



Python 활용 EDA수행 프로젝트

글로벌 이커머스 최적화: 고객 행동 분석 을 통한 수익 증대 전략

프로그래머스 데이터 분석 데브코스

2팀 1조

김시현 서승범 선영훈 이태경 정정룡

목차

1. 서론.....	1
1.1 주제 선정 이유 및 분석 목적.....	1
1.2 활용 기술 및 프레임 워크.....	1
1.2.2 NumPy.....	2
1.2.3 Matplotlib.....	2
1.2.4 Seaborn.....	2
1.2.5 Plotly.....	2
1.3 참여자 정보 및 역할.....	3
2. 데이터 및 문제 정의.....	3
2.1 데이터 설명.....	3
2.2 전처리.....	5
2.2.1 중복열 삭제.....	5
2.2.2 필요없는 열 삭제.....	5
2.2.3 특정 시간 이후 데이터 삭제.....	5
2.3 문제 정의 및 KPI 설정.....	9
2.3.1 문제 정의.....	9
2.3.2 KPI 설정.....	10
3. 심층 분석 및 결과.....	10
3.1 할인율 및 취소율 분석.....	10
3.1.1 최적 할인율 분석.....	10
3.1.2 고수익 제품 집종을 통한 수익 향상.....	13
3.1.3 지연 배송에 따른 주문 취소율 분석.....	14
3.2 재구매율 분석.....	17
3.3 지역별 판매 분석.....	21
3.4 사기 의심 정황 분석.....	26

3.5 RFM 분석.....	29
3.5.1 Kmeans 를 활용한 RFM 고객 군집화.....	30
3.5.2 RFM 고객 군집 해석 및 시각화.....	32
3.5.3 RFM 군집화를 활용한 고객 행동 시계열 분석 및 CLTV.....	39
4. 결론.....	45
4.1 분석 요약 및 제한점.....	45
4.2 마케팅 전략 수립.....	46

1. 서론

1.1 주제 선정 이유 및 분석 목적

본 프로젝트는 공급망 최적화를 통해 기업의 운영 효율성과 수익성을 향상시키는 데 중점을 두고 있다. 공급망 관리는 원자재의 조달부터 최종 제품의 소비자에 이르기까지 전체적인 제품 흐름과 관련 정보, 자금의 흐름을 효율적으로 관리하는 시스템을 말한다. 이는 제품의 제조 및 배송 과정에서 발생할 수 있는 다양한 문제를 식별하고 해결함으로써, 최종 제품의 가치를 극대화하고 고객 만족을 보장하는 데 필수적인 요소이다.

특히 실제 배송일과 예정 배송일, 매출뿐만 아니라 실제 수익을 나타내는 데이터와 배송에 관련된 상세한 변수들, 할인율 등을 포함한 데이터를 활용한다. 이 데이터를 기반으로 다양한 분석을 수행하여 배송 정확성, 할인의 효과 등이 기업의 수익성에 미치는 영향을 깊이 있게 파악하고자 한다.

본 프로젝트의 분석 방법 중 하나인 RFM(Recency, Frequency, Monetary) 모델을 사용하여 고객을 세분화하고, 이를 통해 얻은 고객 군집 정보를 활용하여 고객의 행동 패턴과 장기 가치(CLTV: Customer Lifetime Value)를 분석한다. 이러한 분석은 공급망 관리와 직접적으로 연관되어 있으며, 기업은 가장 가치 있는 고객을 식별하고, 이들에게 맞춤형 서비스와 제안을 제공하여 고객 충성도와 장기적인 수익성을 높일 수 있다.

이러한 분석은 공급망의 효율성을 극대화하고, 고객 만족도를 높이며, 기업의 장기적인 수익성을 보장하는 데 중요한 역할을 할 것이다. 본 연구는 공급망의 복잡성을 관리하고, 전략적 의사결정을 지원하는 실질적인 방안을 제공하고자 한다.

1.2 활용 기술 및 프레임 워크

본 분석은 Python 프로그래밍 언어를 기반으로 수행되었다. Python의 다양한 라이브러리를 활용하여 데이터의 조작, 분석 및 시각화를 포괄적으로 수행하였다. 본 프로젝트에서 사용된 주요 라이브러리는 다음과 같다

1.2.1 Pandas

데이터 조작 및 분석에 필수적인 라이브러리로, 고성능 데이터 구조인 DataFrame과 Series를 제

공한다. 이는 데이터의 로딩, 결측치 처리, 특성 선택 및 변환, 데이터 그룹화 및 집계 등 복잡한 데이터 작업을 용이하게 한다. 본 프로젝트에서는 Pandas를 활용하여 원시 데이터를 로드하고, 필요한 데이터 변환을 수행하였다.

1.2.2 NumPy

수치 계산을 위한 라이브러리로, 고성능의 다차원 배열 객체와 이를 다루는 도구를 제공한다. NumPy는 복잡한 수치 계산을 효율적으로 처리할 수 있게 하며, 특히 대규모 데이터 집합에 대한 빠른 배열 연산을 지원한다. 본 프로젝트에서는 데이터 배열을 생성하고 조작하는 데 NumPy를 사용하였다.

1.2.3 Matplotlib

데이터 시각화를 위한 기본적인 라이브러리로, 다양한 그래프와 차트를 생성할 수 있다. 본 프로젝트에서는 Matplotlib를 사용하여 데이터의 분포와 패턴을 시각적으로 표현함으로써 초기 데이터 분석 단계에서 유용한 인사이트를 제공하였다.

1.2.4 Seaborn

Matplotlib을 기반으로 하는 시각화 라이브러리로, 보다 복잡한 시각화를 보다 쉽게 구현할 수 있다. 본 프로젝트에서는 Seaborn을 사용하여 고급 데이터 시각화를 수행하였고, 특히 복잡한 구조의 데이터를 효과적으로 시각화하여 분석 결과를 명확하게 전달하였다.

1.2.5 Plotly

인터랙티브한 그래프와 시각화를 생성할 수 있는 라이브러리로, 사용자에게 데이터 탐색의 향상된 경험을 제공한다. 본 프로젝트에서는 Plotly를 활용하여 동적인 차트와 대시보드를 제공함으로써, 분석 결과를 보다 직관적으로 탐색할 수 있게 하였다.

1.3 참여자 정보 및 역할

이름	역할
이태경	재구매율 분석, ppt작성
서승범	지역별 판매 분석, ppt작성
선영훈	할인률 및 취소율 분석, 발표
김시현	사기 의심 주문 분석, 보고서 작성
정정룡	전처리, RFM 분석, 보고서 작성

2. 데이터 및 문제 정의

2.1 데이터 설명

사용한 데이터는 상기 설명한 것처럼 공급망과 관련된 Kaggle에서 제공되는 데이터 셋이다.¹ 원본 데이터 셋에서는 총 53개의 열과 180519개의 행으로 구성되어 있다. 주문 시 실제 매출, 거래방법, 예정 배송일, 배송 상황, 주문 날짜 등 상세한 열들로 구성되어 있다. 다만 마스킹 처리 및 중복 열 등 분석에 왜곡을 줄 수 있는 문제가 발견되었다. 이는 전처리에서 자세히 설명한다.

원본 데이터의 간단한 정보를 확인한 후 전처리 이후에 상세 열을 기술한다.

표 1 원본 데이터 info

Number of columns	53
Number of rows	180519
Total number of data-points	9567507
Count of Numerical Features	29
Count of Categorical Features	24

결측은 전체 데이터 셋에서 3.51% 이며 4가지 열에서 발생한다. 추후 전처리가 진행될 시 해당

¹ https://www.kaggle.com/datasets/shashwatwork/dataco-smart-supply-chain-for-big-data-analysis?select=tokenized_access_logs.csv

결측은 사라지므로 결측치 처리는 따로 진행하지 않는다.

그림 1 원본 데이터 Heatmap

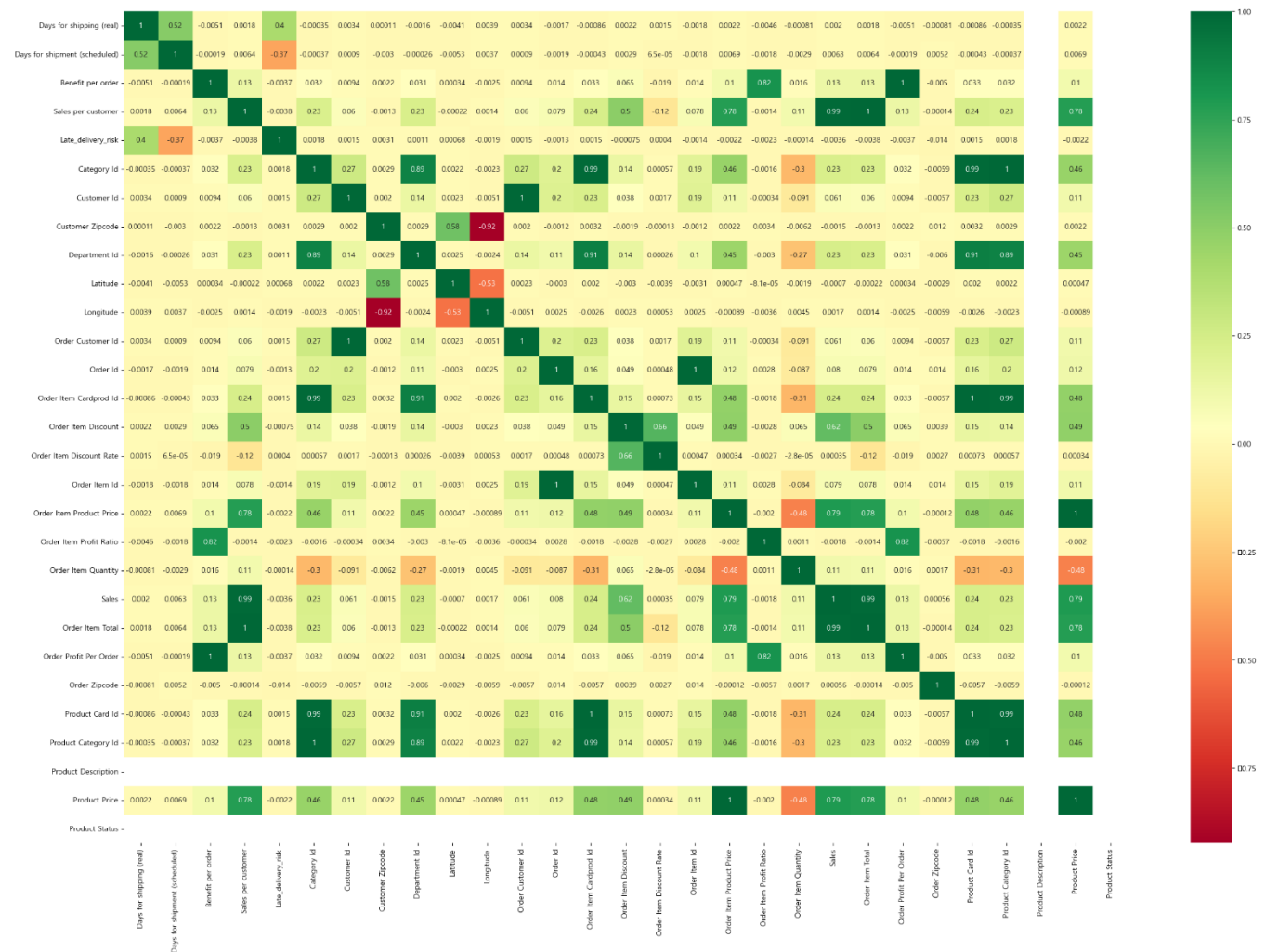


그림 1을 살펴보면 다른 열임에도 상관관계수가 매우 높은 값을 가지는 것을 확인할 수 있다. 이는 중복열로 판단될 수 있으며 자세히 살펴본 결과 중복열을 찾을 수 있었다.

2.2 전처리

2.2.1 중복열 삭제

표 2

사용하는 열	중복열
Order Profit Per Order	Benefit per order
Sales per customer	Order Item Total
Category Id	Department Id, Order Item Cardprod Id, Product Category Id, Product Card Id
Customer Id	Order Customer Id
Product Price	Order Item Product Price
Order Id	Order Item Id

표 2에서 오른쪽은 왼쪽에 해당하는 열과 중복되는 열을 나타낸다. 효율성과 분석에 혼란을 방지하기 위해 중복열은 삭제 후 진행했다.

2.2.2 필요없는 열 삭제

Product Description, Order Zipcode, Product Image, Customer Email 등 마스킹 처리되었거나 분석에 사용하지 않는 열 총 8개를 삭제 후 진행했다.

중복열 및 필요없는 열 삭제로 결측치는 처리되며 남은 총 37개의 열로 분석을 진행했다.

2.2.3 특정 시간 이후 데이터 삭제

해당 전처리는 RFM 분석을 진행하다 문제를 발견되어 전처리가 진행되었다. 전처리 과정을 간략하게 소개한다.

RFM 테이블, 즉 고객별 최근 구매 날짜, 구매 빈도, 구매 총액을 구한 후 테이블을 살펴본 결과 아래쪽에 해당하는 Customer Id의 구매빈도가 전부 1이어서 확인을 진행했다. 확인 결과 Customer Id 가 12435부터 전부 1인 것을 확인했고, Id를 할당할 때 날짜 순으로 했을 가능성을 내포한다고 판단했다. 이 후 연도별 구매횟수 등 자세하게 확인한 결과 표 3에 해당하는 사실을 발견했다.

표 3

Customer Id 1 ~ 12435	2015~2017년 다양한 구매 분포, 2018년 구매 이력 없음
Customer Id 12436 ~ 18634	2018년 단 한 번의 구매이력 존재
Customer Id 18635 ~	2017년 단 한 번의 구매이력 존재

표 3을 확인함으로써 Customer Id가 날짜순으로 할당되었음을 확인했다. 이는 2017년부터 기존 고객의 Id를 재할당했을 가능성 혹은 데이터 오류로 판단된다. 이에 Customer Id가 12436을 기준으로 자세히 확인했다. 해당 고객의 구매 날짜인 2017-10-02 12:46:00를 기준으로 데이터를 분리 후 해당 기점 이전 고객이 이후 고객에 포함되었는지 확인하여 Id를 재할당했는지에 대한 가설을 확인했다. 결과적으로 재할당하지 않음을 확인했으며 이는 해당 시간대 이후 기록된 모든 고객은 이전 고객과 동일한 경우는 없는 것으로 판단되었다.

이에 해당 기점 이후의 데이터는 이전 고객의 거래 기록과 연관 지을 수 없으므로 제거 후 분석을 진행했다.

전처리 이후 간단한 데이터 정보는 아래와 같다.

표 4 전처리 이후 Data info

Number of columns	37
Number of rows	172199
Total number of data-points	6371363
Count of Numerical Features	19
Count of Categorical Features	18

그림 2 전처리 이 후 Heatmap

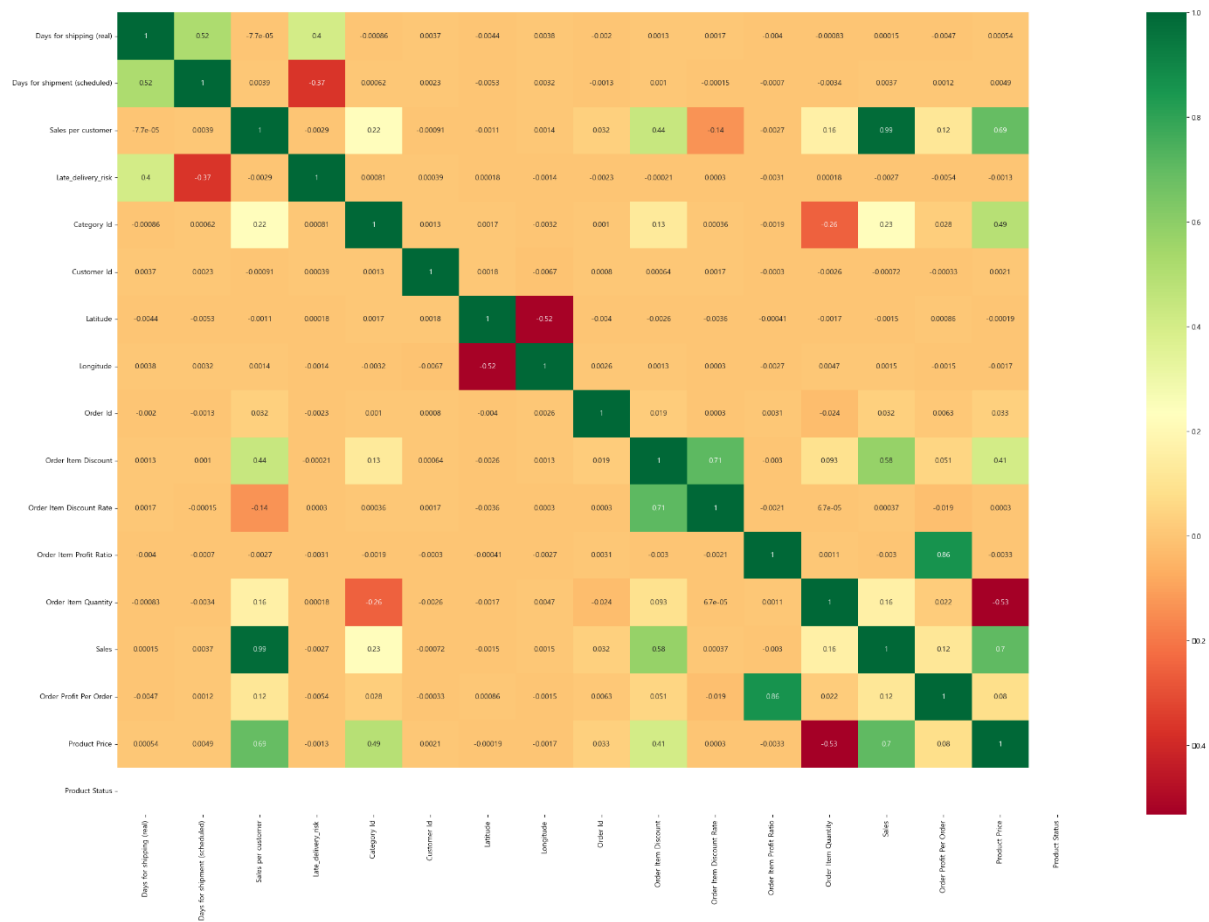


그림 2를 살펴보면 어느 정도 상관관계수가 높은 값을 가지는 열이 존재하나 중복열은 아닌 것으로 판단되었다. 이를 설명하는 열 계산식과 사용된 열을 아래에 기술한다.

