

ZOMATO 데이터 기반 리텐션 & 성장 전략

고객 세그먼트를 중심으로 (for 마케팅팀)

STARTS

CONTENTS

1 기업 소개

2 프로젝트 개요

3 데이터 소개 및 전처리

4 모델링 & 클러스터링

5 대시보드

6 인사이트 및 전략 제안

7 전략 수행

8 결론 및 기대 효과



기업소개



주요 서비스

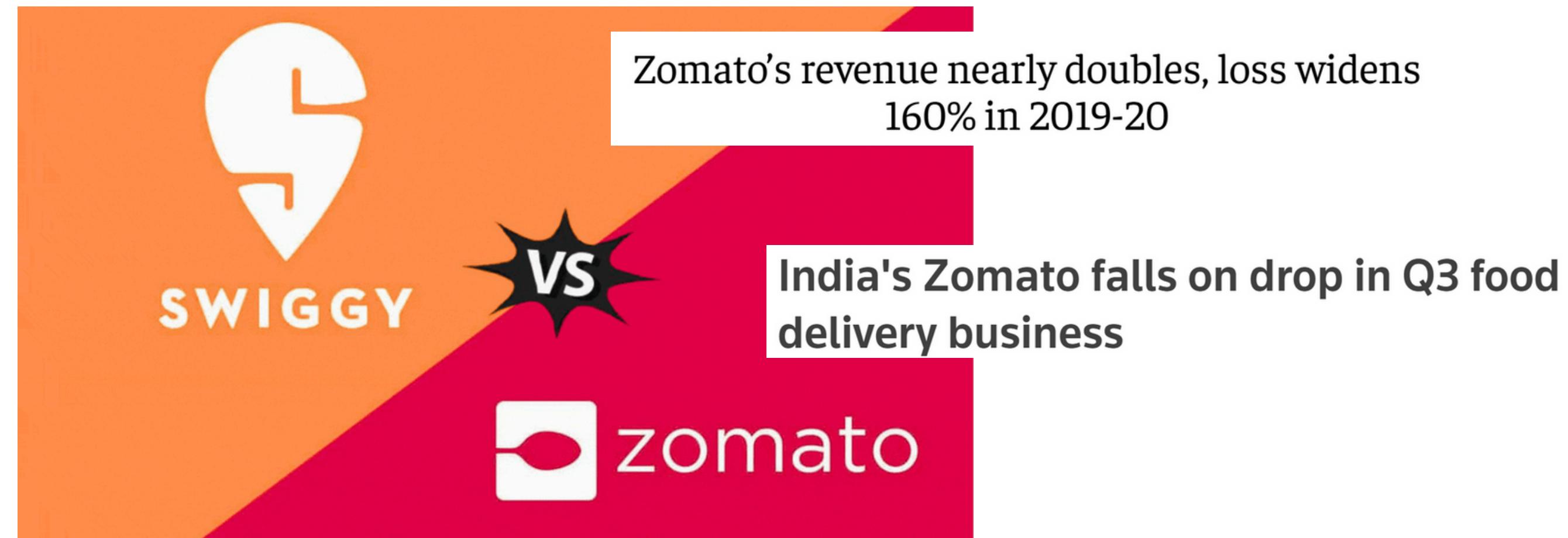
- 음식 배달(앱 기반)
- 레스토랑 리뷰 및 예약
- Zomato Pro 멤버십 혜택
- Blinkit을 통한 퀵커머스(즉시 배달) 서비스 운영

다양한 사용자 **data** 보유

프로젝트 개요



배경 및 문제정의



경쟁사와의 과도한 가격 경쟁

인도 내 Swiggy, Uber Eats 등과의 경쟁 심화로
2019-20년도 손실 160%

고객 확보 비용 급등

고객 유입 중심 전략만으로는
플랫폼의 장기 지속 가능성을 담보하기 어려운 한계

목표 및 기대성과

고객 행동 및 주문 패턴 분석



이탈 가능성이 높은 행동군 식별 및
리텐션 인사이트 도출

매출 decomposition



매출 손실 주요 요인 파악

고객 군집화



세그먼트별 맞춤 전략 수립을 위한
고객 유형 정의

데이터 소개 및 전처리



데이터셋 선정 기준과 평가

5가지 평가 항목을 생성 후 점수를 매겨, 총점 가장 높은 데이터셋을 프로젝트에 사용함

평가항목



1. 분석 목적 명확성

재구매 여부 예측, 유저 세분화 등
명확한 분석 방향 설정 가능



3. 변수 다양성과 품질

orders, users, restaurant 등
5개 테이블 보유, 유저 행동 변수 풍부



5. 데이터 규모 적절

5개 테이블 모두
수만 건 이상 데이터 확보



2. 분석 확장성

분류, 회귀, 군집 등
다양한 기술 적용 가능



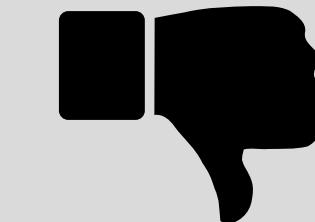
4. 실무 연관성

고객 행동 분석, 고객 이탈 전략 등
외식 산업 내외 실무 적용 가능

원본 데이터 중 일부 테이블 사용

[Content](#)[Contact](#)

사용 테이블



미사용 테이블

[users](#)[orders](#)[restaurant](#)

menu

menu_id
r_id
f_id
cuisine
price

food

f_id
item
veg_or_non_veg

여러 개 메뉴가 한 주문으로 연결
→ 누가 무엇을 주문했는지 추적 불가

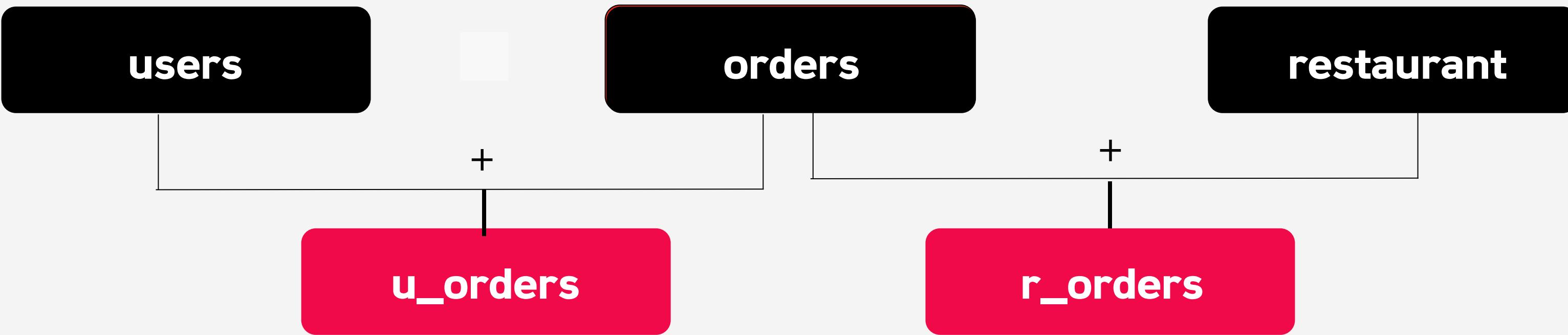
orders에 menu_id가 없음
→ 메뉴 식별 불가

--> 사용자 행동 분석 적합

--> 사용자 행동 분석 부적합

<데이터 1차 병합>

user_id(PK)	user_id(FK)	id(PK)
name	r_id(FK)	name
email	order_date	city
password	sales_qty	rating
Age	sales_amount	rating_count
Gender	currency	cost
Marital Status		cuisine
Occupation		lic_no
Monthly Income		link
educational qualification		
family size		



u_orders 전처리(1)

[Content](#)[Contact](#)

u_orders (148664 rows, 16 col)	
user_id	사용자 고유 id
r_id	레스토랑 고유 id
name	사용자 이름
email	drop 사용자 이메일 주소
password	암호화된 비밀번호
age	사용자 나이
gender	male, female
marital status	혼인 여부
occupation	직업
monthly income	월 소득 수준
educational Qualification	교육 수준
family size	가족 구성원 수
order_date	주문일자
sales_qty	주문수량
sales_amount	주문금액
currency	drop 부제에 사용된 통화

■ 개인 식별 정보

name, email, password 컬럼 삭제

■ 결측치 제거

15만행 중 결측치 개수 약 1,700개
→ 결측치 있는 행 삭제

■ 데이터 타입 변환

user_id, r_id: float → int → category
order_date: object → datetime

■ sales_amount

음수값 존재 → 결제 취소 및 환불 건수 예상
전체 행 중에 단 2행 존재 → 삭제

■ currency

USD 단 2행 존재하고 나머지는 모두 INR(루피)
→ USD 제거하고 컬럼 삭제

u_orders (148664 rows, 12 col)			
user_id	사용자 고유 id	monthly income	월 소득 수준
r_id	레스토랑 고유 id	educational Qualification	교육 수준
age	사용자 나이	family size	가족 구성원 수
gender	성별	order_date	주문일자
marital status	혼인 여부	sales_qty	주문수량
occupation	직업	sales_amount	주문금액

■ One-Hot Encoding

- **marital status:**
'Single', 'Married', 'Prefer not to say'
→ one-hot-encoding
- **occupation:**
Student, Employee, self Employeed, House wife
→ one-hot-encoding
- **monthly income:**
'No Income', 'Below Rs.10000', '10001 to 25000',
'25001 to 50000', 'More than 50000',
→ one-hot-encoding
- **educational qualifications:**
'Post Graduate', 'Graduate', 'Ph.D',
'Uneducated', 'School'
→ one-hot-encoding

r_orders 전처리

[Content](#)[Contact](#)

r_orders (150281 rows, 16 col)	
order_date	주문일자
sales_qty	주문수량
sales_amount	주문금액
currency	drop <small>현재에 사용된 통화</small>
user_id	사용자 id
r_id	식당 id
city	식당이 위치한 도시
cuisine	식당 요리 종류
rating	식당 평균 평점
rating_count	식당 리뷰 수
cost	식당 평균 결제 금액
name	식당 이름
address	식당 주소
lic_no	drop <small>식당 면허 번호</small>
link	식당 웹사이트 링크
menu	메뉴(json)

불필요한 컬럼 삭제

name, address, lic_no, link, menu
컬럼 삭제

결측치 제거

15만행 중 결측치 개수 약 1,700개
→ 결측치 있는 행 삭제

데이터 타입 변환

user_id, r_id: float → int → category
order_date: object → datetime
cost: 통화 기호 제거 후 object → int

