

E-Commerce

데이터 분석



이중훈

DATA INTRODUCTION

문제 정의 및 활용

문제 정의

1) Olist 데이터 분석을 통해 이커머스의 매출 감소, 지역별 판매 불균형, 시간대별 매출 집중, 제품 카테고리 간 매출 편차 등 주요 비즈니스 문제를 파악.

2) RFM 분석을 통해 고객을 세분화하고, 각 세그먼트의 행동 및 매출 기여도를 분석하여, 충성 고객 유지 및 이탈 고객 회복이 필요.

활용 목표

데이터 기반으로 고객 행동과 매출 성과를 분석하여, 판매 성과의 주요 요인을 파악하고 의사결정을 지원하는 인사이트를 도출하여 고객 유지, 매출 증대, 및 운영 효율성 개선을 달성.



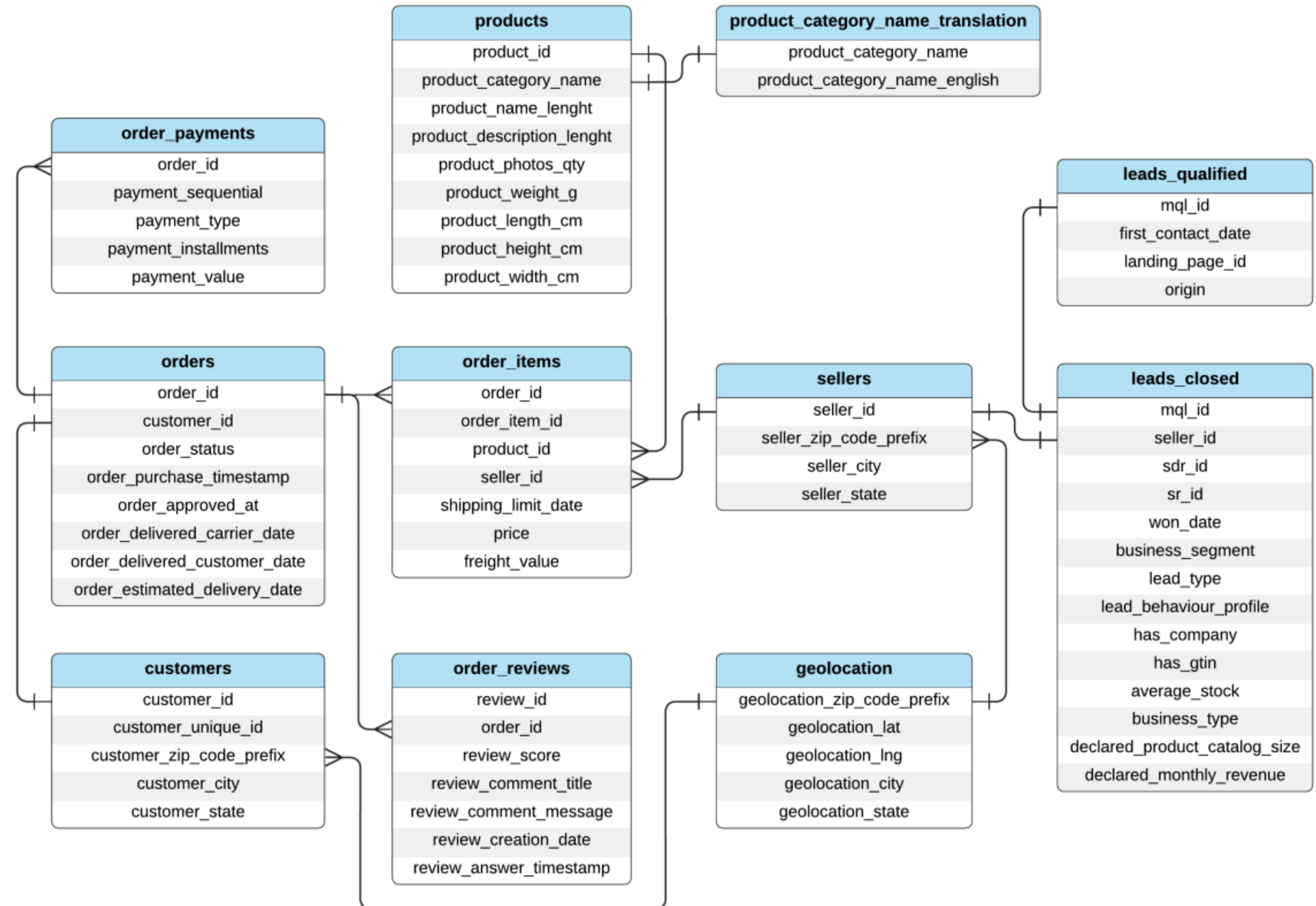
DATA INTRODUCTION

데이터 소개

활용 데이터 셋은 Kaggle에 있는 Olist E-Commerce 데이터 셋으로 브라질 전자상거래 데이터입니다.

