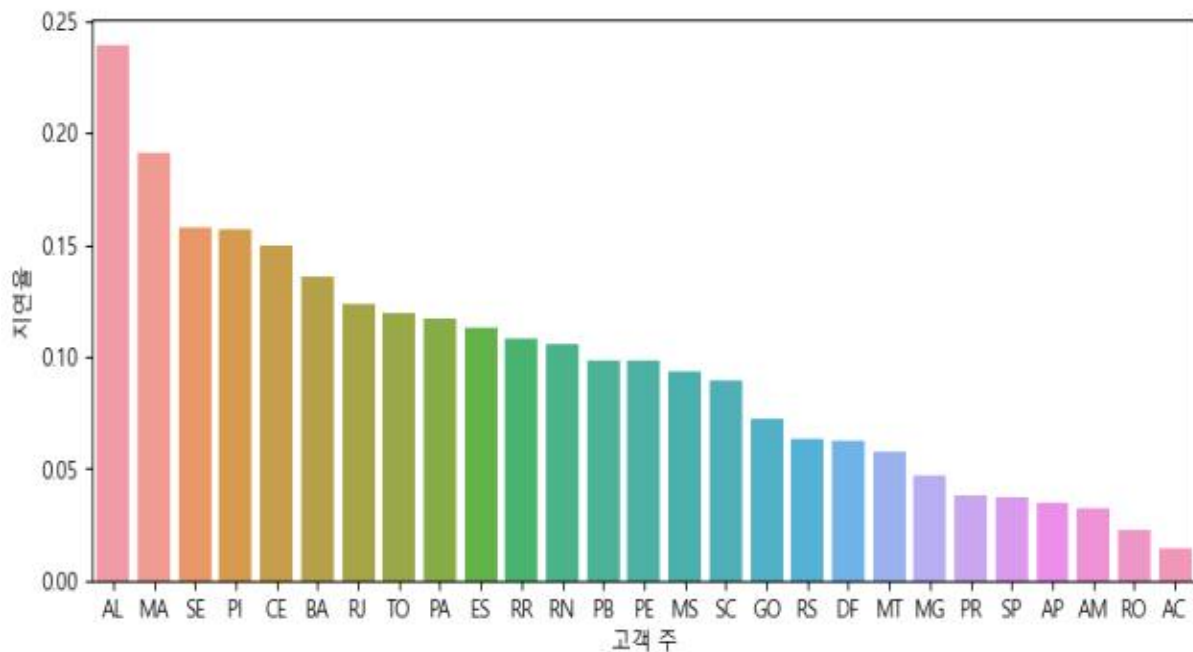
<표 1-5. 27개 주에서의 고객 만족도 상관분석>

| | 실제 배송 기간 | 예상 배송 기간 | 실제 수령 총 기간 | 예상 수령 총 기간 | 수령차이 | 거래 횟수 |
|----|-------------|-------------|---------------|---------------|--------|-------|
| 점수 | -0.487 | -0.141 | -0.514 | -0.164 | -0.300 | 0.298 |

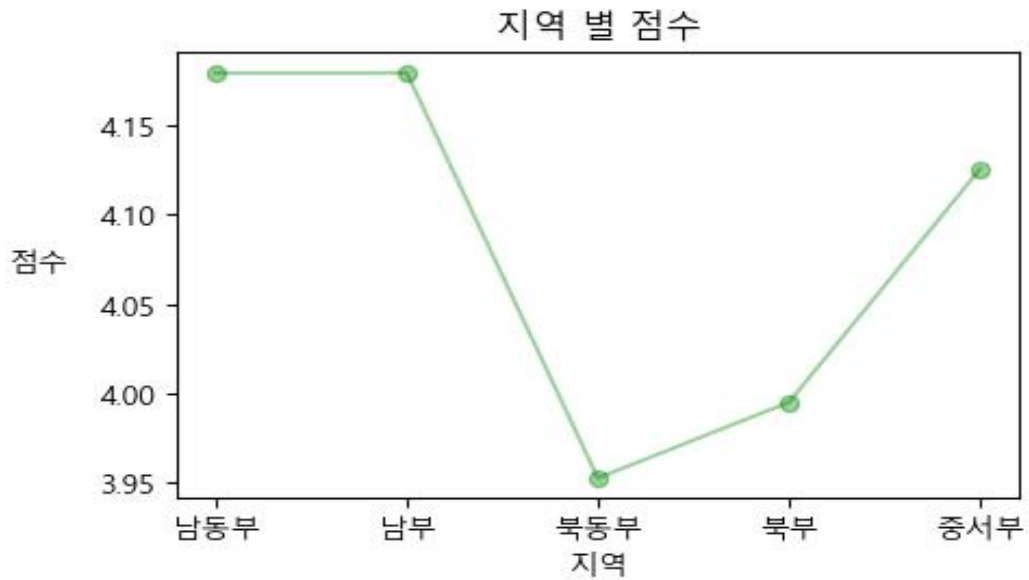
<표 1-6. 지연율이 평균 이상인 주>

| | |
|------------------|--|
| 고객 만족도가 평균 이하인 주 | RR, AL, MA, SE, PA, BA, CE, RJ, PI, ES |
| SP 근처에 위치한 주 | RJ, ES |
| 북동부 지역에 위치한 주 | AL, MA, SE, PI, CE, BA, RN, PB, PE |
| 북부 지역에 위치한 주 | TO, PA, RR |