Sales = Product Price * Order Item Quantity

Order Item Discount = Sales * Order Item Discount Rate

Sales per customer = Sales - Order Item Discount

Order Profit Per Order = Sales per customer * Order Item Profit Ratio

표 5

열 명	열 설명
Type	거래 방법
Days for shipping (real)	실제 배송 날짜
Days for shipment (scheduled)	예정 배송 날짜
Sales per customer	할인 적용 후 총 가격
Delivery Status	배송 상태
Late_delivery_risk	배송 지연 위험성(0, 1로 구성)
Category Id	상품 카테고리 ID
Category Name	상품 카테고리 이름
Customer City	고객 구매가 진행된 매장 도시
Customer Country	고객 구매가 진행된 매장 국가
Customer Fname	First Name
Customer Lname	Last Name
Customer Id	고객 ID
Customer Segment	고객 분류(Consumer, Corporate, Home Office)
Customer State	고객 구매가 진행된 매장 주
Department Name	매장 이름
Market	주문이 배송되는 시장(Africa, Europe, LATAM, Pacific Asia, USCA)
Order City	배송되는 도시
Order Country	배송되는 국가
order date (DateOrders)	주문 날짜
Order Id	배송 ID
Order Item Discount	할인 가격
Order Item Discount Rate	할인율
Order Item Profit Ratio	주문 당 실제 수익 비율
Order Item Quantity	주문 수량
Sales	할인 적용 전 총 가격
Order Profit Per Order	주문 당 실제 수익
Order Region	배송되는 지역
Order State	배송되는 주
Order Status	배송 상태(COMPLETE, PENDING, CLOSED, PENDING_PAYMENT, CANCELED, PROCESSING, SUSPECTED_FRAUD, ON_HOLD, PAYMENT_REVIEW)
Product Name	상품 이름
Product Price	상품 가격

Product Status	재고 상태(0: 재고있음, 1: 재고없음)
shipping date (DateOrders)	배송된 날짜
Shipping Mode	배송 모드(Standard Class, First Class, Second Class, Same Day)

2.3 문제 정의 및 KPI 설정

2.3.1 문제 정의

본 프로젝트에서는 단순히 높은 판매량이 수익성과 동일하다고 가정하지 않는다. 이는 제품의 판매량이 높더라도 해당 제품의 마진이 낮거나 기타 비용이 높을 경우 실질적인 수익 기여도가 낮을 수 있기 때문이다. 따라서 본 연구에서는 판매량 대비 높은 수익을 창출하는 요인을 식별하고, 이를 기반으로 수익 극대화 전략을 수립하고자 한다.

고객 만족도 역시 중요한 요소로 고려된다. 고객 만족도는 재구매율과 직접적인 관련이 있으며, 이는 장기적인 고객 관계 유지에 핵심적인 역할을 한다. 만족도가 낮은 제품군에 대한 개선이 필요하며, 이는 주문 취소율과 재구매율의 분석을 통해 이루어진다.

지역별로 매출 및 카테고리 분석은 구체적인 마케팅 전략에 효과적이다. 각 지역마다 중요한 문화적 특성이 존재하기 때문에 계절성 및 지역별 수요가 좋은 상품을 분석할 수 있다. 이를 통해 세부적인 제품 프로모션을 수립하고 효율적인 수익 창출에 도움이 될 수 있다.

또한, 이커머스 사업에서의 배송 효율성도 중요한 고려 사항이다. 제품의 배송 속도 및 정확성은 고객 만족도에 큰 영향을 미치며, 이는 고객 충성도와 직결된다. 배송 인프라와 로지스틱 시스템의 최적화는 기업의 경쟁력을 높이는 결정적 요소이다.

RFM (Recency, Frequency, Monetary) 분석을 통한 고객 세분화 및 CLTV (Customer Lifetime Value) 계산을 추가하여 고객 가치의 장기적 관리를 강화하는 것도 본 분석의 주요 목표 중 하나이다. RFM 분석은 고객의 최근 구매 활동, 구매 빈도, 구매 총액을 기반으로 고객을 여러 군집으로 분류하고, 각 군집의 특성에 맞는 마케팅 전략을 수립한다. CLTV 모델을 통해 고객의 미래 가치를 예측하고, 이를 기반으로 고객 관리 및 유치 전략을 최적화하는 것이 중요하다. 이는 고객 행동의 시계열 분석을 통해 지속적으로 모니터링되며, 고객 관계를 깊이 있게 관리할 수 있는 전략을 수립하는 데 도움을 준다.

2.3.2 KPI 설정

수익 최적화 KPI: 판매량 대비 수익이 높은 요인과 구간을 찾고 이를 활용하여 수익을 극대화할 전략을 구상한다.

고객 만족도 KPI: 제품의 직접적인 수익성 외에도 고객 만족도는 장기적인 수익성에 중요한 영향을 미친다. 만족도는 주문 취소율 및 재구매율로 측정하며, 낮은 취소율과 높은 재구매율을 달성하기 위한 전략을 모색한다.

배송 효율성 KPI: 배송 속도 및 효율은 고객 만족도에 직접적인 영향을 미친다. 따라서 지역별 및 배송 방법별 배송 효율성을 분석한다. 이를 통해 배송 인프라에 대한 투자 우선순위를 결정하고, 최적의 배송 네트워크 구축을 위한 전략을 제안한다.

RFM 분석 KPI: 고객 세분화를 위해 RFM (최근 구매 일자, 구매 빈도, 구매 금액) 모델을 활용한다. 고객을 여러 군집으로 분류하여 각 군집의 특성을 파악하고, 맞춤형 마케팅 전략을 개발하는 것이 목표이다. 특히, 고가치 고객의 유지 및 가치 증대 전략을 수립하는 데 중점을 둔다.

고객 평생 가치 (CLTV) KPI: 고객의 평생 가치를 계산하여, 장기적인 고객 관계 관리 및 마케팅 전략의 효율성을 높인다. CLTV 모델을 통해 고객의 미래 가치를 예측하고, 이를 기반으로 고객 관리 및 유치 전략을 최적화한다. 이 과정에서 고객 행동 시계열 분석을 통해 고객의 구매 패턴과 변화를 살펴본다.

3. 심층 분석 및 결과

3.1 할인율 및 취소율 분석

3.1.1 최적 할인율 분석

할인은 고객에게 상품을 저렴하게 구매할 기회를 제공하고, 기업에게는 판매량을 증가시키는 효과가 있다. 그러나 할인율이 너무 높거나 낮으면 고객이나 기업 중 한쪽이 만족하지 못하는 상황이 발생할 수 있다. 따라서 양측 모두에게 이득이 되는 최적의 할인율을 찾아 제안하고자 한다.

최적 할인율을 분석하기 전에 데이터의 정확성을 확보하기 위해 이상치를 제거하는 전처리 작업을 수행했다. 이 과정에서 'Order Profit Per Order' 컬럼을 기준으로 IQR(Interquartile Range) 방식을 적용해, 25분위수와 75분위수 사이에 있지 않은 값들을 제거했고, 이로 인해 총 18,032개의 행이 제거되었다.

그림 3. 전처리 수행 전 'Order Profit Per Order'의 히스토그램

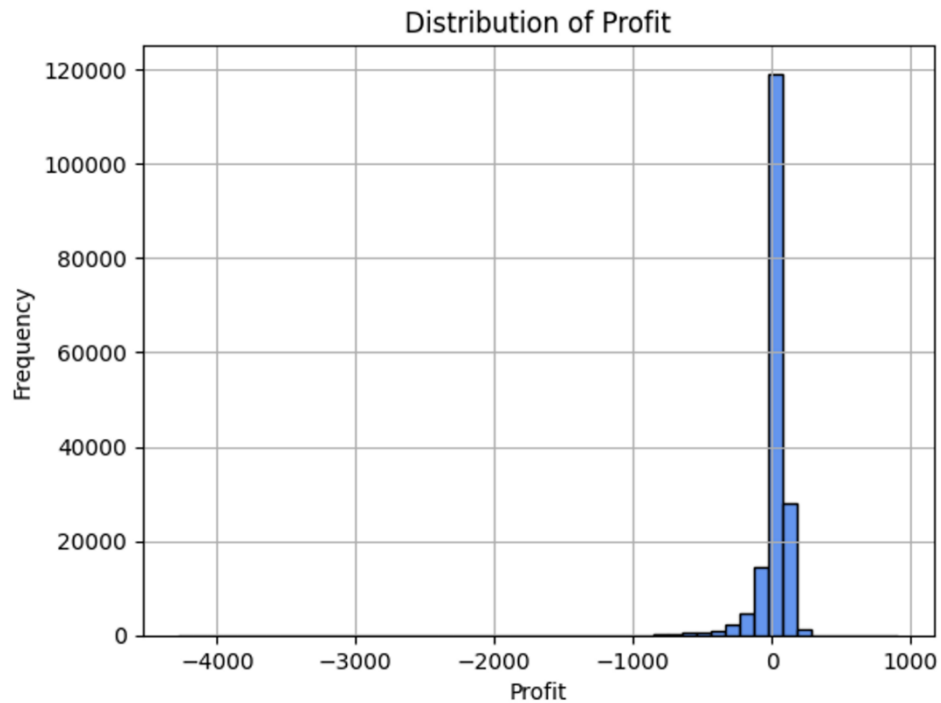


그림 4. 전처리 수행 후의 'Order Profit Per Order' 히스토그램

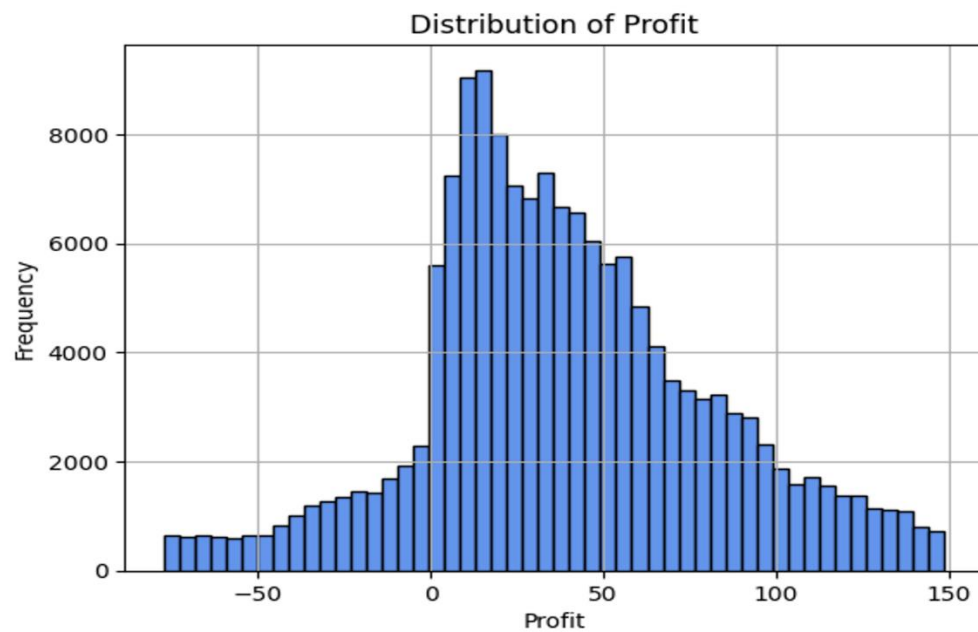
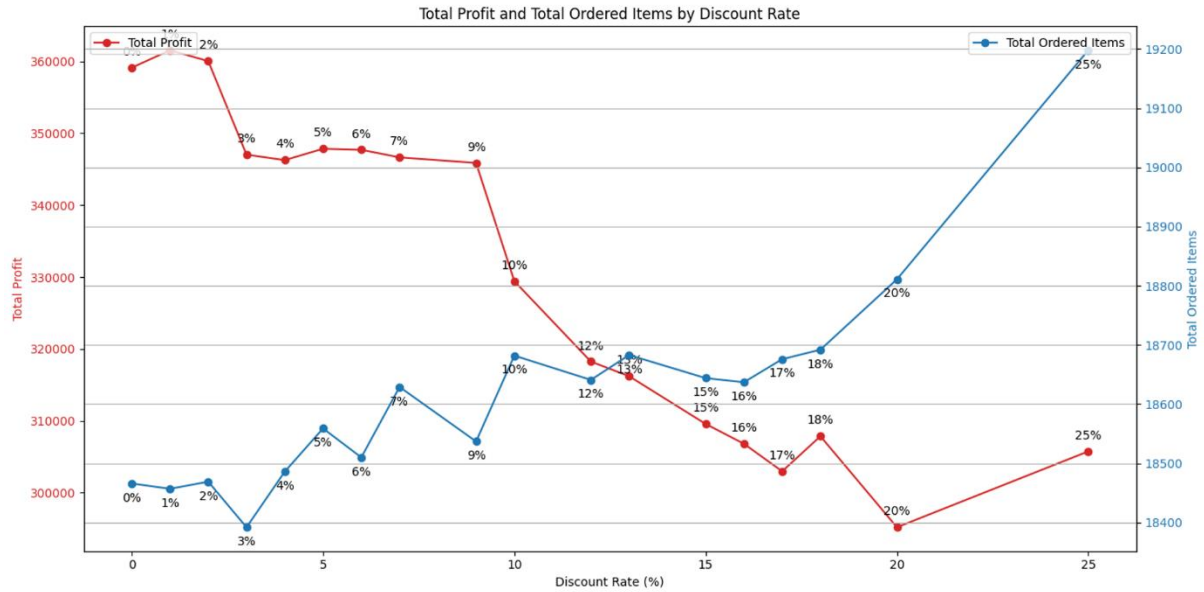
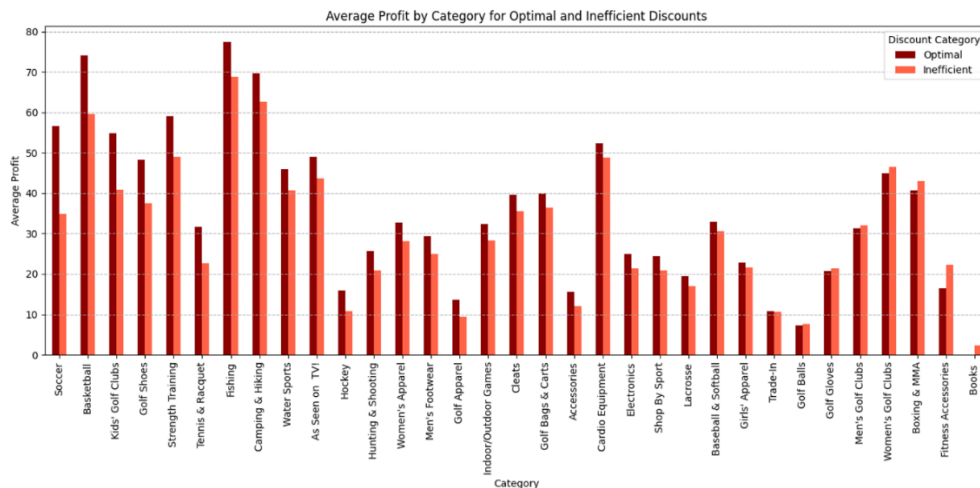


그림 5. 할인율에 따른 전체 이익과(빨간 그래프), 할인율에 따른 품목 주문수(파란색 그래프)



그래프 분석 결과, 할인이 반드시 이익 증대로 이어지지 않는다는 결론을 도출할 수 있다. 할인율이 증가함에 따라 주문 수는 증가하지만 이익은 오히려 감소하는 경향을 보인다. 그러나 일정 구간에서 할인율이 높아져도 주문 수 증가에 비해 이익 감소가 크지 않은 효율적인 '최적 할인율' 구간이 존재한다. 이 구간은 3%에서 9% 사이로 나타났다. 이 최적 할인율 구간을 바탕으로, 특히 이 할인율이 효과적인 상품 카테고리를 파악했다.

그림 6. 최적 할인율과 비효율적 할인율을 적용했을 때의 카테고리별 평균 이익

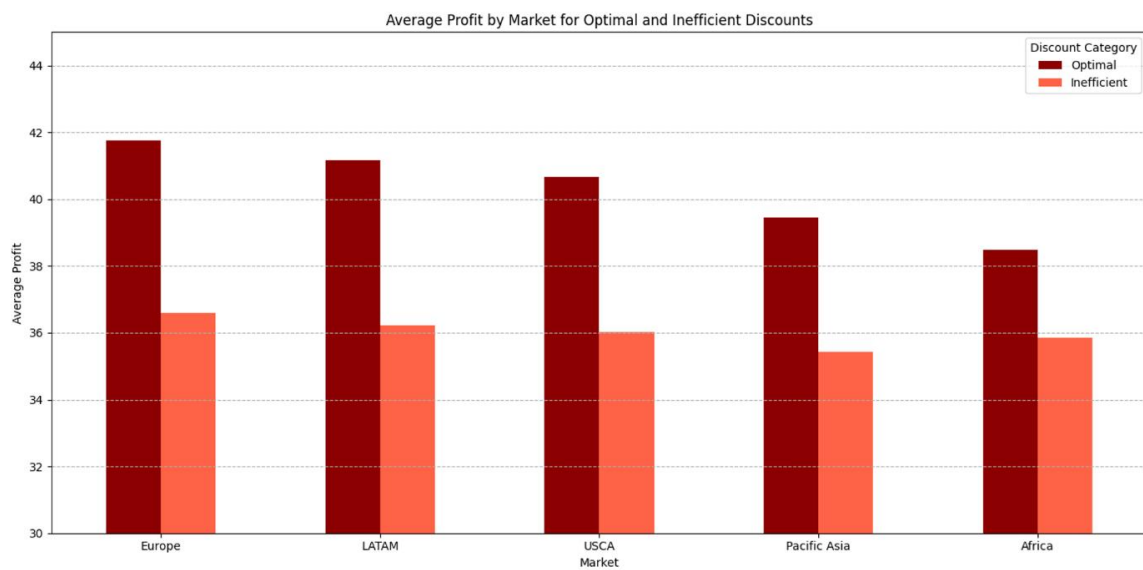


그래프를 보면, 빨간 그래프는 최적 할인율을 적용했을 때의 평균 이익을 나타내며, 주황 그래프는 비효율적 할인율을 적용했을 때의 결과를 보여준다. 이를 통해 최적 할인율을 적용했을 때 가장 이익 증가가 두드러진 상품 카테고리는 Soccer, Basketball, Kid's Golf Clubs, Strength Training,

Tennis & Racquet 순으로 확인되었다. 특히 스포츠 용품 카테고리에서 이 할인율의 효과가 높게 나타나는 경향을 보인다.