2016년부터 2018년까지 Olist에서 이루어진 약 10만 건의 주문 정보가 포함된 데이터로 데이터 스키마는 다음과 같습니다.

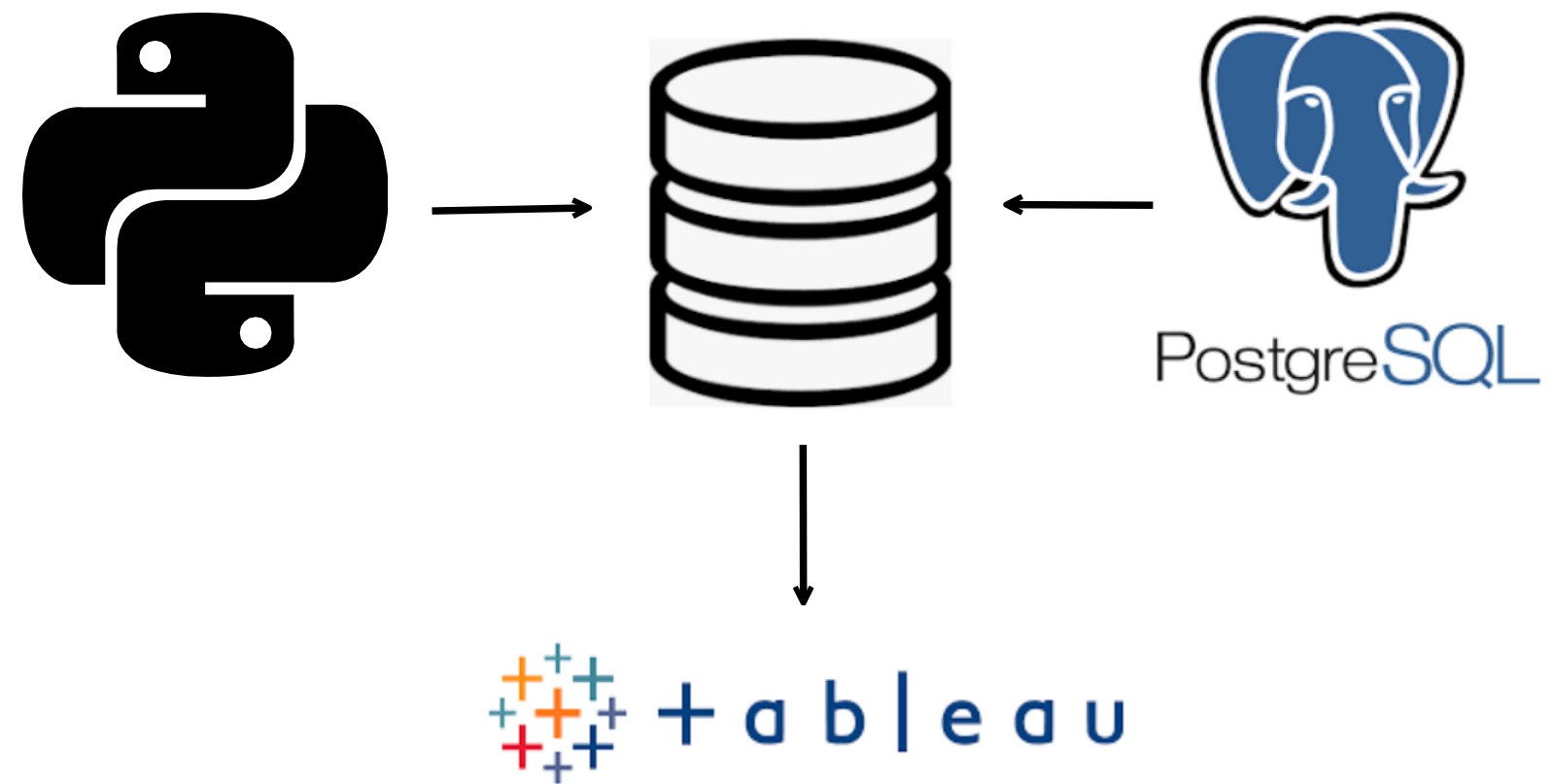
DATABASE SCHEMA



ANALYSTIC METHOD

분석 방법

결측치 및 오류값은 파이썬으로 전처리한 후 DB에 데이터를 적재하고 SQL문으로 분석에 필요한 지표 및 수치 데이터를 가공. 태블로와 연동하여 시각화 및 대시보드 생성 진행.



ANALYSTIC METHOD

분석 목적

개요 :

Olist 데이터를 분석하는 가장 중요한 목적 중 하나는 고객 행동을 이해하는 것입니다.

