



# OLTP REPORT

TEAM 3팀 1조

전도균, 오현지, 방지희, 정정룡, 장동현

# 2016~2022 OLTP 분석: 고물가 시대 생존을 위해

2024년 3월 소비자물가지수가 전년 동월 대비 3.1%, 전월 대비 0.5% 증가하는 등 물가 상승이 계속되고 있다.. 이러한 물가 상승은 일반 소비자들의 구매력에 직접적인 영향을 미치고 있으며, 이는 가격에 대한 민감도가 높아지는 현상으로 이어지고 있다.

이러한 경제 환경 하에서, 이커머스 시장은 특히 더 큰 영향을 받고 있다. 소비자들이 더 신중하게 지출을 결정함에 따라, 이커머스 시장의 매출도 줄어들고 있는 상황이다. 이에 따라, 많은 이커머스 기업들은 시장에서의 경쟁력을 유지하고, 소비자들의 변화하는 요구에 대응하기 위해 다양한 전략을 모색하고 있다. 이런 상황 속에서 한 가지 접근 방식은 과거의 데이터를 분석하여 소비자 행동의 변화를 이해하고, 이를 바탕으로 더 나은 서비스를 제공하려는 노력이다. 특히, 2016년부터 2022년까지의 온라인 트랜잭션 처리 데이터(OLTP) 분석은 기업들이 과거 소비자 행동의 패턴을 파악하고, 시장 동향을 이해하는데 큰 도움이 될 것이다.

이러한 분석을 통해, 기업들은 소비자 만족도를 향상시키는 데 필요한 조치들을 취할 수 있을 뿐만 아니라, 기업 이윤을 증대시키는 전략도 마련할 수 있을 것이다. 소비자 만족도를 높이기 위한 조치로는 개인화된 쇼핑 경험 제공, 가격 정책의 최적화, 고객 서비스의 질 향상 등이 있을 수 있다. 또한, 효과적인 재고 관리와 물류 최적화를 통해 운영 비용을 절감하고, 이윤 증대에 기여할 수 있는 방안도 모색될 수 있다.

## 활용기술

### MySQL

SQL 쿼리를 작성하여 데이터 분석 및 패턴 발견, 비즈니스 인사이트 도출

### Python

결측치 대체 및 데이터 형 변환 등 데이터 전처리 수행

### Excel

python 전처리 수행 데이터 확인 및 MySQL 작업 데이터 추출하여 확인

### Tableau

SQL로 추출한 CSV 데이터와 연결 및 시각화 요소를 조합하여 고객과 기업측면으로 데이터 요약 정보를 제시하는 대시보드 작성

## 정정룡 - 기여도 20%

데이터 전처리(python) 및 코드 통합

평점, 리뷰수, 긍정 리뷰 비율, 재구매 데이터 분석 및 상품 추천 대시보드 작성

지역별, 상품별, 이유별 환불 및 이유 분석 및 시각화

캠페인 주차별 구매수량 및 매출 분석 및 시각화

## 방지희 - 기여도 20%

국가별 매출, 카테고리별 판매량 많은 상위 3개 국가 분석 및 시각화

연도/ 분기/ 월별 매출 분석 및 시각화

각각고객용, 기업용으로 나누어 대시보드 통합

## 오현지 - 기여도 20%

캠페인 종류에 따른 할인율, 카테고리별 인기상품, 판매량 상위 10개 국가 인기상품 분석 및 시각화

프로젝트 결과물 소개 PPT 제작

활용 기술 및 프레임워크, 참여자 정보 및 역할 등 보고서 작성 보조

## 장동현 - 기여도 20%

결제수단별 사용 빈도, 제품별 환불 경향 데이터 분석 및 시각화

주제 선정 이유, 프로젝트 세부 결과&결론, 기대효과 등 보고서 전반 작성

## 전도균 - 기여도 20%

마케팅 캠페인 유무에 따른 판매량과 매출액 비교 분석 및 시각화

마케팅별 판매량이 높은 서브 카테고리 분석 및 시각화

프로젝트 결과물 영상제작



Part One

## 4 데이터 소개 및 EDA

- 1.1 스키마맵 파악
- 1.2 데이터 전처리
- 1.3 데이터 EDA

Part Two

## 17 대시보드 분석

- 2.1 고객용 대시보드
- 2.2 기업용 대시보드

Part Three

## 31 기대효과&결론

- 3.1 기대효과
- 3.2 결론

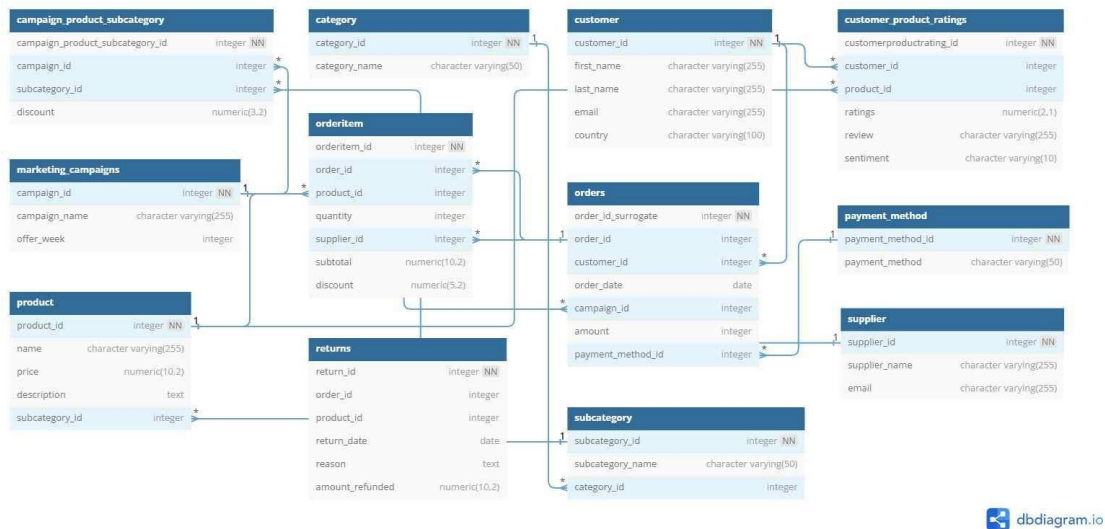


Part One

# 데이터 소개 및 EDA



## 1.1 스키마맵 파악



<https://www.kaggle.com/datasets/sharangkulkarni/oltp-ecommerce-data?select=orders.csv>

의 2016~22년 OLTP 데이터를 활용하였다.

스키마맵을 활용하여 데이터들의 관계를 파악하였고, 파악한 데이터를 통해 고객용 및 기업용 대시보드를 개발하여, 고객들과 기업들이 보다 넓은 범위의 정보를 이용할 수 있도록 지원하고자 한다.

고객용 대시보드는 일반 사용자들이 쉽게 접근하여 활용할 수 있는 형태로 구성될 것이다. 이 대시보드는 사용자들이 제품 또는 서비스와 관련된 데이터를 쉽게 시각화하고 해석할 수 있도록 설계될 것이다.

기업용 대시보드는 기업 내부에서 사용되며, 전략적인 의사결정을 지원하기 위해 설계될 것이다. 이 대시보드는 다양한 부서 및 팀이 사용할 수 있는 형태로 구성될 것이고 기업의 전반적인 건강 상태 및 성과를 파악하고, 전략을 개선하는 데 필요한 인사이트를 제공할 것이다.

## 1.2 데이터 전처리

특정 테이블의 필드 값에 관련된 문제에 직면하게 되었다. 'orders'라는 테이블에서, 'campaign\_id'라는 필드가 있는데, 이 필드에는 정수형(INT) 데이터가 저장되어야 한다. 그런데, 이 과정에서 'campaign\_id' 필드에 NULL 값이 존재하는 것을 발견했다. 보통 NULL 값은 데이터가 없거나 알 수 없는 경우를 나타내지만, 이 상황에서는 NULL 값이 실제로는 '캠페인이 진행되지 않은 기간'을 의미한다는 것을 파악했다.

데이터의 일관성과 정확성을 유지하기 위해, 'marketing\_campaigns'라는 테이블에 'NoCampaign'이라는 특별한 행을 추가하기로 결정했다. 이는 캠페인이 진행되지 않은 기간을 명시적으로 표시하기 위한 조치였다. 이 새로운 행에 할당된 'campaign\_id'는 17로 설정했다. 이는 데이터베이스에서 'AUTO\_INCREMENT' 속성이 설정된 필드에 0 값을 직접 할당할 수 없기 때문에, 자동으로 다음 숫자인 17을 사용한 것이다.