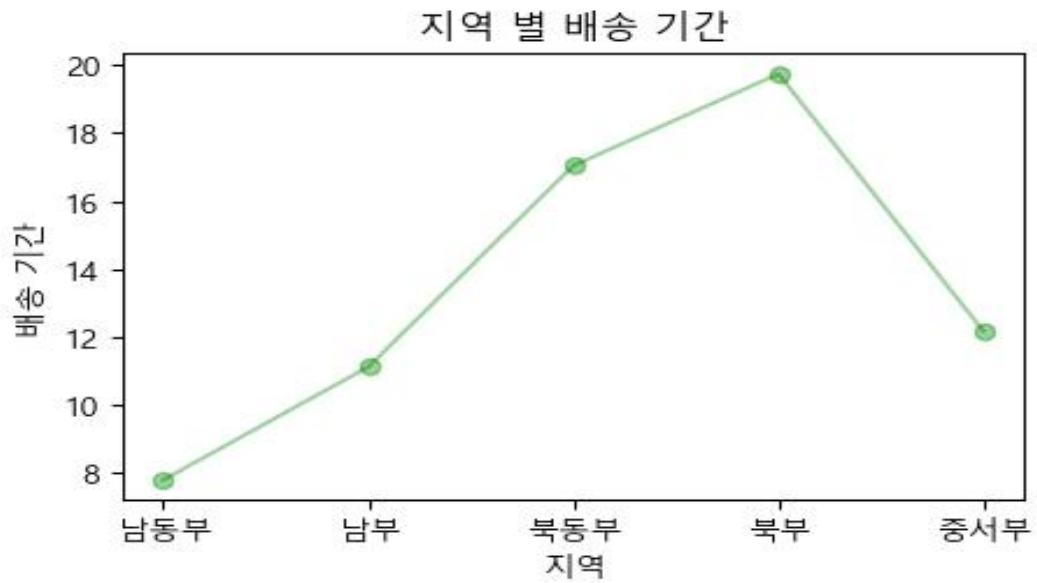
<그림 1-9. 주에 대한 배송 지연율 시각화>



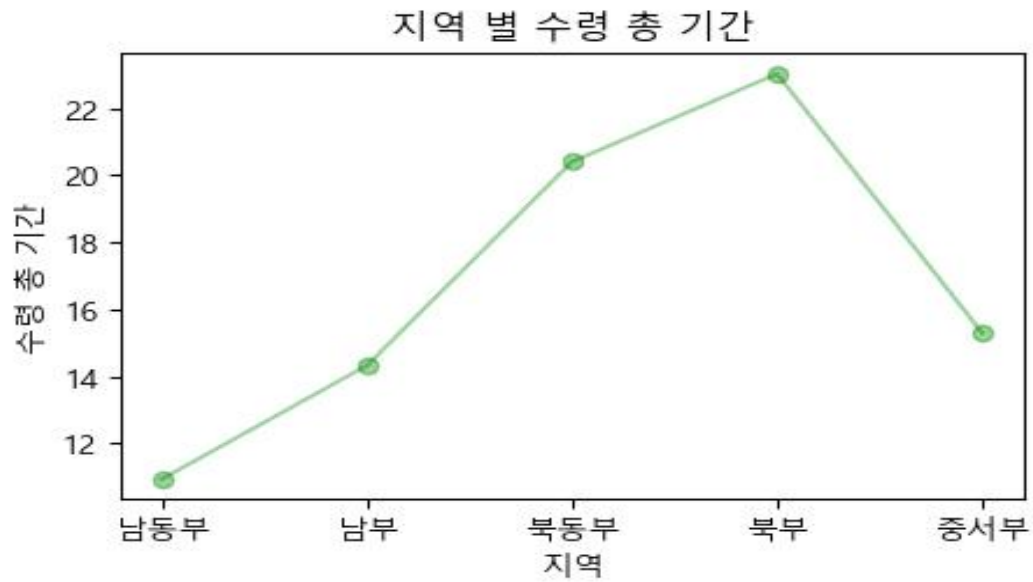
<그림 1-10. 지역별 고객 만족도 시각화>



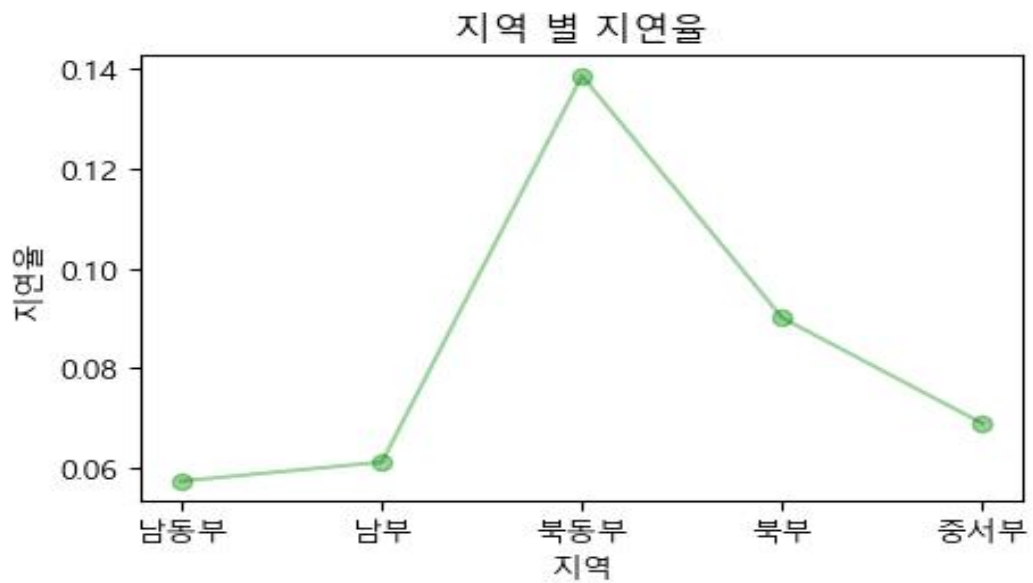
<그림 1-11. 지역별 배송 기간 시각화>



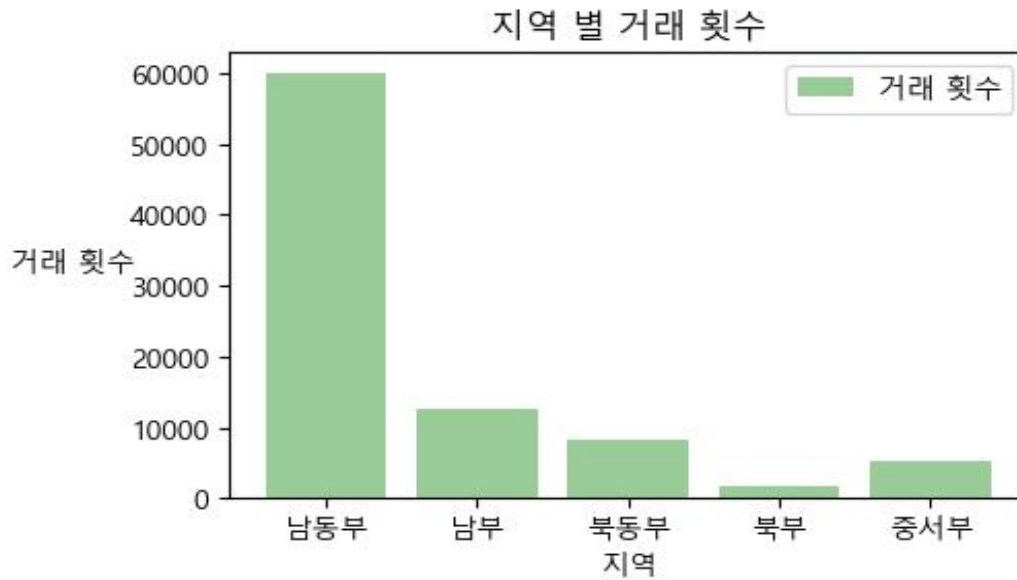
<그림 1-12. 지역별 수령 기간 시각화>



<그림 1-13. 지역별 지연율 시각화>



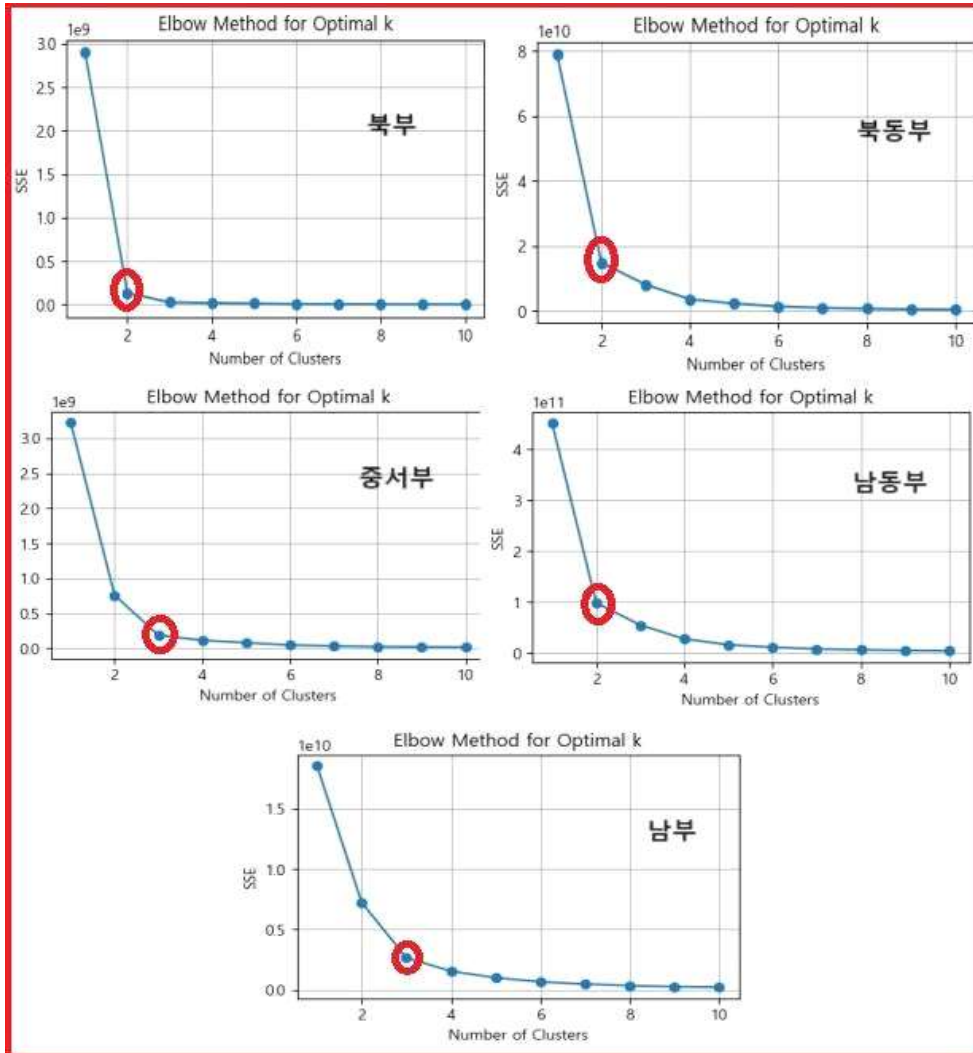
<그림 1-14. 지역별 지연된 거래 횟수 시각화>



<표 1-7. 지역별 비교 분석>