sales_amount

음수값 존재 → 결제 취소 및 환불 건수 예상
전체 행 중에 단 2행 존재 → 삭제

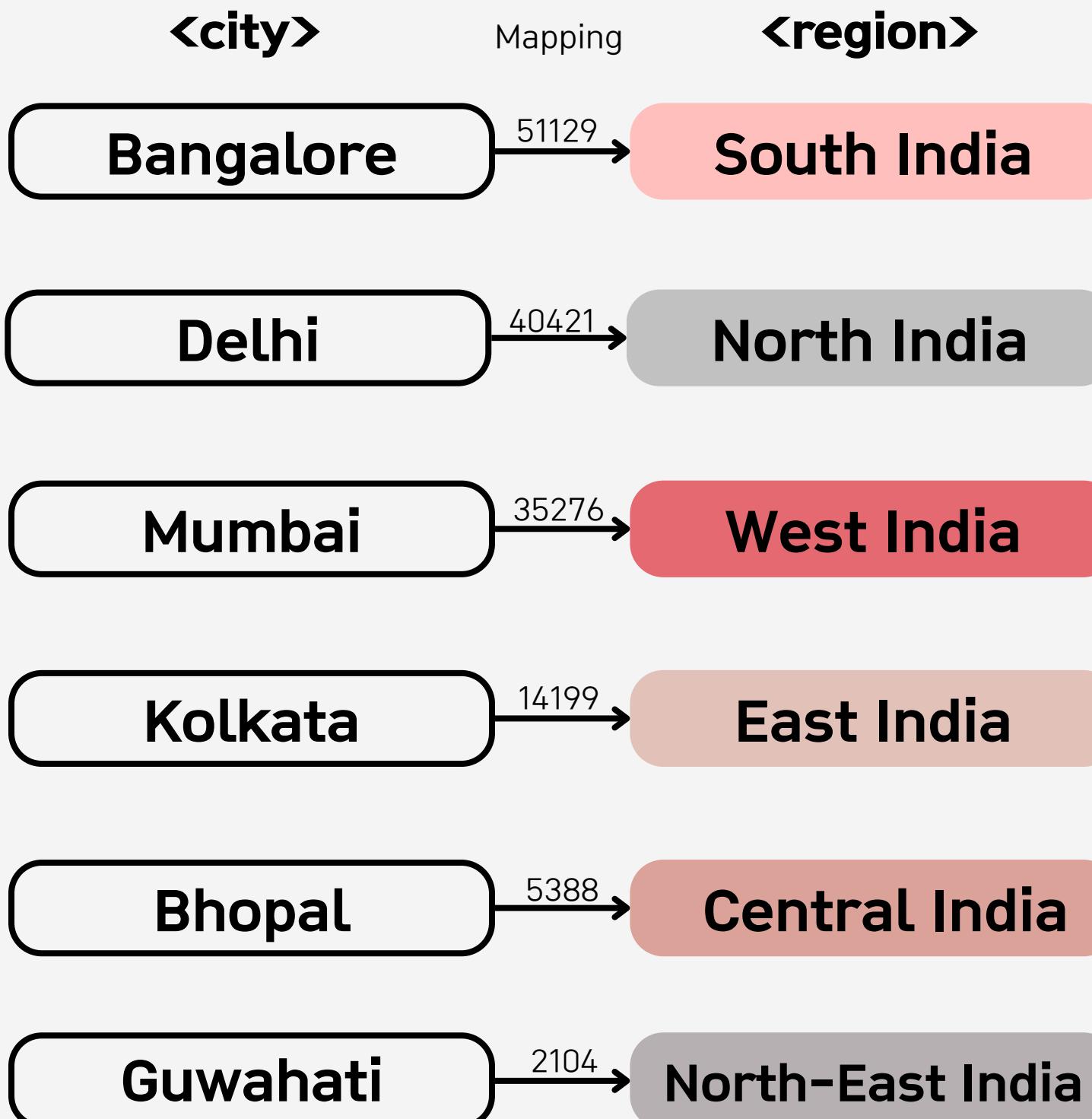
currency

USD 단 2행 존재하고 나머지는 모두 INR(루피)
→ USD 제거하고 컬럼 삭제

rating

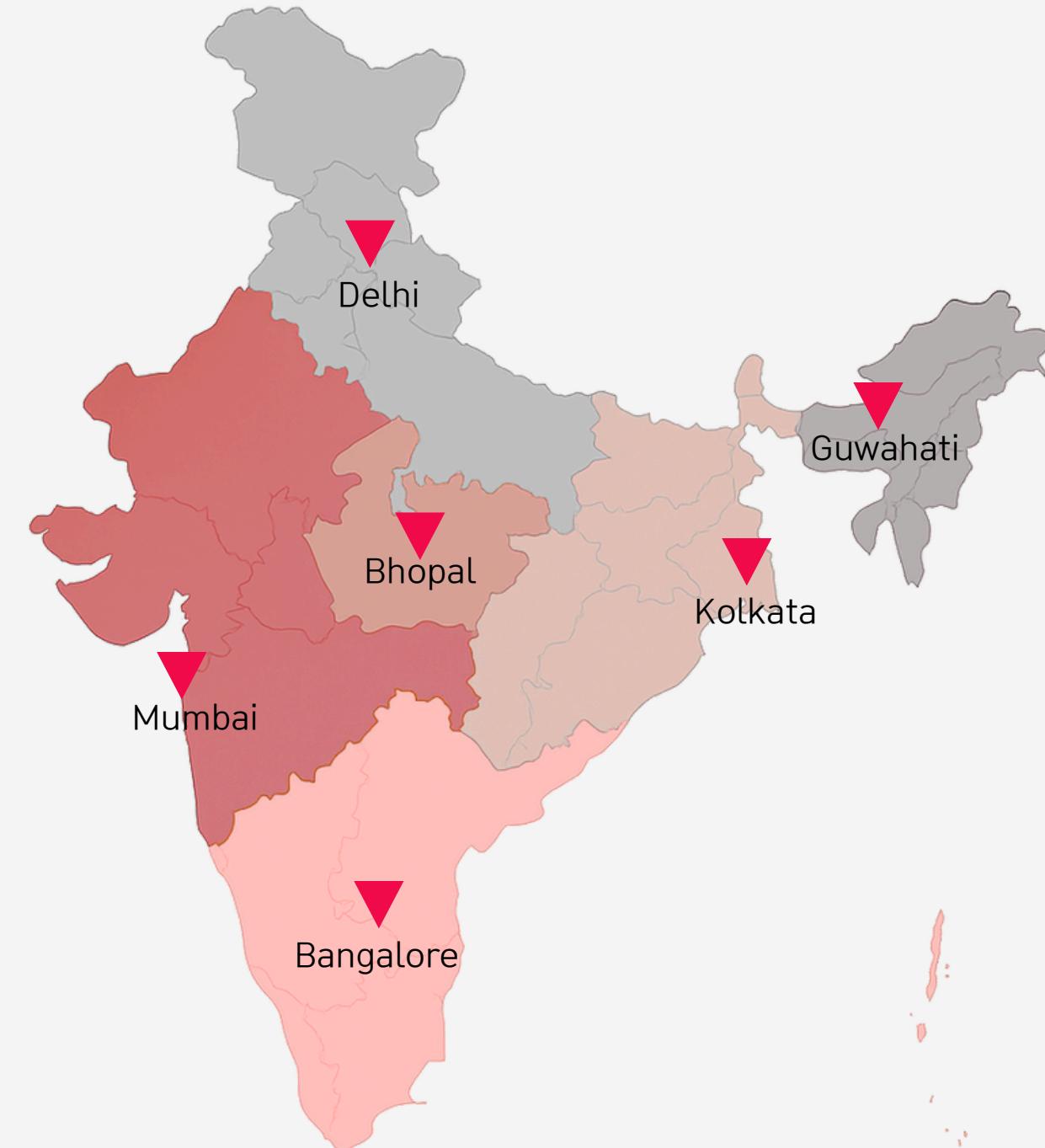
'--' → 'Unrated'로 대체

city(도시) → region(지역) 그룹화

[Content](#)[Contact](#)

city 컬럼에서 실제 도시명만 추출
(예: 'GOTA,Ahmedabad' → 'Ahmedabad')

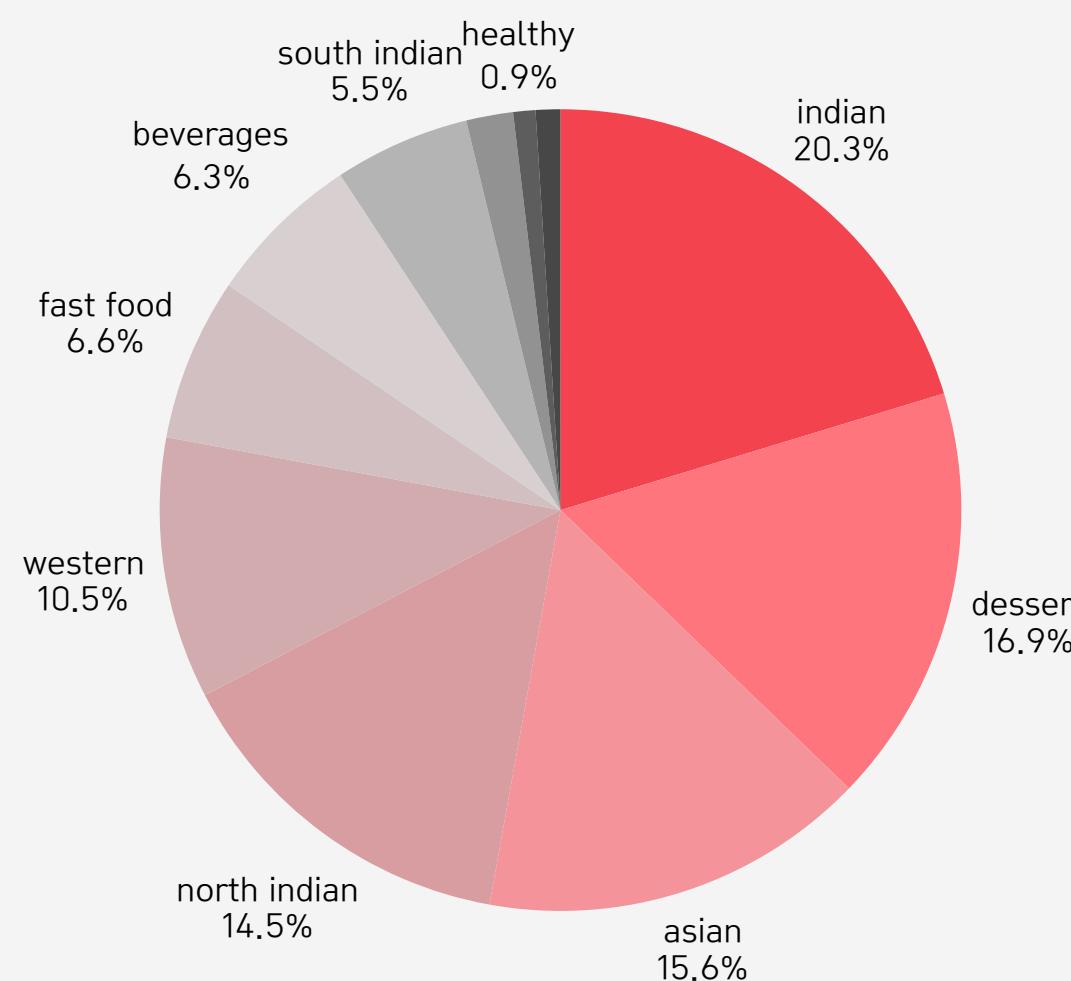
parsed_city 기준으로 region 매핑
초기 매핑률: 62.6% → 최종 매핑률: 100%



cuisine 분류 요약

[Content](#)[Contact](#)**cuisine****cuisine_list****cuisine_categorized****Chinese, Pizzas****Chinese****Pizzas****Asian****Western**

콤마 기준 리스트 변환,
explode() 사용하여 세로로 늘리기



cuisine_categorized (12개 카테고리)

1차 분류
Indian
North Indian
South Indian

2차 분류
Asian
Western
Middle-Eastern

3차 분류
Fast food
Dessert
Beverages

4차 분류
Healthy
Seafood
Meat/Grill

+ cuisine과 관계 없는 행 제거(code valid on bill 등)

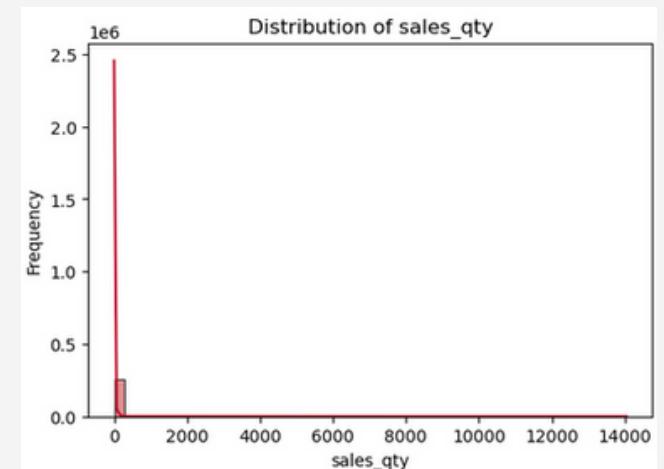
+ 마지막까지 분류에 포함되지 않는 african, tribal, combo 제거

이상치 제거(u_orders, r_orders 공통)

[Content](#)[Contact](#)

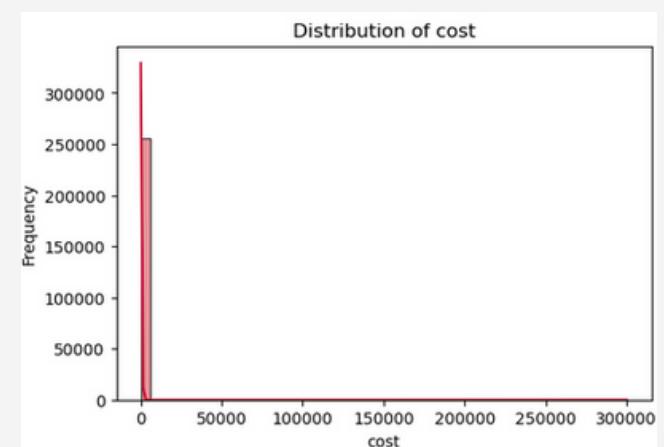
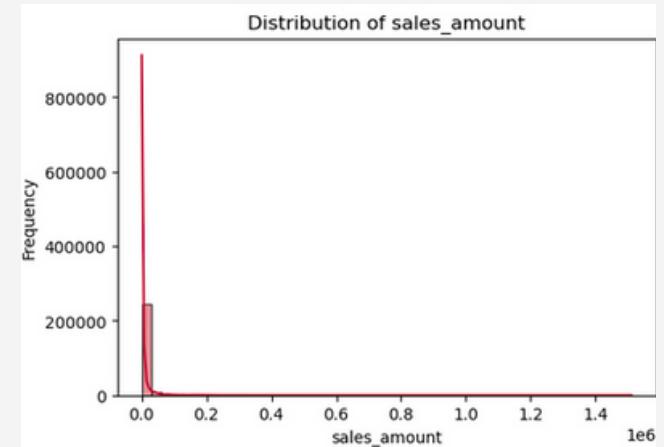
수치형 변수 분포 확인

(sales_qty, sales_amount, cost)



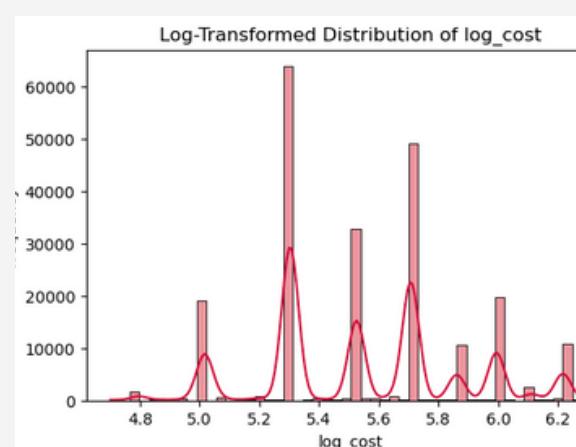
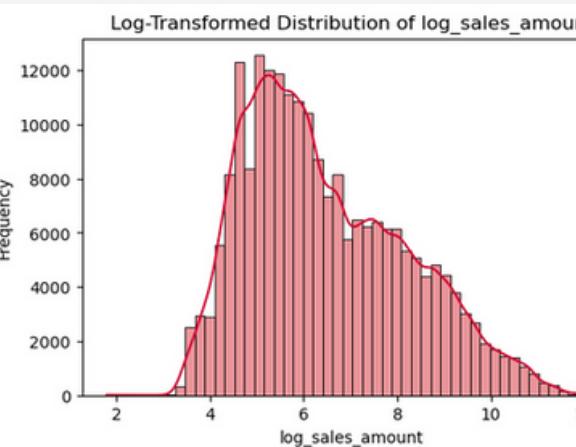
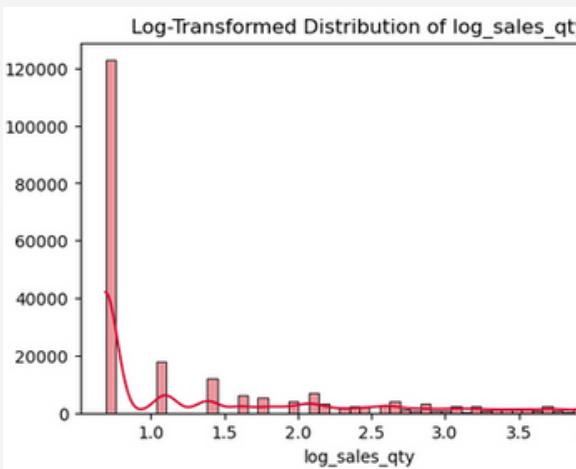
로그 변환

(정규성 없음)



수치형 변수 분포 재확인

(sales_qty, sales_amount, cost)



이상치 제거 (IQR)

원본 행 수: 255576
제거된 이상치 수: 39375

	sales_qty	sales_amount	cost
count	216201	216201	216201
mean	5.93	3357.90	270.37
std	10.63	9513.58	91.74
min	1.00	5.00	109.00
25%	1.00	167.00	200.00
50%	1.00	449.00	250.00
75%	5.00	2194.00	300.00
max	62.00	214870.00	550.00



파생변수 리스트

total_sales_amount	총 주문 금액
total_order_count	총 주문 수
total_quantity	총 수량
avg_order_amount	평균 주문 금액
avg_quantity_per_order	주문당 평균 수량
first_order_date	첫 주문 날짜
last_order_date	마지막 주문 날짜
active_days	활동 기간
days_since_last_order	마지막 주문 후 경과일

avg_days_between_orders	평균 주문 간격
most_active_weekday	가장 자주 주문한 요일
most_visited_restaurant	가장 자주 방문한 레스토랑 ID
order_frequency	하루 평균 주문 빈도
monetary_per_day	하루 평균 소비 금액
recency_ratio	최근 비활성 비율
revisit_count	재방문 수
revisit_rate	재방문율
is_churned	이탈여부



파생변수 리스트

cuisine_variety	레스토랑이 제공하는 요리 종류 수 (0/1 벡터 합계)
indian_focus	인도 전통 요리 비중 (indian,north indian,south indian)
global_ratio	세계 요리 비중 (asian,western, middle eastern)
sweet_ratio	디저트 및 음료 비중 (dessert,beverages)
snack_focus	간편식 요리 비중 (fast food)
single_cuisine	단일 요리 제공 1, 그 외 0
main_cuisines	전체 요리 종류 리스트
cost_per_rating_level	가성비 여부

order_month	주문 월
order_day	요일 번호 (0=월~6=일)
is_weekend	주말 여부 (토~일: 1)
is_unrated	rating =='Unrated' 이면 1 else 0
cost_level	평균 비용 분위수 구간
avg_price_per_item	주문 단가 (결제금액/결제수량)
cost_diff	단가 - 평균비용(cost) 차이
spending_level	평점 구간화
restaurant_review_scale	리뷰 수 기준 레스토랑 규모

<모델링 및 클러스터링용 최종 데이터셋 확정>



🎯 EDA 목표 설계

프로젝트 목표

조마토의 매출 향상을 위한 '리텐션 및 마케팅 전략' 수립

'마케팅팀 전용 대시보드' 생성 → 매출 & 리텐션 & 이탈 지속적 관리

EDA의 목표

1차 EDA로 마케팅의 핵심 타겟 고객군 선별

2차 EDA 분석 방향 파악

🎯 EDA 활용 전략

접근 방식으로 RFM 응용

Recency, Frequency, Monetary 관점으로 인사이트 도출

예상한 EDA 과정

1차 EDA (R, F, M 관점으로 인사이트 확인) →

2차 EDA (필요한 파생변수 생성 & 필요한 추가 분석) →

타겟 고객군 확보

왜 RFM?