그 후, 'orders' 테이블에서 'campaign\_id'가 NULL인 모든 레코드를 찾아서 이들의 'campaign\_id'를 17로 변경했다. 이 변경 작업을 통해, 이전에 NULL로 표시되었던 값들이 이제는 '캠페인이 진행되지 않았음'을 명확하게 나타내는 17로 대체되었다. 이 과정에서 해당 필드의 데이터 타입을 정수형(INT)으로 변경하여 모든 데이터가 올바르게 저장되도록 했다.

'campaign\_product\_subcategory' 테이블에도 유사한 조치를 취했다. 여기서는 'subcategory\_id' 별로 'campaign\_id'가 17인 경우, 'discount' 값이 0이 되도록 총 100개의 행을 추가적으로 삽입했다. 이렇게 함으로써, 캠페인이 진행되지 않은 기간 동안의 할인 정보도 명확히 표현할 수 있게 되었다.

## 1.3.1 고객용 데이터

### 1. 평점 평균, 리뷰수, 긍정리뷰비율

#### 1) 상품별

product_id	AVG_ratings	review_count	good_sentiment_ratio
6	4.20000	5	1.0000
315	4.14286	7	1.0000
62	4.00000	4	1.0000
343	4.00000	6	1.0000
228	3.95455	11	0.9091
453	3.95455	11	0.9091
157	3.88889	9	0.8889
208	3.88889	18	0.8333
146	3.87500	8	0.7500
87	3.83333	3	0.6667

평점, 리뷰 수, 그리고 '좋아요' 비율 같은 요소들은 상품의 인기와 고객 만족도를 파악하는데 매우 유용한 지표로 활용될 수 있다. 이러한 데이터를 종합적으로 분석하여, 각 상품의 성능을 종합적으로 평가함으로써 우수한 상품을 식별하고 추천할 수 있다.

평가 기준을 바탕으로 우수한 상품을 선별한 후, 이를 재구매율 분석에 활용한다면, 재구매 가능성이 높은 우수 상품을 효과적으로 추천할 수 있게 된다. 재구매율은 고객의 충성도와 상품에 대한 지속적인 만족도를 나타내는 중요한 지표이기 때문에, 이를 통해 상품의 장기적인 성공 가능성을 예측할 수 있다. 상품의 초기 평가에서부터 재구매율 분석에 이르기까지, 이러한 데이터 기반 접근 방식은 고객에게 가장 만족스러운 상품을 제안하고, 최종적으로는 기업의 매출 증대에 기여할 수 있는 전략적인 방법이 될 수 있다.



## 2) 카테고리별

category_id	category_name	avg_ratings	review_count	good_sentiment_ratio
9	Automotive	3.19030	268	0.5709
13	Pet Supplies	3.11778	225	0.5644
15	Movies & TV	3.07865	178	0.5618
2	Clothing	3.07386	264	0.4924
17	Jewelry	3.07115	260	0.5500
6	Toys & Games	3.06597	288	0.5174
3	Home & Kitchen	3.04545	253	0.5336
12	Grocery	3.04102	256	0.5039
8	Health & Household	3.02402	229	0.5109
11	Office Products	3.02390	251	0.4861
19	Shoes	3.01073	233	0.5064
5	Sports & Outdoors	3.00840	238	0.5168
16	Video Games	3.00628	239	0.5063
20	Watches	3.00581	258	0.4845
1	Electronics	3.00160	312	0.5128
18	Computers	2.97596	208	0.5481
14	Musical Instruments	2.94378	249	0.4538
7	Beauty & Person...	2.94106	246	0.4512
4	Books	2.91697	277	0.4549
10	Tools & Home Im...	2.89366	268	0.4552

상품 리뷰 수의 분포는 특정 카테고리에 편중되어 있는 경우, 리뷰의 질과 그에 따른 긍정 리뷰 비율의 해석에 있어 주의를 기울여야 한다. 만약 한 카테고리의 상품들이 대다수의 리뷰를 독점 한다면, 그 카테고리의 인기나 상품의 질이 다른 카테고리의 상품보다 반드시 높다고 단정 지을 수 없다. 이는 시장의 현재 수요나 소비자의 선호도가 특정 카테고리에 집중되어 있음을 반영할 수 있으며, 이로 인해 다른 카테고리의 상품들이 상대적으로 적은 주목을 받게 되어 그 가치가 과소평가될 위험이 있다.

이러한 상황에서 긍정 리뷰의 비율을 평가할 때, 특정 카테고리에 리뷰 수가 집중되어 있다면, 그 카테고리의 상품들만이 좋은 평가를 받고 있는 것처럼 보일 수 있다. 그러나, 실제로는 다른 카테고리의 상품들도 높은 품질과 고객 만족도를 가질 수 있음에도 불구하고, 리뷰 수가 적기 때문에 이러한 긍정적인 평가가 충분히 반영되지 않는 문제가 발생할 수 있다. 따라서, 리뷰의 수가 특정 카테고리에 집중되지 않고 고르게 분포되어 있음이 중요하다. 이를 통해 모든 카테고리의 상품들이 공정하게 평가받을 수 있으며, 긍정 리뷰 비율의 의미 또한 각 카테고리에 걸쳐 일관적으로 유지될 수 있다.

이 경우에 해당하지 않았다는 점은, 리뷰 수가 모든 카테고리에 걸쳐 상당히 균등하게 분포되어 있음을 의미한다. 이는 다양한 카테고리의 상품들이 모두 소비자로부터 주목받고 있으며, 각 카테고리의 상품들이 공정한 기회를 통해 평가받고 있음을 나타낸다. 따라서, 이런 상황에서 긍정 리뷰의 비율을 평가할 때, 이 비율은 상품의 질과 고객 만족도를 더욱 정확하게 반영할 수 있다.

### 3) 국가별

country	avg_ratings	review_count	good_sentiment_ratio
Cambodia	3.75000	6	0.8333
United Arab Emirates	3.75000	6	0.6667
Bermuda	3.63636	11	0.8182
Kenya	3.57692	13	0.6923
Somalia	3.56250	8	0.7500
Barbados	3.52941	17	0.7059
American Samoa	3.51020	49	0.7143
United States of America	3.50000	13	0.7692
Monaco	3.48077	26	0.6923
Comoros	3.44737	19	0.7895
Sao Tome and Principe	3.43750	16	0.6875
Niue	3.42000	25	0.6800
Dominican Republic	3.40909	11	0.6364
Tokelau	3.40741	27	0.6667

204개의 다양한 국가가 존재하는데, 이 모든 국가들을 대상으로 분석을 진행하는 건 실질적으로 매우 비효율적인 방법이 될 수 있다. 모든 국가를 일일이 분석하기 위해서는 엄청난 시간과 자원이 필요하게 될 것이므로 가장 먼저 해야 할 일은 분석의 타겟을 명확히 정의하는 것이다.

이렇게 타겟을 선정함으로써, 분석의 효율성을 크게 높일 수 있고, 더 깊이있고 의미 있는 인사이트를 얻을 수 있게 될 것이다. 물론, 이 과정에서는 분석의 목적과 사용할 수 있는 데이터의 종류, 그리고 자원의 한계 등을 고려해야 한다. 분석을 위해 선택된 국가들이 실제로 분석 목적에 부합하는지, 그리고 충분한 데이터를 확보할 수 있는지 등을 면밀히 검토해야 해야할 것이다.

#### 4) 소비자별, 인당

customer_id	avg_ratings	review_count	good_sentiment_ratio
497	4.20000	5	1.0000
486	4.18182	11	0.8182
349	4.10000	5	1.0000
33	4.00000	7	1.0000
39	3.92857	7	0.8571
194	3.87500	4	1.0000
162	3.83333	6	1.0000
439	3.83333	6	0.6667
99	3.80000	5	0.6000
149	3.80000	10	0.8000
136	3.78571	7	0.8571
174	3.78571	7	0.7143
52	3.75000	8	0.8750

avg_ratings	avg_review_count	avg_good_sentiment_ratio
3.012921520	10.0000	0.50612040

인당 평균적으로 10개의 리뷰를 작성한다는 것을 보면, 모든 고객이 자신의 경험을 공유하려는 의지가 있다는 것을 알 수 있다.

즉, 고객들은 단순히 소비자로서의 역할을 넘어서, 자신들의 경험을 통해 다른 사람들에게 정보를 제공한다는 것이다. 이러한 과정에서 모든 고객이 평점을 매기거나 리뷰를 남기는 것을 통해 그들의 적극성을 명확히 파악할 수 있고 그들이 제품이나 서비스에 대해 얼마나 깊이 관여하고 있는지를 보여주는 지표가 될 수 있다.