이어서, 그림 7은 최적 할인율이 각 대륙에서 어떤 효과를 나타내는지 시각화를 진행한 그래프이다.

그림 7. 최적 할인율과 비효율적 할인율을 적용했을 때의 대륙별 평균 이익

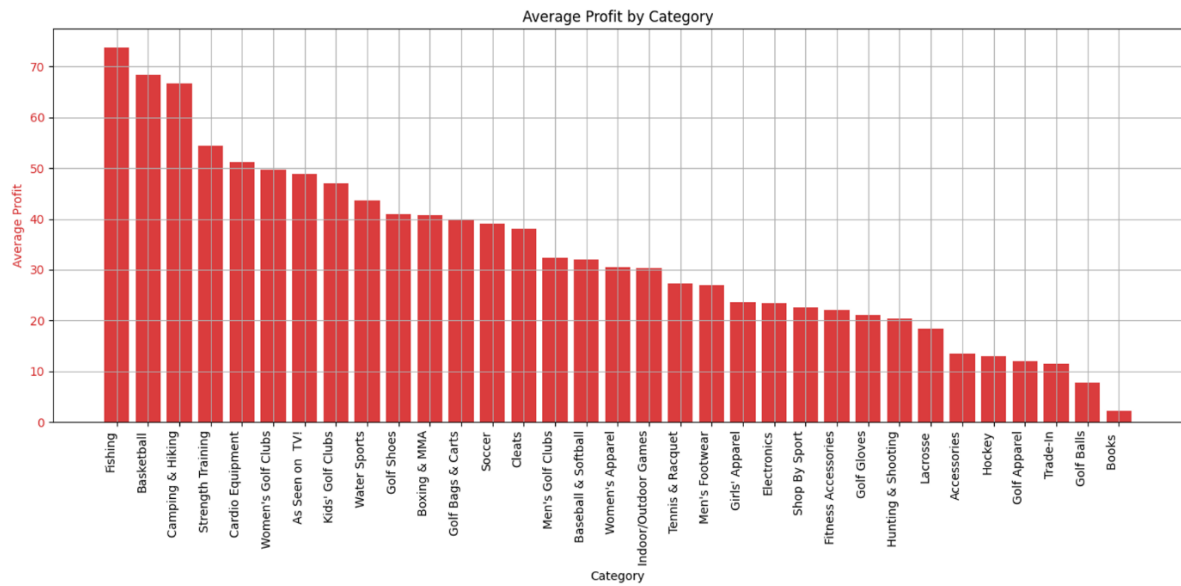


거의 같은 양상을 보이고 있으나, 이 중 가장 평균 이익이 높은 대륙은 유럽으로 나타난다.

3.1.2 고수익 제품 집종을 통한 수익 향상

제품 판매 시 어떤 카테고리에서 수익성이 높은지 확인하기 위해, 각 카테고리별 평균 수익률을 나타내는 그래프를 생성하였다.

그림 8. 카테고리 별 평균 이익 그래프



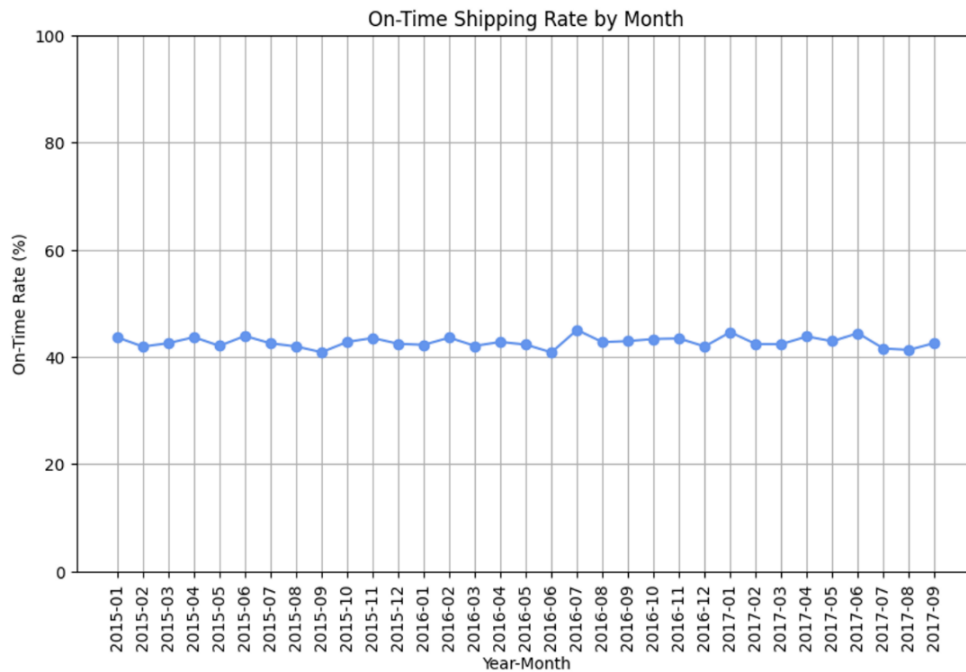
고수익을 창출하는 TOP5 카테고리인 Fishing, Basketball, Camping & Hiking, Strength Training, Cardio Equipment가 확인되었다.

3.1.3 지연 배송에 따른 주문 취소율 분석

이커머스 환경에서 주문 후 배송 지연은 흔히 발생할 수 있으며, 이로 인한 주문 취소율에 미치는 영향을 분석하고자 한다. 이 분석을 통해 주문 취소율을 줄일 수 있는 방안을 찾고자 한다.

먼저, 정시에 도착하는 배송의 비율을 그래프로 시각화하여 나타내었다.

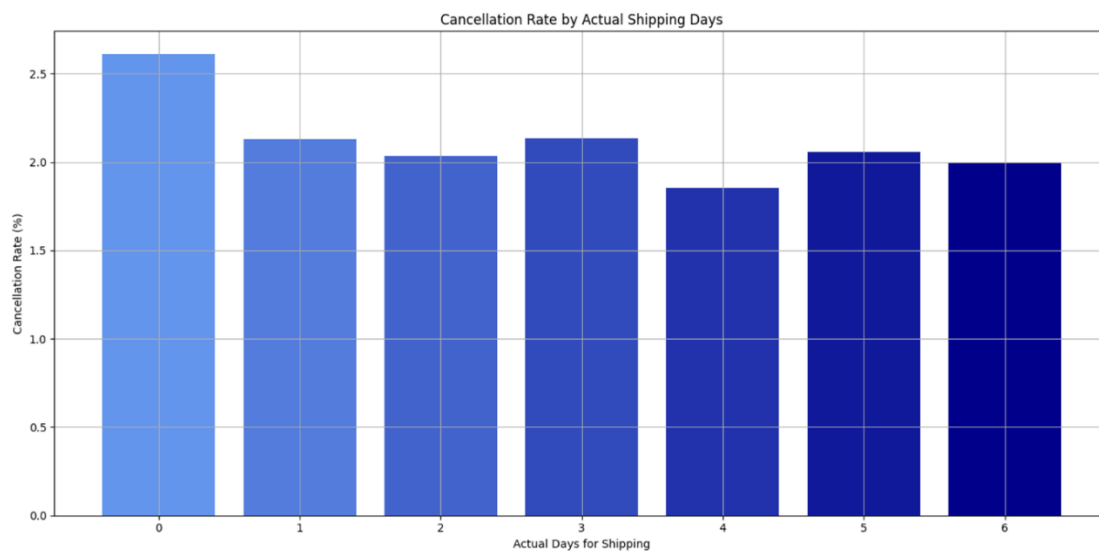
그림 9. 월별 제시간(on-time)에 도착하는 주문 비율



정시 배송이 이루어지는 주문의 비율이 전체적으로 40% 수준으로 낮음을 확인할 수 있다. 이는 큰 변동 없이 비교적 안정적인 배송 서비스가 제공되고 있음을 시사한다.

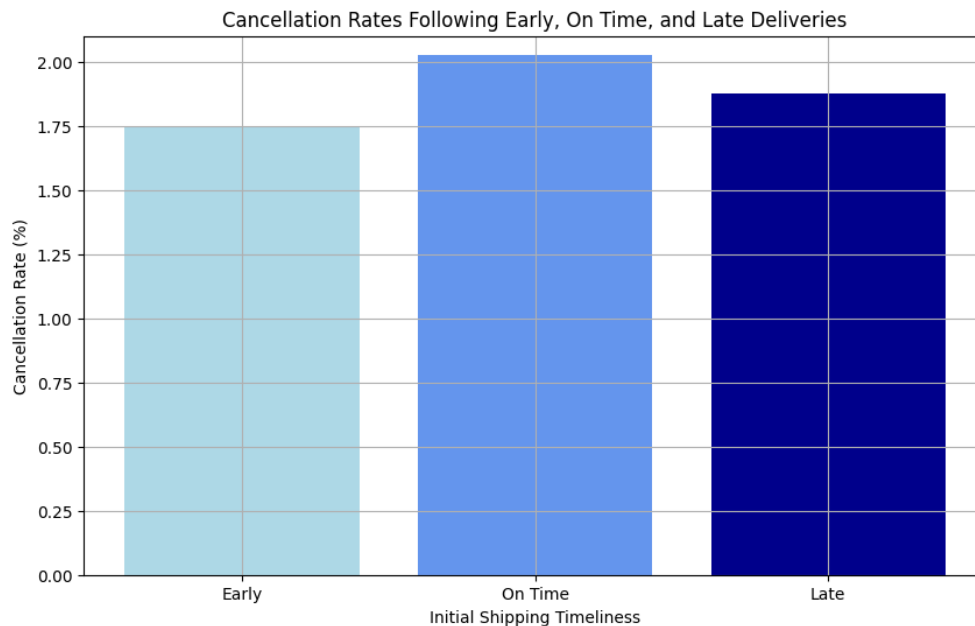
또한, 예정된 배송 시간을 초과하는 것이 아니라 절대적인 배송 지연이 발생할 경우, 이에 따른 주문 취소율의 증가 여부를 분석하였다.

그림 10. 배송일수에 따른 주문 취소율



절대적인 배송 지연이 주문 취소율을 증가시키는 명확한 경향은 나타나지 않았다. 이에 더 나아가, 예정된 배송일보다 실제 배송이 늦어지는 경우 고객이 주문을 취소하는 경향이 있는지 확인하기 위해 관련 데이터를 시각화하여 분석했다.

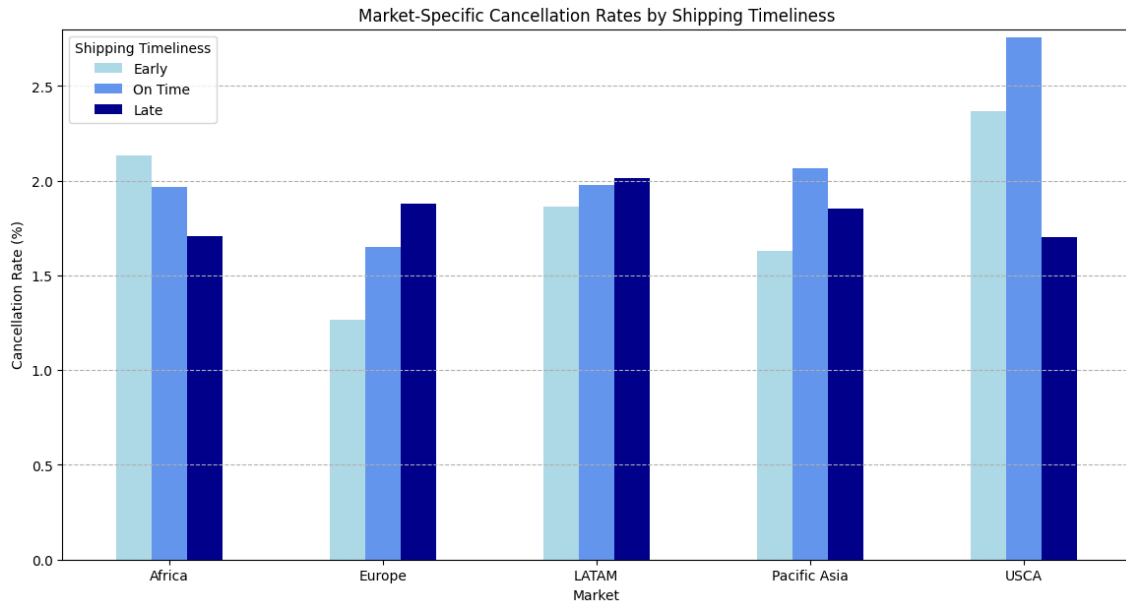
그림 11. 배송 지연도(일찍, 제시간, 늦음)에 따른 주문 취소율



예정된 배송일보다 상품이 일찍 도착하는 경우, 즉 배송이 조기에 완료되면 주문 취소율이 감소하는 경향을 보인다. 이는 고객들이 절대적인 배송 기간보다는 예상 배송일자와 비교하여 상대적으로 빠른 배송에 더 긍정적으로 반응한다는 것을 시사한다. 반면, 배송이 지연될 때 주문 취소율이 증가하는 현상은 관찰되지 않았다. 이러한 결과를 통해 배송 지연을 줄이기보다는 예상 배송일보다 빠르게 상품을 배송하는 것이 주문 취소율을 감소시키는 데 더 효과적일 수 있음을 알 수 있다. 따라서 배송일자를 고객에게 보수적으로 예측하여 알려주는 것이 취소율을 줄이는 데 도움이 될 것으로 판단된다.

위에서는 전체적인 취소율을 확인하였다면, 이번에는 대륙별로 차이가 나타나는지 살펴해보았다.

그림 12 대륙별 배송 지연도에 따른 주문 취소율



상대적 배송일에 따른 고객 반응은 대륙별로 차이를 보인다. 특히 유럽에서 가장 민감한 반응을 보인다. 반면, 라틴 아메리카와 아프리카에서는 상대적 배송일에 대한 반응이 덜 민감하다. 따라서, 유럽 시장에는 배송 관련 투자 증가시켜 고객만족도를 높이고 주문 취소율을 감소시키는 전략이 효과적일 것이다. 반면, 라틴 아메리카와 아프리카에는 상대적으로 배송에 대한 투자를 상대적으로 줄여도 주문 취소율에 큰 영향을 미치지 않을 것으로 보인다.

유럽 배송 전쟁에 대한 기사:

<https://www.chosun.com/economy/weeklybiz/2021/04/09/AJUNL2AZGFFPDGUKNCNXGU6EZA/>

3.2 재구매율 분석

재주문하는 고객과 그렇지 않은 고객 간의 차이를 분석하였다. 기존 데이터를 활용하여 재주문을 하는 고객의 데이터 프레임과 재주문을 하지 않는 고객의 데이터 프레임을 각각 생성했다. 이후, 이 두 데이터 프레임을 상세히 비교하며 재구매율에 영향을 미치는 요인들을 분석하였다.

그림 13. 재구매가 있는 데이터 프레임의 배송 모드 추이

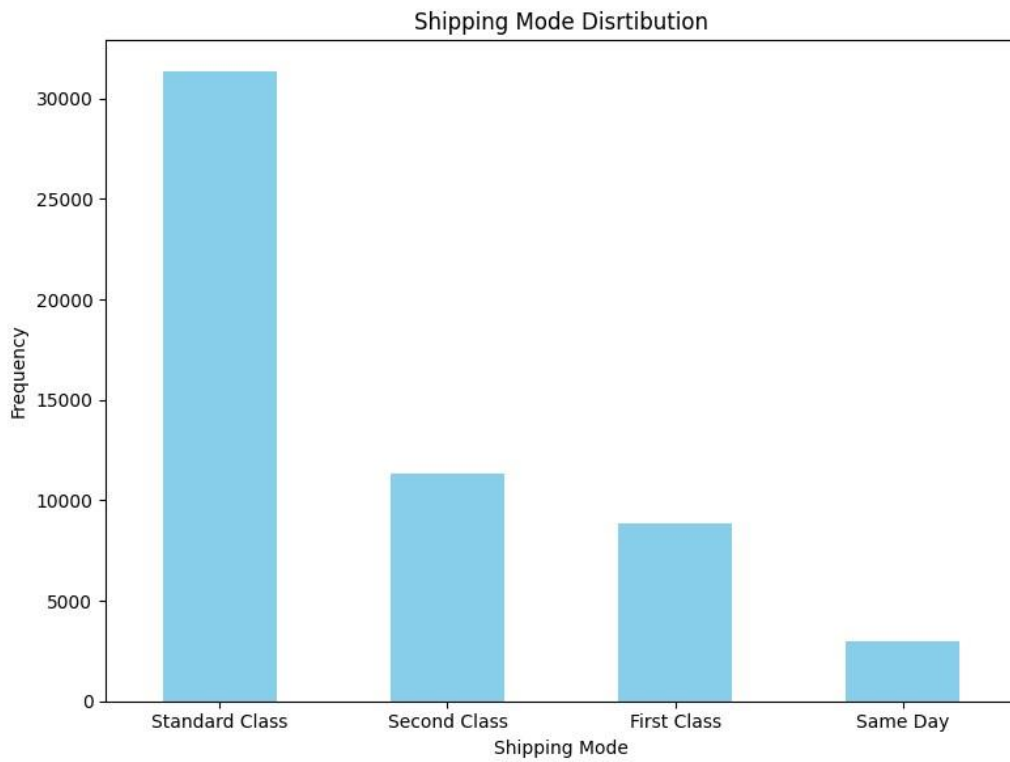


그림 14. 재구매가 없는 데이터 프레임의 배송 모드 추이

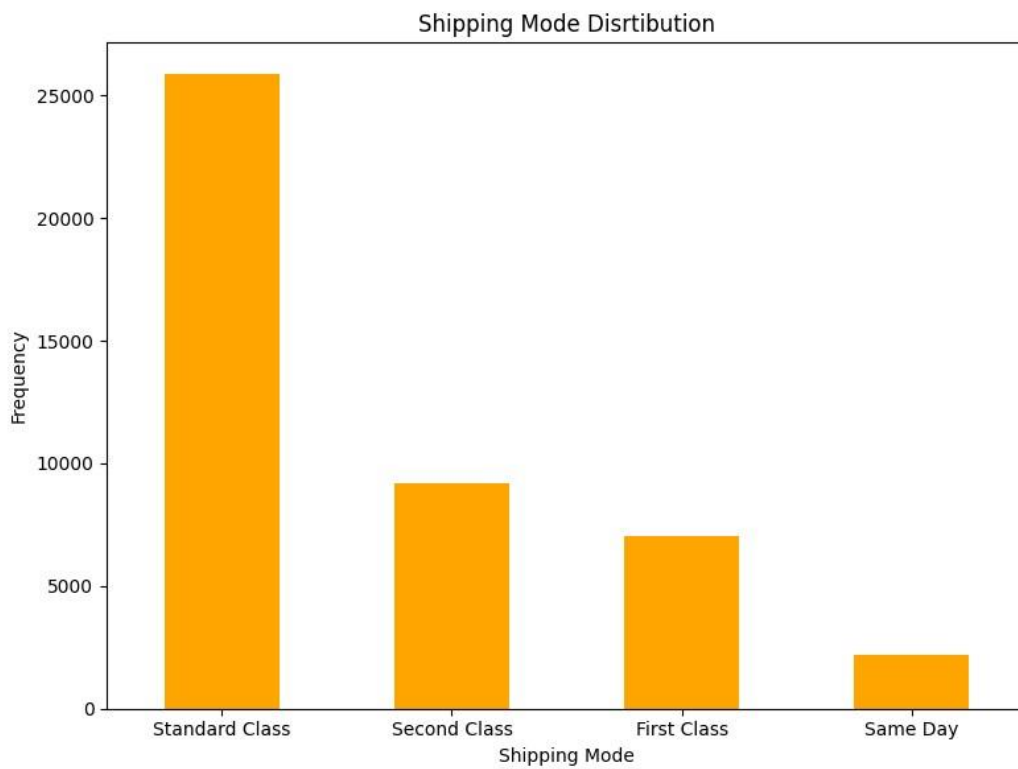


그림 13은 재주문이 있는 데이터 프레임에서 배송 모드의 추이를 나타낸 것이고, 그림 14는 재주문이 없는 데이터 프레임에서의 배송 모드를 보여준다. 두 그래프 사이에 뚜렷한 차이는 없지만, 첫 번째 그래프에서 'standard class'와 'first class' 배송 옵션이 조금 더 많은 비율을 차지하는 것을 확인할 수 있다. 이는 배송 모드가 재구매율에 어느 정도 영향을 미칠 수 있음을 시사한다.