| | |
|------------|---------------------------|
| 고객만족도 | 북동부 < 북부 < 중서부 < 남동부 < 남부 |
| 배송기간, 수령기간 | 남동부 < 남부 < 중서부 < 북동부 < 북부 |
| 지연율 | 남동부 < 남부 < 중서부 < 북부 < 북동부 |
| 지연된거래횟수 | 북부 < 중서부 < 북동부 < 남부 < 남동부 |

<그림 1-15. 지역별 Elbow Method를 사용한 최적 클러스터 수 도출>



<그림 1-16. 솔루션에 대한 적용 시각화>



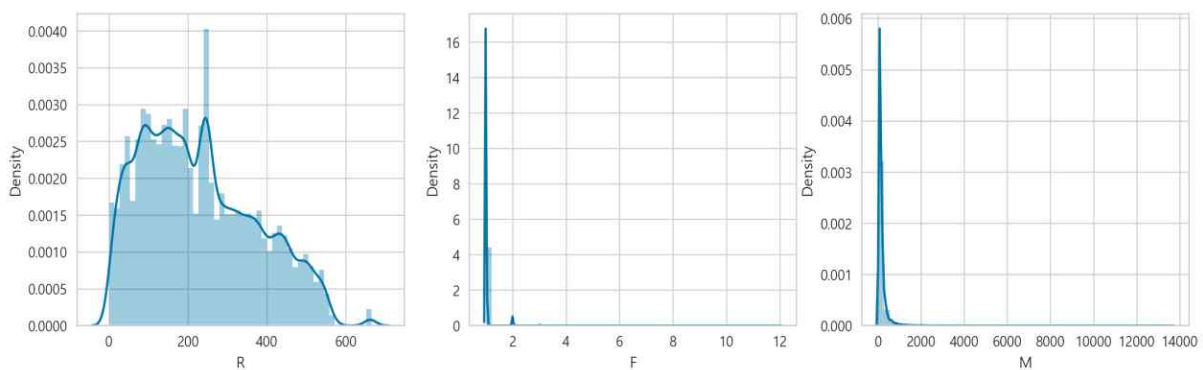
<그림 1-17. 솔루션에 대한 적용 시각화- 북부, 북동부, 중서부, 남부, 남동부, 아마존 특수 물류 운송
거점센터 후보 위치>