타겟이 될 고객 구체화에 적합한, 직관적이고 유의미한 프레임이라 판단하여 RFM 분석을 하지는 않았으나 EDA 단계에서 개념을 활용하고자 함.

1차 EDA (R, F, M 관점으로 인사이트 확인) →

2차 EDA (필요한 파생변수 생성 후 필요한 추가 작업) →

타겟 고객군 확보

실제 과정

1차 EDA 1. 매출과 고객 행동 기초 분석

1-1. 인구통계 확인(연령, 성별, 직업, 소득 등)

1-2. 매출과 리텐션 확인

1차 EDA 2. 행동 기반 지표 분석 → 1차 인사이트

2-1. Recency: 마지막 주문일 관련

2-2. Frequency: 주문 빈도, 주문 간격, 재방문 관련

2-3. Monetary: 누적 소비 금액, 매출 기여도 순위 관련

2-4. 1차 인사이트 확인 → 타겟 군집 확보를 위해 추가 작업 필요 판단

2차 EDA: 클러스터링 결과 + 2차 시각화 → 최종 인사이트

3-1. 클러스터링 결과 확보

3-2. 2차 시각화 (클러스터링 기반 고객 세그먼트 등의 태블로 필드 추가)

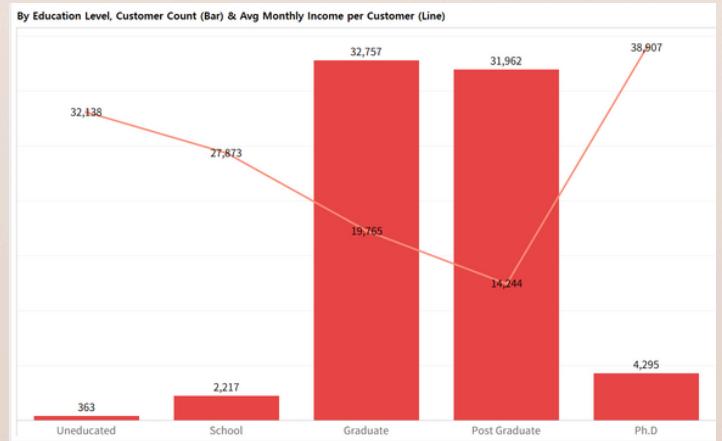
3-3. 인사이트를 전략과 액션 플랜에 반영 (마케팅 전략, 대시보드)

1-1. 인구통계 확인(연령, 성별, 직업, 소득 등)

📌 고객군 인사이트

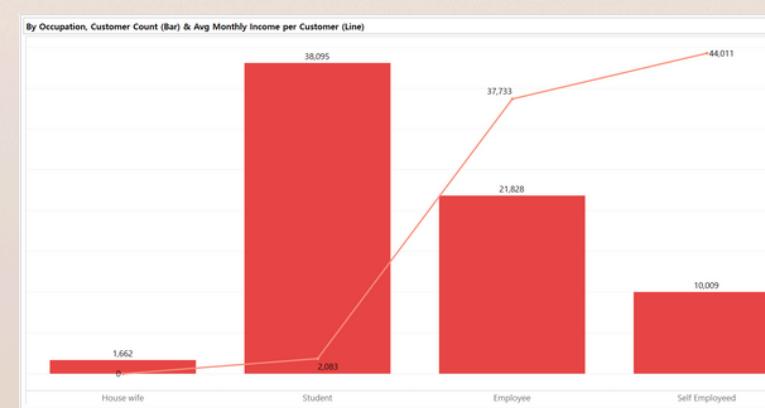
- 고객 대부분은 인도에 사는 10~30대로 연령, 국적 다양성 낮음
- 직업에서는 학생이, 소득에서는 무소득자(용돈 등의 원인으로 월 소득은 상위권)이 절반 가량을 차지
- 무소득을 제외하면 중산층이 다음으로 많음
- 고학력자 비율이 절반 이상 차지
- 자영업자의 소득이 가장 높게 나타남 (Uneducated와 관련 있을 가능성 추측)
- 직장인과 자영업자는 비중은 비교적 적으나 구매력이 높아 마케팅 전략 대상이 될 가능성 확인 필요

By Education Level, Customer Count (Bar) & Avg Monthly Income per Customer (Line)



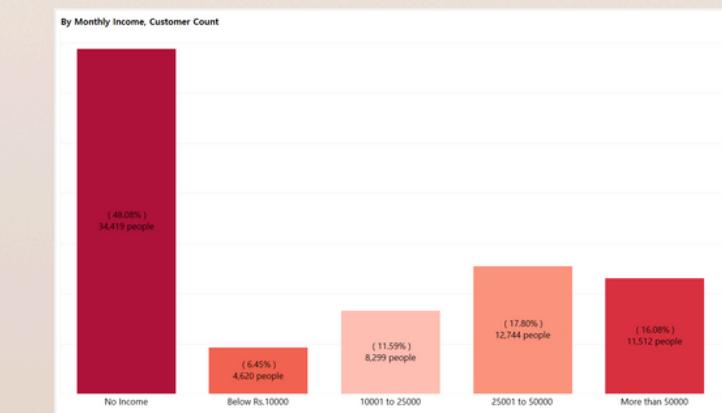
Uneducated 고등교육 받지 않음(소득 상위)
School 재학중
Graduate 학사
Post Graduate 석사 이상
Ph.D 박사 이상(소득 상위)

By Occupation, Customer Count (Bar) & Avg Monthly Income per Customer (Line)



처음부터 범주형으로 존재하던 직업 컬럼: 4고객군으로 나누어 있음.
여기서의 Student는 비중은 가장 크나 구매력이 낮음.
직장인과 자영업자는 비중은 비교적 적으나 구매력이 높음.
이 두 고객군을 합하면 Student 군의 수와 비슷해지기에
LTV 높은 고객군, 즉 마케팅 전략 타겟이 될 가능성 확인.
보다 구체적인 고객별 매출 기여도 확인 필요하다 판단한 근거.

By Monthly Income, Customer Count



No Income 무소득
0 ~ Rs. 5,000b 극빈층 (poorest)
Rs. 5,001 ~ 10,000 저소득층 (lower income)
Rs. 10,001 ~ 25,000 중하층 (lower-middle class)
Rs. 25,001 ~ 50,000 중산층 (middle class) - 무소득 다음으로 많은 분포.
Rs. 50,000 이상 상위 소득층 (upper class)

PLFS 공공통계와 CPHS 민간조사 등에 따르면
2019년 기준 인도의 개인 월 소득 중앙값은 약 8,000~10,000 INR 사이임.

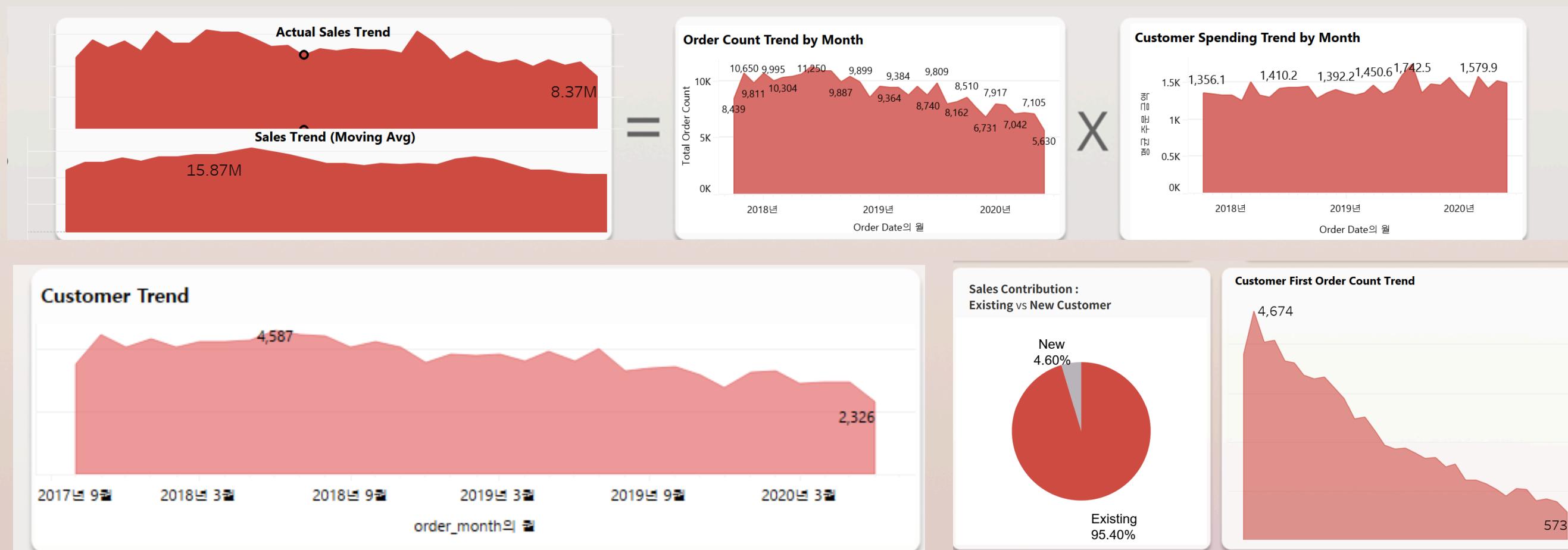
1-2. 매출과 리텐션 확인

📌 매출 하락세 확인:

고객의 이용 금액은 늘어나는 추세이나 주문 횟수가 줄어들어 전체 매출이 줄어드는 것으로 추정됨.
→ 주문 횟수를 늘리는 전략이 우선 필요하다고 판단

📌 고객 수 감소 확인:

그 원인은 신규 고객 줄어드는 폭이 커서인 것으로 추정됨, 그러나 신규 고객의 매출 기여도가 낮음.
→ 기존 고객을 대상으로 한 리텐션 전략이 더 우선이라고 판단



1차 EDA 2에서 1차 인사이트 확인 결과, 타겟 군집 확보를 위해 추가 작업 판단

2차 EDA로 클러스터링 결과를 반영한 2차 시각화 진행→ 최종 인사이트 확보

📌 1차 EDA 후 1차 인사이트

1. 고객군 간 유의미한 차이 포착

구매금액, 주문횟수, 이용기간, 최근 주문일 등에서 의미 있는 차이 관측 → 고빈도 고액 고객군을 식별할 수 있다는 가능성 확인

2. But 일부 지표에서 변별성 부족

주문간격, 직업, 연령 등의 컬럼 분석 결과 → 고객군 간 뚜렷한 차이 미발견

원인: 데이터의 다양성 부족? or 분석 단계의 얕음? → 판단 위한 심화 분석 필요

📌 2차 EDA 후 프로젝트 방향 설정

3-1. 1차 인사이트를 보완하기 위한 클러스터링 진행

세그먼트 확보 → ROI 높은 고객군 식별 가능성 높아짐, 타겟 마케팅 설계 기반 확보

3-2. 클러스터링 결과를 더한 2차 시각화 진행

클러스터링 결과 확보된 ‘유저 세그먼트’를 데이터셋에 반영 후 태블로에서 (복합 필드 생성 통한 입체 분석)

1차 EDA는 Tableau 기반 단일 조건 분석 위주

2차 EDA는 다중 조건 필터링, 계산 필드 조합 분석으로 확장→ 패턴의 교차점 발견, 세부 고객 행동 구조 파악

3-3. 2차 인사이트 확인 후 액션너블 전략형 대시보드 제작 계획 수립

보강된 고객 세그먼트로 분석 고도화 → 마케팅팀이 전략 실행 가능한 actionable 대시보드 제작 계획 수립

모델링 & 클러스터링

고객 이탈 예측 모델링 과정

목적 : **zomato 고객 이탈 여부를 분류하는 이진 분류 모델 개발**

사용 모델 :

RandomForest

XGBoost

MLP

목적 : **총 5가지 이탈 정의 방식**에 따라 비교 

고객 이탈 예측 모델링 개선 과정

이탈(=Is_churned) 정의 기준을 바꾼 것이 개선의 핵심

- 1차** $\text{days_since_last_order} - \text{last_order_date} > 100$
+random split
마지막 주문일로부터 100일 초과
- 2차** $\text{days_since_last_order} - \text{last_order_date} > 100$
+ 시계열 정렬 후 70% split
1차 기준 그대로 사용하되, 시계열 분할 적용
(train : 최근 이전) (어떤어떤 문제 발생)
- 3차** $\text{last_order_date} - \text{before_last_order} > 100$
마지막 주문일과 마지막 주문 이전의 간격이
100일 초과
- 4차** $\text{last_order_date} < 2020-03-01$ and $\text{order_count} \geq 2$
가장 최근 주문일자가 '2020-03-01' 이전
이면서 2회 이상 주문
- 5차** $2020-03-01 - \text{last_order_date} > 100$
& $\text{order_count} \geq 2$
'2020-03-01'기준으로 100일간 주문 없음
& 총 주문 2회 이상

고객 이탈 예측 모델링 개선 과정



정의1의 경우,

문제 :

단 1회 주문자까지 이탈자로 포함 →
단발성 고객 왜곡 / 클래스 불균형



정의3의 경우,

문제 :

이탈자 과다 추출 →
모델이 이탈자만 예측



정의5의 경우,

문제:

전체 고객 중 잔존자가 너무 적음→
데이터 불균형 극심 → 모델 편향



정의2의 경우,

문제 :

시계열 분할은 보완됐으나
불균형 문제 여전



정의4의 경우,

문제 :

기준을 늦춰도 클래스 불균형 지속 →
예측 타겟이 너무 협소해짐

고객 이탈 예측에서 전략 고객 식별로

1 처음엔 고객 이탈이 문제라고 생각했다.

- 유저 수 감소와 매출 하락을 보고,
“이탈 예측 모델을 통해 이탈 가능성이 높은 고객을 선제 대응하자”고 판단
→ is_churned 변수 정의 및 모델링 시도

하지만..

- 예측 결과가 실행 가능한 인사이트로 이어지기 어려웠음
- 이탈 예측 보다는 의미 있는 고객군 정의와 인사이트 도출이 우선적이라고 판단

2 그렇다면, EDA를 통해 본질적인 문제를 다시 정의하자.