그림 15. 배송 지연에 따른 재구매율

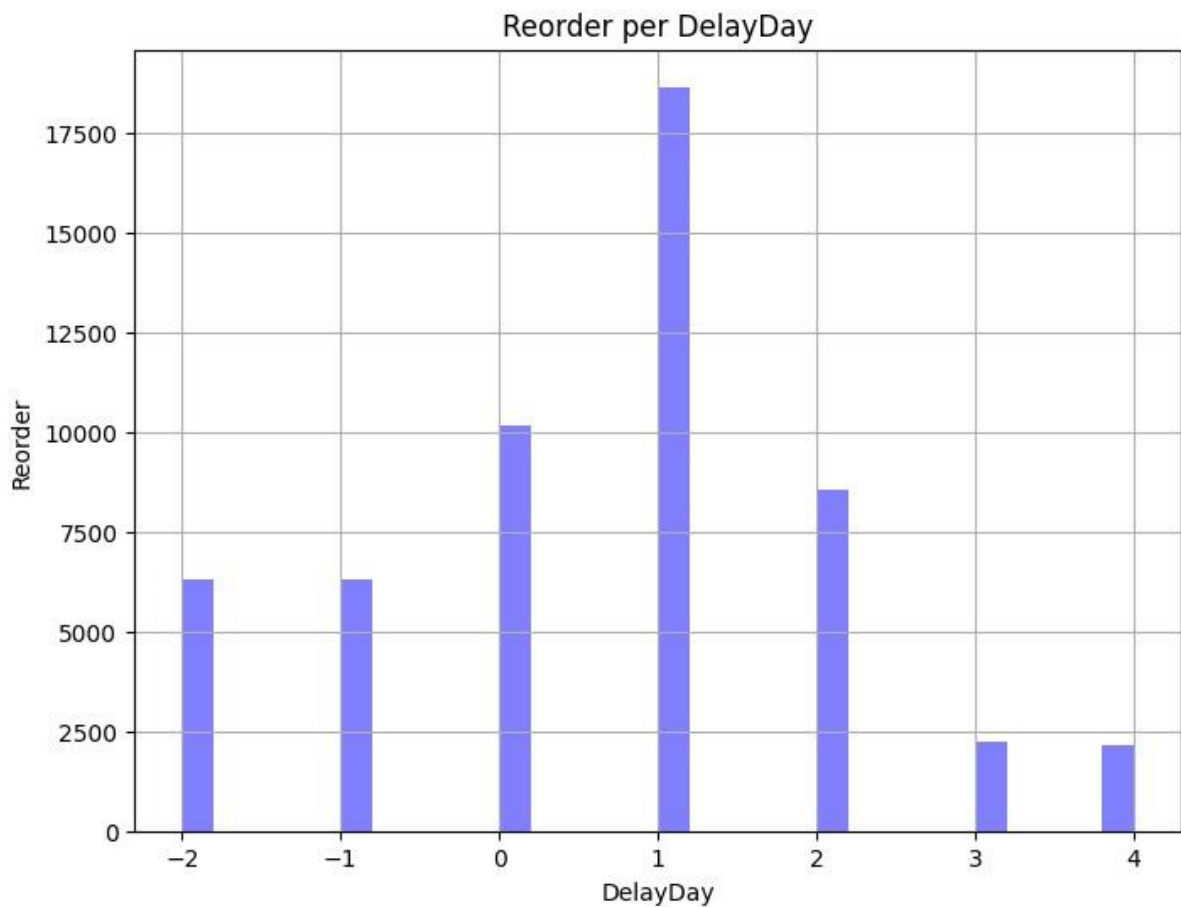


그림 15는 배송 지연에 따른 재구매율을 나타낸 그래프이다. 3일 이상 배송이 지연될 경우, 재구매율의 감소가 나타날 수 있음을 알 수 있다.

그림 16. 할인율에 따른 재구매 비율

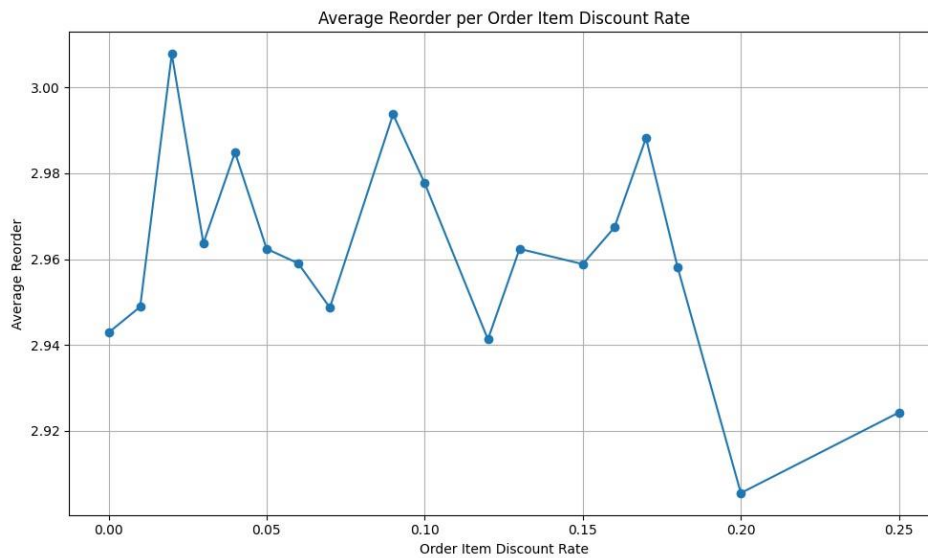


그림 16은 할인율에 따른 재구매율을 보여준다. 높은 할인율이 재구매율을 증가시키지 않는다는 점을 확인할 수 있다. 실제로, 할인율이 높은 제품의 재구매율은 오히려 감소하는 경향을 보인다. 이는 할인율이 높은 제품을 구매하는 고객들이 대부분 일회성 구매로 끝나는 경우가 많다고 추측할 수 있다.

그림 17. 각 품목에 대한 기업의 이익률에 따른 재구매율

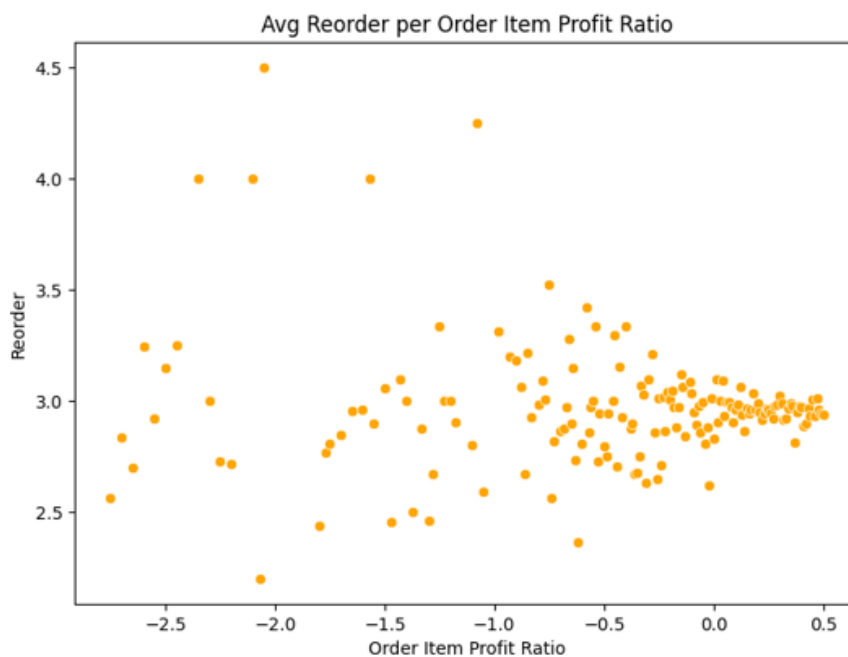


그림 17은 제품별 기업 이익률과 재구매율 사이의 관계를 나타낸다. 분석 결과, 기업의 이익률이 낮을수록 재주문율이 높게 나타나는 경향을 보인다. 이는 낮은 이익률로 인해 제공되는 할인이나 가격 경쟁력이 재구매를 촉진할 수 있음을 시사한다.

그림 18. 재주문율이 높은 나라 10 순위

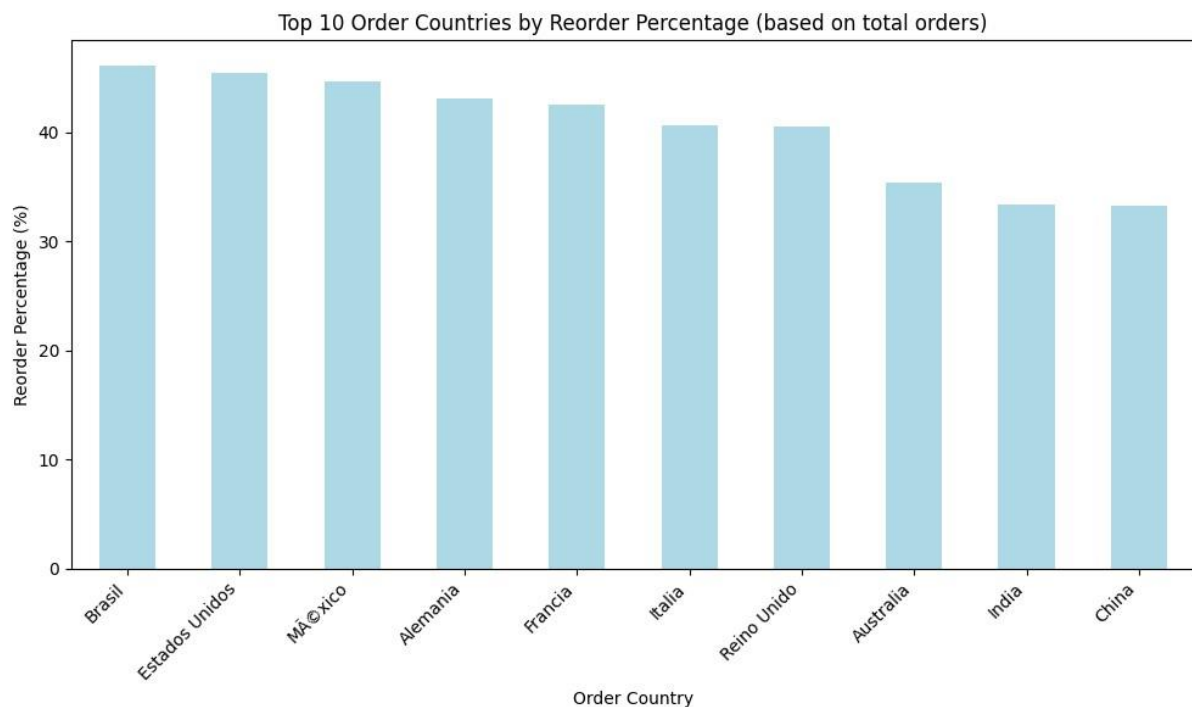


그림 18은 재주문율이 높은 나라를 전체 재주문율에 대한 비율로 나타내는 그래프이다. 각각 브라질, 미국, 멕시코가 1,2,3순위를 차지했다

3.3 지역별 판매 분석

지역별로 다양한 사회적 및 문화적 요인이 존재함에 따라, 선호하는 제품군이나 계절에 따라 매출이 변동될 수 있다. 이러한 차이를 고려하여 효과적인 제품 마케팅 전략을 개발하기 위해 다양한 시각화를 통해 분석을 진행했다. 우선, 각 지역별 매출과 실제 이익 분포를 검토하였다.

그림 19 지역별 매출 분포

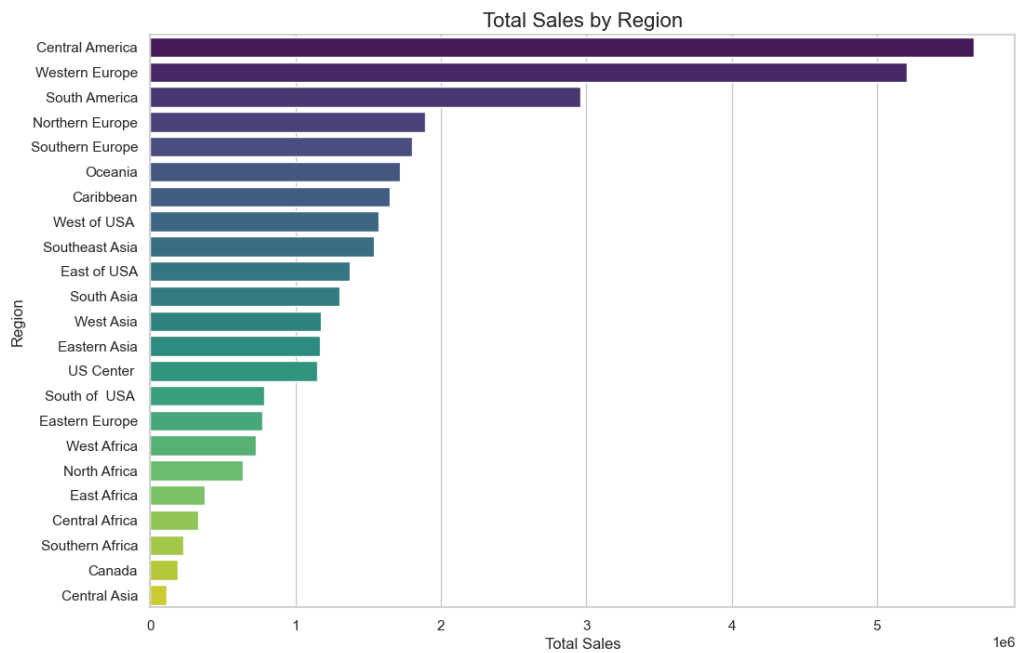


그림 20 지역별 실제 이익 분포

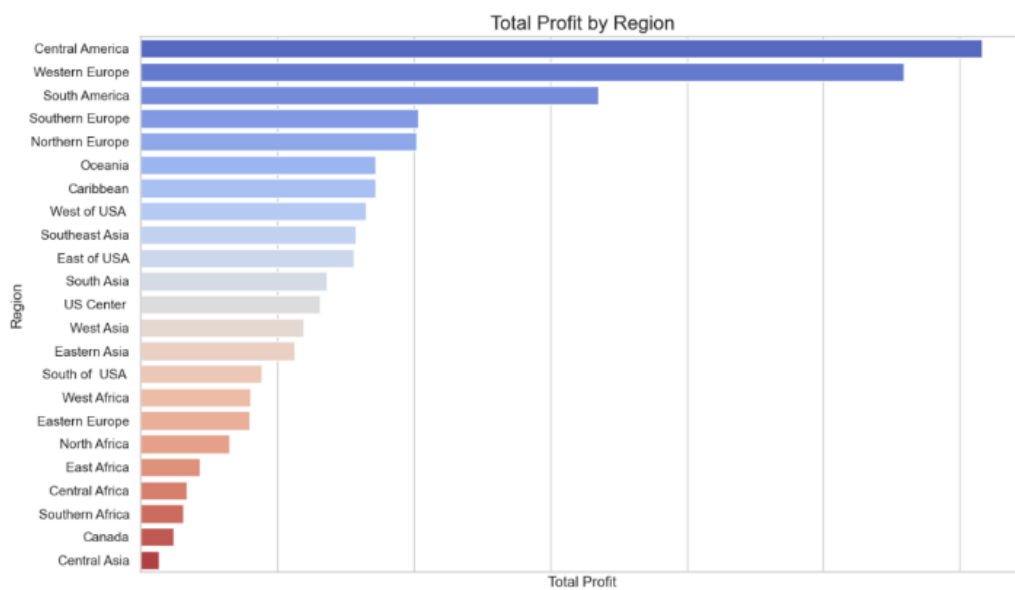


그림 19, 20을 확인해보면 지역별 매출과 실제 이익 간 차이가 거의 없다. 더 자세히 살펴보기 위해 분기별로 시각화를 진행해보았다.

그림 21 분기, 지역별 매출 및 실제 이익 분포



그림 21을 분석한 결과, 지역별로 중요한 문화적 요인이 존재하며, 이러한 요인들이 계절적 행사와 연결되어 매출에 큰 영향을 미치는 것을 확인할 수 있었다. 특히 미국에서는 4분기에 있는 블랙 프라이데이와 크리스마스가, 중국에서는 춘절이 매출 증가에 중요한 역할을 하는 것으로 나타났다.