## 2. 재구매 분석

customer_id	country	category_name	subcategory_name	product_name	product_id	purchase_count	repurchase_count	avg_repurchase_time	total_purchase_count
1	Saint Vincent and the Grenadines	Electronics	Smartphones	Smartphones - Product 1	1	1	0	0.0000	3082
2	Malaysia	Electronics	Smartphones	Smartphones - Product 1	1	1	0	0.0000	2669
3	Lao People's Democratic Republic	Electronics	Smartphones	Smartphones - Product 1	1	1	0	0.0000	2524
4	United Kingdom	Electronics	Smartphones	Smartphones - Product 1	1	1	0	0.0000	2960
5	Martinique	Electronics	Smartphones	Smartphones - Product 1	1	2	1	1585.0000	3086
6	Faroe Islands	Electronics	Smartphones	Smartphones - Product 1	1	1	0	0.0000	2818
7	Saint Kitts and Nevis	Electronics	Smartphones	Smartphones - Product 1	1	2	1	797.0000	3316
8	Guatemala	Electronics	Smartphones	Smartphones - Product 1	1	2	1	2228.0000	3557
9	Burundi	Electronics	Smartphones	Smartphones - Product 1	1	2	1	920.0000	3314
10	China	Electronics	Smartphones	Smartphones - Product 1	1	2	1	1399.0000	3616

### 고객에게 어떤 상품을 어떤 근거로 추천할 것인가?

고객에게 상품을 추천할 때는 다양한 요소들을 고려해야 한다.

첫 번째 단계로, 몇 개의 카테고리를 추천하는 것이 적절할지 고려해야 한다. 이를 위해서는 소비자가 한 번 구매할 때 평균적으로 몇 가지 카테고리의 상품을 구매하는지의 데이터를 분석하는 것이 중요하다. 이 데이터는 고객의 구매 패턴과 선호도를 이해하는 데 필수적이며, 이를 바탕으로 고객에게 가장 관련성 높은 카테고리를 선별하여 추천할 수 있다.

두 번째로, 선별된 카테고리 내에서도 고객이 서브 카테고리를 자유롭게 재선택할 수 있도록 하는 것이 중요하다. 이 과정에서 고객의 편의를 높이기 위해 우수한 서브 카테고리는 크기를 더 크게 구성하여 시각적으로 강조할 수 있다. 여기서 '우수함'은 서브 카테고리의 평점, 재구매율 등의 지표로 평가할 수 있다. 또한, 고객이 커서를 해당 서브 카테고리에 올렸을 때, 평균 평점, 재구매율 등 서브 카테고리의 전반적인 정보를 쉽게 확인할 수 있도록 정보 제공 방식을 구성하는 것이 좋다. 이는 고객이 보다 정보에 기반하여 의사 결정을 할 수 있도록 돕는다.

마지막으로, 고객이 서브 카테고리를 선택한 후에는 해당 서브 카테고리 내의 상세한 상품 정보를 확인할 수 있도록 구성하는 것이 중요하다. 이 단계에서는 상품의 상세 설명, 사용자 리뷰, 평점, 가격 등을 포함한 상품의 모든 정보를 제공하여 고객이 자신의 필요와 선호에 가장 잘 맞는 상품을 선택할 수 있도록 돕는다. 이렇게 단계별로 고객의 선택을 유도하고 정보를 제공하는 것은 고객의 쇼핑 경험을 개선하고, 만족도를 높이는 데 크게 기여할 수 있다.

customer_id	avg_category_per_order
1	4.56757
2	4.86905
3	4.29348
4	4.82828
5	5.01020
6	4.64894
7	4.40870

소비자 별로 각 구매 건마다 평균적으로 몇 가지의 카테고리를 구입하는지 확인하는 과정을 통해, 보통 4~5가지 카테고리에 해당하는 상품을 구매한다는 사실을 알 수 있었다. 이는 소비자들이 단일 카테고리에만 국한되지 않고, 여러 카테고리에 걸쳐서 상품을 선택함으로써 다양한 욕구와 필요를 충족시키려고 한다는 것을 시사한다.

이 데이터를 통해 계산된 전체 평균이 4.75라는 수치는 소비자들이 상당히 광범위한 카테고리에 서 상품을 선택하고 있음을 나타내고, 이러한 평균치는 소비자들이 한 번의 쇼핑 세션에서 비교 적 다양한 종류의 상품에 관심을 가지고 있음을 보여준다. 또한, 이 정보는 상품을 추천하거나 마케팅 전략을 세울 때 매우 중요한 지표가 될 수 있다. 예를 들어, 상품을 추천할 때는 소비자 가 관심을 가질 만한 여러 카테고리를 고려하여, 그들의 다양한 관심사와 필요를 만족시킬 수 있는 상품을 제안하는 것이 유리할 것이다.

customer_id	category_name	purchase_count	repurchase_rate	avg_ratings	review_count	good_sentiment_ratio	ranking
1	Health & Household	42	0.73810	3.02402	229	0.51092	1
1	Toys & Games	38	0.71053	3.06597	288	0.51736	2
1	Clothing	36	0.69444	3.07386	264	0.49242	3
1	Books	34	0.79412	2.91697	277	0.45487	4
1	Pet Supplies	34	0.76471	3.11778	225	0.56444	5
2	Jewelry	36	0.75000	3.07115	260	0.55000	1
2	Clothing	33	0.75758	3.07386	264	0.49242	2

각 소비자별로 가장 선호하는 상품 카테고리를 파악하는 것은 맞춤형 상품 추천 시스템을 구축하는 데 있어 핵심적인 접근 방법 중 하나다. 이 과정에서 소비자가 가장 많이 구입하는 상위 5 개의 카테고리 데이터 세트를 생성하는 것은 매우 중요한 첫걸음이다. 이 데이터 세트를 통해 우리는 각 소비자가 어떤 유형의 상품에 대해 높은 관심을 가지고 있는지, 그리고 그들의 구매 행태가 어떠한지를 더욱 정밀하게 이해할 수 있다.

구매 횟수를 기준으로 상위 카테고리를 선별하는 것 외에도, 구매 횟수가 동일한 경우에는 재구매율과 평균 평점을 추가적인 기준으로 사용하여 우선 순위를 정한다. 이는 단순히 많이 구매된 상품이 좋은 상품이라는 것을 넘어서, 소비자들이 얼마나 그 상품에 만족하고 있는지, 그리고 그 들이 해당 상품을 지속적으로 구매할 의향이 있는지를 종합적으로 평가하는 것이다.

이러한 분석 방법을 통해 생성된 데이터는 소비자 개인별 맞춤형 추천 시스템의 정확성과 효율 성을 대폭 향상시킬 수 있다.



## 1.3.2 기업용 데이터

### 1. 환불 분석

모든 상품에서 환불이유를 찾고 해결하는 것은 중요하나, 특히 빠르게 대처해야 하는 상품은 무엇 일지, 정말 문제가 있는 상품이 있는 건지 등을 나타낼 수 있는 타겟을 설정한 분석 진행을 꼭 표로 한다.

#### 1) 환불 이유별 횟수

reason	total_refund_count
Changed my mind	278
Color doesn't match	274
Found a better deal	266
Product didn't meet expectations	257
Defective product	252
Ordered by mistake	250
Size doesn't fit	247
Item not as described	234
Product damaged during shipping	232
Received wrong product	210

환불 요청 과정에서 이용 가능한 환불 이유가 오직 10가지로 제한되어 있다는 점은, 고객이 환불 요청 시 직접 글을 작성하여 이유를 제시하는 대신, 사전에 정의된 몇 가지 이유 중에서 선택해야 한다는 것을 의미한다. 이러한 방식은 환불 처리 과정을 표준화하고 간소화하는데 도움을 줄 수 있으며, 고객에게는 보다 명확하고 간편한 환불 요청 절차를 제공한다.

## 2) 지역별 환불 분석

country	reason	refund_count	total_refunds_per_country
Lao People's Democratic Republic	Changed my mind	10	50
Lao People's Democratic Republic	Product didn't meet expectations	9	50
Lao People's Democratic Republic	Received wrong product	7	50
Lao People's Democratic Republic	Ordered by mistake	6	50
Lao People's Democratic Republic	Color doesn't match	5	50
Poland	Product didn't meet expectations	6	37
Poland	Found a better deal	6	37
Poland	Size doesn't fit	5	37
Poland	Item not as described	4	37
Poland	Product damaged during shipping	4	37
Mozambique	Defective product	8	34

204개국에 걸쳐 물건을 판매하는 것은, 매우 복잡한 물류 및 배송 체계를 요구한다. 이렇게 광범위한 지역에 상품을 배송하는 과정에서는, 각 나라의 물류 구조와 관련된 다양한 문제점이 발생할 수 있다.

환불 이유별 횟수를 분석했을 때, 전반적으로 배송 오류에 관련된 항목이 낮은 비율을 차지한다고 할지라도, 환불이 많이 발생하는 특정 국가들에서 상위 환불 이유가 주로 배송 오류와 관련 되어 있다면, 이는 우리의 가설이 실제로 맞아떨어진다는 것을 의미한다. 실제로 확인 결과, 이 가설은 정확하게 들어맞았다.