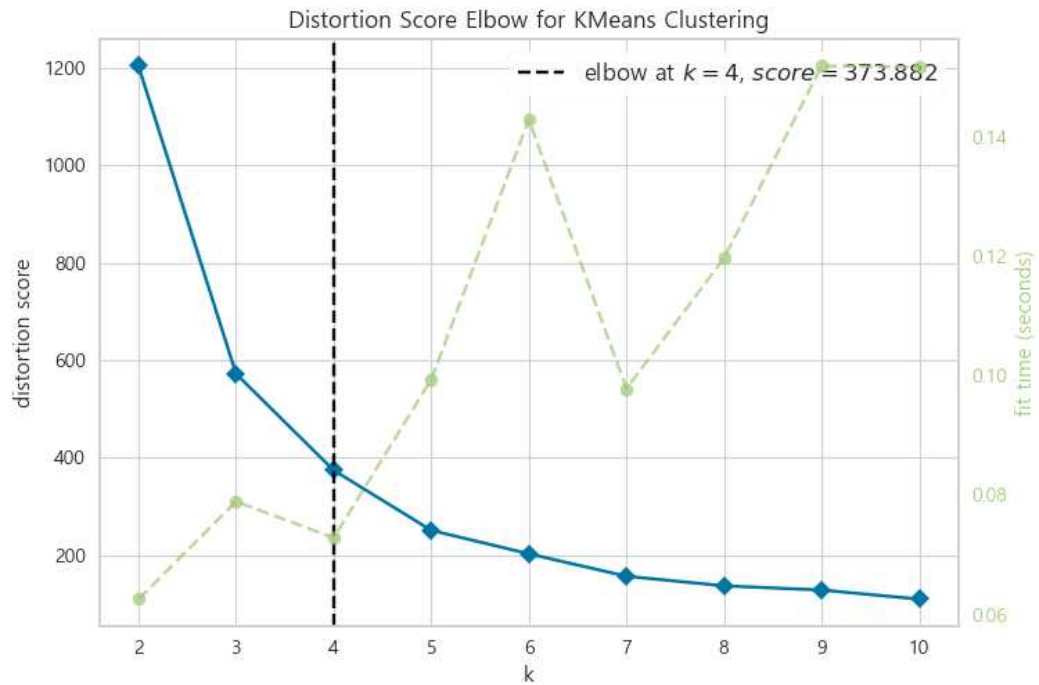
<그림 2-1. RFM 점수화>

```
=====
R-Score
1      0.198825
2      0.196005
3      0.204442
4      0.198931
5      0.201798
Name: proportion, dtype: float64
-----
F-Score
1      0.969943
2      0.030057
Name: proportion, dtype: float64
-----
M-Score
1      0.200000
2      0.200035
3      0.200059
4      0.199906
5      0.200000
Name: proportion, dtype: float64
=====
```

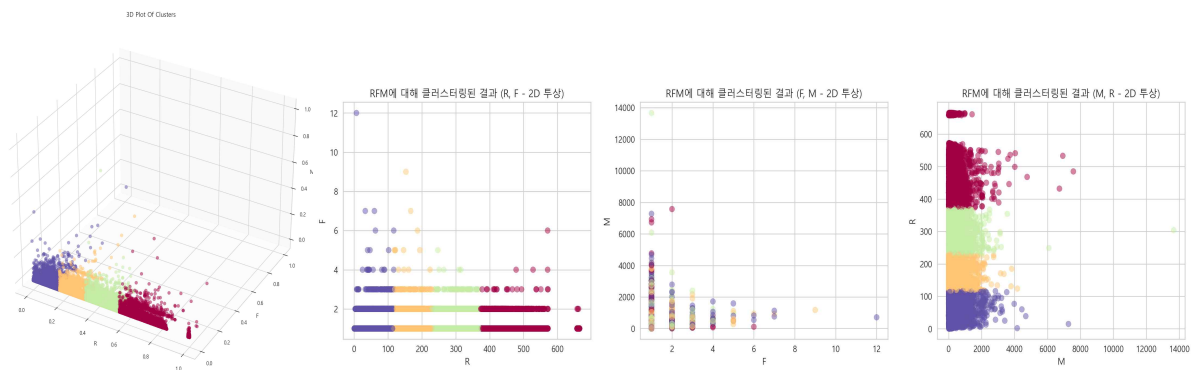
<그림 2-2. RFM 분포>



<그림 2-3. Elbow method>



<그림 2-4. 클러스터링 상태 확인>



<표2-1. 클러스터 그룹 별 R, F, M 평균 및 표준편차>

| 클러스터 그룹 | 평균 | | | 표준편차 | | |
|---------|-------------|-------------|-------------|-------------|-------------|-------------|
| | $R(x_{rk})$ | $F(x_{fk})$ | $M(x_{mk})$ | $R(s_{rk})$ | $F(s_{rk})$ | $M(s_{rk})$ |
| 0 | 457.554563 | 1.024398 | 161.090622 | 57.120831 | 0.169177 | 242.921027 |
| 1 | 66.547730 | 1.039249 | 168.657148 | 33.144510 | 0.239897 | 234.537221 |
| 2 | 174.495632 | 1.034733 | 155.744728 | 31.674353 | 0.240163 | 188.632863 |
| 3 | 293.873586 | 1.032282 | 163.026391 | 42.639385 | 0.191738 | 228.423436 |

<표2-2. 클러스터 그룹 별 R, F, M의 변동계수>

| 클러스터 그룹 | 변동 계수 | | |
|---------|--|--|--|
| | $w_1(CV_{rk} = \frac{s_{rk}}{x_{rk}})$ | $w_2(CV_{fk} = \frac{s_{fk}}{x_{fk}})$ | $w_3(CV_{mk} = \frac{s_{mk}}{x_{mk}})$ |
| 0 | 0.124839 | 0.165147 | 1.507977 |
| 1 | 0.498056 | 0.230837 | 1.390615 |
| 2 | 0.181519 | 0.202826 | 1.211167 |
| 3 | 0.145094 | 0.185742 | 1.401144 |