표 6 지역별 매출 및 실제 이익 표준편차 상위 5

	Order Region	Sales		Order Profit Per Order	
		mean	std	mean	std
10	Northern Europe	205.595188	118.731880	21.933037	108.563692
22	Western Europe	203.720660	116.602284	21.874952	100.243940
5	East Africa	203.150593	115.881716	23.308709	100.679549
17	Southern Europe	202.695842	115.322655	22.925470	96.843498
11	Oceania	197.472362	112.867420	19.728281	98.877484

표 6을 통해 매출 변동이 큰 지역을 분석한 결과, 북유럽, 남유럽, 서유럽, 동아시아, 동남아시아 순으로 나타났다. 이 지역들은 계절에 따른 소비 패턴의 큰 변화를 보여주며, 이는 시장 변화에 신속하게 대응하는 마케팅 전략의 필요성을 시사한다.

그림 22 지역별 카테고리 실제 이익 분포

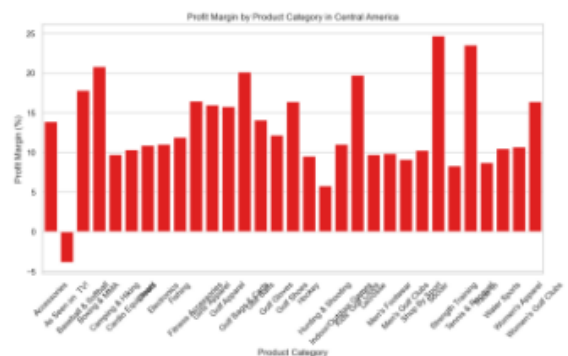
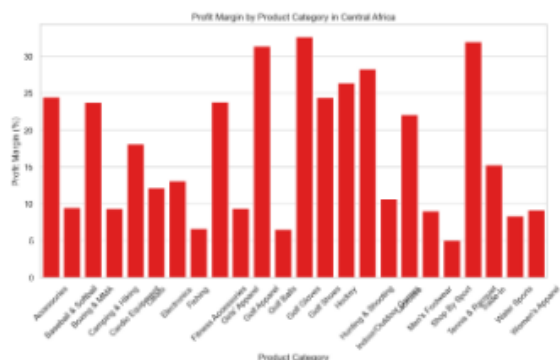
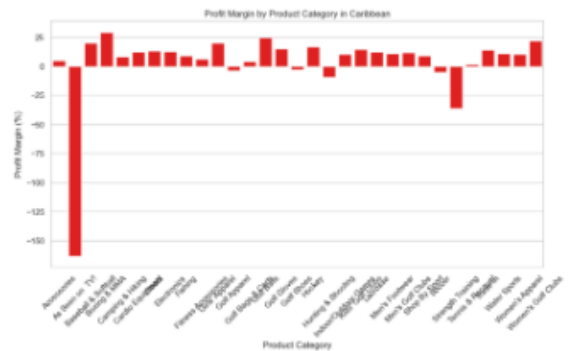


그림 22는 네 개의 지역을 선택하여 각 지역별로 카테고리별 실제 이익 분포를 보여주는 그래프다. 이 그래프를 통해 각 지역에서 선호하는 카테고리가 다르며, 심지어 같은 지역 내에서도 문화적 및 사회적 요인에 따라 선호하는 카테고리가 다를 수 있다..

표 7 지역별 카테고리 상위 마진율

	Order Region	Category Name	Sales	Order Profit Per Order	Profit Margin (%)
126	Central Asia	Tennis & Racquet	224.949997	90.379997	40.177817
424	Southern Africa	Hunting & Shooting	149.949997	58.500000	39.013005
17	Canada	Lacrosse	124.949997	42.730000	34.197680
67	Central Africa	Golf Gloves	424.830000	138.850002	32.683662
402	Southeast Asia	Lacrosse	774.689993	245.870002	31.737857

추가적인 분석을 위해 지역별 카테고리 분석을 실시했다. 각 지역별로 카테고리의 매출과 실제 이익을 바탕으로 '마진'이라는 열을 생성하여 마진율이 높은 상위 지역을 분석했다. 분석 결과, 중앙아시아에서는 테니스 및 라켓 제품의 마진율이 높았으며, 이는 스포츠에 대한 관심 증가와 테니스의 인기가 높아지고 있음을 반영한다. 남부 아프리카에서는 사냥 및 총기 관련 제품의 마진율이 높는데, 이는 야생 동물과 사냥 관련 관광이 활발한 지역적 특성 때문으로 보인다. 캐나다에서는 라크로스 관련 제품의 마진율이 높으며, 이는 캐나다에서 인기 있는 스포츠 중 하나이며 꾸준한 수요가 있는 것으로 파악된다. 중앙 아프리카에서는 골프 장갑의 마진율이 높는데, 이는 골프가 부유층 사이에서 인기 있는 스포츠로 고품질 골프 장비에 대한 수요가 있는 것으로 나타났다. 동남아시아에서는 라크로스 제품의 마진율이 높은 것을 보아, 이 스포츠의 인기가 높아지는 추세를 파악할 수 있다.

그림 23 지역별 가장 마진율이 높음 카테고리

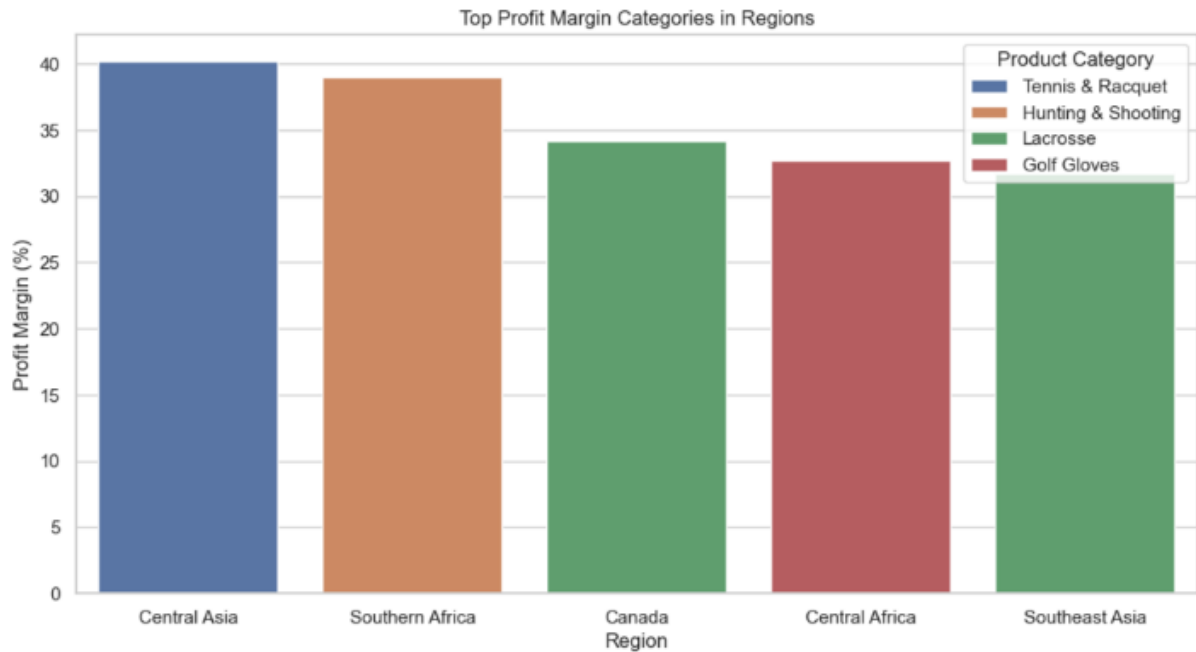


그림 23은 각 지역에서 마진율이 가장 높은 카테고리를 보여준다. 만약 각 지역별로 카테고리의 제조사 데이터가 확보된다면, 추가적인 분석을 통해 더 깊이 있는 인사이트를 얻을 수 있을 것이다. 가격, 제품, 서비스 간의 경쟁 분석을 실시하여 시장 점유율을 높일 수 있는 전략을 개발하고, 이를 통해 더욱 효과적으로 시장에 대응하는 다양한 전략을 수립할 수 있을 것으로 예상된다.

3.4 사기 의심 정황 분석

주문 처리 상태에서 일부 행이 'suspected fraud'로 표시된 것을 확인했다. 주문 상태는 COMPLETE, PENDING, CLOSED, PENDING_PAYMENT, CANCELED, PROCESSING, SUSPECTED_FRAUD, ON_HOLD, PAYMENT_REVIEW 등 다양하다. 운영적 관점에서 사기 의심 거래를 최소화하는 것이 중요하므로, 사기 의심 거래가 나타나는 특정 특징들을 분석했다. 사기 의심 정황이 포착된 주문은 총 3,878 건으로, 이는 전체 주문의 약 2.25%에 해당한다. 또한, 이러한 주문의 결제 방식은 모두 계좌 이체로 확인되었다.

그림. 24 사기 의심 정황이 나타나는 나라별 주문량 그래프

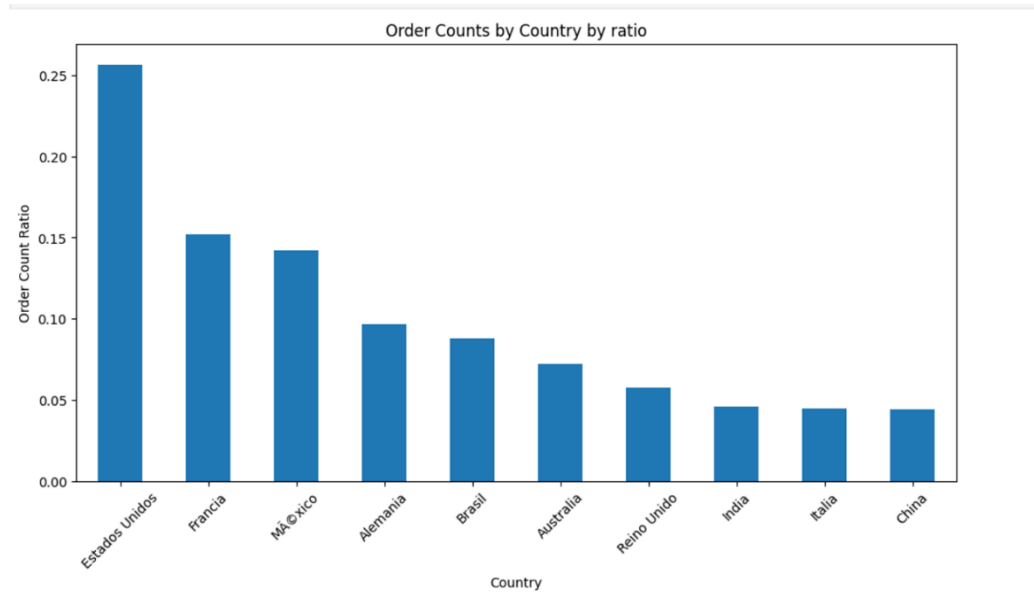
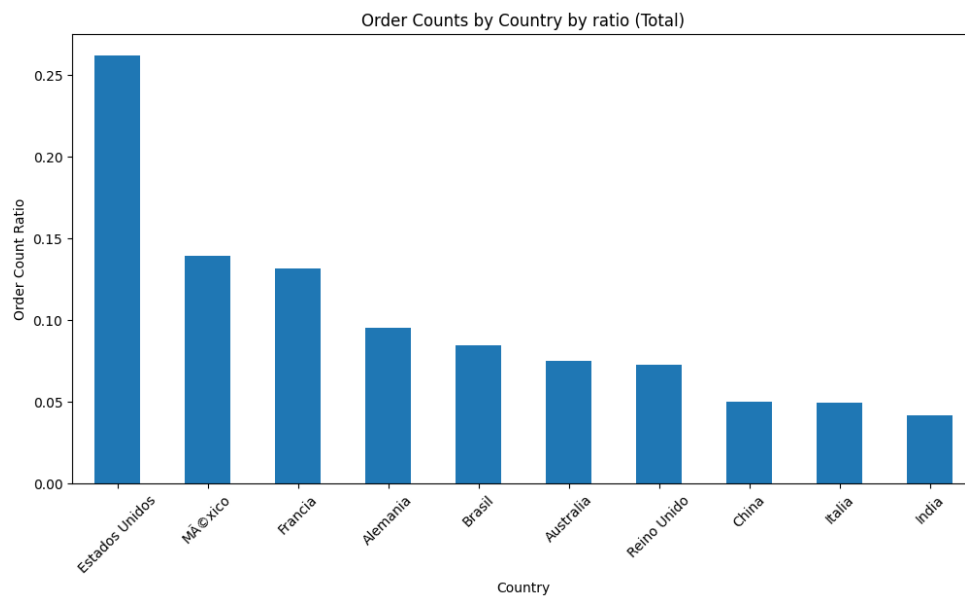


그림. 25 모든 주문 처리 상태의 나라별 주문량 그래프



사기 의심 정황이 가장 많이 발견된 나라는 주문량이 가장 많은 미국이었다. 주문량에서 두 번째로 많은 멕시코는 사기 의심 정황에서는 세 번째로 많았고, 두 번째로는 프랑스가 차지했다. 이런 결과를 통해 사기 의심 정황이 빈번히 발생하는 나라의 결제 시스템에 추가적인 보안 강화가 필요함을 알 수 있다.

이후 사기 의심 정황 거래의 판매 카테고리 중 어떤 카테고리가 가장 많은 비율을 차지하는지 살펴보았다.

그림. 26 사기 의심 정황이 나타나는 카테고리별 주문량 그래프

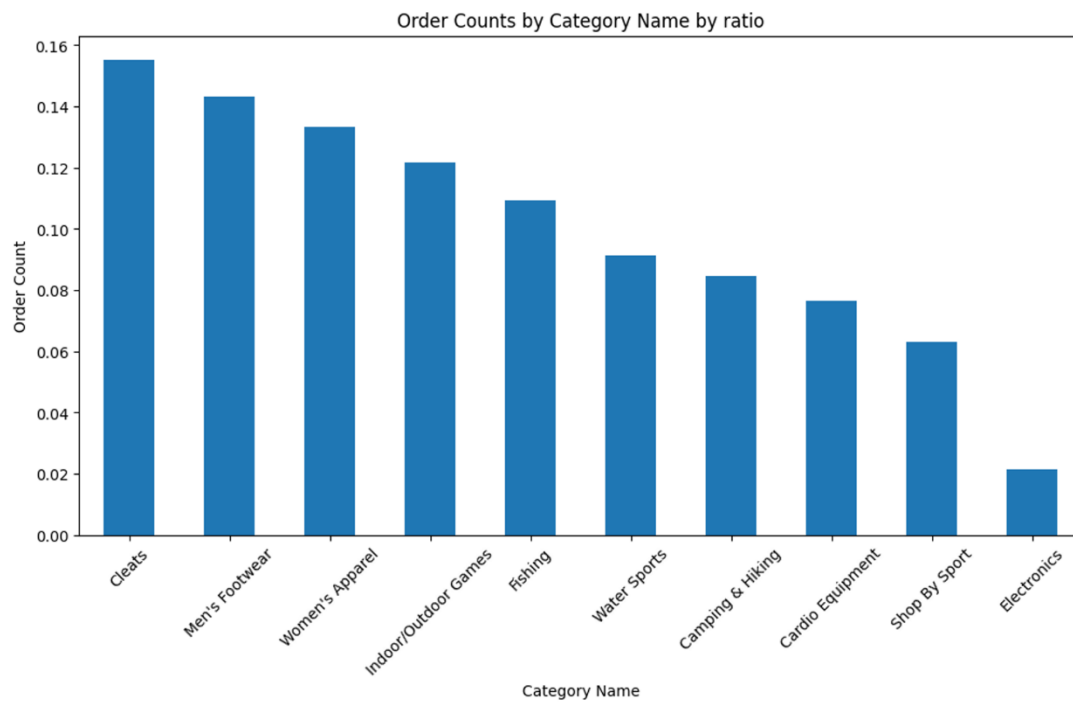
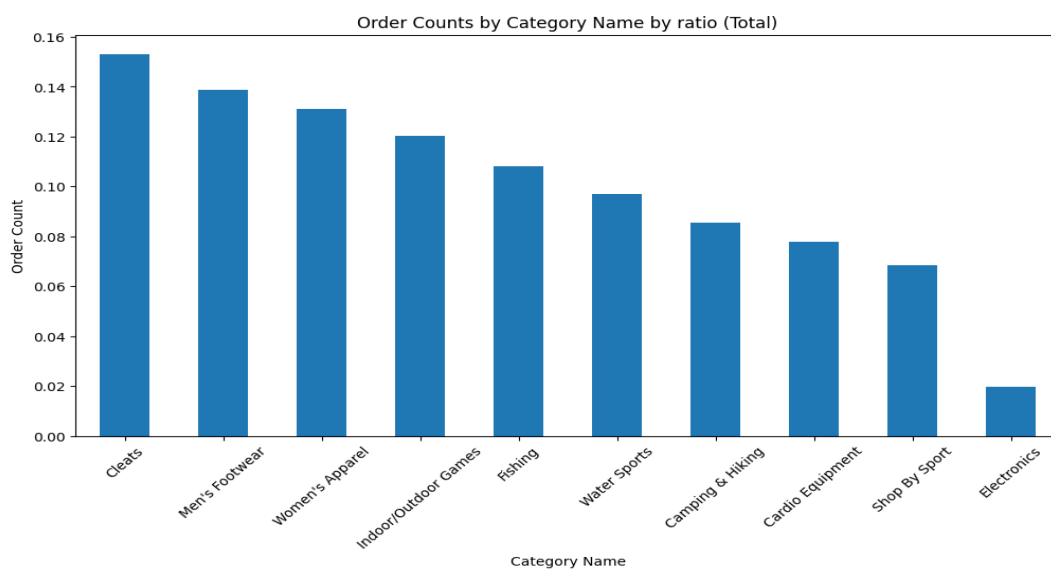


그림. 27 모든 주문 처리 상태의 카테고리별 주문량 그래프



cleats와 men's footwear는 온라인 쇼핑몰에서 가장 많이 판매되는 상위 두 제품이며, 사기 의심 정황에서도 이 두 항목의 구매 비율이 가장 높았다.

이에 따라, 사기 의심 정황이 가장 많이 보고된 상위 10개국에서는 결제 시 추가 보안 절차의 도입이 필요하다고 볼 수 있다. 특히 계좌 이체를 사용할 때 본인 인증 절차를 강화함으로써 사기 의심 거래를 줄일 수 있을 것이다.