따라서, 이는 해당 국가에서 담당하는 물류업자나 그 국가의 물류 구조 자체에 문제가 있을 가능성이 높음을 시사한다. 이러한 인사이트는 우리에게 각 나라별로 물류 체계를 면밀히 파악하고, 필요한 경우에는 물류 업체를 교체하거나 배송 프로세스를 개선할 필요가 있음을 알려준다. 또한, 이는 우리가 물류 및 배송 시스템의 효율성을 향상시키기 위해 어떤 조치를 취해야 할지 에 대한 구체적인 방향성을 제공한다.

국가별로 상이한 물류 환경을 고려하여, 배송 프로세스를 개선하고 최적화한다면, 배송 오류로 인한 환불 건수를 현저히 줄일 수 있을 것이다.

### 3) 환불율 분석

product_id	product_name	subcategory_id	subcategory_name	category_id	category_name	total_sales	total_refunds	product_refund_rate	avg_refund_rate	stddev_refund_rate
151	Skincare - Product 1	31	Skincare	7	Beauty & Personal Care	516	11	0.02132	0.011572000	0.005222861910782273
478	Analog Watches - Product 3	96	Analog Watches	20	Watches	568	12	0.02113	0.011502000	0.004917358799236941
256	Stationery - Product 1	52	Stationery	11	Office Products	515	9	0.01748	0.008680000	0.004771793699093111
454	Running Shoes - Product 4	91	Running Shoes	19	Shoes	541	9	0.01664	0.010510000	0.0033029141648552958
385	Adventure - Product 5	77	Adventure	16	Video Games	555	9	0.01622	0.008098000	0.004253555608478328
261	Printers - Product 1	53	Printers	11	Office Products	501	8	0.01597	0.009388000	0.0035550794715752507
146	Educational Toys - Product 1	30	Educational Toys	6	Toys & Games	564	9	0.01596	0.009198000	0.0035587483606598206
410	Necklaces - Product 5	82	Necklaces	17	Jewelry	542	8	0.01476	0.009506000	0.002743168709366057
344	Wind Instruments - Product 4	69	Wind Instruments	14	Musical Instruments	545	8	0.01468	0.004768000	0.005390418451760101
277	Fresh Produce - Product 2	56	Fresh Produce	12	Grocery	561	8	0.01426	0.011506000	0.001428917590094348

각 상품의 총 판매 대비 환불 횟수를 비율로 나타낸 `product_refund_rate`를 통해 어떤 상품들이 상대적으로 높은 환불률을 보이는지 파악할 수 있다. 이는 제품의 품질 문제, 고객의 기대치 미충족, 또는 물류 과정 중의 문제 등을 나타낼 수 있다.

서로 다른 카테고리(`category_name`) 및 서브카테고리(`subcategory_name`)에 대한 환불율을 비교하여, 특정 범주에서의 이슈를 파악할 수 있다. 예를 들어, 'Skincare - Product 1'은 환불률이 상대적으로 높게 나타나 이에 대한 추가 조사가 필요할 수 있다.

`stddev_refund_rate`는 환불율의 변동성을 나타내며, 이를 통해 어떤 상품이 다른 상품에 비해 환불율이 일정하지 않은지(즉, 불규칙한 패턴을 보이는지)를 확인할 수 있다.

높은 환불율을 보이는 상품(`product_refund_rate`가 높은 상품)에 대해서는 추가적인 품질 검사, 고객 피드백 수집, 혹은 설명 및 마케팅 전략의 재검토가 필요할 수 있다.

높은 판매량을 보이는 상품들에서의 환불률을 분석하여, 판매량이 많음에도 불구하고 환불률이 낮은 상품들을 벤치마킹할 수 있다.





Part Two

# 대시보드 분석



총 매출액 \$781,526,428

1. Lao People's Democratic Republic - \$14,960,696
2. Mozambique - \$12,500,798
3. United States Virgin Islands - \$11,749,117
4. Iraq - \$10,791,790
5. Sudan - \$8,943,343

...

## 2.1 고객용 대시보드



이 시스템은 고객별 맞춤 상품 추천을 위해 세심하게 구성된 데이터 시각화 도구를 포함하고 있다. 사용자는 최상단의 인터페이스를 통해 특정 고객을 선택할 수 있으며, 이를 통해 해당 고객에 대한 맞춤형 상품 추천 과정이 시작된다.

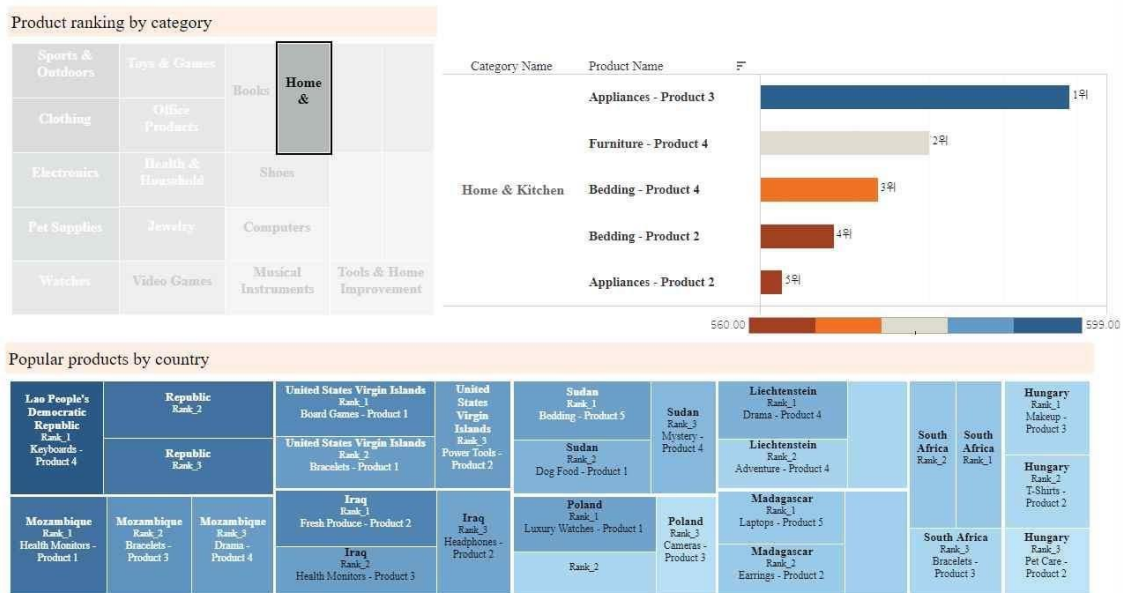
선택된 고객에 대한 분석은 좌측 상단에서 시작되는데, 여기서는 고객이 가장 많이 구입한 상위 5개 카테고리 크기를 변화시켜 시각적으로 표현한다. 이 크기 변화는 각 카테고리의 구매횟수를 반영하여, 고객의 구매 선호도를 명확하게 드러내준다.

우측 상단에서는 좌측 상단에서 선택된 카테고리에 대한 더 세부적인 분석을 제공합니다. 여기서는 해당 카테고리 내의 서브 카테고리를 우수성 순서대로 배열하며, 우수성은 재구매율, 평점, 그리고 긍정적 리뷰 비율의 순위 총합에 따라 결정된다. 서브 카테고리의 중요성은 크기와 음영을 통해 강조되어, 고객에게 더 우수한 제품을 추천할 수 있는 기반을 마련한다.

마지막으로, 하단 부분에서는 우측 상단에서 선택된 서브 카테고리 내의 특정 상품에 대한 상세한 정보를 제공한다. 이 정보에는 상품의 평점, 리뷰 수, 긍정적 리뷰 비율, 그리고 재구매율이 포함되어 있어, 추천하려는 상품이 고객의 기대와 요구를 충족시킬 수 있는지에 대한 평가를 가능하게 한다.

이렇게 구성된 시스템은 고객의 과거 구매 이력과 선호도를 기반으로 한 맞춤형 상품 추천을 가능하게 하며, 고객 만족도를 높이고 재구매를 촉진하는데 중요한 역할을 한다. 상품의 세부적인 특성을 깊이 분석하고, 고객별 맞춤 추천을 통해 더 개인화된 쇼핑 경험을 제공함으로써, 기업은 고객 충성도와 브랜드 가치를 동시에 높일 수 있다.