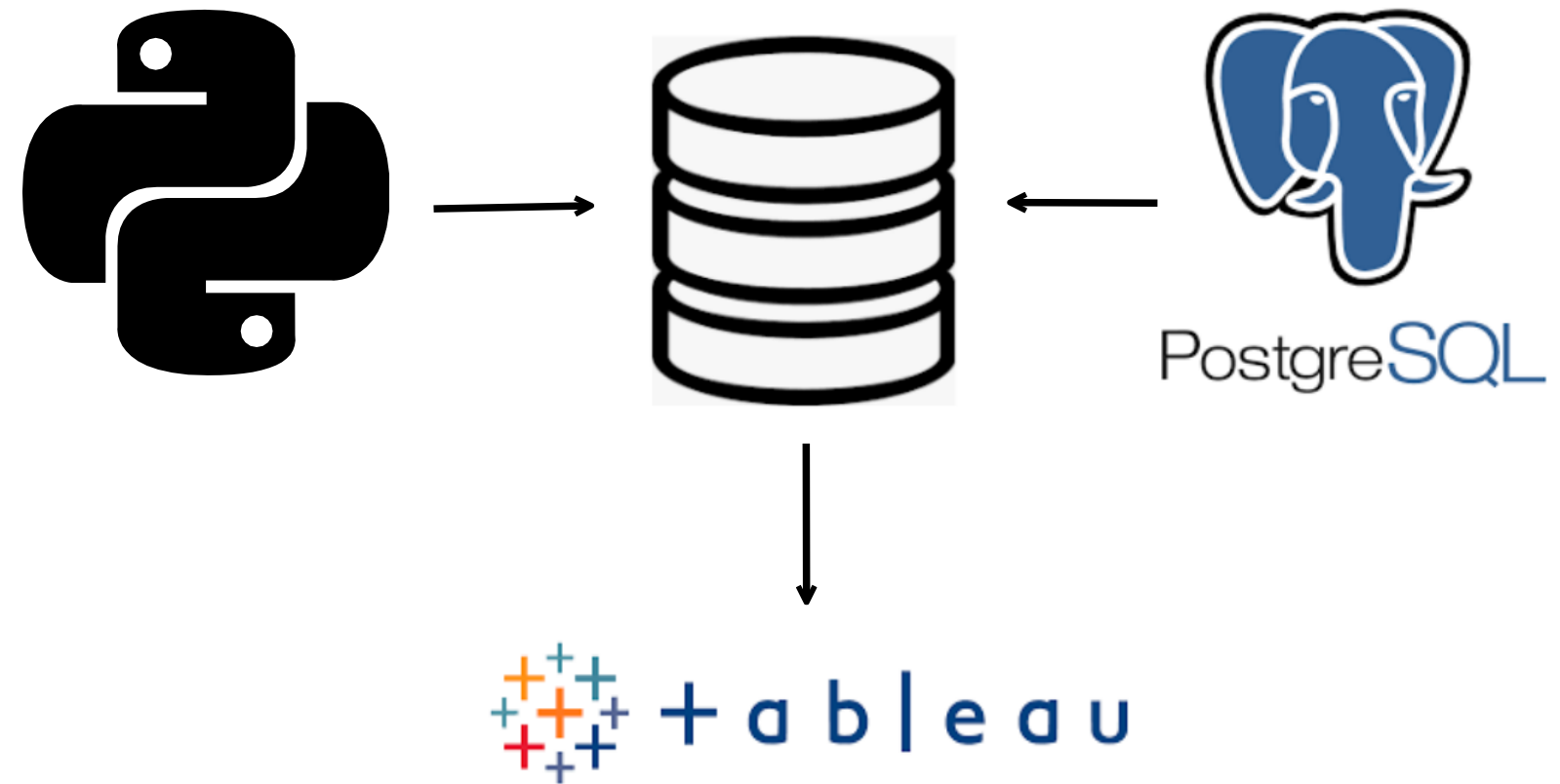
목적 :

고객의 구매 패턴 및 행동 이해.

고객 세그먼트 정의 및 맞춤형 마케팅 전략 개발.

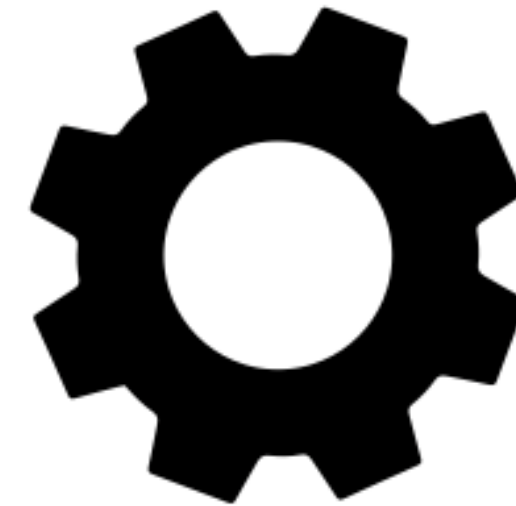
활용 지표 :

- 지역별 구매금액, 구매건수, 구매유저수, 건당 구매 금액 추세
- 특정 시간대나 요일별 구매 트렌드
- 고객의 행동 및 가치를 평가할 수 있는 지표
- RFM 분석을 통한 고객 세분화 지표