“떠나는 고객을 보기보다, ‘매출에 기여하는 고객이 누구인가?’에 주목”

- 고객 대다수 평균 구매액이 낮고, 신규 유입도 줄고 있음
- 그런데 상위 3% 고객이 전체 매출의 15% 기여
- 즉, 매출을 유지하는 핵심 고객은 소수에 집중되어 있음



3 전략 전환: ‘누가 매출에 기여하는가’

- 단순 평균 고객 분석 -> ROI 중심 고객군을 식별하고 집중하는 전략
- 이 전략을 구현하려면,
사전 정의 없이 행동 데이터 기반으로 그룹을 나눌 수 있는 방법이 필요
- 비지도 학습 = 클러스터링이 최적

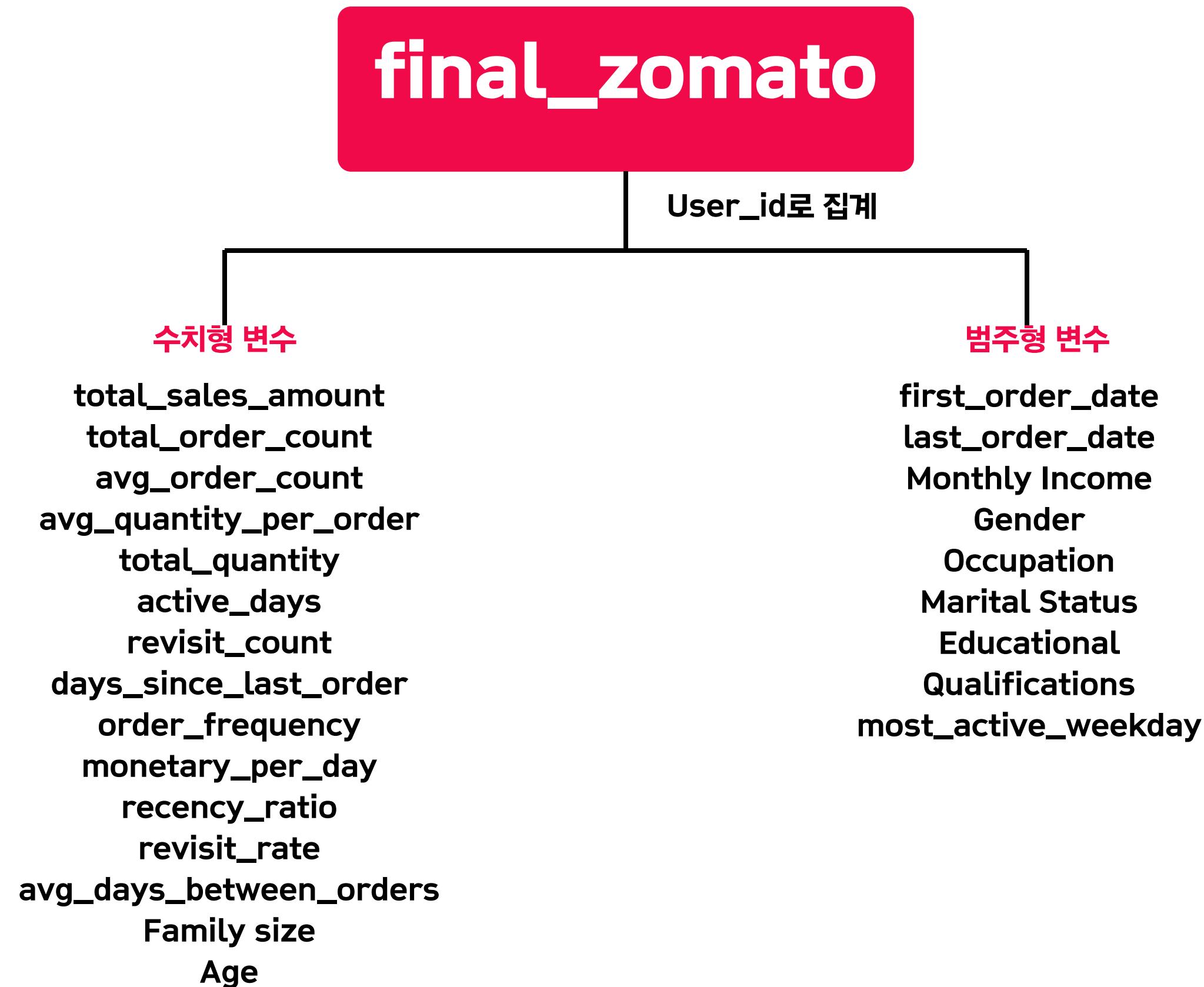
4 그래서, 클러스터링을 선택했다.

- 다양한 고객 행동 지표(금액, 빈도, 선호도 등)를 기반
→ 자연스럽게 형성된 고객 세그먼트를 식별
- 이를 통해 ROI가 높은 고객군을 명확히 타겟팅 가능
- 클러스터 수($k=3$) 설정도 고액-중간-저액 고객군 구조에 맞춰 설계

전략 요약

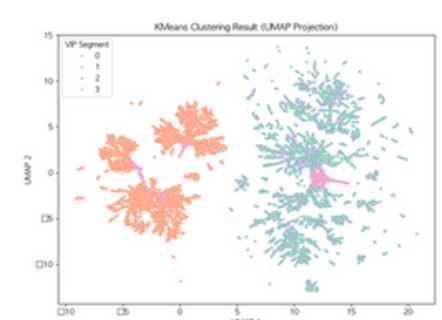
- 고객 이탈 예측은 실효성 낮음
- 데이터 기반 재정의: ‘누가 기여하는가’에 주목
- ROI 높은 고객을 식별하기 위해 클러스터링 도입
- 클러스터별 마케팅 타겟 전략 수립 기반 마련

클러스터 진행



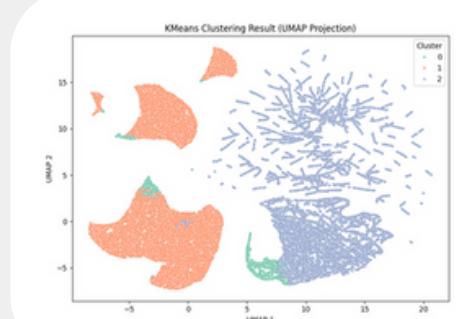
클러스터링

K-Means



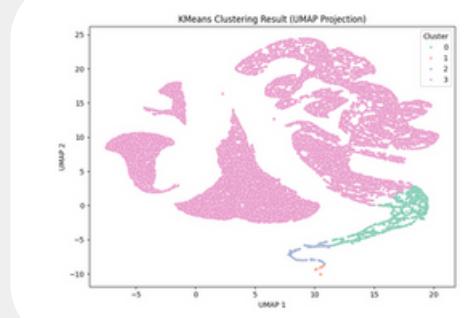
1

Silhouette score : **0.1749**
Davies-Bouldin Score : **1.708**



6

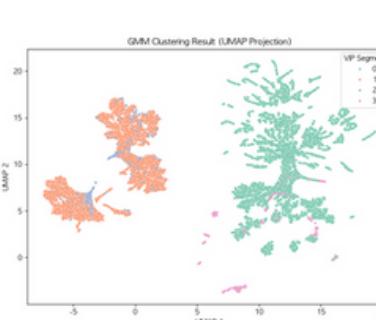
Silhouette score : **0.4806**
Davies-Bouldin Score : **1.0339**



26

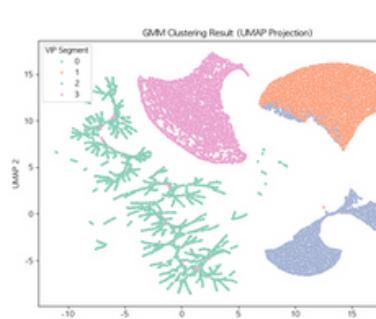
Silhouette score : **0.7798**
Davies-Bouldin Score : **0.539**

GMM



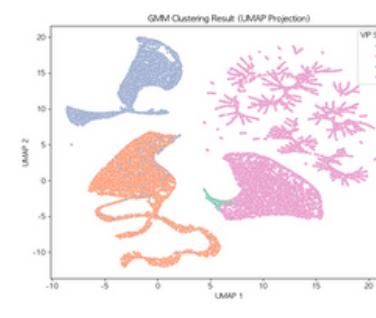
8

Silhouette score : **0.1942**
Davies-Bouldin Score : **3.7141**



16

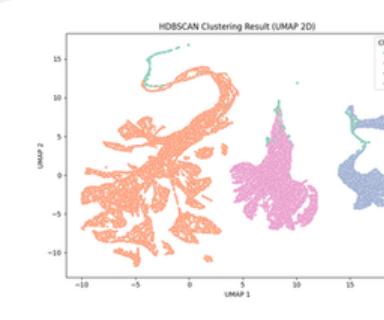
Silhouette score : **0.174**
Davies-Bouldin Score : **2.1432**



17

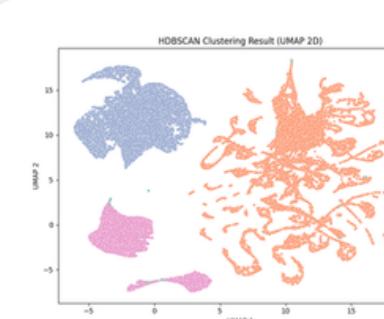
Silhouette score : **0.2197**
Davies-Bouldin Score : **1.6674**

DBScan



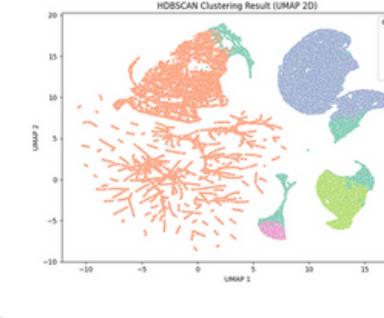
25

Silhouette score : **0.0096**
Davies-Bouldin Score : **2.0017**



17

Silhouette score : **0.3115**
Davies-Bouldin Score : **1.5543**



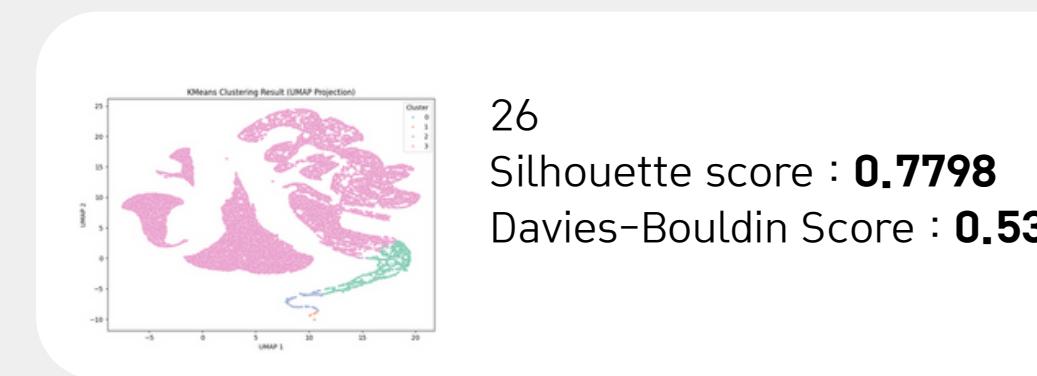
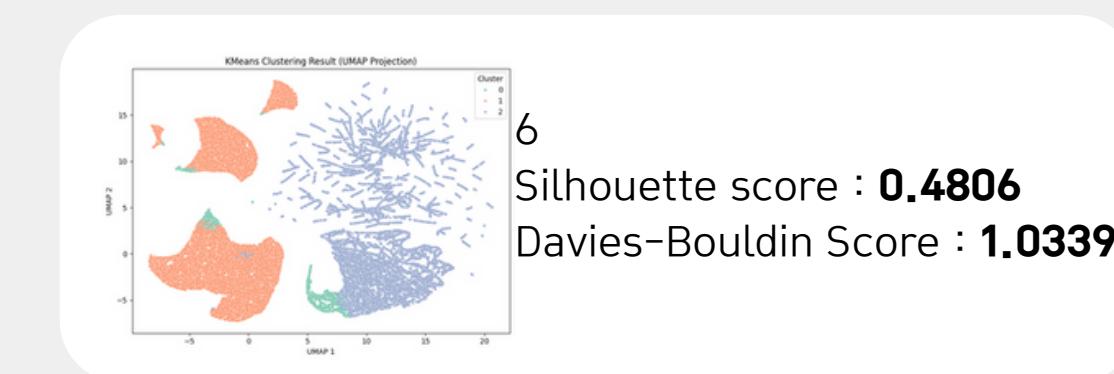
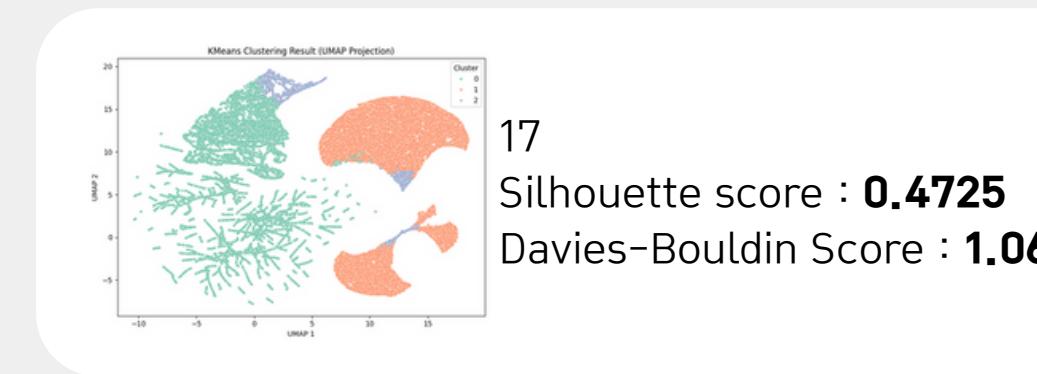
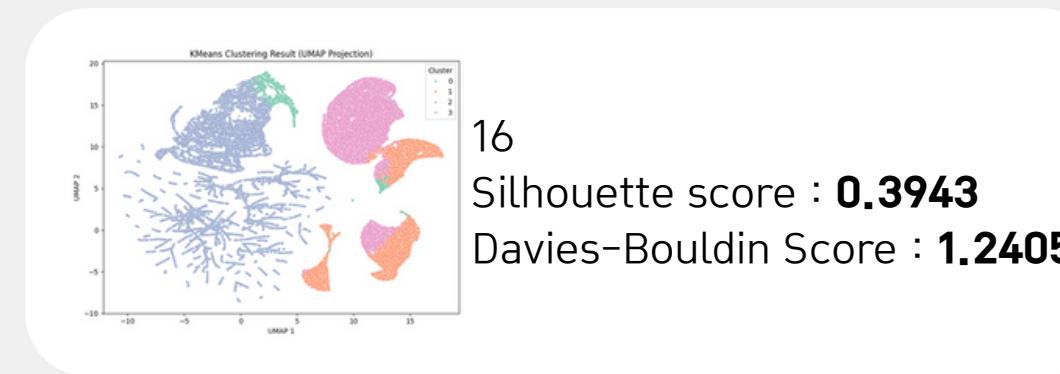
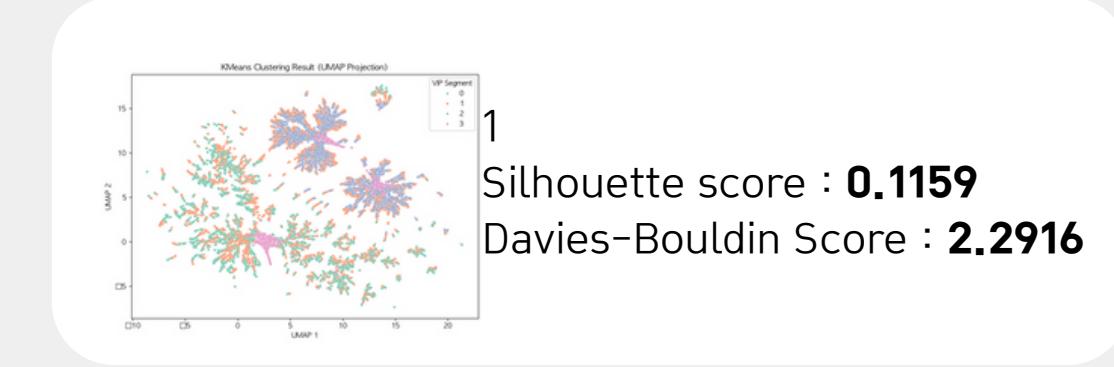
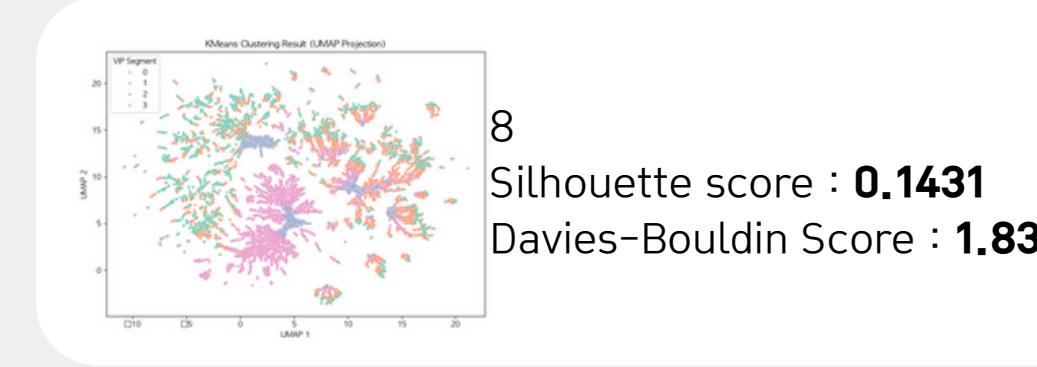
19