<표2-3. RFM 점수 중 R, F, M 가중치>

| 정규화 여부 | 가중치 | | |
|--|----------|----------|----------|
| | W_1 | W_2 | W_3 |
| 정규화되지 않음 $(w_n = \frac{\min(CV_{xn})}{CV_{x1} + CV_{x2} + \dots + CV_{xk}} (x: r, f, m))$ | 0.131478 | 0.210499 | 0.219776 |
| 정규화됨($W_n = \frac{w_n}{w_1 + w_2 + w_3}$) | 0.234049 | 0.374718 | 0.391233 |

<표2-4. RFM 점수 분포에 따른 고객 등급>

| 점수 분포(%) | ~ 50% | 50% ~ 20% | 20% ~ 5% | 5% ~ |
|----------|---------|-----------|----------|-----------|
| 경계 점수 | (0, 56) | (56, 69) | (69, 81) | (81, 100) |
| 등급 | Silver | Gold | Platinum | Diamond |

<표2-5. RFM 점수의 유의성 검증 결과>

| ANOVA 대상 | F-score | P-value |
|-------------|---------|---------|
| 고객등급과 거래 가치 | 5699 | 0.0 |
| 고객등급과 R | 7483 | 0.0 |
| 고객등급과 F | 1442 | 0.0 |
| 고객등급과 M | 8328 | 0.0 |

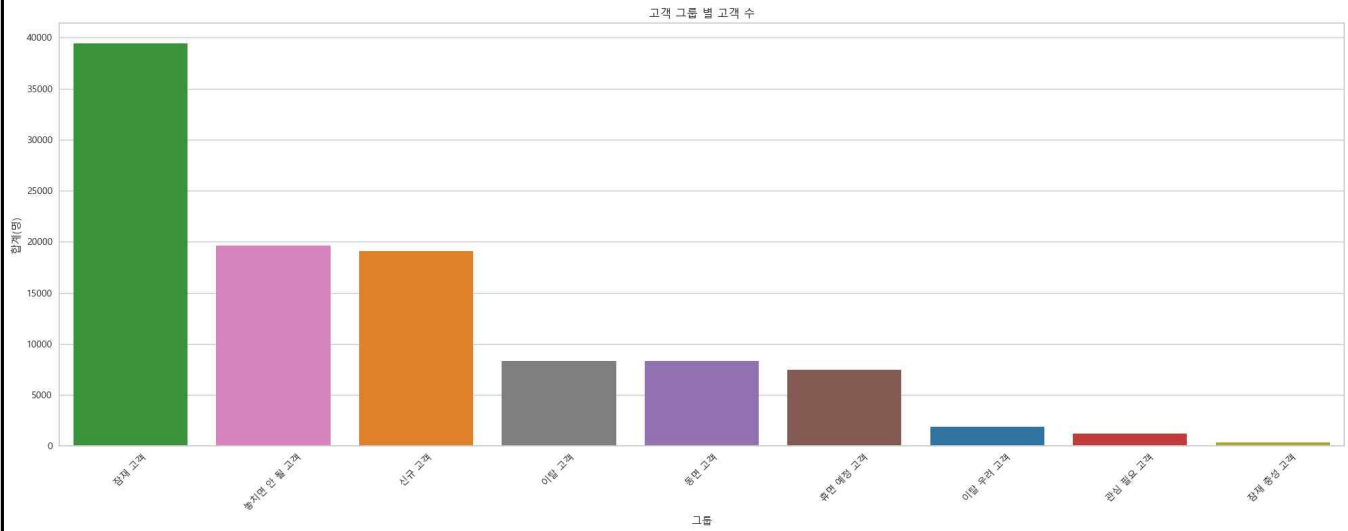
<표2-6. 등급에 따른 매출 비중 및1인당 평균 거래액>

| 등급 | 인원 수 | 매출 비중(%) | 1인당 평균 거래액 |
|----------|-------|----------|------------|
| Silver | 40239 | 17.36 | 68.10 |
| Gold | 29757 | 41.08 | 208.17 |
| Platinum | 10083 | 24.16 | 319.92 |
| Diamond | 5026 | 17.40 | 318.41 |