DATA PREPROCESSING

데이터 전처리



01 중복값 제거 및 주요 변수 생성

데이터 예시

order_id	customer_id	order_status	order_purchase_timestamp	order_approved_at	order_delivered_carrier
e481f51cbdc54678b7cc49136f2d6af7	9ef432eb6251297304e76186b10a928d	delivered	2017-10-02 10:56:33	2017-10-02 11:07:15	2017-10-04 19:3
e481f51cbdc54678b7cc49136f2d6af7	9ef432eb6251297304e76186b10a928d	delivered	2017-10-02 10:56:33	2017-10-02 11:07:15	2017-10-04 19:3
e481f51cbdc54678b7cc49136f2d6af7	9ef432eb6251297304e76186b10a928d	delivered	2017-10-02 10:56:33	2017-10-02 11:07:15	2017-10-04 19:3

		customer_unique_id
is_revenue	order_status	
y n	shipped	1008
	invoiced	300
	delivered	94488
	approved	2
	unavailable	6
	processing	284
	canceled	428

중복값 제거

customer_id와 order_id의 조합은 고유해야하므로 중복제거.

데이터 수 : 115609 -> 96516

order_status를 통해 매출 포함여부에 따라 변수 생성

1. 고객의 주문이 실제 매출로 이어진 상태

상세 설명

approved (승인): 주문이 승인되어 매출로 연결될 가능성이 높음.

delivered (배송 완료): 고객에게 상품이 전달되었으므로 확정 매출로 간주됨.

invoiced (송장 발행): 결제가 완료되고 송장이 발행된 상태. 대부분 매출로 기록됨.

shipped (배송 중): 매출로 집계할 가능성이 있음(특히 배송 시점에 매출로 기록하는 경우).

2. 주문이 최종 결제에 포함되지 않거나, 매출이 취소된 상태.

상세 설명

canceled (취소): 고객 요청 또는 기타 사유로 주문이 취소되어 매출에서 제외됨.

created (생성): 주문이 생성되었지만 승인이나 결제가 진행되지 않은 초기 단계. 매출 확정 아님.

processing (처리 중): 주문 처리 단계로, 아직 매출 확정 전 상태.

unavailable (처리 불가): 주문이 불가능하거나 실패한 상태로 매출에 포함되지 않음.

최종 데이터 : 96516개



COUNTRY ANALYSIS

성과 지표 분석

목적

특정 기간 동안의 변화와 추이를 이해하여 성장 또는 감소 요인을 분석

전년 동월 대비 및 전월 대비 성과를 비교

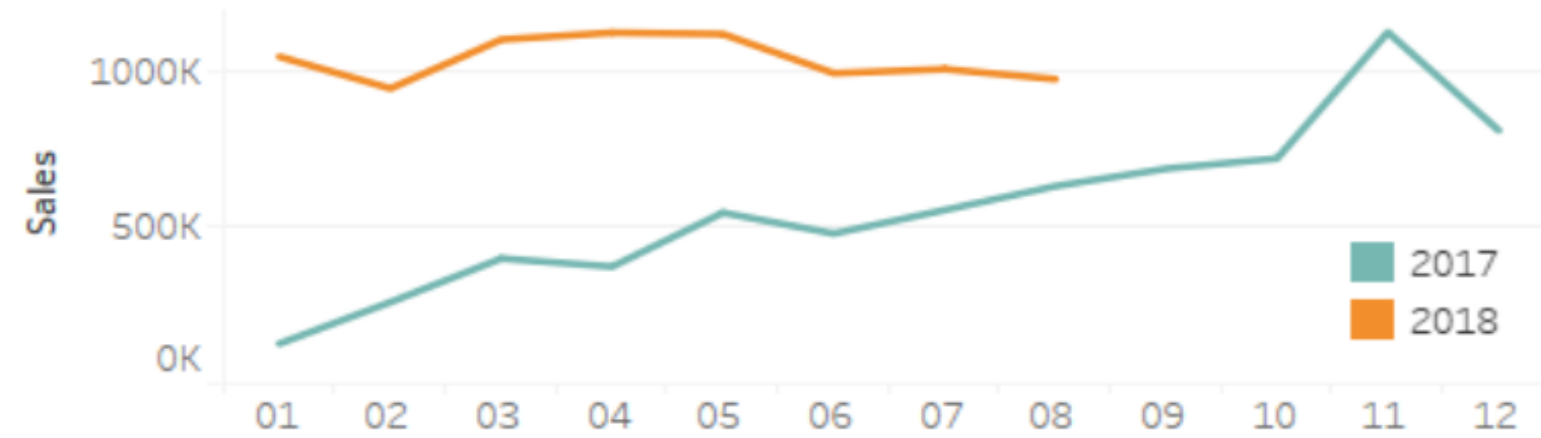
카테고리별 판매 데이터를 기반으로 시장 트렌드 및 수요 파악

대시보드 링크:



01 기간별 지표 추이

Sales Trend By Month

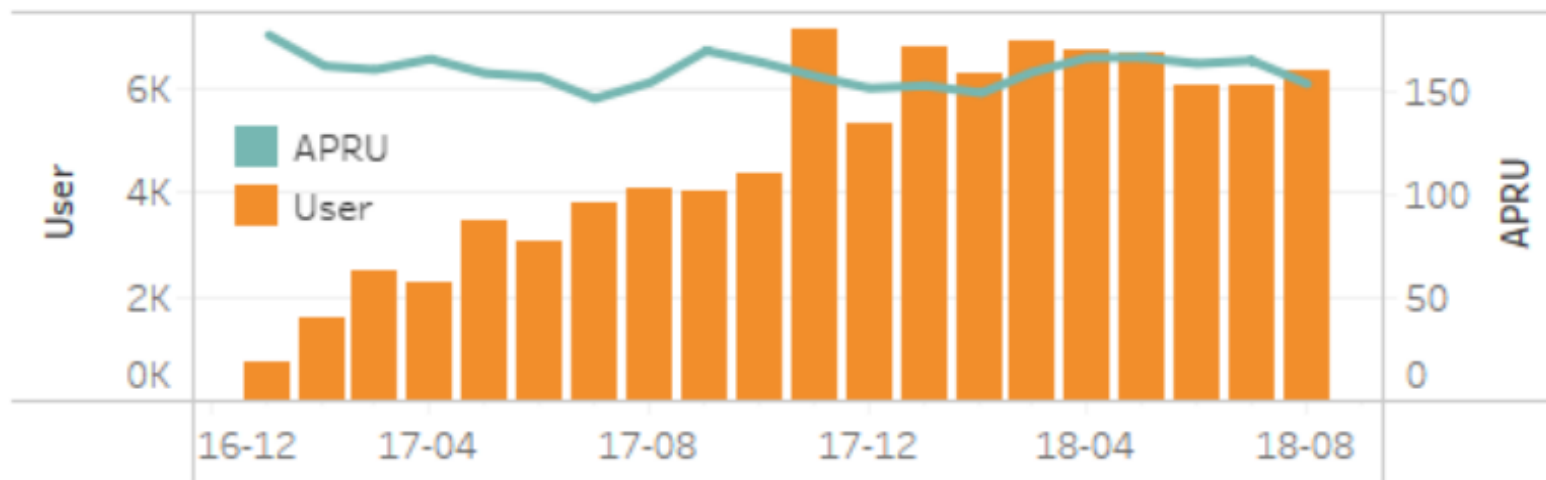


월별 매출 추이

2017년은 11월까지 꾸준히 매출이 상승하다가 12월에 소폭 감소함.(11월에 블랙프라이데이가 포함되어 매출이 급상승한것으로 나타남.)

2018년은 전체적으로 매출이 2017년 보다 높지만 2분기에 상승하다가 3분기들어 감소하는 추세를 보임.

User & APRU



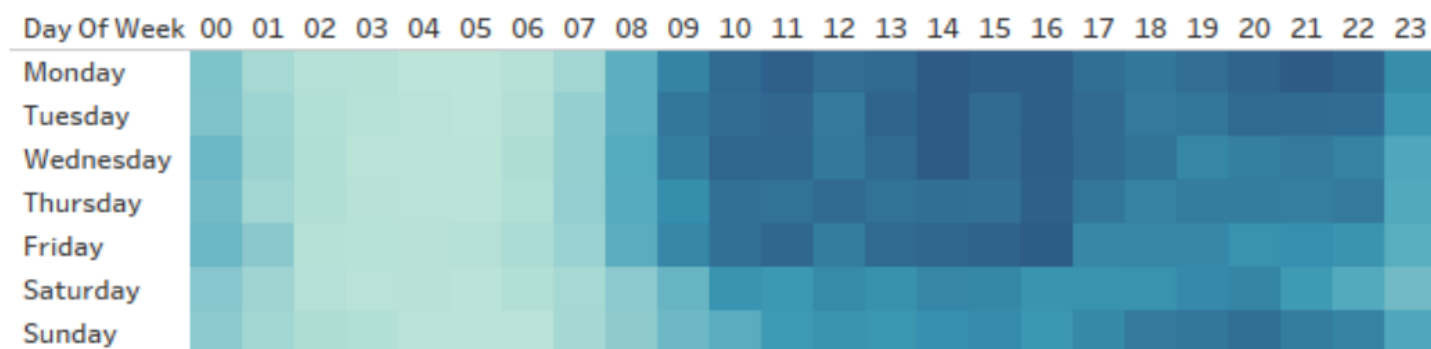
유저수 X APRU별 매출

유저수는 꾸준히 상승 추세지만, APRU가 감소 추세를 보이는 것으로 보아 신규 유저의 대부분이 저수익 고객일 가능성이 있음.

단순히 유저 수를 늘리는 것보다, 높은 지출을 하는 유저의 비율을 늘리는 전략이 필요.