이 시각화 도구는 사용자에게 효과적인 데이터 탐색 방법을 제공하기 위해 두 개의 히트맵과 상호작용하는 막대그래프로 구성되어 있다. 좌측 부분에 위치한 첫 번째 히트맵은 다양한 카테고리에 걸쳐 상품의 인기를 색상의 강도로 나타낸다. 이 히트맵의 각 항목을 클릭하면, 해당 카테고리 내 상품 순위를 나타내는 막대그래프가 오른쪽 부분에 동적으로 표시된다. 이 막대그래프는 사용자가 특정 카테고리에서 가장 인기 있는 상품이 무엇인지를 한눈에 볼 수 있게 해주며, 상품의 인기를 기반으로 한 순위 정보를 제공한다.

하단 부분에 위치한 두 번째 히트맵은 나라별로 잘 팔리는 상품을 보여준다. 이 히트맵은 각 나라와 상품 간의 판매 성공률을 색상의 강도로 시각화 하여, 특정 지역에서 어떤 상품이 인기 있는지를 쉽게 파악할 수 있게 한다. 이러한 방식으로, 사용자는 글로벌 시장에서의 상품 선호도와 판매 동향을 효과적으로 이해할 수 있으며, 시장 진입 전략이나 마케팅 계획을 수립하는 데 있어 귀중한 인사이트를 얻을 수 있다.

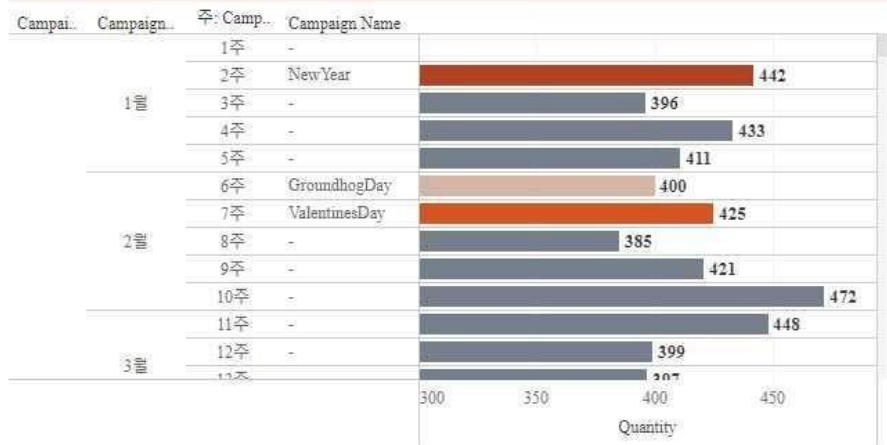
이 시각화 도구의 통합된 접근 방식은 복잡한 데이터를 쉽게 탐색하고 분석할 수 있게 해주며, 다양한 관점에서 시장과 소비자의 행동을 이해하는 데 도움을 준다. 사용자는 카테고리별 인기 상품의 순위를 확인하는 것에서부터 나라별로 잘 팔리는 상품의 트렌드를 파악하는 것까지, 맞춤형 정보를 신속하게 얻을 수 있게 된다. 이는 기업이 보다 정확한 타겟팅과 전략적인 의사결정을 할 수 있게 지원하는 동시에, 시장의 변화와 소비자의 요구에 민첩하게 대응할 수 있는 기반을 마련해 준다.

# Campaign

## Comparison of Sales by Campaign



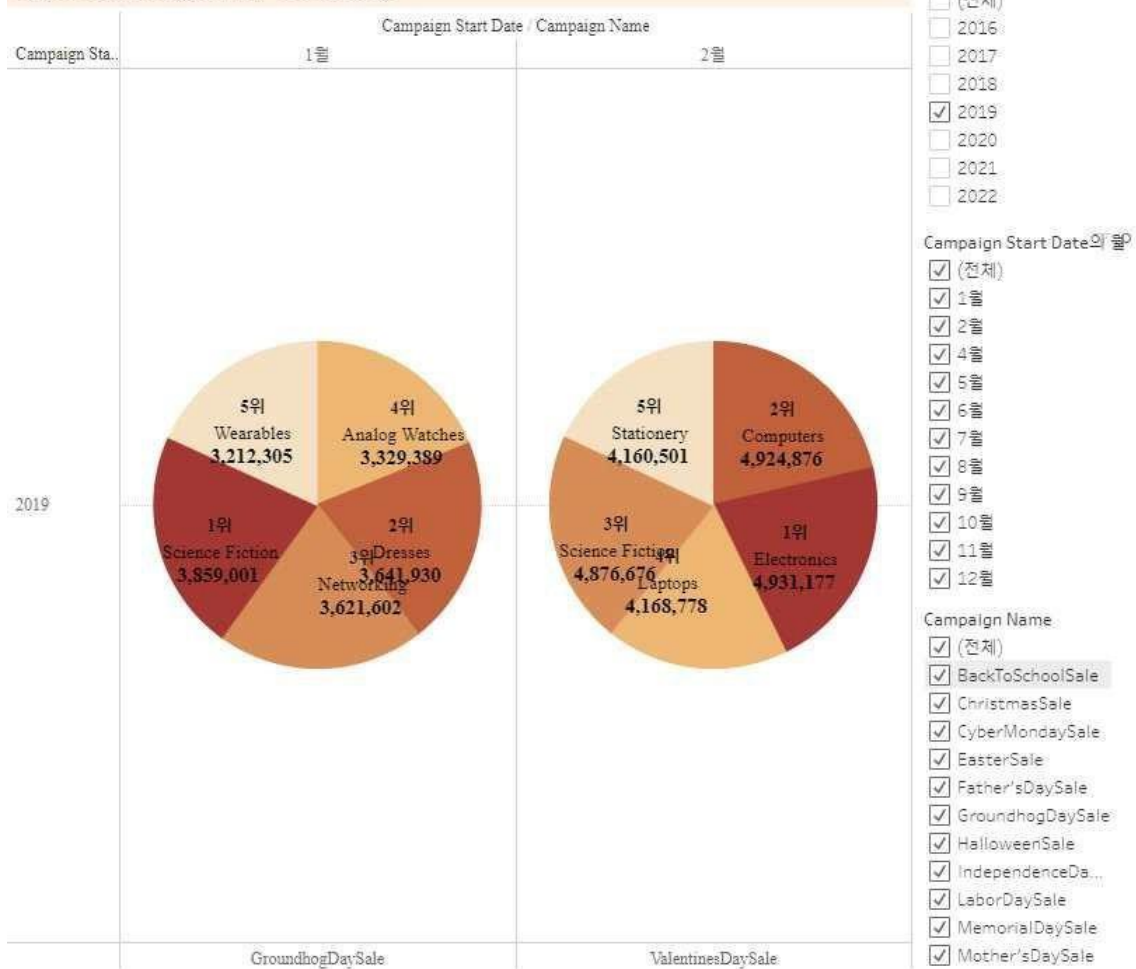
## Comparison of Sales Volume by Campaign



캠페인 기간 동안 일반적으로 관찰되는 현상 중 하나는 매출의 감소와 판매량의 증가가 동시에 발생하는 것이다. 이러한 패턴은 일반적으로 세일 행사와 같은 프로모션 활동과 연관되어 있다고 볼 수 있다. 즉, 할인된 가격으로 인해 제품 단위당 이익은 감소하지만, 저렴한 가격이 더 많은 수의 구매를 유도함으로써 판매량은 상승하는 것을 의미한다.

이와 같은 캠페인의 영향은 단기적으로 매출 감소라는 형태로 나타날 수 있으나, 이는 또한 상품의 대량 유통을 가능하게 하여 장기적인 고객 기반 확대와 재고 순환을 촉진시키는 기회를 제공한다. 할인이라는 인센티브는 신규 고객의 유입을 늘리고, 기존 고객에게는 추가 구매를 장려할 수 있다. 이는 고객 충성도와 브랜드 인지도를 높이는 데 기여하며, 추후 가격 정상화 시 보다 높은 매출을 기대할 수 있는 기반을 마련한다.

## Top5 Subcategory by Campaign



캠페인 기간 동안 특히 랩톱과 시계와 같은 고가의 전자제품들이 높은 판매량을 기록한 것으로 나타났다. 이러한 상품들은 일반적으로 높은 단가를 자랑하기 때문에, 캠페인 기간 중에 실시된 할인과 프로모션은 소비자들에게 큰 유인이 된다. 이는 이러한 고가 품목에 대한 구매 결정이 일반적으로 더 많은 고민과 비교를 요구하기 때문에, 할인이 고객들이 구매를 결정하는데 중요한 요소가 되기 때문이다.

이런 현상은 소비자들이 평소에는 구매를 주저했던 비싼 상품을 할인된 가격에 구매할 수 있는 기회로 인식하기 때문에 발생한다. 이는 잠재적 구매자들에게 매력적인 가격 제안을 통해 그들의 구매를 유도하며, 결과적으로 더 높은 판매량을 이끌어낼 수 있다. 특히, 랩톱과 시계와 같은 상품들은 사용자의 일상생활 및 업무에 중요한 역할을 하므로, 구매자들은 이러한 품목에 대해 더 민감하게 반응할 가능성이 크다.