Silhouette score : **0.3249**
Davies-Bouldin Score : **1.4393**

클러스터링

K-Means

(K-Means가 다른 방식보다 전반적으로 성능이 우수하다고 판단하여 집중적으로 시도)



최종 클러스터 k=3

모델

K-Means

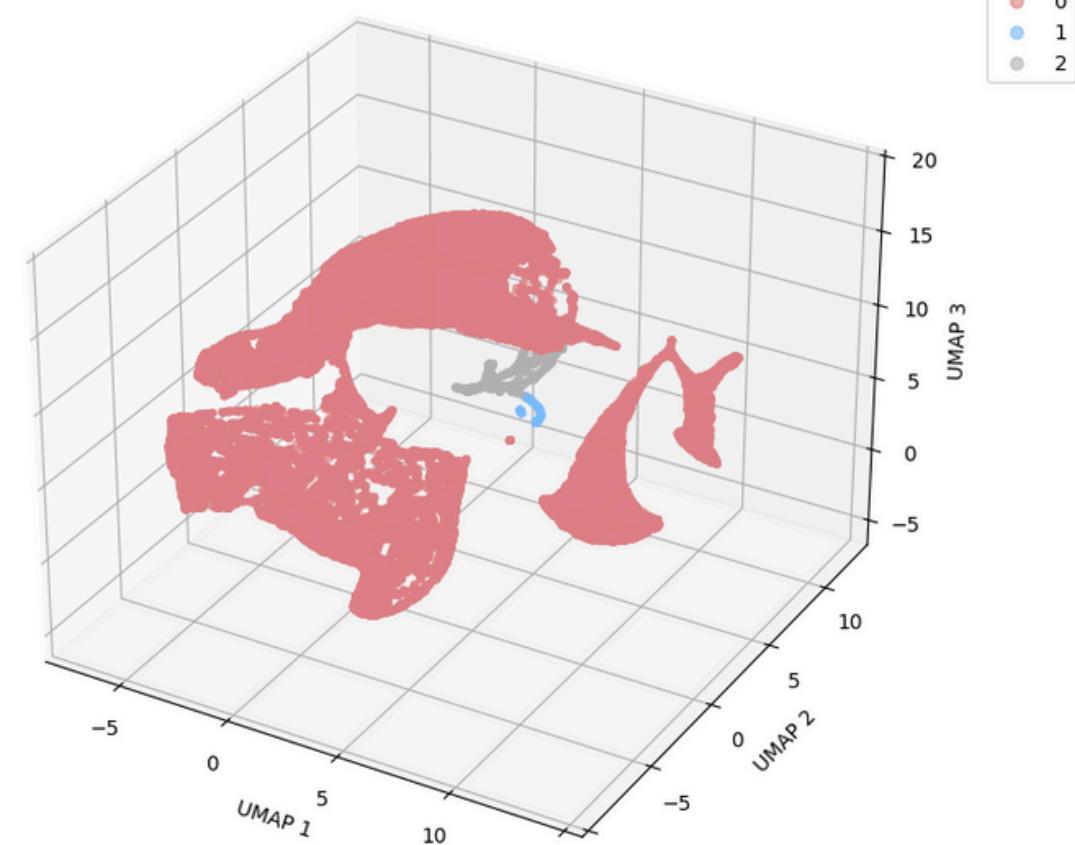
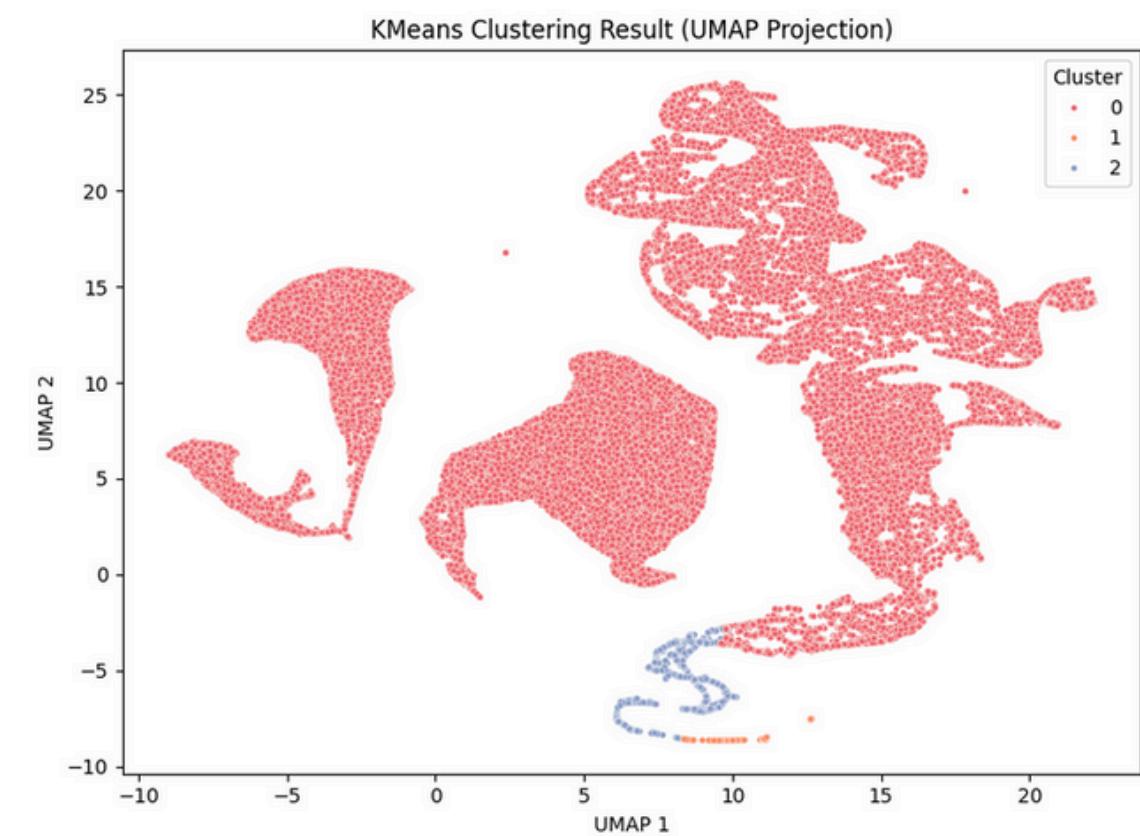
전처리

결측치 삭제
RobertScaler

Silhouette Score : 0.8809
Davies-Bouldin Score : 0.4931

사용 변수

- total_sales_amount
- total_order_count
- avg_order_count
- avg_quantity_per_order
- total_quantity active_days
- days_since_last_order
- order_frequency
- monetary_per_day
- recency_ratio
- avg_days_between_orders



Regular User(Cluster 0) 69,569명

1. 오래, 자주 방문하는 핵심 충성고객층

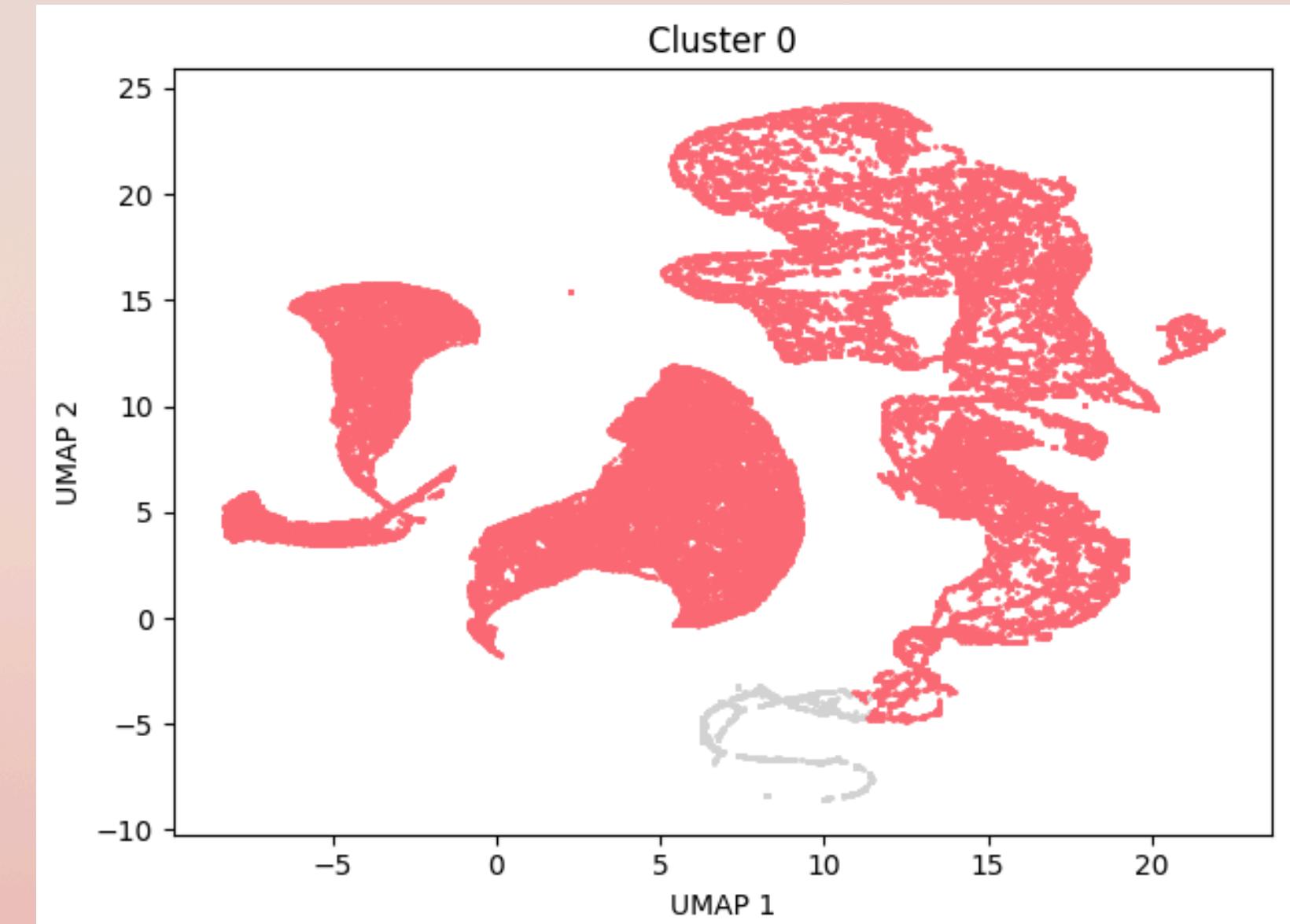
- 플랫폼을 장기간에 걸쳐 꾸준히 이용하는 고객들로 구성
- 다른 그룹에 비해 서비스 이용기간이 길고, 주문 빈도 역시 높아 핵심 충성고객층으로 분류
- 기업의 전체 활성 사용자 수 유지에 중요한 역할

2. 1회 구매액은 낮지만, 활동 기간이 긴 고객층

- 개별 주문의 결제 금액이나, 구매 수량은 비교적 낮음
- 서비스를 이용하는 기간이 길기 때문에 전체적으로 높은 누적 매출 기여

3. 서비스 재구매/재방문률이 높은 고객층

- 다른 군집에 비해 서비스에 대한 재구매율과 재방문 빈도가 높게 나타남
- 정기적으로 서비스를 이용하면서 브랜드에 대한 신뢰와 친밀도가 높은 편



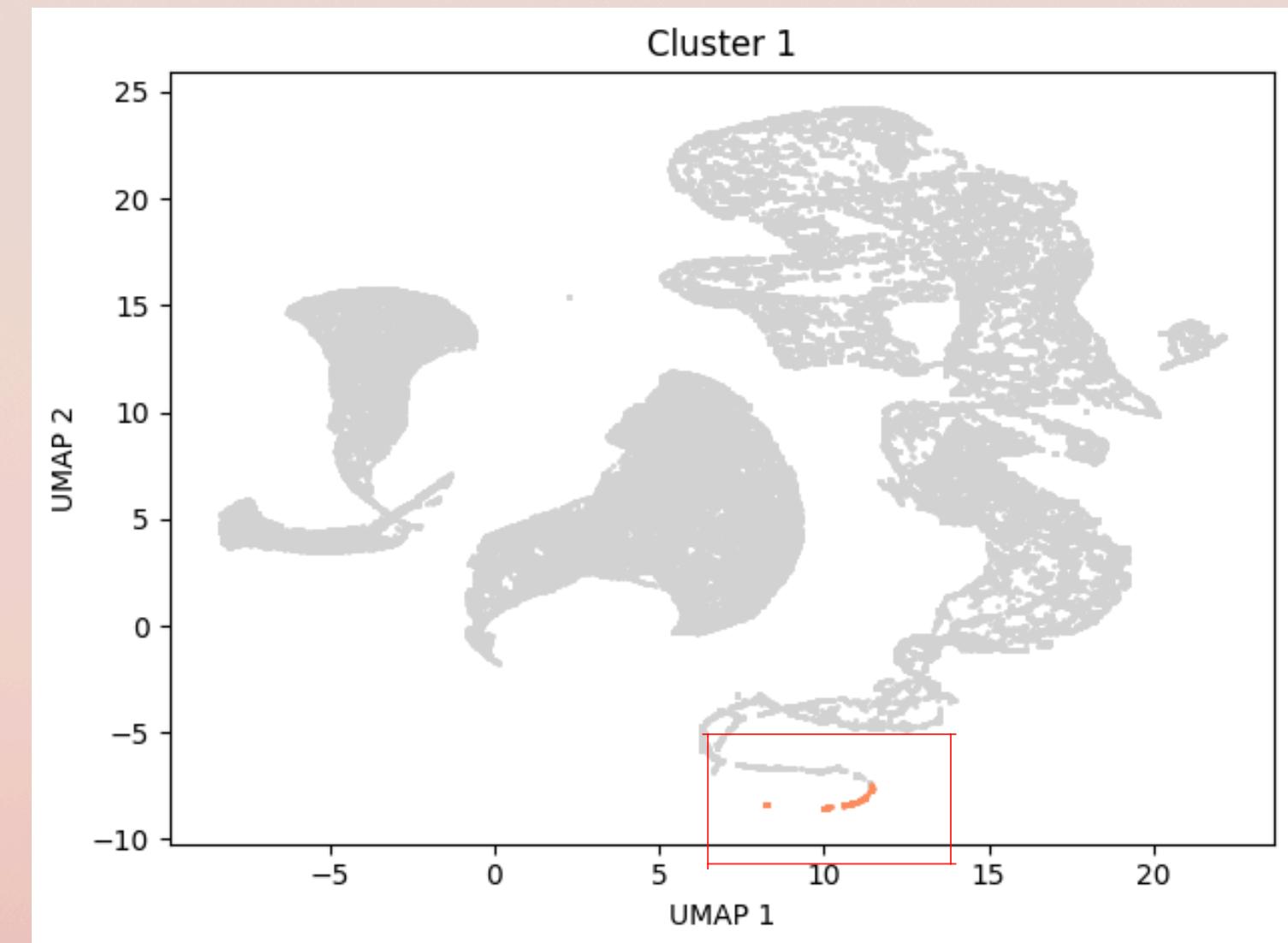
Premium One-Timer(Cluster 1) 206명

1. 주문 빈도는 매우 낮지만, 단일 거래 규모는 압도적인 고객군

- 플랫폼을 이용하는 횟수와 전체 활동기간은 극히 짧지만,
한 번 주문할 때마다 결제 금액과 구매 수량이 큰 고객
- 일상적인 소액 반복 구매와는 반대의 경향