(ARPU를 늘리기 위해 고가 상품이나 업셀링(업그레이드 구매) 기회를 제공할 수 있어야함.)

Sales By Hour



요일 X 시간대별 매출

전체적인 매출은 월~금의 매출이 토,일 매출보다 대체로 높았음.

평일에는 약 10시부터 16시까지 매출이 가장 높은 것으로 나타남.

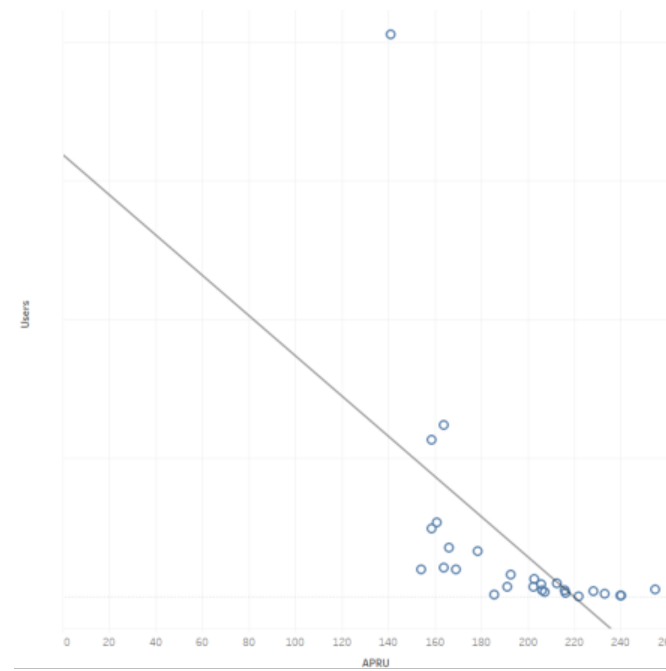
주말에는 약 18시부터 22시까지 매출이 가장 높은것으로 나타남

02 주별 매출 및 카테고리 분석

Sales By State



User&APRU



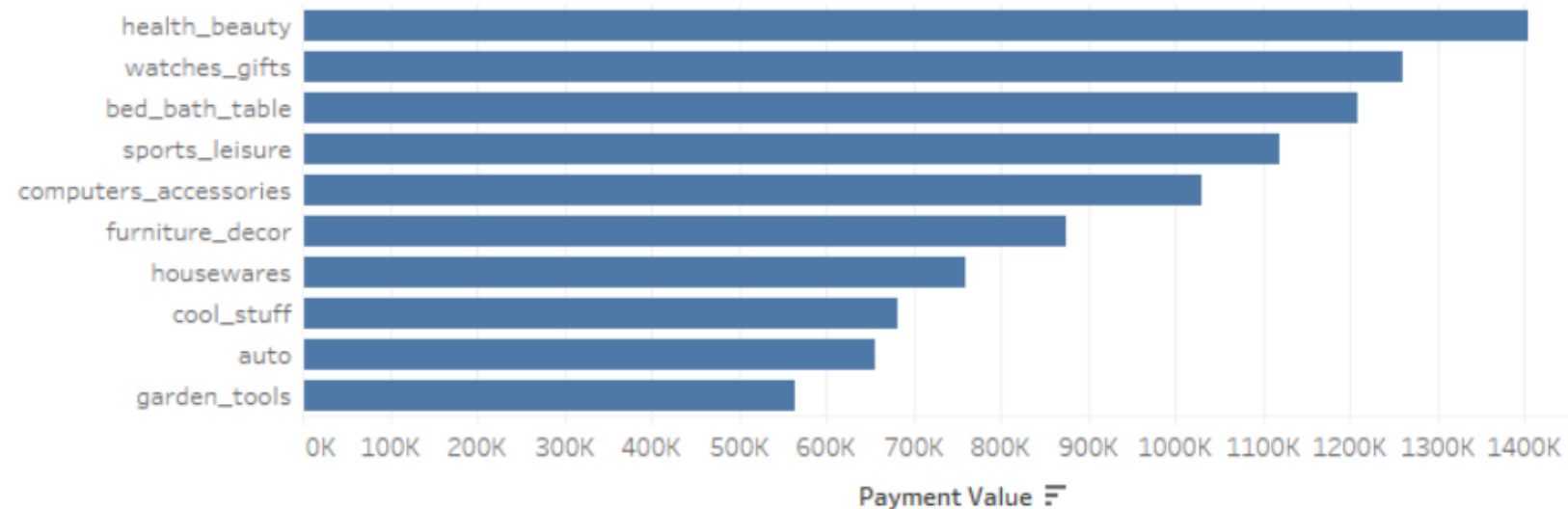
주별 매출

주별 매출은 상파울로주가 5,655,665헤알로 가장 높았고, 호라이마주가 9,647헤알로 가장 낮았음.

그러나 상파울로주는 APRU가 가장 낮아 고객당 구매 금액이 140.5헤알로 가장 낮은 것으로 나타남.