# Payment

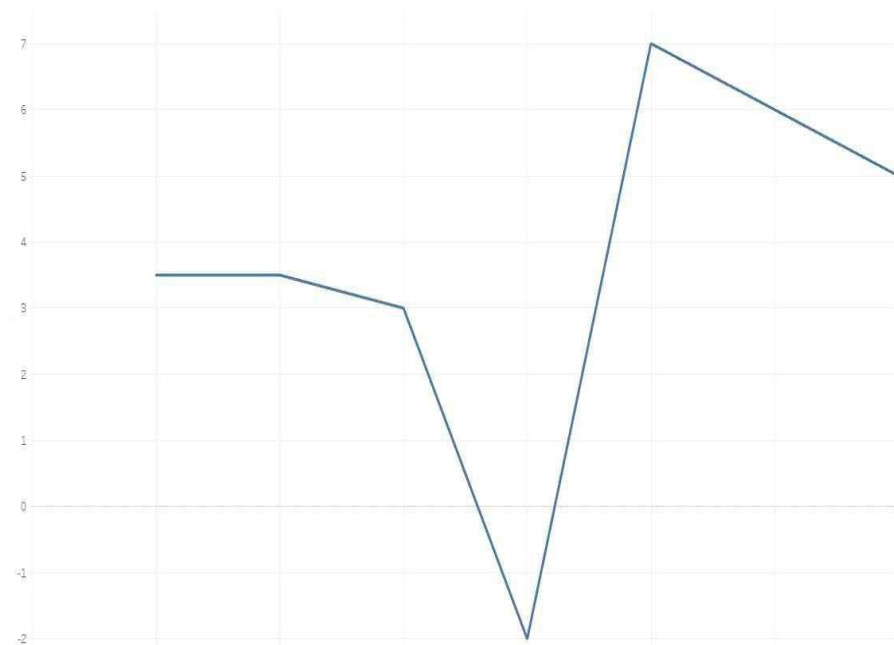
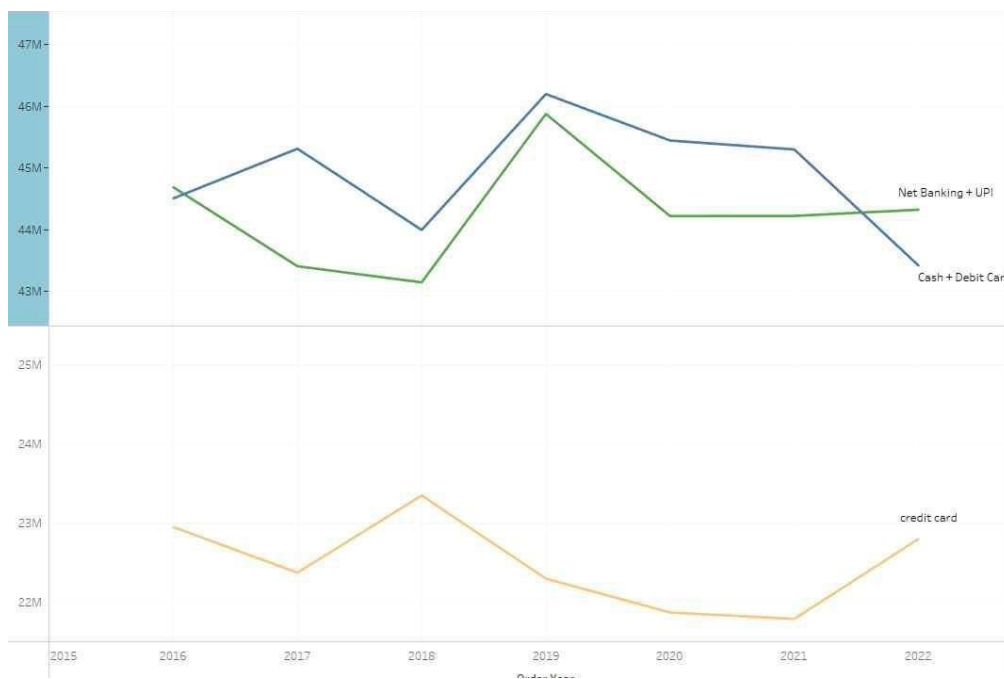
## Patterns by income group

Payment Method Type	Income Group		
	High Income	Lower Middle Income	Upper Middle Income
Cash + Debit Card	51,356,210	164,948,012	97,881,516
credit card	24,841,255	81,882,823	50,725,284
Net Banking + UPI	50,641,215	163,579,688	95,670,708

세계은행의 소득 기준을 적용하여 고소득, 중소득, 저소득 국가로 시장을 세분화한 결과,, 저소득 국가에서의 결제액이 가장 많다는 데이터가 나타났다. 이는 일반적인 시장 경향과 다소 상반된 현상으로, 저소득 국가들이 소비재에 대한 상당한 지출을 하고 있음을 시사한다. 이러한 정보는 기존의 시장 가정에 도전하며, 경제적 수준에 상관없이 특정 상품이나 서비스에 대한 강력한 수요가 존재할 수 있음을 나타낸다.

이 발견은 기업이 저소득 국가들을 대상으로 한 마케팅 전략을 재고할 필요가 있음을 알려준다. 특히, 이러한 시장에서는 가격에 민감하면서도 필요와 욕구를 충족시키는 제품에 대한 수요가 클 수 있음을 반영하고 있다. 또한, 저소득 국가들에서의 높은 결제액은 현지 소비자들의 구매 행태, 분배 채널, 결제 수단 등에 대한 추가적인 이해를 통해 더욱 깊이 파악할 필요가 있다.





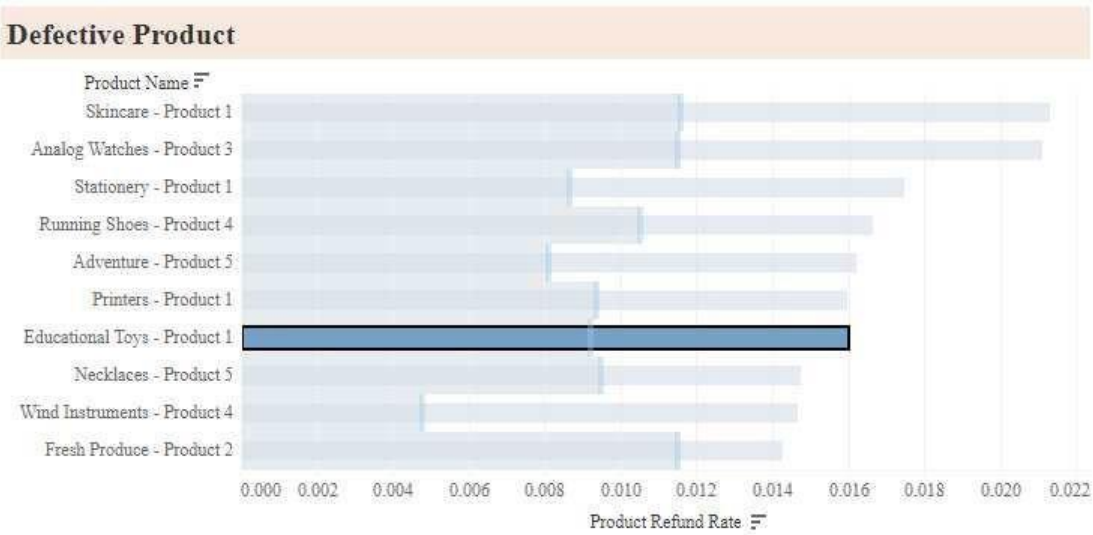
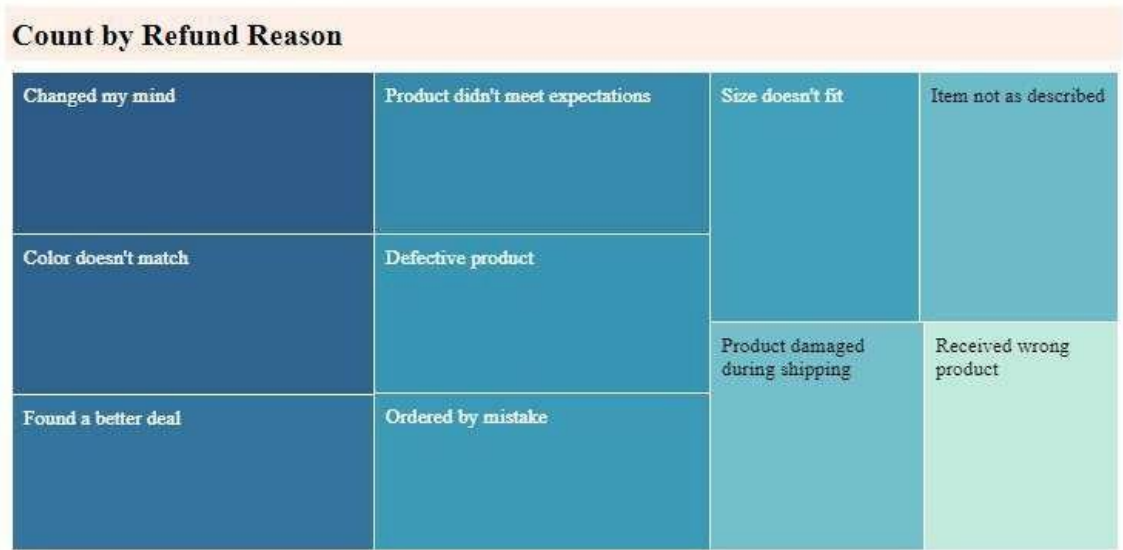
결제 수단별 사용률의 변화를 분석한 결과, 현금 및 직불카드와 Net Banking 및 UPI(통합 결제 인터페이스)의 사용률이 유사한 패턴으로 변화하는 반면, 신용카드의 사용률은 이와 반대 방향으로 움직이는 현상을 확인할 수 있었다. 이러한 명확한 음의 상관관계는 두 결제 방식 그룹이 같은 외부 요인의 영향을 받고 있음을 시사하며, 특히 경제 성장률과 신용카드 사용률의 반대 방향 움직임은 이 외부 요인 중 하나가 경제 성장률일 가능성을 나타낸다.