2. 재방문 및 장기 이용은 거의 없는 고객군

- 서비스에 재방문하거나 오랜 기간 이용하는 비율이 낮고,
단 기간 1회성 소비하는 고객
- 단체 주문이나, 대량 구매(법인, 행사, 특별목적)를 하는 것으로
추측



At-Risk Potential(Cluster 2) 1,819명

1. 높은 구매 단가와 수량을 보이는 고객군

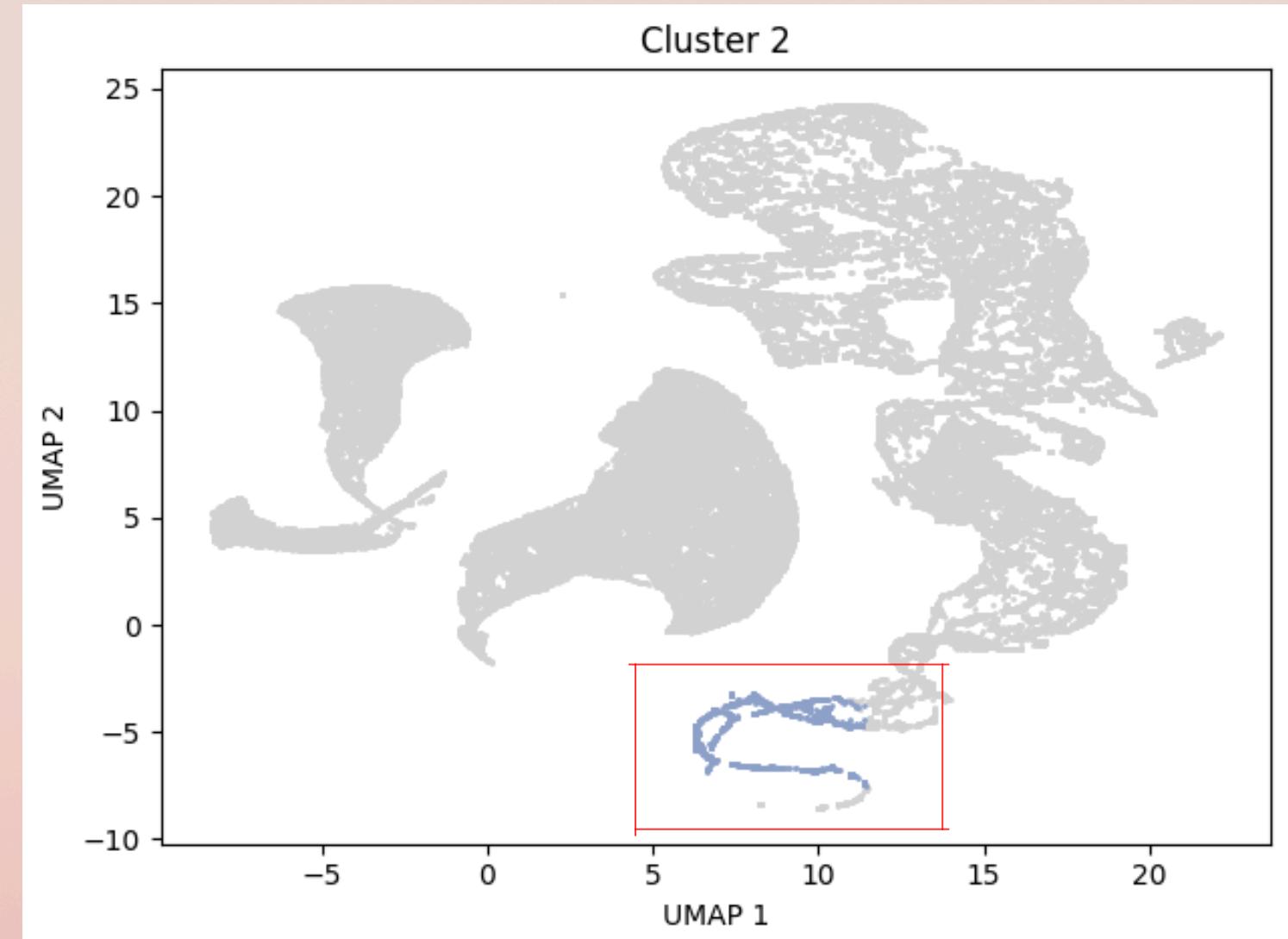
- 주문 시 결제 금액과 구매 수량이 Loyal Regular보다 월등히 높지만, Premium one-Timer고객만큼 크지는 않은 고객

2. 재방문 및 장기 이용은 낮으나, 그렇지 않은 소수 고객 존재

- 서비스 이용 기간이 길지 않으며, 주문 횟수와 빈도도 낮은 편
- 재구매를 하는 고객도 소수 존재

3. 성장 가능성이 큰 미래 후보군

- 현재도 고가치 고객이지만, 서비스 만족도가 높아질수록 점진적으로 이용 빈도가 더욱 커질 수 있음



프로젝트 목표

조마토의 매출 향상을 위한 ‘리텐션 및 마케팅 전략’ 수립

‘마케팅팀 전용 대시보드’ 생성 → 매출 & 리텐션 & 이탈 지속적 관리

대시보드 소개

대시보드 목표

매출 & 리텐션 & 이탈 추적 + 전략 확인(타겟 고객, 자동화) 용이

대시보드 사용자 정의

조마토 마케팅팀

매출 대시보드

마케팅팀을 위한 활용 TIP

- 전사 매출 흐름을 빠르게 파악하고,
시기별/지역별/고객세그먼트별
매출 기여도를 분석하여
마케팅 전략 수립과 예산 배분
에 활용 가능

Guide 1. 필터 사용법

- **기간 필터**
연도 및 월 선택 가능
→ 원하는 시점의 매출 추이 확인 가능
- **분석 기준 필터**
성별 / 연령대 / 직업군 중 선택
→ 세그먼트별 매출 비교 분석 가능
- **통화 단위 안내**
INR(인도 루피) 기준 제공
(2019년 기준 환율:
1USD ≈ 70INR)



Guide 2. 상단 KPI(핵심 지표)

- **Total Revenue (총 매출액)**
전체 매출 규모를 빠르게 파악
→ 시즌별 캠페인 효과 및 연간
매출 성과 비교 지표로 활용
- **Total Orders (총 주문 수)**
→ 고객 유입 변화나 캠페인 반응
모니터링에 유용한 지표
- **Total Food Quantity (총 음식 수량)**
→ 고객이 주문한 전체 음식 수량의
합계
→ 메뉴 구성 전략 및 유형별
(단품 vs 세트) 판단
- **Avg Order Value (평균 주문 금액)**
1회 주문당 평균 결제 금액
→ 객단가 트렌드 확인 및
고소득층 분석 시 활용
- **Avg Order Quantity (평균 주문 수량)**
1회 주문 시 포함된 평균 음식
수량
→ 주문 유형(1인/다인) 구분에
활용

매출 대시보드

마케팅팀은 이렇게 활용해주세요!

Guide 3. 항목

- Monthly Avg Sales

Amount(월별 평균 매출액)

→ 특정 월의 급증/급감 확인
프로모션 반응 분석 및 연간 캠페인

스케줄링에 활용

- Avg Sales Amount by Day of Week(요일별 평균 매출)

→ 요일별 소비 트렌드 확인 및
요일 마케팅 전략 수립에 활용

Guide 4. 항목

- Order Frequency & Order Amount

(주문 빈도에 따른 매출)

→ 고빈도 문 고객군의 매출 기여도
확인 가능

→ 세트 구성 전략에 우선 타겟 도출
유용

- Customer Count by Order Value(금액 구간대별 고객 분포)

평균 주문 금액 구간별 고객 분포

→ 고객이 선호하는 가격대 탐색
및 타겟 가격 전략 수립

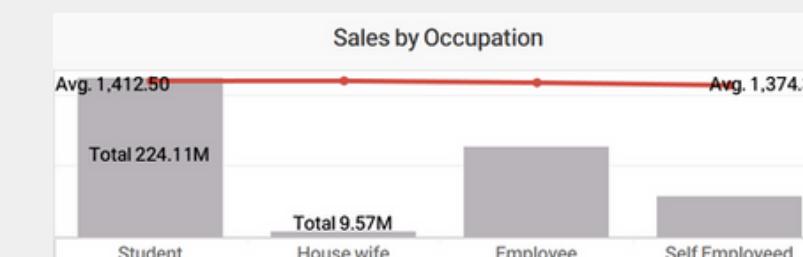
- Customer Orders & Total Sale(주문 수 대비 총매출 분포)

→ 고매출 고객군 탐색

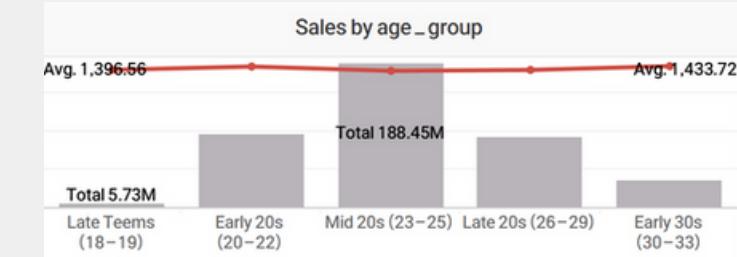
→ VIP 및 리텐션 타겟 설정에 활용



Sales by Occupation



Sales by age group



Guide 5. 항목

- Sales by Gender/Occupation/Age group

(선택 기준별 총 매출 및 평균 주문 금액 비교)

※좌측 필터(성별/직업군/연령대)
선택에 따라 해당 항목별 매출비교로
자동 변경됨
→ 매출 기여 높은 세그먼트 식별 가능

- Regional Sales & Repurchase Rate

(지역별 매출 & 재구매율)

→ 충성 고객 밀집 지역 파악
→ 지역별 맞춤 오프라인 캠페인,
광고 예산 우선 배정 등
로컬 마케팅 전략 수립에 활용

이탈률 대시보드

마케팅팀은 이렇게 활용해주세요!

마케팅팀을 위한 활용 TIP

- 마케팅팀이 이탈률 대시보드를 보고 어떤 내용을 전반적으로 확인할 수 있고 이 대시보드를 통해서 어떤 걸 할 수 있을지 포괄적으로 담는 느낌 = 이게 결국 용도 = 목적

Guide 1. 필터 사용법

- **기간 필터**
연도 및 월 선택 가능
→ 원하는 시점의 매출 추이 확인
- **통화 단위 안내**
INR(인도 루피) 기준 제공
(2019년 기준 환율:
1USD ≈ 70INR)



Guide 2. 항목

- **Actual Sales Trend (실제 매출 추이), Sales Trend(Moving Avg)**

시간 흐름에 따라 매출 추이 변화 식별 가능
→ 연간/분기별 실적 계획 수립

- **Order Count Trend Month(주문 건수)**

매출 변화를 주문 건수와 대비시켜 파악
→ 활성 고객 확보 시기 파악 후 리마인드 마케팅 필요성 판단

- **Customer Spending Trend(월별 평균 소비 금액 트렌드)**

주문 건수와 매출 그래프 함께 분석시 원인 분석 가능→ 고객 단가(ARPU) 증감 파악→ 프로모션 전략 조정

이탈률 대시보드

마케팅팀은 이렇게 활용해주세요!

Guide 3 항목

- Sales Amount by Month (2018 & 2019 Combined)**

월별로 매출 트렌드를 파악
→ 매출 감소 시점에 이탈 요인 있는지 진단, 개선 전략 도출

- Unique Buyers Count by Month (2018 & 2019 Combined)**

월별 실제 구매를 한 고객 수 변화
→ 고객 기반 성장을 모니터링 지표로 활용

- Avg. Sales Amount by Month (2018 & 2019 Combined)**

고객이 월별 평균적으로 소비한 금액 → 월별 객단가 변화 추적 (효과 높은 프로모션 방식 최적화)



Guide 4 항목

- Total Sales by region (지역별 총매출 현황)**

각 지역별 누적 매출 규모를 비교
→ 고매출 지역, 저매출 지역 별 특성에 맞는 캠페인 기획

- Sales Contribution: Existing vs New Customer**

매출에서 신규고객과 기존 고객이 차지하는 비율
→ 고객별 맞춤형 CRM 전략설계

- Customer First Order Count Trend**

신규 고객 유입 여부 확인
→ 신규 유입을 트렌드별 확인 후 저조한 시기에 전략 기획

고객 대시보드

마케팅팀은 이렇게 활용해주세요!

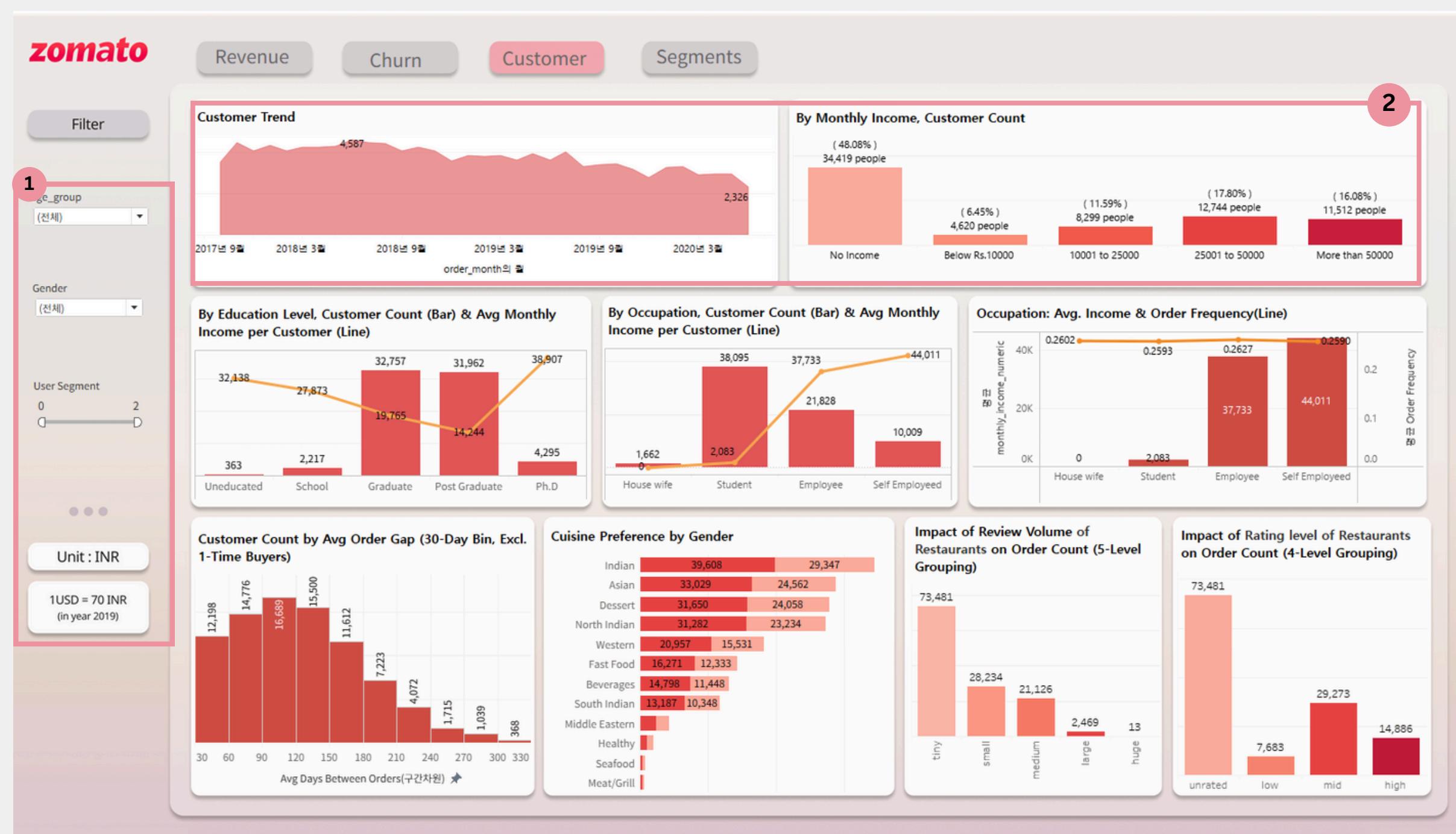
마케팅팀을 위한 활용 TIP

- 핵심 고객군의 연령, 성별, 소득, 직업 특성을 파악하고, 구매력, 주문 주기, 음식 선호도 기반의 타겟 마케팅을 기획하여 고객 세그먼트별 맞춤 전략 수립에 활용 가능

Guide 1. 필터 사용법

- 연령대 필터(18~33세)**
연령대에 따른 소비 성향 세분화
→ 세대별 선호 메뉴, 소비 규모 파악
- 성별 필터**
성별에 따른 고객 특성 비교 가능
→ 남여별 주문 성향 확인 가능
- 세그먼트별 필터**
(클러스터링 분석을 기반으로
분류된 3가지 세그먼트)
0 = Loyal Regular
1 = Premium One-timer
2 = At-risk Potential)

- 통화 단위 안내**
INR(인도 루피) 기준 제공
(2019년 기준 환율:
1USD ≈ 70INR)



Guide 2. 항목

- Customer Trend**
(전체 고객 수의 월별 변화 추이)
→ Zomato의 전체 고객 기반이 어떻게 변화해왔는지를 확인 가능
→ 특정 시점에 고객 수 급감/증가 원인 탐색에 활용
- Monthly Income Customer Count**
(고객 월 소득 분포)
고객군의 경제적 여력 파악
→ 고소득/저소득 고객 비율 확인
→ 가격 정책, 할인/프리미엄 전략 방향 결정에 유용

고객 대시보드

마케팅팀은 이렇게 활용해주세요!