APRU가 가장 높은 주는 파라이바주로 254.5헤알.

Top 10 Sales Category



카테고리별 매출

Health & Beauty 카테고리가 가장 높은 매출을 기록하고 있으며, 소비자들에게 가장 인기 있는 제품군임을 나타냅니다.

Watches & Gifts와 Bed, Bath & Table 카테고리가 그 뒤를 이어 높은 매출을 차지하고 있으며, 선물용 제품이나 생활용품의 수요가 강세를 보이고 있음을 보여줍니다.

상위 카테고리의 매출 기여도를 보면, 개인적인 소비(뷰티, 패션)와 생활 필수품(침구 및 가구)에 대한 수요가 강하다는 것을 확인할 수 있습니다.

CUSTOMER ANALYSIS

고객 세분화 분석

목적

각 고객 그룹에 맞춤형 캠페인을 제공함으로써 마케팅 비용을 최적화

Recency가 낮은 고객을 식별하여 이탈 원인을 파악해 고객 이탈을 방지

세분화된 데이터를 활용하여 고객 가치를 극대화할 수 있는 전략을 수립



01 지표 및 세그먼트 분류 기준

지표&점수	1점	2점	3점	4점	5점
최근성 (Recency)	180일 초과	91~180일	31~90일	16~30일	0~15일
구매빈도 (Frequency)	1회	2회	3회	4회	5회 이상
구매금액 (Monetary)	하위 25%	25%~40%	40%~70%	70%~90%	상위 10%

최근성은 2018/08/29
(2017/09~2018/08 중 마지막 구매일 기준)

Scores	Segment (Kor)
555, 545, 455, 445	챔피언 고객
554, 544, 454, 553, 543, 533, 443, 433, 355, 345, 335	충성 고객
551, 552, 541, 532, 531, 431, 432, 423, 352, 351	잠재적 충성 고객
521, 515, 522, 422, 421	신규 고객
525, 524, 523, 515, 514, 415, 315, 314, 313	잠재 고객
534, 424, 344, 334, 324, 323	관심 필요한 고객
312, 232, 331, 231, 222, 223	잠재 휴면 고객
215, 254, 245, 244, 234, 224	이탈 방지 필요 고객
311, 325, 243, 242, 141, 142, 113, 124	이탈 위험 고객
223, 222, 123, 122, 212, 211	휴면 고객
111, 112, 121, 131, 141, 151	이탈 고객

02 고객 세그먼트 결과

지표&점수	1점	2점	3점	4점	5점
최근성 (Recency)	46.76%	27.43%	16.19%	5.86%	3.76%
구매빈도 (Frequency)	97.53%	2.31%	0.13%	0.02%	0.01%
구매금액 (Monetary)	25.01%	14.99%	30.01%	19.99%	10%

RFM Segment

이탈고객 26,444 (36.86%)	휴면고객 15,981 (22.27%)	잠재휴면고객 6,431 (8.96%)	신규고객 5,494 (7.66%)
	이탈우려고객 10,409 (14.51%)	잠재고객 3,702 (5.16%)	놓치면안될고객 3,098 (4.32%)

지표별 비율 해석

최근성(Recency)

1점 고객이 46.76%로 가장 높은 비중을 차지하며, 이는 많은 고객이 최근 구매 활동이 없음을 의미합니다.

상위 점수(4~5점)의 고객 비율은 각각 5.86%, 3.76%로 매우 낮아, 최근 구매를 유도하는 캠페인이 필요합니다.

구매빈도(Frequency)

대부분의 고객이 1점(97.53%)으로, 구매 활동이 거의 없는 고객이 많습니다.

상위 빈도 고객(4~5점)의 비중은 0.02% 이하로, 충성 고객 확보가 필요합니다.

구매금액(Monetary)

3점(30.01%)이 가장 많은 비중을 차지하며, 중간 수준의 구매력이 있는 고객이 다수입니다. 1점(25.01%) 및 2점(14.99%) 비율도 높아, 상위 고객군으로 전환하기 위한 업셀링 전략이 요구됩니다.

RFM 점수변환 후 세그먼트 결과

이탈 고객(36.86%)이 가장 큰 비중을 차지하며, 이들의 재참여를 유도하는 전략이 필요함.

휴면 고객(22.27%)과 이탈 우려 고객(14.51%) 또한 큰 비중을 차지해 주요 관리 대상이며, 이들을 재활성화하기 위한 맞춤형 캠페인이 요구됨.

잠재 고객(5.16%)과 신규 고객(7.68%)은 구매 증대 전략으로 활성화를 유도하여 상위 고객군으로 전환이 필요합니다.