3.5 RFM 분석

RFM (Recency, Frequency, Monetary) 분석은 고객의 구매 패턴을 이해하고, 이를 기반으로 고객 세분화를 수행하여 효율적인 마케팅 전략을 수립하는 데 중요한 도구이다. RFM 모델을 활용하여 고객들을 여러 군집으로 분류하고, 각 군집의 특성을 파악하여 맞춤형 마케팅 전략을 개발한다. 이 과정은 크게 세 단계로 구성된다:

1. K-means 군집화를 통한 고객 세분화: 고객을 Recency(최근 구매 날짜), Frequency(구매 빈도), Monetary(구매 총액) 값에 따라 군집화하고, 이를 통해 고객의 현재 가치와 행동 경향을 분석한다.
2. 군집 결과의 해석 및 시각화: 각 군집의 특성을 분석하고 시각화하여, 이해하기 쉽고 실행 가능한 인사이트를 제공한다. 이는 군집별로 맞춤화된 커뮤니케이션과 프로모션을 계획하는 데 중요한 기초 자료가 된다.
3. 고객 행동 시계열 분석 및 CLTV (Customer Lifetime Value) 계산: 장기적인 관점에서 고객의 가치를 평가하고, 향후 마케팅 및 영업 전략에 적용할 수 있는 고객 생애 가치를 계산한다. 이는 고객 유지 및 장기적인 관계 개선에 중점을 둔 전략 수립에 도움을 준다.

이 분석을 통해 얻은 인사이트는 고객의 구매 경향을 더 잘 이해하고, 효과적인 고객 관계 관리 및 마케팅 전략을 수립하는 데 크게 기여할 수 있다. 이는 최종적으로 고객 만족도를 높이고, 회사의 수익성을 극대화하는 결과로 이어질 것을 기대한다.

3.5.1 Kmeans를 활용한 RFM 고객 군집화

RFM table을 생성하기 전 간단한 전처리로 취소된 배송 건수에 대한 행은 RFM 값을 산출하는데 왜곡을 줄 수 있으므로 삭제하고 진행했다.

실제로 제품이 고객에게 배송된 시점이 구매 완료의 가장 명확한 지표라고 판단하여 Recency를 산출할 때 주문 날짜가 아닌 배송 날짜를 활용했다. Recency의 값은 오늘 날짜(데이터 상 2017년 10월 2일)로부터 가장 최근에 배송된 날짜와의 차이를 뜻한다. Frequency는 총 구매 횟수를 뜻하며 MonetaryValue는 총 지출 금액을 뜻한다. 모두 고객별로 산출된 열이다.

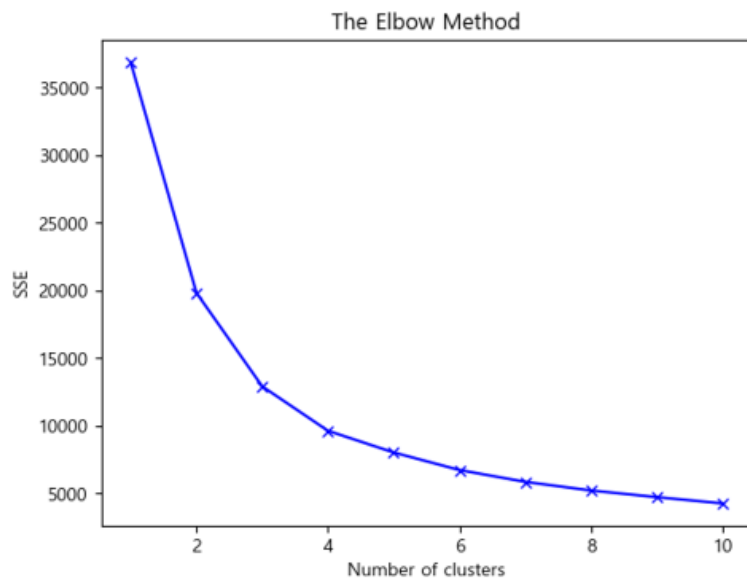
표 8 고객별 RFM Table

Customer Id	Recency	Frequency	MonetaryValue
1	671	1	472.450012
2	17	4	1618.660042
3	114	5	3189.200037
4	262	4	1480.709993
5	340	3	1101.919998
...
12431	88	13	6614.170039
12432	153	9	4622.410011
12433	58	3	1293.489998
12434	107	6	2883.950024
12435	400	1	931.710007

12297 rows × 3 columns

표 8의 RFM Table을 활용하여 군집을 생성하기 전 Elbow Plot과 Silhouette Score를 산출하여 최적의 군집 수를 찾았다.

그림 28 Elbow Plot



Silhouette Score 또한 고려하여(표는 생략) 최적의 군집 수를 5로 지정하여 군집을 생성했다. Recency, Frequency, MonetaryValue를 각 축으로 군집이 잘 생성되었는지 확인한다.

그림 29 RFM 3D Plot

3D RFM Clusters

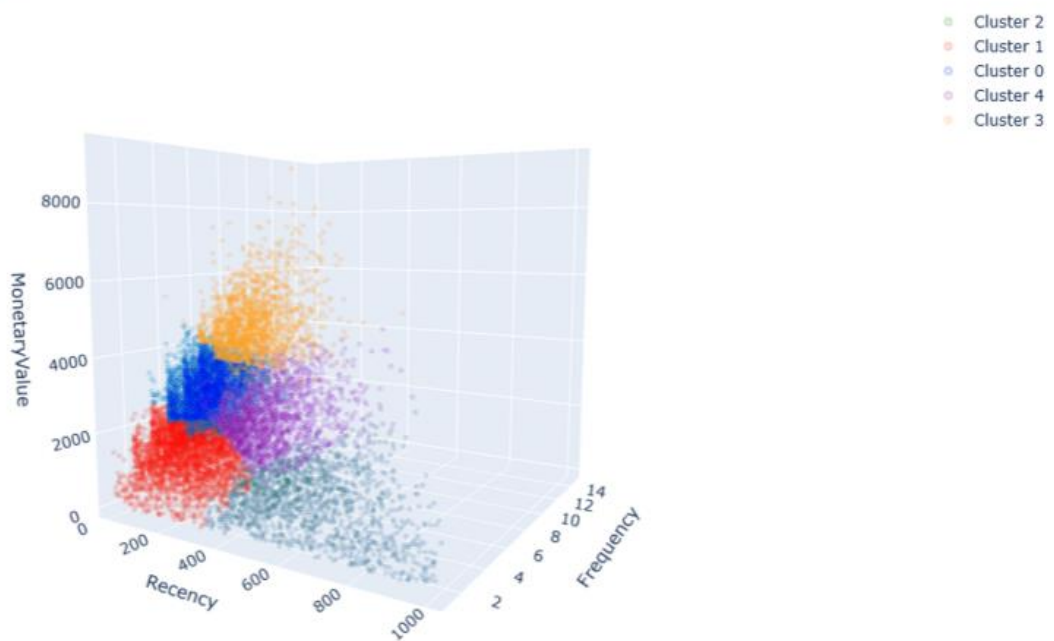


그림 29에서는 군집이 성공적으로 생성된 것을 확인할 수 있었다. 그러나 밀도가 높은 데이터 특성상 random state를 고정시키지 않으면 군집 번호가 변동될 가능성이 있어, 이를 고정하여 일관된 결과를 얻을 수 있도록 조치하였다. 추가적으로 DBSCAN 등의 밀도 기반 군집화 모델을 시도하였으나, 데이터의 높은 밀도로 인해 군집화 과정이 어려웠다는 점을 밝힌다. 군집의 자세한 해석 및 결과는 추후 기술한다.

3.5.2 RFM 고객 군집 해석 및 시각화

군집 해석을 진행하기 전 Recency, Frequency, MonetaryValue의 분포를 간략히 살펴본다.

그림 30 Recency 분포

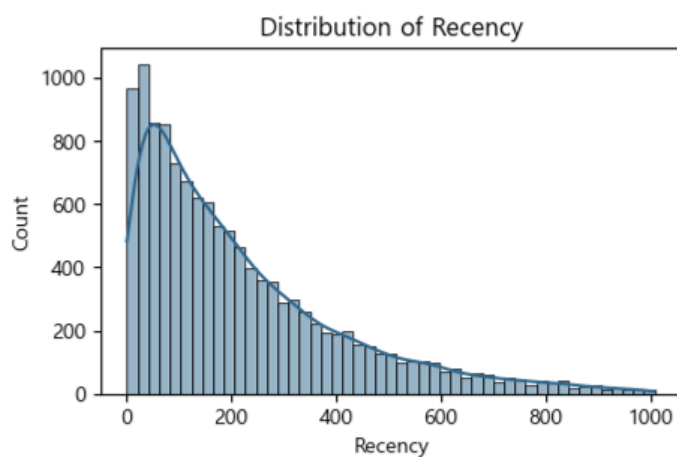


그림 31 Frequency 분포

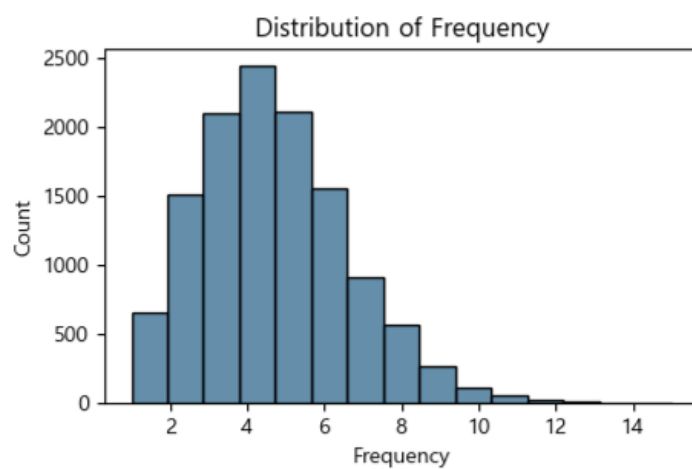


그림 32 MonetaryValue 분포

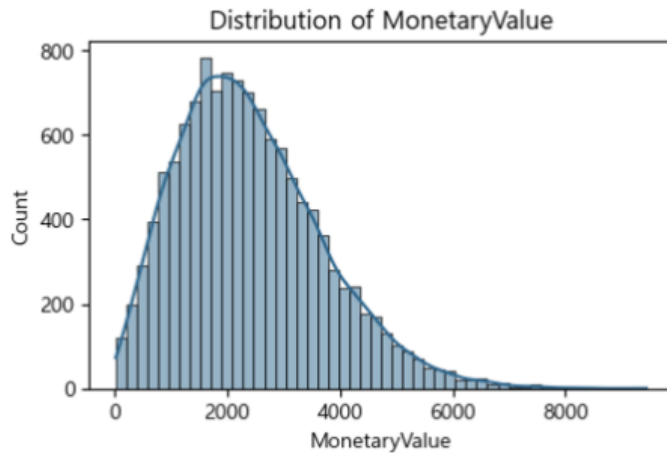


그림 30, 31, 32를 살펴보면 꽤 많은 사람이 최근까지 구매이력이 있는 것으로 파악된다. 또한 절반 이상의 고객이 4번 이상 주문 경험의 있음을 알 수 있다.

3.5.2.1 군집 해석

본격적인 분석에 앞서 표 9를 활용해 각 군집 별로 Recency, Frequency, MonetaryValue 및 고객 수, 주문 당 지출 금액을 활용하여 군집 해석을 진행한다.

표 9 군집별 특징

	Cluster	Recency_mean	Frequency_mean	MonetaryValue_mean	Count	M/F
0	0	99.198380	5.186095	2789.617354	3826	537.903235
1	1	150.697618	2.841478	1351.895271	3274	475.771807
2	2	611.416609	2.049417	1023.303038	1457	499.314309
3	3	115.673118	7.827419	4536.016266	1860	579.503418
4	4	365.948404	4.384574	2394.686080	1880	546.161571

표 9에 따르면, 군집 분석 결과는 다음과 같다. 0번 군집은 최근에 구매한 것은 물론 자주 구매하고 상당한 금액을 지출했음을 볼 수 있다. 1번 군집의 경우 최근 구매한 지 오래되었고, 그 빈도도 낮아 잠재적으로 이탈할 위험이 있는 고객으로 분류된다. 2번 군집의 경우 구매한지 굉장히 오래되었으며 구매 빈도와 지출 금액 또한 매우 낮다. 이들은 불만족한 고객일 확률이 높으며 이미 이탈했거나 이탈할 가능성이 높은 군집으로 해석된다. 3번 군집은 최근에 구매가 이루어졌으며,

구매 빈도와 지출 금액이 가장 높다. 이들은 VIP 고객 혹은 단골 고객으로 해석된다. 4번 군집은 과거에는 충성스러운 구매 패턴을 보였을 수도 있으나 최근 구매가 없어 잠재적으로 이탈 위험이 있는 고객으로 해석된다. 이러한 분석을 통해 군집별로 다음과 같은 명칭을 부여하고 직관적인 시각화를 진행했다.

0: Loyal Customers

1: At Risk Customers

2: Dissatisfied Customers

3: VIP Customers

4: Loyal Customers (At Risk)

3.5.2.2 군집별 RFM 및 이익 분포

각 군집별로 고객 수, Recency, Frequency, MonetaryValue, 실제 총 이익의 분포를 살펴봄으로써 군집이 잘 생성되었는지 집중하면 좋을 군집을 탐색한다.

그림 33 군집별 고객 수

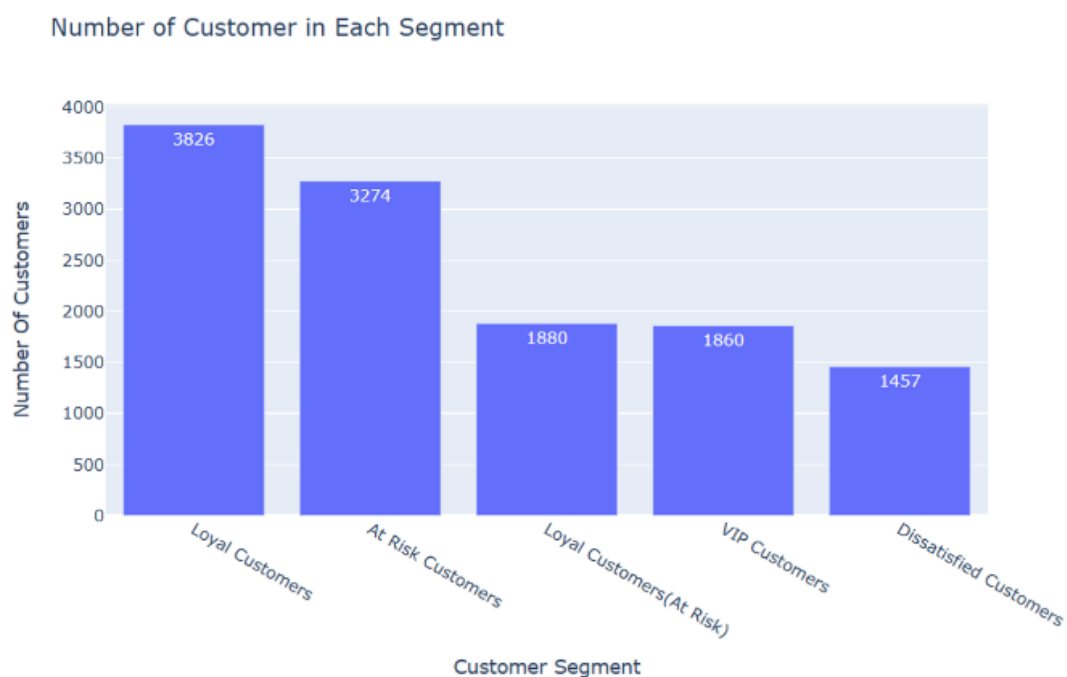


그림 33을 통해 확인할 수 있는 바에 따르면, 충성 고객이 전체 고객 중 가장 큰 비율을 차지하고 있어, 회사의 효율적인 운영을 반영한다. 그러나 이탈 위험 고객과 충성 고객(이탈위험) 역시 상당수 존재하여 구매 빈도를 증가시키기 위한 전략이 요구된다. 이는 회사가 현재 운영하는 마케팅 전략을 재검토하고, 특히 이러한 위험군에 속하는 고객들의 유지 및 활성화를 위한 구체적인 방안을 마련해야 함을 시사한다.

그림 34 군집별 Recency 분포

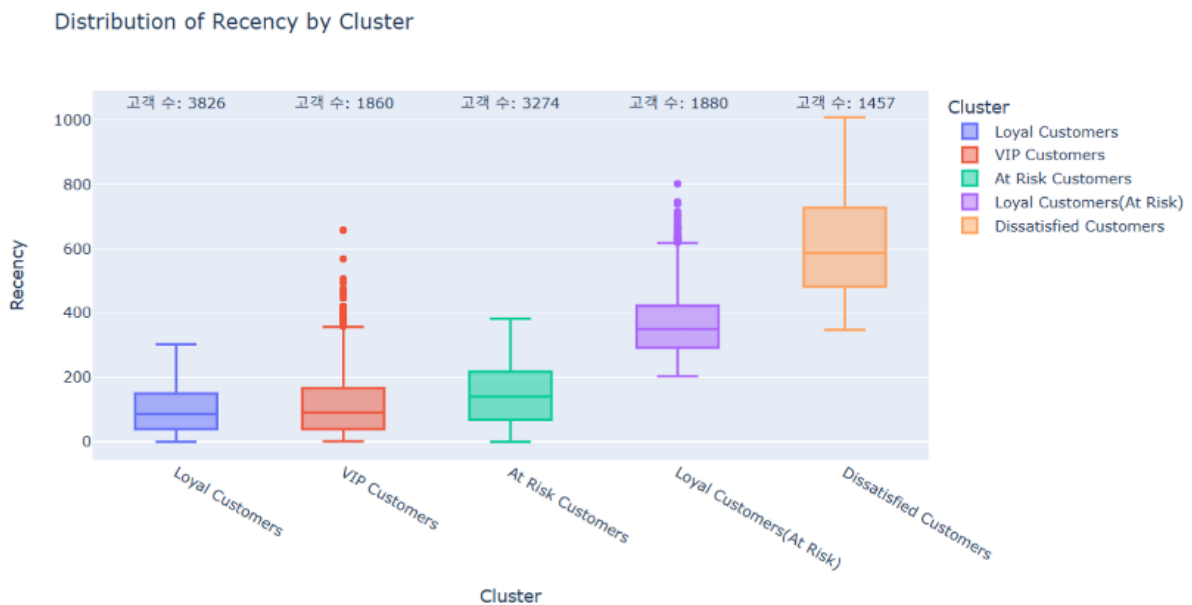


그림 34를 분석해보면, 일부 VIP 고객들이 높은 Recency 값을 가지고 있는 것을 확인할 수 있다. 이는 VIP 고객이라 할지라도 이탈 위험을 갖고 있음을 나타낸다. 또한, 불만족 고객군에서는 최근 구매한 사람조차도 마지막 구매 이후 거의 1년이 지난 사례가 관찰되어, 이들이 이미 이탈했을 가능성이 높다는 점을 시사한다. 이는 고객 유지 전략을 재정비할 필요가 있음을 보여주며, 특히 고객 충성도가 높지 않은 군에 대한 추가적인 마케팅 및 관리 전략이 요구된다.