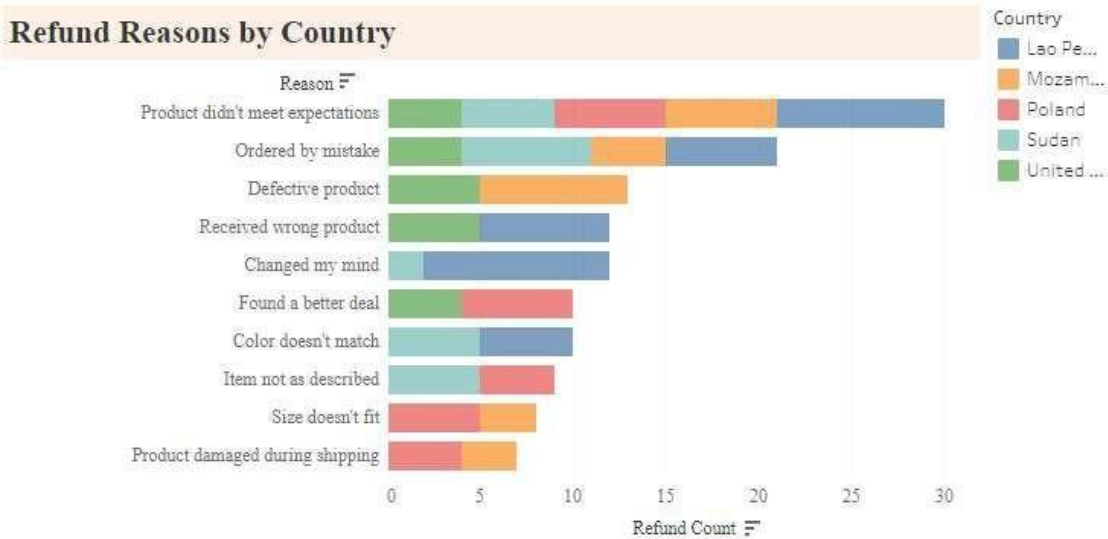
경제 성장률의 변화는 소비자의 구매력, 소득 수준, 그리고 소비 심리에 직접적인 영향을 미치며, 이는 다시 소비자들이 상품 및 서비스 구매를 위해 선택하는 결제 수단에 영향을 준다. 경제가 성장하면서 소비자의 구매력이 증가하고, 소득이 늘어날 때, 사람들은 보다 편리하고 빠른 결제 수단을 선호할 수 있으며, 이는 Net Banking이나 UPI와 같은 디지털 결제 수단의 사용 증가로 이어질 수 있다. 반면, 신용카드 사용률의 감소는 소비자들이 미래의 불확실성에 대비해 즉시 결제할 수 있는 수단을 선호하거나, 경제 성장에 따른 소득 증가로 신용카드 대신 직불카드나 디지털 결제 수단을 사용할 수 있는 여유가 생겼기 때문일 수 있다.

또한, 경제 성장이 소비자들의 신용 접근성을 개선시키면서 신용카드에 대한 의존도가 줄어들고, 대신 신속하고 간편한 디지털 결제 수단이 선호되는 현상으로 이어질 수 있다. 이는 특히 젊은 세대와 기술에 익숙한 사용자들 사이에서 더욱 두드러질 수 있으며, 이러한 세대는 전통적인 신용카드보다는 모바일 결제와 같은 현대적인 결제 수단을 선호하는 경향이 있다.

따라서, 이러한 경향을 이해하고, 소비자의 변화하는 요구와 경제적 환경 변화에 적응하는 결제 솔루션을 제공하기 위한 전략을 수립하는 데 이 정보를 활용해야 한다.

# Refund







이 분석은 네 가지 주요 그래픽 요소를 통해 복잡한 데이터를 시각화하고 있다. 첫 번째, 트리맵은 환불 이유 별 총 횟수를 나타내며, 환불 횟수가 많은 이유일수록 색상이 더 진하게 표시되어 있다. 이를 통해 가장 주된 환불 이유를 빠르게 식별할 수 있으며, 시각적으로 각 이유의 상대적 중요성을 파악할 수 있다.

두 번째, 좌측 하단에는 환불 횟수가 가장 많은 상위 5개 국가를 중심으로 그들의 주요 환불 이유를 나타내는 막대그래프가 있다. 전체적으로 보았을 때 배송 오류 관련 환불 횟수가 적었음에도 불구하고, 이 막대그래프에서는 가장 환불이 많은 국가들에서 배송 오류가 압도적으로 많은 환불 이유로 나타난다. 이는 특정 국가에서의 물류 구조나 배송 업체에 문제가 있을 수 있음을 시사하며, 해당 업체나 구조에 대한 재검토와 개선이 필요함을 알려준다.

세 번째, 서브 카테고리 내에서 평균보다 높은 표준편차를 가진 환불율을 보이는 상품을 탐색하기 위한 막대그래프가 있다. 각 막대의 중간에 위치한 하늘색 라인은 해당 상품이 속한 서브 카테고리의 평균 환불율을 나타내며, 이를 통해 어떤 상품이 그 서브 카테고리 내에서 유난히 높은 환불율을 보이는지 식별할 수 있다.

마지막으로, 우측 하단에는 선택한 특정 상품에 대한 환불 이유와 각각의 횟수 정보를 제공하는 섹션이 있다. 이는 상품별로 구체적인 환불 원인을 깊이 분석할 수 있는 기회를 제공하며, 제품 개선이나 고객 서비스 개선에 있어 실질적인 인사이트를 제공한다.

이렇게 네 가지 분석 도구는 함께 활용될 때, 환불 데이터의 복잡한 면모를 효과적으로 분석하고, 국가별, 상품별, 그리고 환불 이유별로 깊이 있는 인사이트를 제공한다. 이를 통해 기업은 보다 정보에 기반한 의사결정을 할 수 있으며, 특정 문제에 대한 해결책을 보다 명확히 도출할 수 있다.

# Revenue

Sales by Year, Quarter, Month

	2016	2017	2018	2019	2020	2021	2022
	111,203,760	109,982,200	109,554,686	113,241,076	110,585,381	110,284,028	109,604,640

	2016	2017	2018	2019	2020	2021	2022
1	28,097,062	26,949,411	26,894,153	27,788,574	27,641,965	27,747,829	26,777,815
2	28,451,749	28,799,500	27,795,538	27,410,830	26,613,289	27,104,689	27,452,515
3	27,789,561	27,168,551	27,559,522	29,716,457	27,609,885	27,709,824	28,736,258
4	26,863,388	27,064,737	27,305,473	28,325,216	28,720,241	27,721,687	26,618,052