03 최종 세그먼트별 전략 제안

Segment	특징	목표	전략
우수 고객	최근 구매, 자주 구매하며 높은 금액을 소비하는 핵심 고객	충성도 유지 및 브랜드 홍보 촉진	SNS 및 추천 프로그램, 멤버십 혜택 강화, 리뷰 요청 및 보상 제공
충성 고객	자주 구매하고, 높은 금액을 소비하며 브랜드 충성도가 높은 고객	유지 및 구매 확대 유도	업셀링 및 크로스셀링 제안, 친구 초대 프로그램, 정기구독 상품 홍보
잠재적 충성 고객	최근 구매 이력이 있으며, 잠재적으로 충성 고객으로 전환 가능성이 있는 고객	자주 구매하도록 유도	재구매 추가 혜택 제공, 구매이력 기반 제품 추천, 멤버십 가입 유도 캠페인
신규 고객	최근 첫 구매를 한 고객	충성 고객으로 전환	웰컴 이메일 캠페인, 다음 구매 할인코드 제공, SNS 팔로우 추가혜택
잠재 고객	구매 빈도가 낮지만 구매 가능성이 있는 고객	구매 행동을 유도	첫 구매 할인 제공, 한정시간 프로모션, 방문기록 활용 리타겟팅
관심 필요한 고객	구매 이력이 있으나 관심과 추가적인 관리를 필요로 하는 고객	재방문 유도 및 브랜드 관심 증대	할인 및 제품 체험행사 초대, 포인트 소멸 알림, 재구매 쿠폰 발송
잠재 휴면 고객	구매 빈도가 점차 감소하여 휴면 상태로 전환될 가능성이 있는 고객	재참여 유도	재구매 혜택 캠페인, 고객 복귀 프로모션, 무료 배송 프로모션
이탈 방지 필요 고객	이탈 가능성이 있어 구매 복귀를 유도해야 하는 고객	이탈 방지 및 재활성화	복귀 할인 제공, 포인트 추가적립 캠페인, 최저가 보상제
이탈 위험 고객	이탈 위험이 높으며, 브랜드와의 연결성이 약화된 고객	브랜드와의 연결성 회복	재구매 쿠폰 제공, 재방문 유도 이메일, 소규모 상품 추천
휴면 고객	장기간 활동하지 않아 관계 회복이 필요한 고객	관계 회복 및 구매 재활성화	과거 관심상품 광고노출, 휴면 복귀 캠페인, 이탈이유 설문조사 후 상품제공
이탈 고객	이미 이탈한 고객	고객 복귀 유도	복귀 이벤트 초대, 신제품 수시 정보제공, 이탈이유 설문조사 후 상품제공



03 최종 세그먼트별 전략 제안

Segment	특징	목표	전략
우수 고객	최근 구매, 자주 구매하며 높은 금액을 소비하는 핵심 고객	충성도 유지 및 브랜드 홍보 촉진	SNS 및 추천 프로그램, 멤버십 혜택 강화, 리뷰 요청 및 보상 제공
충성 고객	자주 구매하고, 높은 금액을 소비하며 브랜드 충성도가 높은 고객	유지 및 구매 확대 유도	업셀링 및 크로스셀링 제안, 친구 초대 프로그램, 정기구독 상품 홍보
잠재적 충성 고객	최근 구매 이력이 있으며, 잠재적으로 충성 고객으로 전환 가능성이 있는 고객	자주 구매하도록 유도	재구매 추가 혜택 제공, 구매이력 기반 제품 추천, 멤버십 가입 유도 캠페인
신규 고객	최근 첫 구매를 한 고객	충성 고객으로 전환	웰컴 이메일 캠페인, 다음 구매 할인코드 제공, SNS 팔로우 추가혜택
잠재 고객	구매 빈도가 낮지만 구매 가능성이 있는 고객	구매 행동을 유도	첫 구매 할인 제공, 한정시간 프로모션, 방문기록 활용 리타겟팅
관심 필요한 고객	구매 이력이 있으나 관심과 추가적인 관리를 필요로 하는 고객	재방문 유도 및 브랜드 관심 증대	할인 및 제품 체험행사 초대, 포인트 소멸 알림, 재구매 쿠폰 발송
잠재 휴면 고객	구매 빈도가 점차 감소하여 휴면 상태로 전환될 가능성이 있는 고객	재참여 유도	재구매 혜택 캠페인, 고객 복귀 프로모션, 무료 배송 프로모션
이탈 방지 필요 고객	이탈 가능성이 있어 구매 복귀를 유도해야 하는 고객	이탈 방지 및 재활성화	복귀 할인 제공, 포인트 추가적립 캠페인, 최저가 보상제
이탈 위험 고객	이탈 위험이 높으며, 브랜드와의 연결성이 약화된 고객	브랜드와의 연결성 회복	재구매 쿠폰 제공, 재방문 유도 이메일, 소규모 상품 추천
휴면 고객	장기간 활동하지 않아 관계 회복이 필요한 고객	관계 회복 및 구매 재활성화	과거 관심상품 광고노출, 휴면 복귀 캠페인, 이탈이유 설문조사 후 상품제공
이탈 고객	이미 이탈한 고객	고객 복귀 유도	복귀 이벤트 초대, 신제품 수시 정보제공, 이탈이유 설문조사 후 상품제공