그림 35 군집별 Frequency 분포

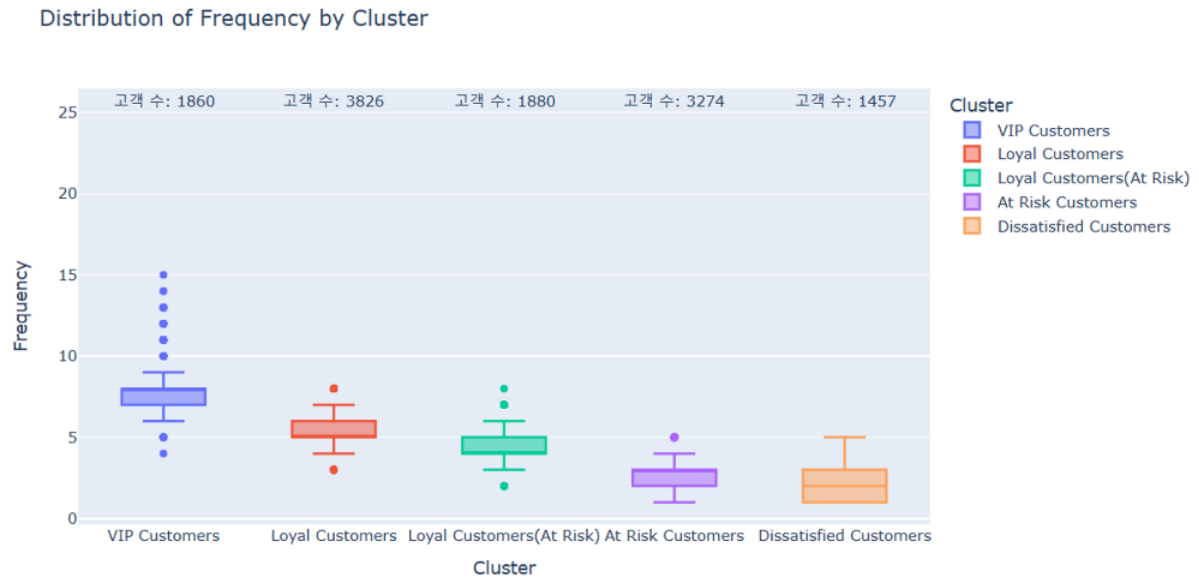


그림 35를 바탕으로 볼 때, VIP 고객 중 일부는 구매 횟수가 특히 높은 것으로 나타났다. 이들은 VVIP 고객으로 분류할 수 있으며, 쿠폰이나 이벤트와 같은 특별한 마케팅 활동을 통해 이들을 장기적으로 유지하는 것이 중요하다. 한편, 이탈 위험의 고객군은 수가 많음에도 불구하고 구매 빈도가 낮은 편이다. 이러한 고객군을 다시 활성화시키고 구매를 유도하기 위한 전략이 필요하다는 것을 알 수 있다. 이를 위해서는 타겟 마케팅과 개인화된 접근이 중요할 것으로 보인다.

그림 36 군집별 MonetaryValue 분포



그림 37 군집별 실제 총 이익 분포

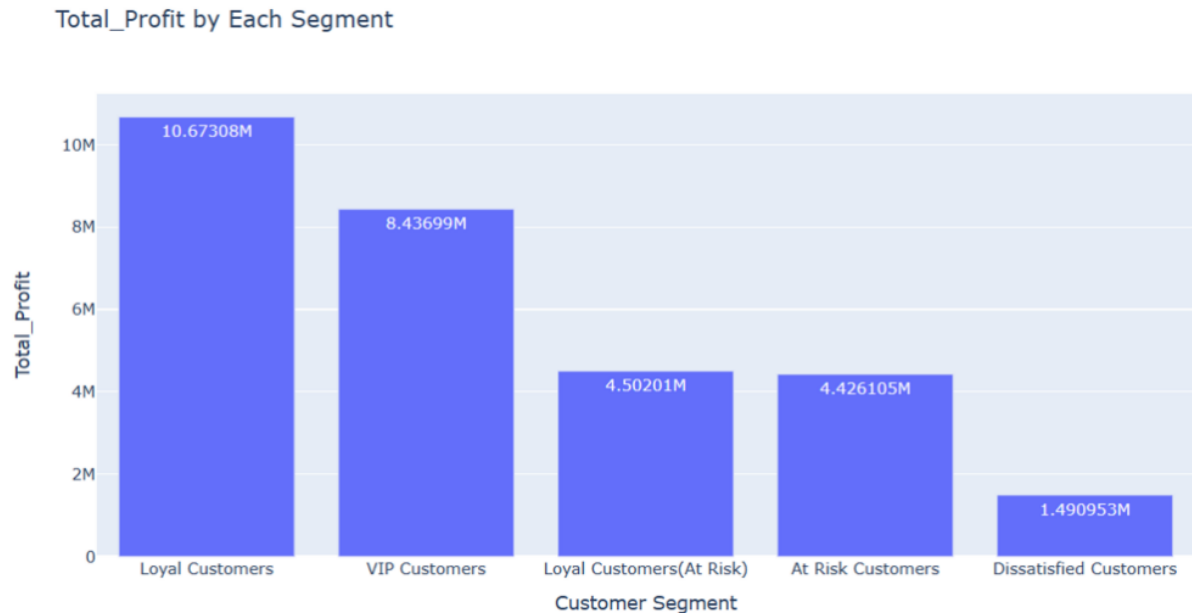


그림 36과 37을 통해 볼 때, VIP 고객은 전체 고객 수가 적음에도 불구하고 총 이익에 상당한 기여를 하고 있음이 명확하다. 따라서 이들의 지속적인 유지는 회사에 있어 필수적인 전략으로 여겨진다. 한편, 충성 고객(이탈 위험)의 수는 이탈 위험 고객과 비교할 때 상대적으로 적지만, 실제로 기여하는 이익은 비슷한 수준임을 확인할 수 있다. 이는 이탈 위험이 있는 충성 고객을 효과적으로 관리하고 유지하기 위한 전략이 절실함을 나타낸다. 이들에 대한 집중적인 관리와 맞춤형 마케팅 전략을 개발하는 것이 중요할 것이다.

3.5.2.3 군집별 특성 파악

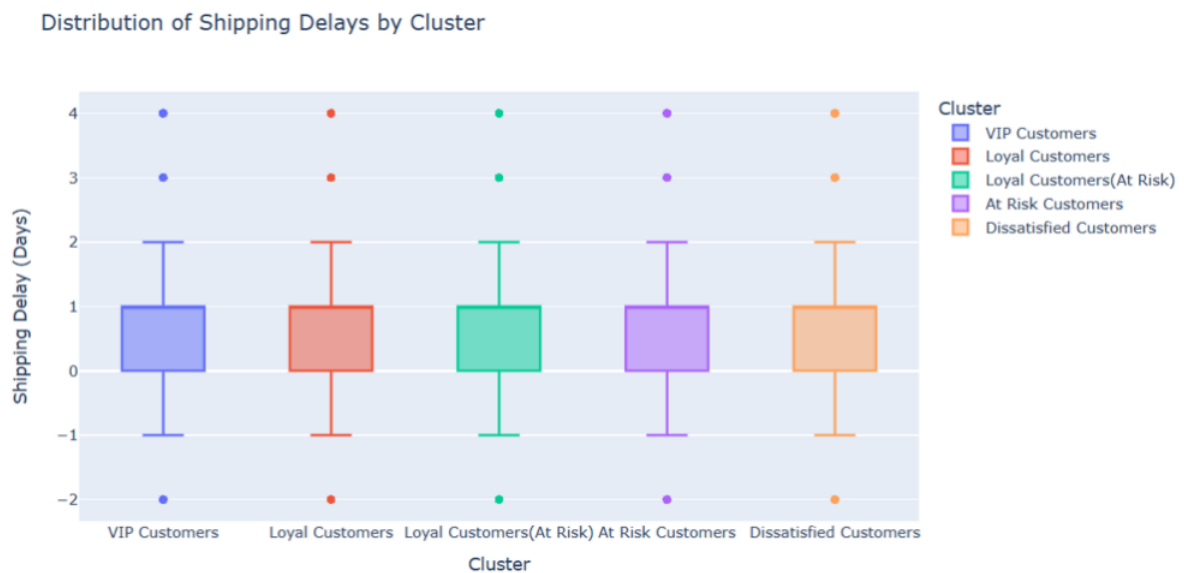
각 군집의 특성을 파악하고 이를 기반으로 군집별 맞춤 전략을 수립하기 위해, 고객 특성, 배송 관련 정보, 할인율 등 다양한 요소를 활용하여 군집별로 단일 열 시각화 분석을 수행하였다.

그림 38 군집별 Customer Segment 분포



군집별로 고객 특성이 다를 것으로 예상되어 군집 내 분포 비율을 살펴보았으나 전혀 차이가 나타나지 않았다. Order Country의 경우도 특징을 찾을 수 없었다.

그림 39 군집별 배송 지연 분포



군집별로 배송 관련 특성에 차이가 있을 것으로 예측했으나, 배송 지연, 배송 모드, 지연 배송 위험성, 배송 상태, 주문 상태 등의 변수에서는 유의미한 차이가 관찰되지 않았다. 이 결과는 고객의 이탈이나 불만이 배송과 관련된 문제 때문이 아님을 시사한다.

추가적으로 할인과 관련해서도 살펴보았으나 전혀 차이가 나타나지 않았다. VIP 고객으로 분류되어도 쿠폰을 잘 사용하거나 할인이 적용되지 않는 것으로 판단됨. 할인된 가격이나 마진율 등의 경제적인 요인을 검토해보아도 유의미한 차이는 나타나지 않았다. 이 결과는 고객의 만족도가 효율성이 좋은 상품을 구매하는 것에 기반한 것이 아님을 나타낸다.

3.5.3 RFM 군집화를 활용한 고객 행동 시계열 분석 및 CLTV

고객 군집별로 다양한 변수를 활용하여 분석을 진행했지만, 뚜렷한 특징을 도출하기 어려웠다. 이에, 군집별 패턴과 특징을 시계열 분석을 통해 찾아보기로 했다. 월별 데이터를 기준으로 각 군집에서의 신규 유입, 활동 중인 고객, 이탈 고객 등의 행동 패턴을 분석하였다. 이를 바탕으로 고객 생애 가치(CLTV)를 계산하여 군집별 마케팅 전략을 마무리 짓는다. 이러한 접근은 시간에 따른 고객 행동의 변화를 포착하여 보다 정밀한 고객 관리 전략을 수립하는 데 도움을 줄 것이다.

3.5.3.1 고객 행동 패턴 시계열 분석

그림 40 월별 군집 신규 유입 분포

New Customer Count by Cluster from 2016 Onwards

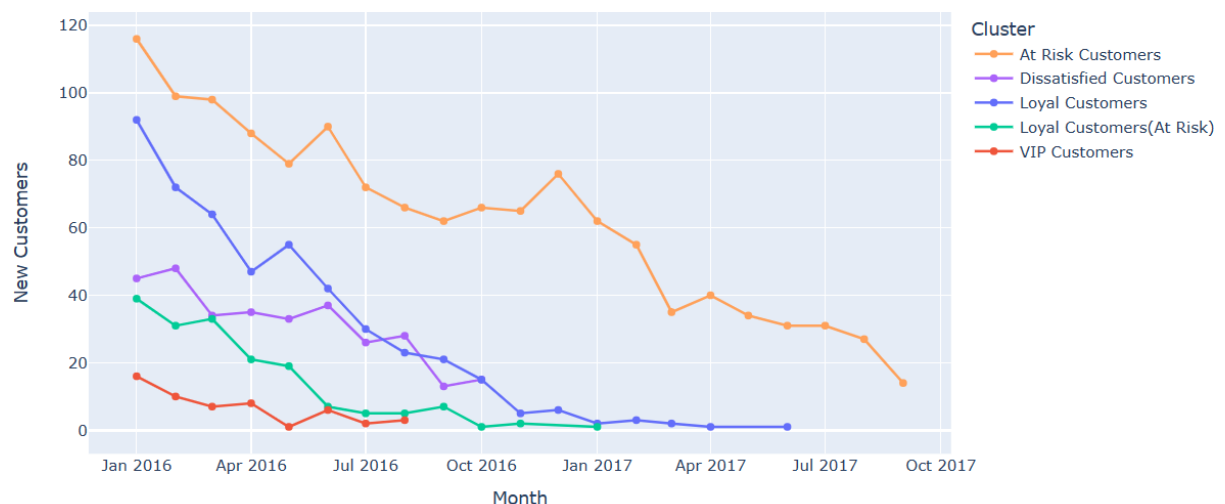
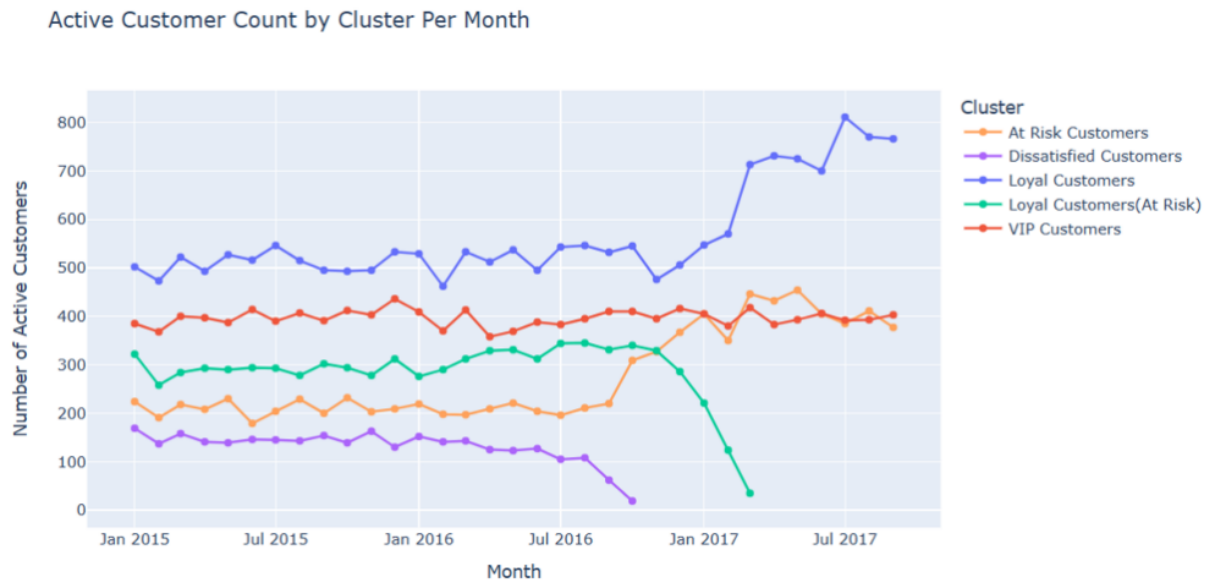


그림 40은 월별로 각 군집에 신규로 유입되는 고객 수의 분포를 보여준다. VIP 고객 군집에는 신규 유입이 거의 없는데, 이는 대부분의 VIP 고객이 처음부터 단골이었을 가능성을 시사한다. 충성 고객은 2016년 초에 유입이 많았지만, 2016년 말 이후로는 유입이 줄어들어 이 군집을 타깃으로 한 마케팅 전략이 필요하다. 이탈 위험 고객은 지속적으로 높은 신규 유입률을 보여주며, 이들

은 일시적으로 활동하는 고객일 가능성이 높다. 이러한 고객군이 지속적으로 존재하는 경우, 그들이 매출에 미치는 영향을 고려할 필요가 있다.

그림 41 월별 실제 활동하는 고객 분포



2016년 중반부터 불만족 고객과 이탈 위험이 있는 충성 고객의 활동이 현저히 줄어들며, 이들이 대부분 이탈했을 가능성이 높다고 볼 수 있다. 그러나 2017년부터 충성 고객의 활동이 급격히 증가하고, 이탈 위험 고객의 활동도 상승세를 보이기 시작했다. 이는 2016년에 활동이 감소한 이후, 2017년에 효과적인 마케팅 전략이 시행되었을 가능성을 시사한다. 이러한 변화는 마케팅의 성공적인 적용이 고객 활동에 어떻게 긍정적인 영향을 미칠 수 있는지를 보여주는 사례로 해석될 수 있다.

그림 42 월별 이탈 고객 분포

Churned Customer Count by Cluster Per Month, Considering Cluster-Specific Purchase Intervals



불만족 고객의 이탈은 예상되었지만, 이탈 위험이 있는 충성 고객이 2016년 중반부터 크게 이탈하기 시작했다는 점은 주목할 필요가 있다. 더욱이 2017년부터는 충성 고객과 VIP 고객까지도 이탈 추세를 보이기 시작했다. 그림 20에서는 충성 고객과 VIP 고객 군집에 신규 유입은 없었지만, 실제 활동하는 고객 수에는 큰 변동이 없음을 볼 수 있다. 이는 휴면 상태였던 고객들의 활동이 증가했음을 시사하며, 이는 2017년에 진행된 특별 이벤트나 프로모션의 영향일 가능성이 높다. 이러한 변화는 회사의 마케팅 전략이 고객 활동에 어떻게 긍정적으로 작용할 수 있는지를 보여주는 중요한 지표이다.