	2016	2017	2018	2019	2020	2021	2022
1	9,300,254	9,606,021	9,293,475	9,457,403	9,664,779	10,340,214	9,170,119
2	8,461,450	8,160,160	8,255,565	8,358,429	9,288,528	7,947,270	8,197,420
3	10,335,358	9,183,230	9,345,113	9,972,742	8,688,659	9,460,345	9,410,276
4	9,274,471	9,466,846	8,949,275	9,229,876	8,745,957	9,255,329	9,193,691
5	9,711,199	9,809,797	9,663,820	9,167,756	9,471,781	9,299,333	9,365,188
6	9,466,079	9,522,858	9,182,443	9,013,198	8,395,551	8,550,028	8,893,637
7	9,439,865	8,783,894	9,096,427	9,711,445	9,264,087	9,336,576	9,265,287
8	8,901,743	9,609,329	9,213,073	10,233,658	9,381,344	9,367,180	9,716,297
9	9,447,954	8,775,328	9,250,022	9,771,354	8,964,454	9,006,068	9,774,675
10	8,838,498	9,041,023	8,708,222	9,720,966	9,961,592	9,610,551	8,485,398
11	8,554,564	8,738,880	9,676,567	9,297,887	9,443,051	9,014,362	9,721,570
12	9,472,327	9,284,834	8,920,684	9,306,363	9,315,598	9,096,773	8,411,084

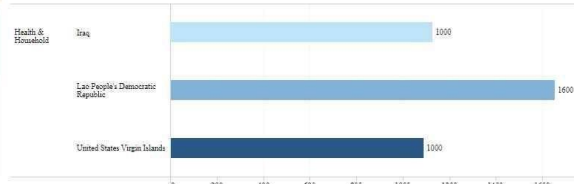
년도별 매출 데이터를 분석함으로써, 우리는 회사의 성장 추이를 파악하고, 그에 따른 성과를 평가하는 데 필요한 중요한 정보를 얻을 수 있다. 이런 정보를 기반으로, 우리는 회사의 과거와 현재의 매출 성과를 비교하여, 경영 성과가 어떻게 변화하고 있는지를 명확히 이해하며, 이를 통해 회사가 성장하는 방향이 예상과 일치하는지, 또는 시장 변화에 맞춰 적절히 대응하고 있는지를 판단할 수 있다.

뿐만 아니라, 과거 매출 데이터의 추세를 분석하여, 미래의 매출에 대한 예측을 수행하고, 이를 바탕으로 향후의 예산 수립과 경영 전략을 계획할 수 있다. 예를 들어, 지속적인 매출 증가 추세를 확인할 수 있다면, 회사는 확장 전략을 고려하거나, 새로운 투자 기회를 모색하는 것이 타당할 수 있다. 반면, 매출이 정체되거나 감소하는 경향을 보이고 있다면, 회사는 비용 구조를 재검토하고, 효율성을 증대 시키거나, 새로운 수익 창출 방안을 모색하는 등 경영 방향을 재조정해야 할 수 있다.

이렇듯 년도별 매출 분석은 단순히 회사의 역사적 성과를 평가하는 것을 넘어서, 앞으로의 성장 전망을 예측하고 그에 따라 회사의 재정적 및 전략적 계획을 세워 나가는 데 꼭 필요한 과정이라 할 수 있다. 이 과정을 통해 기업은 시장의 변화에 능동적으로 대응하고, 장기적인 성공을 위한 기반을 마련하며, 지속 가능한 성장을 도모할 수 있게 된다.

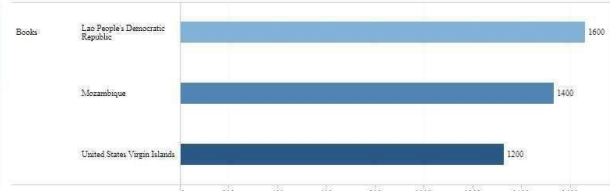
Top3 Countries by Sales Volume per Category

Books	Pet Supplies	Health & Household	Watches	Automotive	Sports & Outdoors	Office Products
Electronics	Jewelry	Toys & Games	Clothing	Video Games	Shoes	
Musical Instruments	Toys & Games	Home & Kitchen	Grocery	Beauty & Personal Care	Movies & TV	



Top3 Countries by Sales Volume per Category

Books	Pet Supplies	Health & Household	Watches	Automotive	Sports & Outdoors	Office Products
Electronics	Jewelry	Toys & Games	Clothing	Video Games	Shoes	
Musical Instruments	Toys & Games	Home & Kitchen	Grocery	Beauty & Personal Care	Movies & TV	



Top3 Countries by Sales Volume per Category

Books	Pet Supplies	Health & Household	Watches	Automotive	Sports & Outdoors	Office Products
Electronics	Jewelry	Toys & Games	Clothing	Video Games	Shoes	
Musical Instruments	Toys & Games	Home & Kitchen	Grocery	Beauty & Personal Care	Movies & TV	



United States Virgin Islands, Mozambique, Laos, and Iraq의 구매 데이터를 면밀히 분석해본 결과, 이들 국가가 모든 카테고리에서 상위 3위 안에 드는 주요 구매국이라는 사실을 확인할 수 있었다. 이는 이들 국가가 이커머스 시장에서 상당한 수요와 매출을 창출하고 있음을 나타내며, 전 세계적인 판매 네트워크 내에서 그들이 차지하는 비중이 매우 큼을 의미한다. 그만큼 이 지역에 대한 경쟁 업체들의 관심도 높아질 가능성이 크며, 이는 기존 고객을 유지하고 추가적인 시장 점유율을 확보하는 데 있어 우리에게 심각한 경쟁을 야기할 수 있다.

이러한 상황에서 우리는 단순히 현 상태를 유지하는 것을 넘어, 각 지역의 구매 패턴과 시장 동향을 철저히 분석하고 이해함으로써, 더욱 혁신적이고 맞춤형 상품 개발에 힘써야 한다. 또한, 각 지역의 문화, 소비자 선호도, 구매력 등 다양한 요소들을 고려한 맞춤형 마케팅 전략과 판매 전략을 수립해야 한다. 이는 고객 충성도를 높이고, 잠재 고객층을 확대하는 데 중요한 역할을 하게 된다.

또한, 경쟁사들과 차별화될 수 있는 고유의 가치 제안을 통해 고객들의 마음을 사로잡을 필요가 있다. 이를 위해 고객 서비스 향상, 지역 사회와의 긴밀한 관계 구축, 지역별 맞춤형 프로모션 및 할인 정책 등을 포함한 다양한 전략이 필요하다. 지속 가능한 관계를 구축하여 장기적으로 해당 지역에서의 비즈니스를 강화해 나가는 것은, 우리 회사의 지속적인 성장과 성공에 있어 필수 불가결한 요소가 된다.



Part Three

# 기대효과 & 결론



## 기대효과

**고도화된 시장 분석:** 카테고리 및 지역별 세분화를 통한 데이터 분석은 기업에게 소비자의 행동, 선호도 및 구매 패턴에 대한 깊이 있는 인사이트를 제공한다. 이는 특정 시장 내에서 어떤 상품이나 서비스가 더 큰 성공을 거둘 수 있는지 예측하는 데 도움을 주며, 결과적으로 시장 진입 및 확장 전략을 보다 명확히 수립할 수 있게 한다.

**개인화 마케팅 전략의 강화:** 고객별 맞춤 상품 추천과 같은 데이터 기반 접근법은 개인화된 마케팅 활동을 촉진한다. 이는 고객 개인의 취향과 요구에 맞는 커뮤니케이션과 제안을 가능하게 하여, 고객 경험을 풍부하게 하고, 장기적인 고객 관계를 구축하는 데 기여한다.

**리스크 관리 및 예측:** 세분화된 데이터 분석을 통해 기업은 특정 시장이나 상품에 대한 리스크를 더 잘 이해하고 예측할 수 있다. 이는 재고 관리, 가격 책정 전략 및 프로모션 활동에 있어 보다 정보에 기반한 결정을 내리게 하여, 불필요한 비용을 줄이고 수익성을 극대화하는 데 도움을 준다.

**경쟁 우위 확보:** 이와 같은 분석을 통해 얻은 통찰력은 기업이 경쟁사 대비 독특한 가치 제안을 개발하고, 시장 내에서 독특한 위치를 차지할 수 있도록 지원한다. 이는 소비자에게 더 매력적인 선택지를 제공함으로써, 시장 점유율을 확대하고 브랜드 인지도를 높이는 결과로 이어질 수 있다.

## 결론

이러한 데이터 분석과 시각화 도구의 활용은 단순한 정보 제공을 넘어서, 기업의 전략적 의사결정 과정에 근본적인 변화를 가져올 수 있다. 시장의 복잡성과 경쟁이 치열해지는 현대 비즈니스 환경에서, 데이터에 기반한 통찰력은 기업이 변화하는 소비자 요구에 민첩하게 대응하고, 지속 가능한 성장을 이루는 데 필수적이다. 따라서, 이와 같은 세분화된 데이터 분석 접근법은 기업이 더 넓은 시야를 가지고 시장을 바라보게 하며, 혁신을 통해 지속 가능한 성공을 달성하기 위한 전략적 기반을 제공한다. 결국, 이는 기업이 더욱 효과적으로 시장을 선도하고, 소비자의 기대를 초과하는 제품과 서비스를 제공할 수 있는 능력을 향상시키는 핵심적인 경쟁력이 된다.