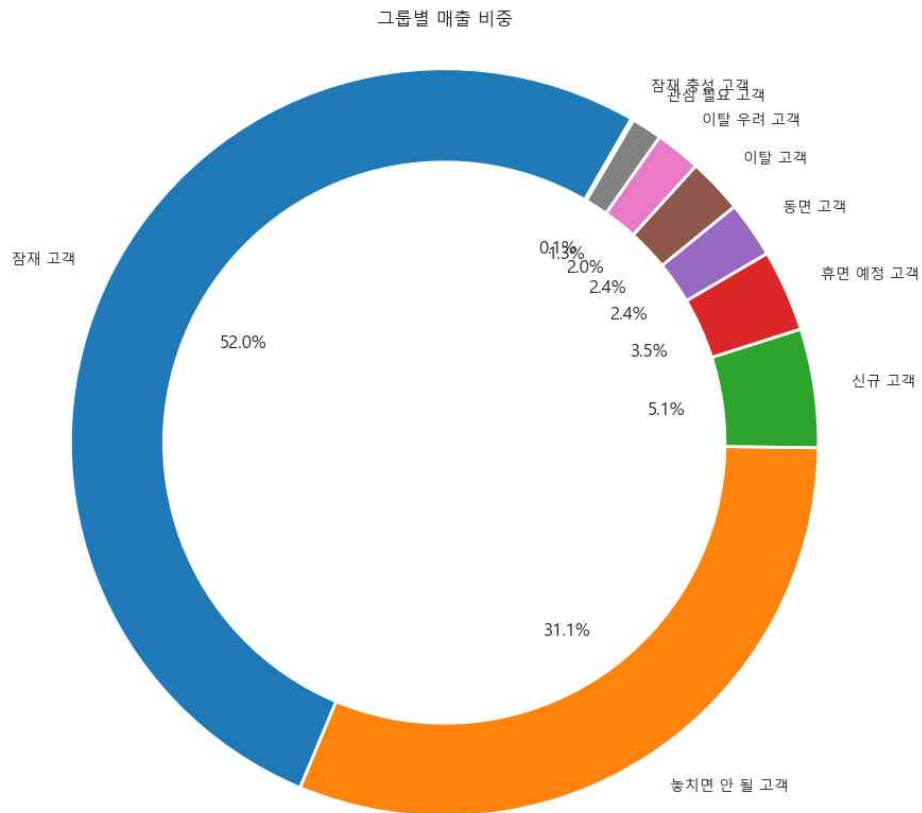
<표2-7. R, F, M 그룹 별 고객 등급>

| 고객 그룹 | 특징 | 범주화된 R, F, M |
|------------|---|---|
| 잠재 고객 | 최근에 거래하고 빈도가 낮으나 큰 금액을 소비 | '525', '524', '523', '515', '514', '513', '425', '424', '413', '414', '415', '315', '314', '313' |
| 놓치면 안 될 고객 | 최근에 거래하지 않았으나 큰 금액을 소비 | '214', '215', '115', '114', '113' |
| 신규 고객 | 최근에 유입되어 빈도가 낮으며 적은 금액을 소비 | '522', '521', '512', '511', '422', '421', '412', '411', '311' |
| 동면 고객 | 오래 전에 거래했고 빈도가 낮으며 적은 금액을 소비 | '322', '223', '222', '123', '122', '212', '211' |
| 이탈 고객 | 최근성, 빈도, 금액이 모두 가장 낮은 고객 | '111', '112', '121' |
| 휴면 예정 고객 | 최근성, 빈도, 금액이 모두 평균 이하인 고객 | '321', '312', '221', '213' |
| 이탈 우려 고객 | 최근에 거래하지 않았으나 빈도가 높고 많은 금액을 소비 | '225', '224', '125', '124' |
| 관심 필요 고객 | 최근성, 빈도, 금액이 모두 평균 이상인 고객 | '325', '324' |
| 잠재 충성 고객 | 최근에 거래했고 1번 이상 거래했으나 금액이 평균 이하인 고객 | '423', '323' |