그림 43 월별 주문 빈도

Monthly Orders and Customer Count by Cluster

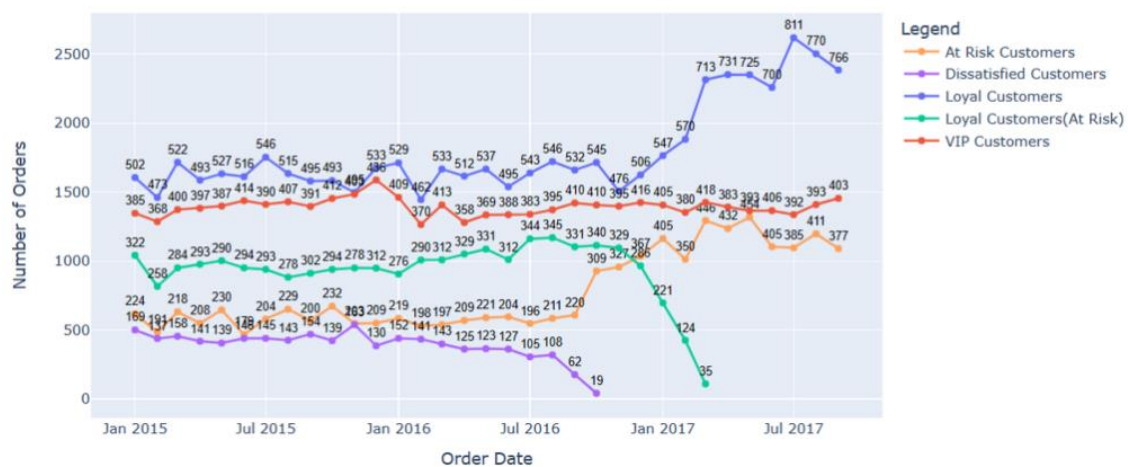
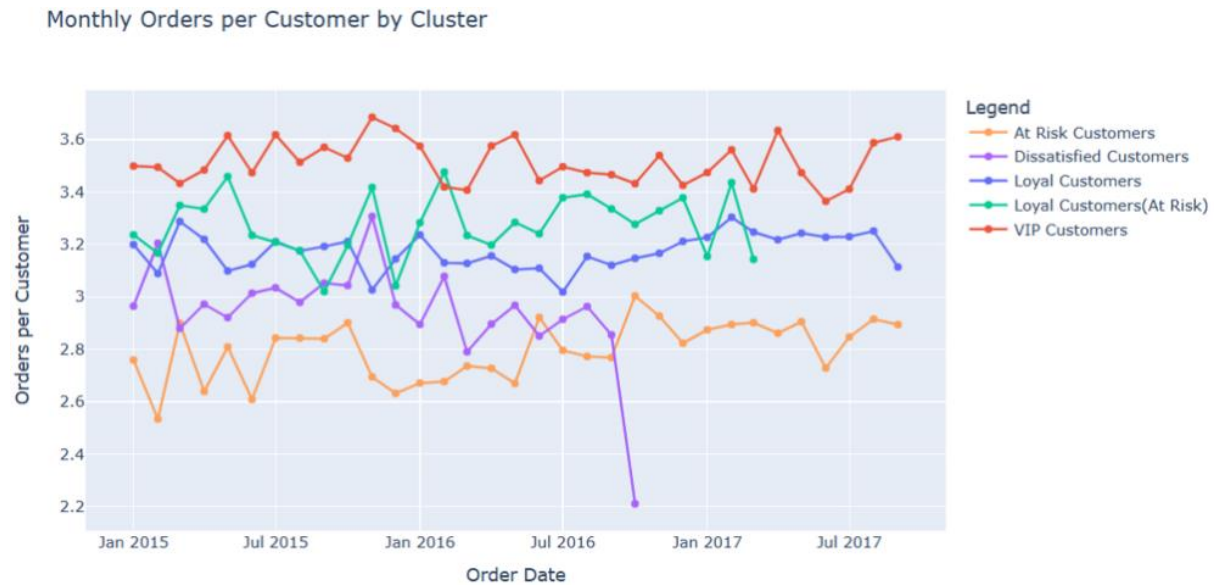


그림 44 월별 고객 당 주문 빈도



충성 고객과 VIP 고객의 높은 주문 빈도는 긍정적인 지표로 평가될 수 있다. 특히, 2017년부터 이탈 위험 고객의 주문 빈도가 증가한 것 역시 주목할 만하며, 이는 해당 연도에 특별 이벤트가 있었을 가능성을 시사한다. 그림 24에서는 VIP 고객과 충성 고객의 주문 수가 꾸준히 높은 것으로 나타났다. 이탈 위험이 있는 충성 고객 역시 주문 빈도가 높으나, 전체 고객 수에 비해 이미 상당수가 이탈한 것으로 보인다.

그림 45 월별 실제 수익 분포

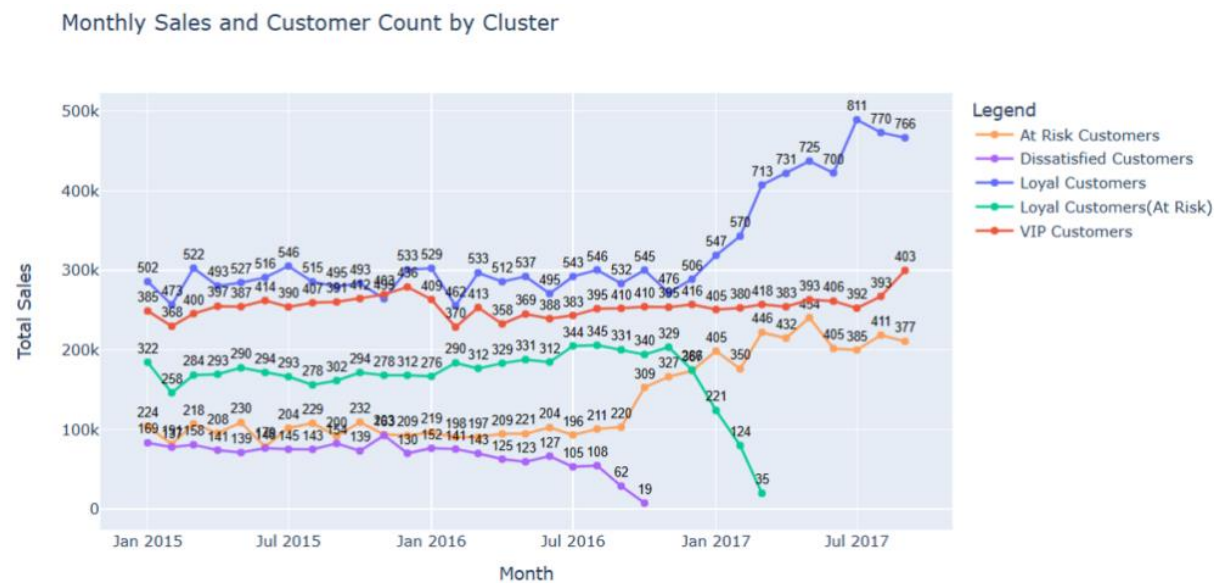


그림 46 월별 고객 당 실제 수익 분포

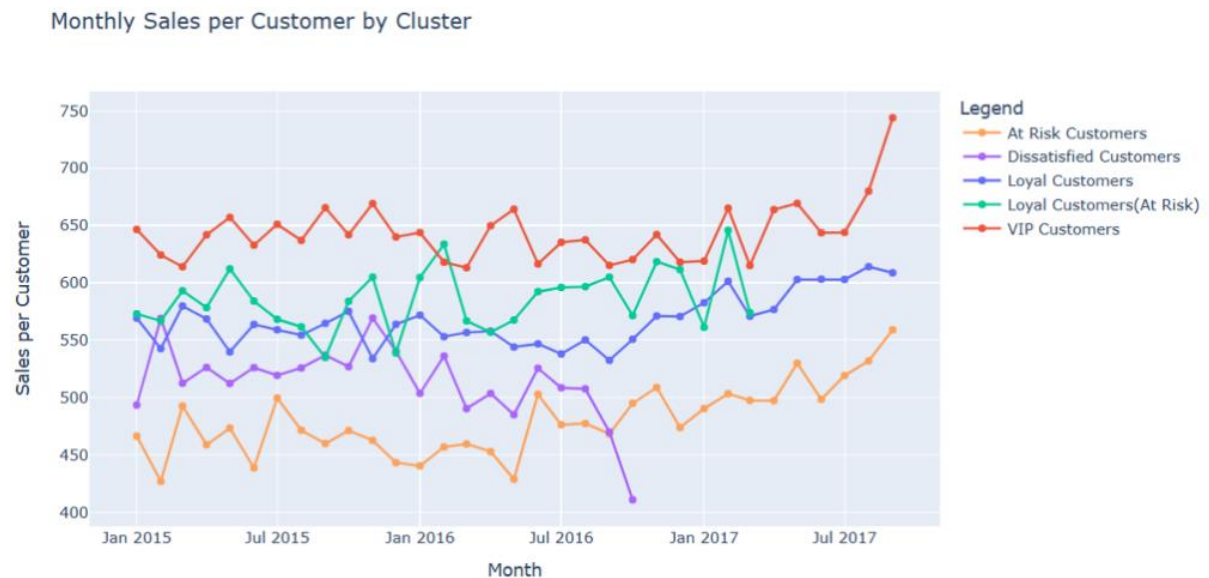


그림 45, 46의 주문 빈도 그래프와 비슷한 양상을 보인다. 2017년도부터 VIP 고객, 충성고객, 이탈위험 고객 모두 고객 당 수익이 증가하고 있는 것을 볼 수 있다. 이 또한 2017년에 무언가 이벤트가 있었음을 시사한다.

3.5.3.2 CLTV

CLTV(Customer Lifetime Value)는 고객 생애 가치를 나타내는 지표로, 고객이 기업에 어느 정도의 경제적 가치를 제공할 수 있는지를 측정한다. 이 지표는 각 고객이 생성할 예상 매출과 수익을 표현한다. 고객 행동 패턴을 분석하여 군집별로 가치 평가를 수행했으며, 이제 이를 보다 체계적이고 확실한 KPI인 CLTV로 설정하여 분석하려 한다.

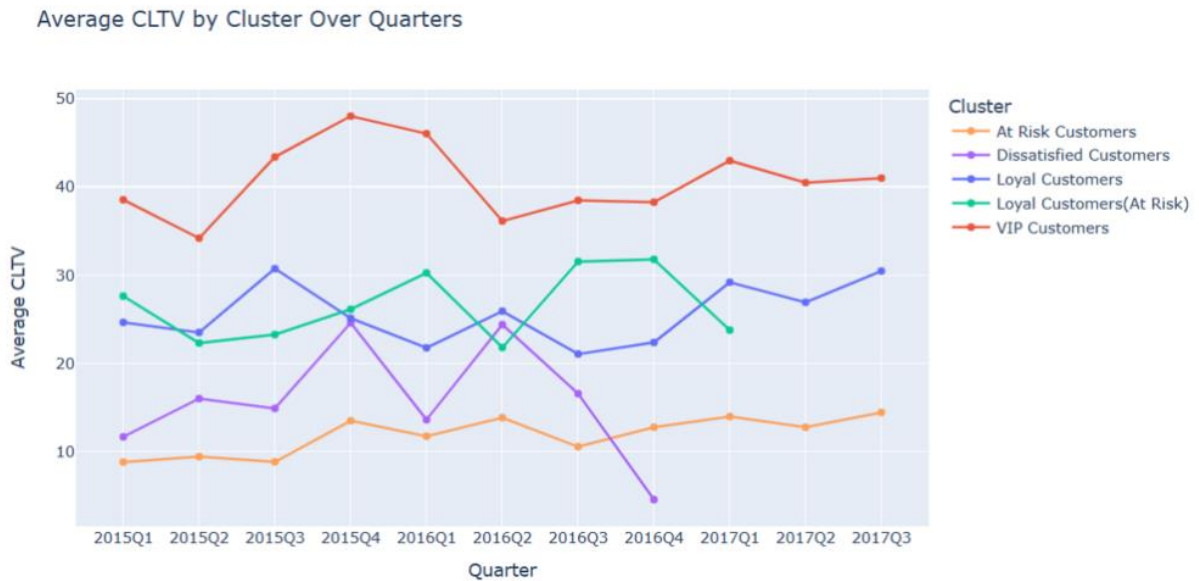
CLTV의 계산 방식은 다음과 같다.

CLTV = 고객 가치 * 평균 고객 수명

- 고객 가치 (=평균 구매 가치 * 평균 구매 빈도)
- 평균 구매 가치 (=특정 기간 내 총 수익 / 특정 기간 내 총 구매 횟수)
- 평균 구매 빈도 (=특정 기간 내 구매 횟수 / 특정 기간 내 총 고객 수)
- 평균 고객 수명 (=고객이 지속적으로 구매한 기간 / 총 고객 수)

여기서 평균 고객 수명 값은 1/평균 구매 간격으로 계산하였으며 군집별로 따로 구하여 진행했다.

그림 47 분기별 CLTV 분포



VIP 고객은 지속적으로 높은 값을 유지하며, 2017년부터 VIP, 충성 고객, 이탈 위험 고객 모두 상승세를 보이고 있다. 이 그룹에 프로모션과 이벤트를 집중하면 수익 창출에 크게 기여할 것으로 예상된다. 반면, 불만족 고객과 이탈 위험 충성 고객은 이미 이탈한 상태이므로 이들을 대상으로 한 이벤트는 진행하지 않는 것이 좋다. 이에 완전히 새로운 신규 고객을 유치하기 위한 마케팅 전략을 개발하는 방법을 추천한다.

4. 결론

4.1 분석 요약 및 제한점

할인율 및 취소율 분석의 결과로 최적의 할인율을 3%에서 9%로 제시했지만, 실제적인 구매 유도와 이익 최대화를 위해서는 더 많은 데이터와 실험을 필요로 한다. 배송 정책에서는 보다 보수적인 배송날짜 고지를 통해 취소율을 감소시킬 수 있는 방법을 제안했지만, 이는 지역별로 다른 배송 관련 투자가 필요함을 시사하며, 특정 대륙에 대한 배송 관련 투자를 늘릴 경우 다른 투자가 줄어들 수 있어 그 균형을 찾는 것이 중요하다.

재구매 분석에서는 높은 할인율이 반드시 재구매율을 높이지 않는다는 점을 확인했다. 이에 최적의 할인율을 유지하는 것이 기업에게 더 유리하며, 배송 지연을 최소화하여 고객의 만족도를 높이는 것이 중요하다. 그러나 '재구매'의 정확한 정의가 모호했고, 재구매를 촉진할 수 있는 획기적인 방법을 찾는 데 한계가 있다.

지역별 판매 분석에서는 지역별로 사회적 요인이나 문화적 차이에 따라 제품 선호도와 계절에 따른 매출이 다르게 나타남을 확인했다. 시장의 변화에 발 빠르게 대응하여 마케팅 전략을 조정해야 하지만, 이를 위해 추가적인 데이터, 즉 지역에 따른 제조사 데이터 등이 필요하다.

사기 의심 주문 분석에서는 특정 지역에서 사기 의심 사례가 자주 발생한다는 점을 파악하고, 계좌 이체 방식에서 특히 이러한 정황이 빈번하게 나타나는 것을 확인했다. 본인인증 절차를 강화하는 방안을 제안하면서, 다른 결제 방식에 비해 계좌 이체 방식에서 사기 의심 사례가 더 많이 발생한다는 점을 감안할 필요가 있다. 그러나 '사기 의심'의 정확한 기준이 명확하지 않으며, 이러한 주문 처리 상태를 정의하는 데 한계가 있다.

RFM 분석을 통해 고객 군집별 행동과 수익 기여도의 차이를 명확히 식별할 수 있었다. 분석 결과, 특히 이탈 위험이 높은 충성 고객 군집에서 이탈율을 조기에 감소시킬 수 있었다면 매출 증대에 크게 기여할 수 있었을 것으로 나타났다. 하지만, 이 군집의 고객들이 이미 이탈한 상태였으므로, 불필요한 투자를 줄일 수 있었다는 점에서 분석이 유용했다. 또한, 이탈 위험이 있는 고객층이 지속적으로 유입되고 있으며, 일시적으로 상품을 구매해보는 '찍먹' 고객층이 유지되고 있음을 확인했다. 이에 따라, 각 군집에 맞춘 맞춤형 전략을 수립하여 고객 만족도를 높이고, 이탈율을 감소시키며, 신규 고객 유입을 촉진할 수 있는 방안을 제안한다.

고객 행동 패턴의 시계열 분석에서는 2016년에 대규모 이탈과 2017년에 특별 프로모션 또는 이벤트가 있었던 것으로 추정되나, 이에 대한 자세한 데이터가 부족하여 확정 짓기 어렵다. 관련 데이터를 추가로 수집할 경우, 더욱 효과적인 마케팅 전략을 수립하는 데 도움이 될 것으로 보인다. 현재 RFM 값으로 고객을 세분화하고 있지만, 성별이나 나이와 같은 추가적인 지표를 통합한다면 더욱 구체적이고 효과적인 맞춤 전략을 개발할 수 있을 것이다.

4.2 마케팅 전략 수립

전략적 투자 및 마케팅 강화

투자 증대: 각 지역에서 수익성이 높은 제품 카테고리에 대한 투자를 확대해야 한다. 이는 해당 제품의 시장 점유율을 확대하고, 경쟁력을 강화하는 데 기여할 것이다.

마케팅 전략: 계절별로 변동하는 수요를 고려한 마케팅 전략을 수립하고 실행해야 한다. 특히, 수요가 높아지는 시즌에 맞춰 적극적인 프로모션과 광고 캠페인을 계획함으로써 판매를 최대화할 수 있다.

가격 재검토

가격 조정: 이익률이 낮은 제품 카테고리에 대해서는 가격 전략을 재검토할 필요가 있다. 이는 비용 효율성을 높이고, 장기적으로 수익성을 개선하는 데 중요하다.

VIP 고객 및 충성고객 관리

멤버십 혜택 강화: 고객의 구매 이력에 따라 차등화된 혜택을 제공하여 고객의 지속적인 구매를 유도한다. 예를 들어, 구매 금액이 높은 고객에게는 추가 할인 혜택 또는 포인트 적립을 제공한다.

설문조사를 통한 피드백 수집: 고객의 만족도를 주기적으로 평가하여 서비스 개선에 반영한다. 참여 유도를 위해 추첨을 통해 상품을 제공하는 등의 인센티브를 마련한다.

이탈 위험이 있는 고객 전략

이탈 위험이 높은 고객을 위한 적극적인 마케팅 캠페인: 개인화된 쿠폰 발행과 프로모션을 통해 재구매를 유도한다. 특히, 구매 빈도를 높이기 위한 전략으로 '단기적 할인 이벤트'를 진행한다.

신규 고객 유입 전략

마케팅 캠페인과 온라인 광고를 통해 브랜드 인지도를 높이고, 특히 소셜 미디어를 활용한 타겟 마케팅을 강화한다.

"찍먹 마케팅" 전략을 통해 소규모 구매를 장려하고, 첫 구매 고객에게는 특별 할인을 제공하여 브랜드에 대한 좋은 첫인상을 제공한다.