Guide 3. 항목

- Education Level:**

Customer & Avg. Income (Line)

(고객 학력 수준별 분포 & 평균 소득)

막대 : 학력 수준별 고객 수

선형 : 해당 학력 그룹의 평균 월 소득

→ 학력 수준에 따른 고객 분포 및

구매력 파악

- Occupation :**

Customer & Avg. Income (Line)

(직업군별 고객 분포 수 & 평균 소득)

막대 : 직업군별 고객 수

선형 : 해당 직업군의 평균 월 소득

→ 어떤 직업군이 Zomato 고객기반에

많이 분포되어 있는지 파악

→ 소득 수준과 결합한 세그먼트화

가능

- Occupation:**

Avg. Income & Order

Frequency (Line)

(직업군별 평균 소득 VS. 주문 빈도)

막대 : 평균 월 소득

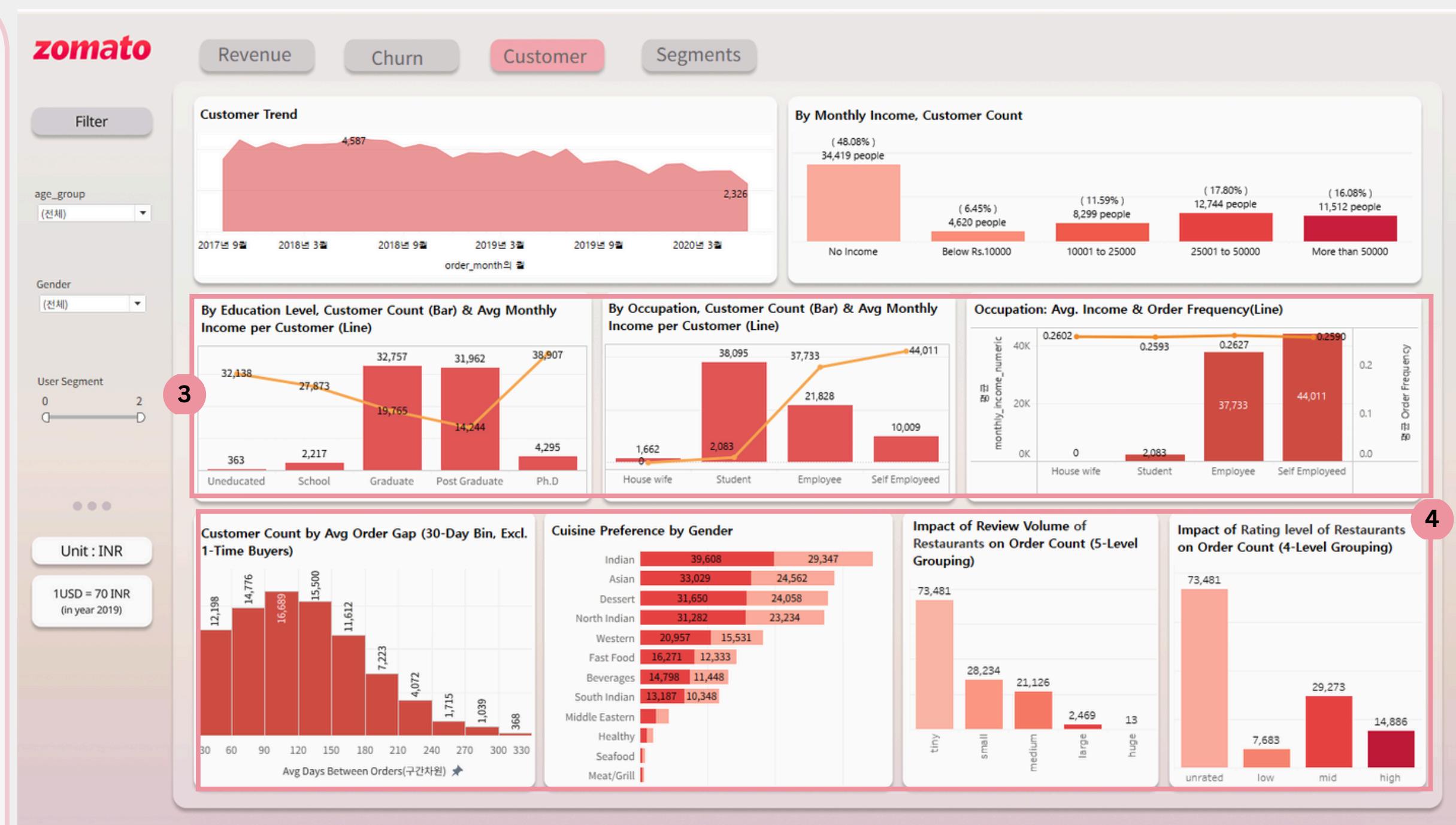
선형 : 해당 직업군의 평균 주문 빈도

→ 소득 대비 실제 구매 활동이

얼마나 활발한지 비교

→ 구매력만으로는 평가하기 어려운

행동 기반 인사이트 제공



Guide 4. 항목

- Avg. Order Gap (30Day-Bin, Excl. 1-Time Buyers)**

(평균 주문 간격 기준 고객 수 분포)

*30일, 단 1회성 고객 제외

→ 단골 고객의 평균 구매 주기 파악

- Cuisine Preference : Gender**

(성별 기반 음식 선호도)

성별에 따른 주요 음식 종류 선호 현황 파악

→ 성별 타깃 마케팅 기획

- Review Scale : Order Count**

(리뷰 단위 Scale 기준 주문 수)

리뷰 단위 별 주문 수 분포

→ 고객의 리뷰 활동 성향 파악,
리뷰 기반 고객 참여도 추정 가능

군집 대시보드

마케팅팀을 위한 활용 TIP

- 마케팅팀이 군집 대시보드를 보고 어떤 내용을 전반적으로 확인할 수 있고 이 대시보드를 통해서 어떤 걸 할 수 있을지 포괄적으로 담는 느낌 = 이게 결국 용도 = 목적

Guide 1. 필터 사용법

• 시즌 필터(문순 이후, 문순, 가을, 겨울)

인도의 계절성에 따라 군집별 주문/매출 변화 확인 → 계절별 주요 고객군과 구매 패턴 파악

• 월 필터

월별 트렌드 변화, 성수기·비수기 군집별 반응 분석 → 마케팅 집중/이벤트 시즌 전략 수립 활용

• 년도 필터

연도별 군집 변화 추이 확인 → 서비스 성과 분석 및 연도별 타겟 전략 활용

• 통화 단위 안내

INR(인도 루피) 기준 제공
(2019년 기준 환율:
1USD ≈ 70INR)



Guide 2 항목

◦ Customer Overview (고객 세그먼트 설명)

전체 고객을 3가지(Royal Regular, Premium one-timer, At-Risk Potential로 구분

• Ratio(customer Count & Promotion)

3가지의 세그먼트별 고객 수
→ 세그먼트 별 규모 파악

• Sales Contribution

각 세그먼트 별 매출 기여도->
마케팅 ROI(효율) 극대화 고
객 파악

• Avg. Spending per Customer(Each Cluster)

각 세그먼트 별 평균 소비 금액
파악 -> 고객 가치별 차별화
전략

• Active Days(Each Cluster)

세그먼트 별 전체 기간동안
활동일수 파악 → 세그먼트별
충성고객 진단

군집 대시보드

마케팅팀은 이렇게 활용해주세요!

Guide 3. 항목

• Customer Deep Dive

고객 세그먼트별로 행동, 소비, 소득, 주문 패턴 등 핵심 특성 분석하는 영역

• Customer Segment 필터

각 세그먼트를 선택해 원하는 데이터/지표 확인

• Customer Spending Amount Compared to Restaurant Average Spending Amount (4-Tier Classification)

고객의 소비 패턴을 4개 등급으로 구분하여 평균 주문 금액이 레스토랑 평균 대비 어느 수준인지 파악
→ 고액 결제층 타겟 전략, 저 소비군 활성화 전략 기획

• Total Sales per Month (Aggregated Across All Year)

세그먼트 별 연도/월 별 매출 트렌드 패턴 파악
→ 세그먼트 별, 월 별 마케팅 전략 기획

• Customer Proportion by Monthly Income Group(5-Tier)

세그먼트 별 소득 분포 파악
→ 소득층 분포 확인 후 가장 분포가 높은 고객들 타겟 전략



Guide 3. 항목

• Relationship between Order Frequency and Total Sales Grouped by Segment

세그먼트 별 주문 빈도와 총 매출의 관계

→ 주문 빈도별 타겟 프로모션, 구간별 차별화 전략 수립

• Order Interval per Customer & Average Order Count by Segment

세그먼트 별 평균 주문 간격과 주문 횟수의 분포

→ 세그먼트별 행동 패턴 파악, 맞춤형 타겟팅 전략

• Churn Risk Distribution Across Customer Cluster(5-Level-Assessment)

세그먼트 별 이탈 위험 등급 (5단계) 분포 파악

→ 5단계 분포 비율 확인 후 Highly at Risk 비중이 높은 세그먼트에 즉각적 리텐션 마케팅

군집 대시보드

Guide 4. 항목

- Targeted & Automated Action Plans**

고객 유형별로 자동 타겟팅 및 맞춤. 액션을. 실행 하는 전략적 추천 영역

- Loyal Regulars**

지속적으로 반복 구매/활동 활발한 고객군
→ 리워드/적립 집중 제공

- At-Risk Potential**

사용한 금액은 많으나, 최근 활동이 적거나, 재구매 주기가 길어진 고객군
→ 리텐션(이탈방지) 캠페인

- Premium One-Timers**

1~2회 단발성으로 고액 결제 후 이용하지 않는 고객군
→ 재구매 유도 전략

예란님!!!

지금 여기 발표대본 짜고계신가
요!??

마케팅팀은 이렇게 활용해주세요!



전략 제안



데이터 기반 : 공통 마케팅 전략 인사이트

문제 인식: 몬순(6-9월) 매출 하락

우기 시즌 (Monsoon Season) 전략

- 강수량이 가장 많은 몬순(6~9월)
- 비 오는 날 맞춤 콘텐츠
→ 비 오는 날, 영화 한 편과 딱 어울리는 음식 세트 추천
 - zomato X OTT 협업 :
→ ‘몬순 무비 나잇’ 세트 메뉴 출시

학생 다수 분포

- 대학교, 대학원 등 시험기간 공략
→ 시험기간 한정 심야 전용 메뉴
→ 학교 인근 배달비 무료 or 특정 레스토랑 제휴 할인
- zomato 대학생 서포터즈
→ 서포터즈가 배달 메뉴 리뷰 및 야식 인증 챌린지 운영
→ 서포터즈 피드백 수렴: 문제점, 프로모션 반응 등 수집하여 브랜드 이미지 개선에 활용
- 커뮤니티/바이럴 마케팅
→ 대학별 커뮤니티 오픈채팅 타겟 홍보
→ 인플루언서와 콜라보: 유튜브/인스타 공부 관련 크리에이터와 협업

문제 인식: (10월 이후) 가을부터 매출 증가

연말 성수기 시즌 전략

- 10~12월 인기 외식 업종/레스토랑 추천 큐레이션
→ 지금 예약 시 10% 할인 쿠폰 + 조마토 등급 안내 포함
- VIP 시크릿 혜택도 포함해 상위 고객 리텐션 강화
- 명절/연말 인사 콘텐츠와 함께 발송
- 전년도 인기 레스토랑 3유형 x 월별 TOP 큐레이션 매거진 발송

사회경제적으로 발달한 지역 위주로 성장

- South India: 인구 약 250M. IT산업 등 경제 중심지
- North India: 인구 약 450M. 인도 인구의 35% 이상
- West India: 인구 약 180M. 금융/산업 중심지

→ 매출 상위 3개 지역(S/N/W)을 ‘고성장 상업권’으로, 농업,관광,내륙 기반의 E/NE/C 지역의 ‘신흥 성장권’과 구분

- 마케팅 전략의 우선순위 명확히 설정 및 지역 특화 접근
- **상업권:** LTV 중심의 프리미엄 전략
 - **성장권:** 유입 확대 및 초기 정착 유도 전략 실행

문제 인식: 주말 매출 저조

주말 주문 활성화 전략

- 주말/야식 : 금~일 저녁 배달 주문 시 10% 할인
- 주말 단골 유도 메시지 & 레스토랑 쿠폰 발송
- 금요일 친구 추천 15% 할인
- 토요일·일요일 배달비 무료, 즉시 10% 할인
- VIP 사전 예약 시 리워드 제공



[데이터 드리븐 프로젝트 제안]

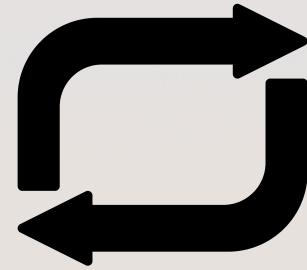
주말 배달 감소의 원인:

1. 문화/종교: 주말엔 가족 단위의 외식 문화 강함
2. 요리 중심: 주말엔 집에서 전통 요리 직접 조리
3. 사회적 활동 증가: 주말 레스토랑 예약 선호

주말 레스토랑 앱 통해 예약 시, 리워드 제공
→ 조마토 앱 친화도 및 이용률 ↑
→ 장기적으로 조마토 충성고객층 확보

Loyal Regular 고객을 위한 전략 제안

자동화된 리텐션 전략과 감성 마케팅을 통해 고객 생애가치 증대와 자발적 홍보를 유도
→ 장기적으로 수익성과 브랜딩을 동시에 확보



반복 구매 유지

자동 리워드 시스템

“5회 주문 시 무료배송 쿠폰 자동 지급”
앱 알림: “1회만 더 주문하면 리워드 도착!”