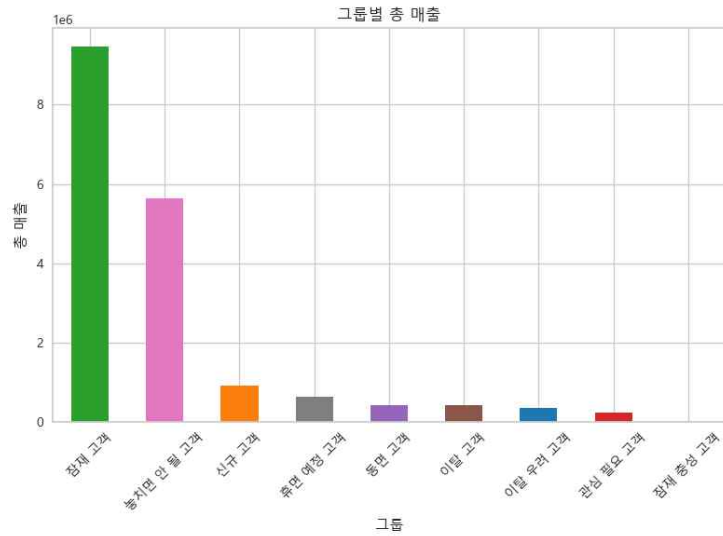
<그림 2-5. 그룹별 고객 수>



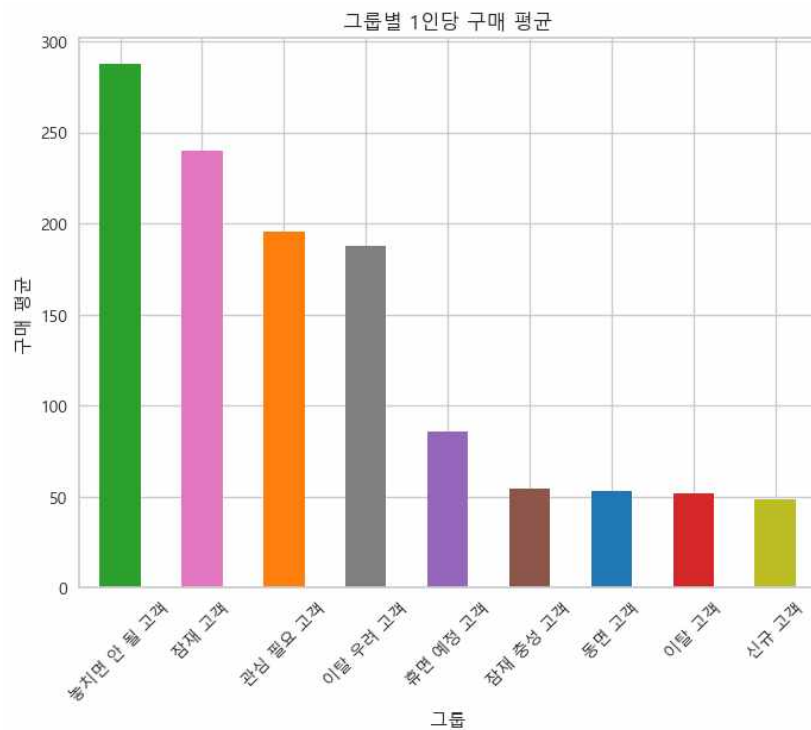
<그림 2-6. 그룹별 매출 비중>



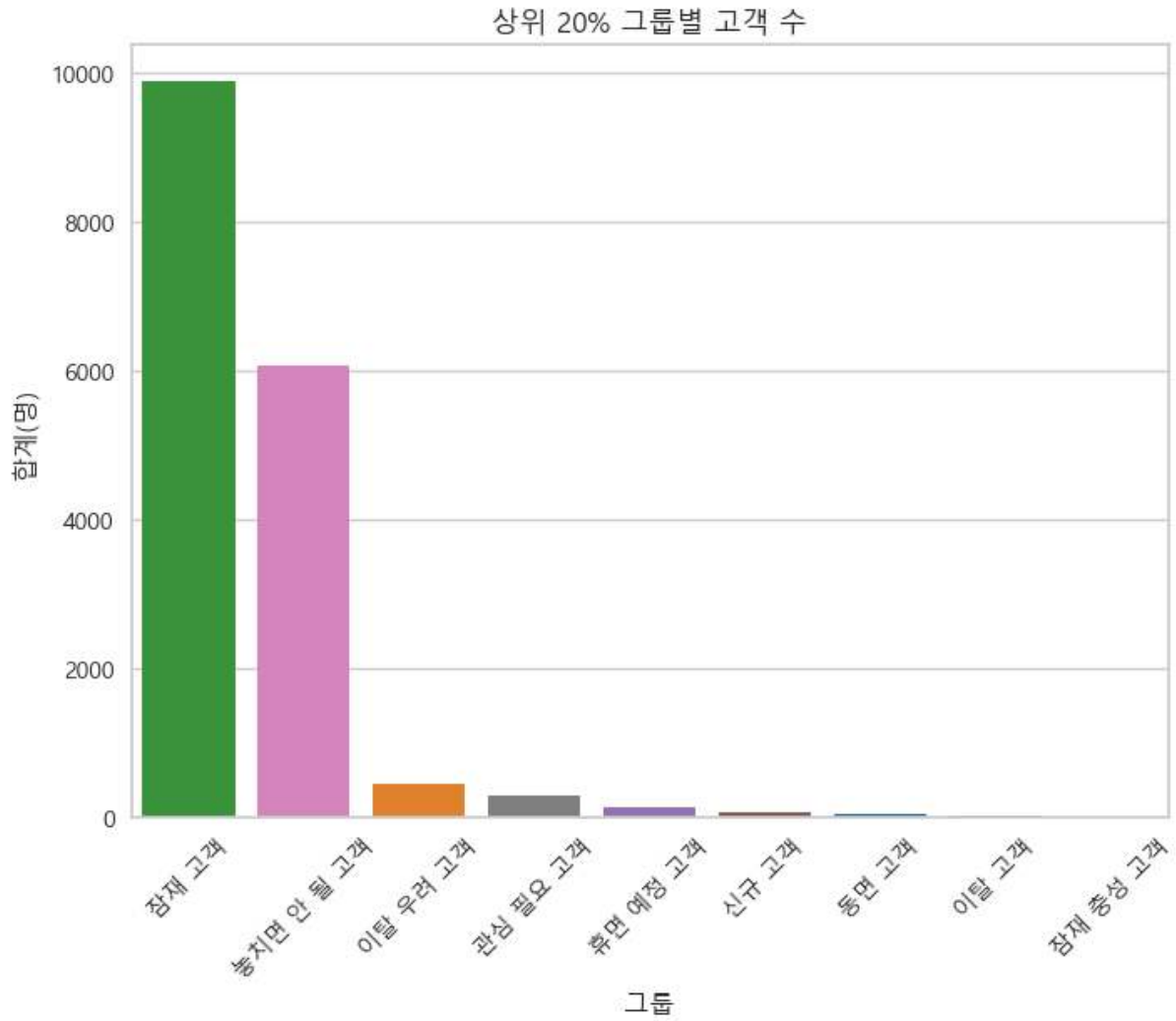
<그림 2-7. 그룹별 총 매출>



<그림 2-8. 그룹별 1인당 거래 평균>



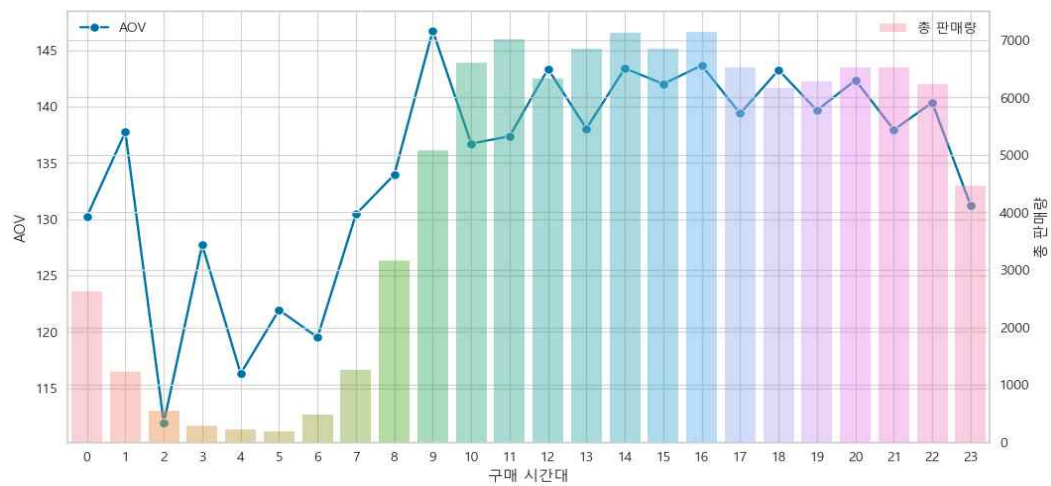
<그림 2-9. 상위20% 그룹별 고객 수>



<표2-8. 그룹별 매출 비중, 1인 거래 평균, 상위20% 고객 수>

| 고객 그룹 | 고객 수 | 매출 비중 | 1인 거래 평균 | 거래 가격 상위20% 고객 수 |
|---------------|-------|-------|----------|------------------------|
| 잠재 고객 | 39423 | 52.0% | 239.6 | 9879 |
| 놓치면 안 될 고객 | 19648 | 31.0% | 287.3 | 6083 |
| 신규 고객 | 19064 | 5.13% | 48.87 | 84 |
| 동면 고객 | 8351 | 2.43% | 53.14 | 56 |
| 이탈 고객 | 8327 | 2.38% | 51.86 | 38 |
| 휴면 예정 고객 | 7466 | 3.53% | 85.76 | 138 |
| 이탈 우려 고객 | 1910 | 1.98% | 188.0 | 455 |
| 관심 필요 고객 | 1231 | 1.32% | 195.2 | 292 |
| 잠재 충성 고객 | 339 | 0.1% | 54.76 | 4 |

<그림3-1. 시간대별 AOV 및 총판매량>



<그림3-2. Power BI>

3번째 KPI : 시간대별 카테고리 AOV(평균 주문 가치)

시간대별 AOV 및 총판매량

● AOV ● 총판매량



선택 시간대로 이동



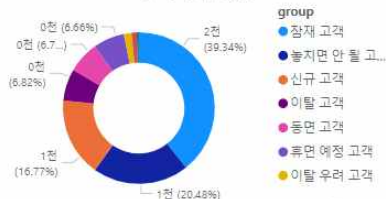
<그림3-3. 각 시간대의 카테고리별 AOV 및 총판매량, 고객 등급 비율, 고객 지역별 AOV, 고객 주별 AOV >

← | 선택한 시간대 : 9시

카테고리 별 AOV 및 총 판매량



고객 등급 비율



group
● 잠재 고객
● 불치면 안 될 고...
● 신규 고객
● 이탈 고객
● 동면 고객
● 휴면 예정 고객
● 이탈 우려 고객

고객 지역별 AOV



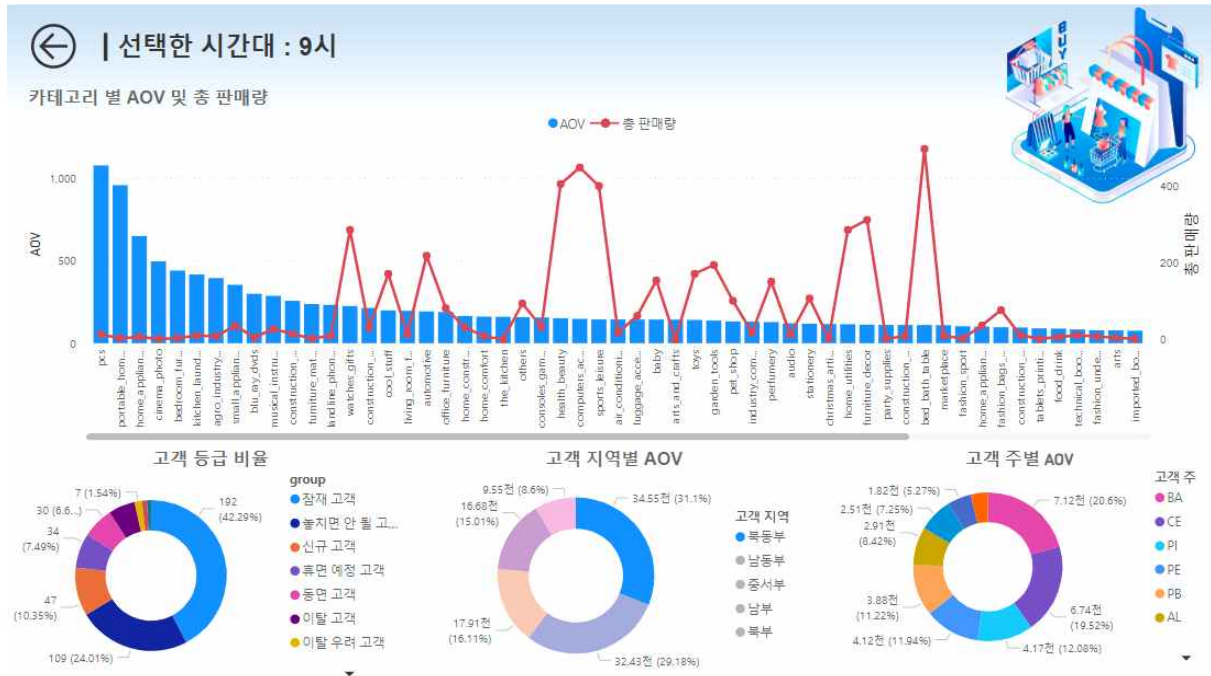
고객 지역
● 북동부
● 남동부
● 북서부
● 남서부
● 북부
● 남부

고객 주별 AOV

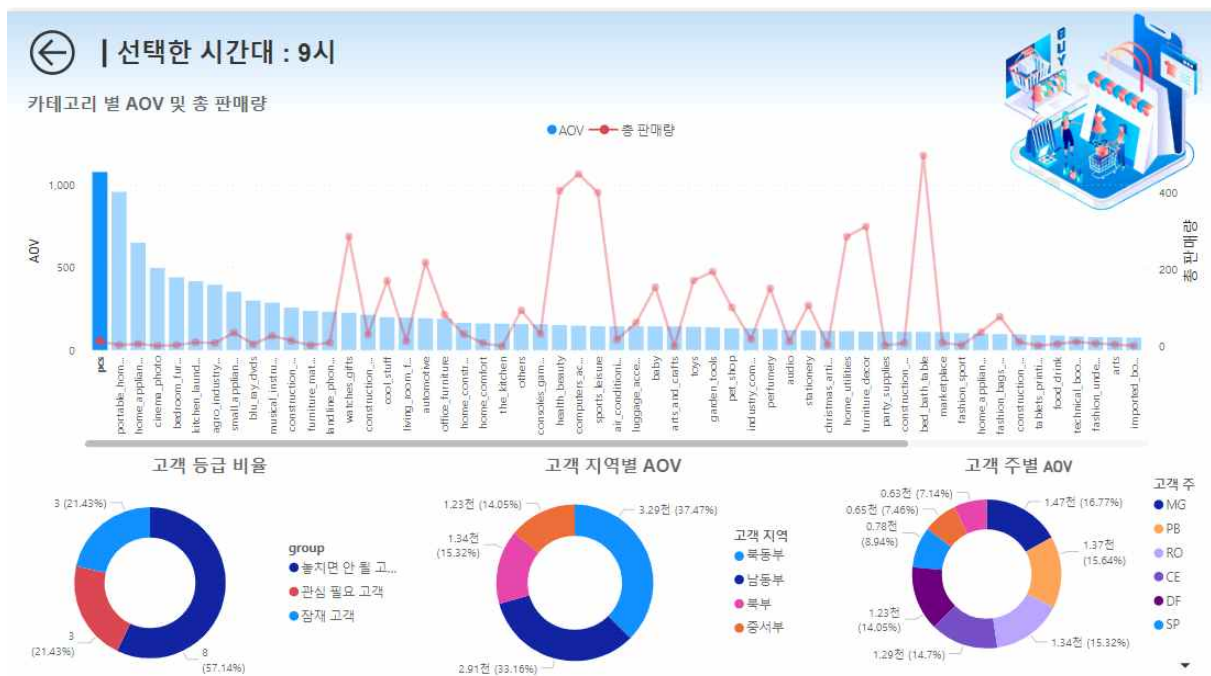


고객 주
● SP
● MG
● RJ
● DF
● BA
● CE

<그림3-4. 해당 시간대의 해당 고객 지역의 고객 등급 비율, 고객 주별 AOV>



<그림3-5. 해당 카테고리의 고객 등급 비율, 고객 지역별 AOV, 고객 주별 AOV>



<그림3-6. 선택한 카테고리의 해당 지역의 고객 등급 비율, 고객 주별 AOV>