주기적 리마인드 캠페인

장기 미주문 고객에게 push 알림 + 쿠폰

선호 메뉴 기반 추천

최근 3회 주문 기준 유사 메뉴 자동 추천
→ 추천 시점: 주말 전날/ 식사 시간대 전후

객단가 상승

콜라보 마케팅

아이돌, 게임, 캐릭터 굿즈 한정 프로모션
→ 특정 금액 이상 주문 시 참여 가능
→ 예: “월간 팬굿즈 + 무료배송”



자연스러운 확산 유도

SNS 인증 프로모션

“Zomato 주문 사진 업로드+해시태그”
→ 고객 주문 시 추가 할인

등급제 & 멤버십 도입

Silver~Diamond 등급 운영

→ 월 1회 정기 혜택

→ “다음 등급까지 N 회 남음” 알림
→ 브랜드 충성도 강화 + 소비 유도

At-risk Potential 고객을 위한 전략 제안

이탈 위험 고객, 단발성 소비로 끝나기엔 아까운 잠재 충성 고객층
→ 재방문 유도와 소비 성향별 맞춤 전략을 통해 ‘로열 고객’으로 전환

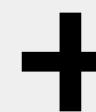


자동 리텐션 시나리오 진행

복귀 유도 시나리오 자동화

- Day 30: 10% 할인 쿠폰 이메일
- Day 37: 선호 기반 메뉴 추천 푸시 알림
- Day 45: 한정 리워드 쿠폰 발송

60, 90일도 있으면 좋을 것 같다.



시간 압박 전략 병행

예: “쿠폰 사용 가능 기간 D-3” 알림



소비 성향별 맞춤 전략

premium/vip_spender

- 레스토랑 평균 금액대 대비 높은 금액을 지불
→ 재방문율 높음
- **준-프리미엄 사용자로 간주**한 혜택 제공
(프라이빗 쿠폰&리마イン더,
프리미엄 큐레이션 메일링,
리텐션 전용 혜택 안내 채널 등)

expected/under_spender

- 레스토랑 평균 금액대 대비 낮은 금액을 지불
→ 재방문율 낮음
- **이탈 집중 관리 필요**

(첫 리텐션 시 혜택 강화: 복귀 첫 주문 시 쿠폰,
심리적 유도 텍스트: 지금 주문하면 맞춤 혜택!,
번들 메뉴 유도: 소액 소비 업셀링 등)



주기적 방문 유도 프로모션

고정 혜택 프로모션 시스템화

- 매주 금요일 고가 음식 할인 쿠폰 발행
- 점심, 저녁 타임(11시/17시)
선착순 타임어택 쿠폰 지급
- 주말에는 ‘장보기’ 카테고리 중심 프로모션 운영
→ 예: 첫 이용 고객 장보기 20% 할인 쿠폰

Premium One-Timer 고객을 위한 전략 제안

구매력은 높지만 단발성인 프리미엄 고객층을 겨냥한 차별화된 경험과 혜택

→ 자동화된 타겟팅 + 프리미엄 브랜딩 전략이 핵심



복귀 유도



미복귀 고객 자동 타겟팅

14일/30일 등 미접속 기준일 초과 고객
→ 쿠폰 + 추천 메뉴 앱 알림/이메일 발송
→ “다음 주문 시 ₹15,000 할인 쿠폰 도착!”

고가 주문자 자동 태깅

Premium One-Timer 평균 소비 금액 초과
→ **자동태깅 후 프리미엄 맞춤 전략**
→ 프리미엄 메뉴 주 1회 큐레이션 발송
→ 프리미엄 고객 전용 할인 제공
→ 고급 레스토랑 위주 추천/ 광고
-프리미엄 테마 세트 메뉴 구성

한정 기간 리미티드 혜택 강조

“프리미엄 고객님만을 위한 48시간 한정 할인”
→ 타임어택 + 한정 메세지 (구매전환율↑)

프라이빗 초대형 이벤트 연계

고액 고객 대상 오프라인 셰프 테이스팅 이벤트
Zomato VIP 디너 초대



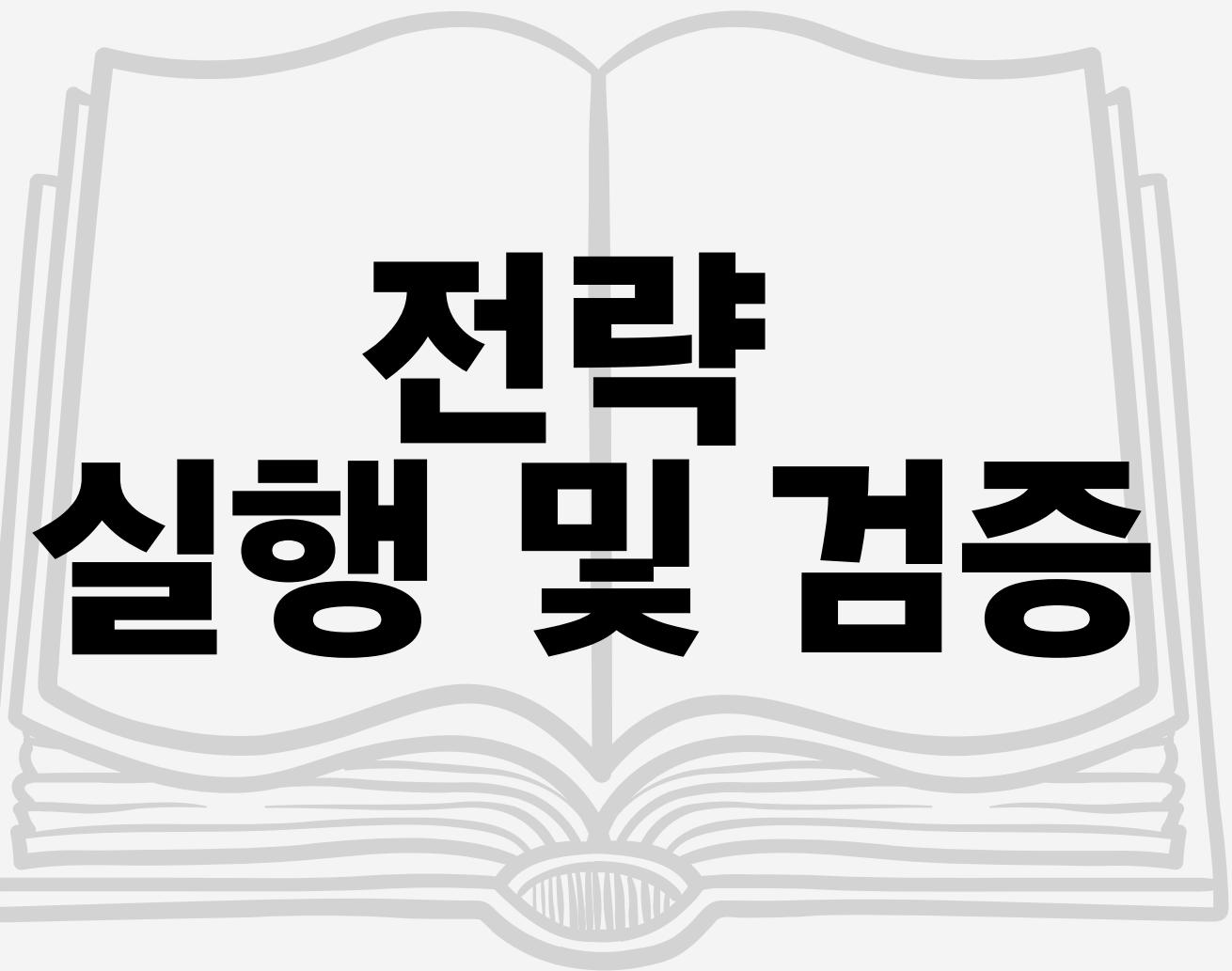
만족도 강화 & 제휴 확장

레스토랑 추가 서비스 제공

- 고급 레스토랑 우선 예약권
- Zomato 프리미엄 고객 위한 명당 우선 배정
- Zomato 프리미엄 고객만 주문 가능한 메뉴
 - 웰컴 드링크 또는 셰프 스낵 제공
 - 기념일 메세지 디저트 플레이트 제공

고급 브랜드와 연동 프로모션

-신용카드사, 뷰티/호텔 브랜드와 연계
예: “Zomato에서 ₹100,000 이상 주문 시
○○호텔 뷔페 1인 이용권 추첨”



전략 중 자동화 메일링 시스템 구현

기존 방식

수동으로 코드를 실행해서, 고객을 추출하고 메일을 직접 발송하는 방식

로컬 Python 실행
고객 데이터 전처리/필터

수동 코드 실행
메일 발송 함수 직접 호출

발송 결과 수동 확인



자동화 방식

고객을 자동으로 선별해 정기적으로 (격주 월요일 09:00)
쿠폰, 레스토랑 추천 등 이메일을 자동 발송하는 “자동화 시스템” 구축

Google Cloud Run
자동 메일 발송 프로그램을 배포

Google Cloud Storage
고객 데이터, 쿠폰 이미지 보관

Google Cloud Scheduler
지정한 시간마다 Cloud Run
서비스에 HTTP 요청

Google Run Service
GCS에서 최신 데이터를 읽고,
조건 필터링 하여 이메일 발송

자동화 메일링 시스템 테스트

쿠폰 발송 대상

최근 2주간 주문 기록이 있고,
At-risk Potential중에서
주문 금액과 주문 수량이
모두 평균 이상인 고객을 선별 ->3명

이메일 설정

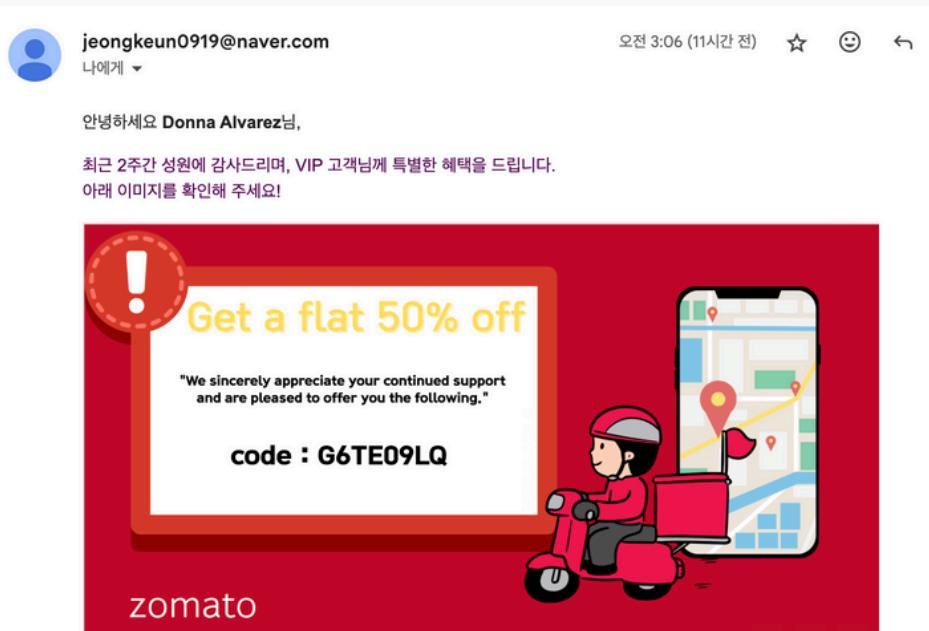
이름은 데이터에서의
user 이름을 그대로 사용하고,
이메일은 수강생 이메일로
변경 후 테스트

시간 설정

Google Scheduler를 사용해서
매주 화요일 03시06분에
메일이 보내지도록 테스트

테스트 검증

3개의 메일 정상작동



한계점

정적 데이터 기반

- 현재 시스템은 저장된 csv 데이터 기준으로만 분석, 발송하므로, 실시간 고객 행동변화는 어렵다. 하지만 데이터 베이스가 구축되어 있고, 데이터가 지속적으로 업데이트 된다면, 보완 가능

실제 행동 변화 검증 불가

- 메일 발송 이후 고객의 실제 반응, 재구매율 증가 등 행동변화에 대한 실증적 검증은 추가적인 추적 데이터와 실험 설계가 필요.

시스템 장애 발생 가능성

- 클라우드 환경 또는 SMTP 서버 장애 시 자동화 프로세스가 일시적으로 중단될 수 있다.

프리미엄 고객 전략 검증 - A/B 테스트 가상 설계안

목적 프리미엄 전략이 고객 행동 변화(재방문, 주문 금액 등)에 미치는 영향을 정량적으로 검증하기 위한 가상의 A/B 테스트 설계

가설 프리미엄 전략을 적용받은 고객군(B그룹)은 기존 고객군(A그룹) 대비 재방문률과 평균 주문 금액이 유의미하게 증가할 것이다.

A그룹(Control)

기존 고객 경험 유지



B그룹(Variation)

프리미엄 전략 적용

큐레이션, 고급 메뉴 추천 등



대상고객 Premium One-Timer 중 무작위 샘플링 (50:50 분할)

테스트 기간 : 2주간

노출 조건: 최초 로그인 시 자동 그룹 지정

전환 기준

2주 내 1회 이상 주문 발생

활용 목적

- 데이터 기반 마케팅 효과 측정 방안 확보

- 실무 관점의 전략 검증 및 향후 실험 기반 최적화 가능성 확보

비교 지표 (Metric)

Recency

days_since_last_order(핵심지표),
recency_ratio, avg_days_between_orders

Frequency

order_frequency,
total_order_count

Monetary

avg_order_amount,
total_sales_amount, monetary_per_day

*제한 사항:

- 정적 데이터 기반 시뮬레이션: 실제 행동 변화 검증까지는 어려움
- 통계적 유의성 분석(T-test 등)은 실제 서비스 도입 후 가능
 - 유의미한 차이 분석은 실제 플랫폼 테스트 필요

프로젝트를 마무리 하며

한계점

○ 데이터 불균형 & 분포 단조로움

- Zomato 데이터의 특성상 고객층이 20~30대 중심으로 편중되고, 직업/소득/연령대 범주 다양성이 낮아 세분화가 어려웠음
- 그 결과, 클러스터링 시 일부 클러스터(0번)에 과도하게 쏠리는 현상이 발생
- 머신러닝 기반 분석에 적합하지 않은 구조로 인해서 군집 간 구분이 뚜렷하지 않음

○ 고객 이탈 예측 모델의 한계

- 시계열 구조 이해 부족으로 초기에 데이터 누수가 발생했고, 다양한 이탈 정의 방식을 총 5차례 시도했지만, 클래스 불균형이 지속되어 실무적 예측에 한계
- 실직적인 예측 성능 개선으로 이어지지 못함
- > 결국 예측 기반 전략이 아닌, 행동 기반 세분화(군집 분석)로 분석 방향을 전환하게 됨

아쉬웠던 점 & 보완할 점

○ RFM 분석 구조화 부재

- RFM 요소와 유사한 의미의 파생변수를 만들어 사용했으나, 'RFM 분석'이라는 방식으로 구조화해보지 않았던 것이 아쉬움.
- **RFM 기반으로 고객 행동을 정량화해보는 시도를 해봤더라면, 군집의 타당성과 전략 제시가 더 설득력 있게 가능했을 것으로 생각되어 추후 추가 진행해보고 싶음**

○ 클래스 불균형 대응 부족

- 이탈 예측 모델에서 클래스 불균형이 지속적으로 문제였음
- 일부 버전에서 test set에서 이탈자가 전혀 존재하지 않는 문제도 발생함
- **SMOTE 및 GAN 기반의 합성 데이터 생성 기법 같은 방식 등의 보완 기법들을 추후 활용해서 불균형 문제를 보다 정교하게 다뤄보고 싶음**

Thank
You